

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

The White House Office of Science and Technology Policy—on behalf of the Fast Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics of the Subcommittee on Networking and Information Technology Research and Development (NITRD) of the National Science and Technology Council, the National Artificial Intelligence Initiative Office, and the NITRD National Coordination Office—released a Request for Information (RFI) seeking public input on how the United States can responsibly advance and adopt privacy-enhancing technologies (PETs) in the United States in a manner that equitably benefits individuals and society. The RFI was published in the Federal Register on June 9, 2022, and the comment period was open through July 8, 2022.

This document contains responses received from academia, the private sector, and civil society. Please note that several submissions are currently under review and will be posted later.

In accordance with the RFI instructions, only the first 10 pages of content were considered for each response.

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Table of Contents

Cover page	
Table of Contents	
Accenture	1
Access 4 Learning Community	7
ACT The App Association	13
Agita Labs	22
AHIP	35
Altman, Micah; Cohen, Aloni; and Vadhan, Salil	42
Amazon Web Services (AWS)	50
Anonos	56
Archer, David W.; Varia, Mayank; Smart, Nigel; Malozemoff, Alex; Darais, David; Baum, Carsten; Rosulek, Mike; Tromer, Eran; and Near, Joe	80
Argonne National Laboratory (ANL)	88
Arm	101
Asemio	108
Austin, Lisa; Lie, David; Nikolov, Aleksandar; and Papernot, Nicolas	117
Bitfount	127
Brave Software	132
BurstIQ	137
Canetti, Ran; Kaptchuk, Gabe; Reyzin, Leonid; Smith, Adam; and Varia, Mayank	146
Carnegie Mellon University	160
Center for AI and Digital Policy (CAIDP)	178
Chu, Bessie	190
Cloudflare	192
Computer & Communications Industry Association (CCIA)	203
Computing Research Association (CRA) Computing Community Consortium (CCC)	210
Confidential Computing Consortium	213
Cybernetica	220
Data Freedom Foundation	226
Datavant	237
Deloitte Consulting LLP	244
Diveplane Corporation	258
Duality Technologies	269

Electronic Privacy Information Center (EPIC)	280
Fortanix Inc.	289
Future of Privacy Forum	296
Georgetown University Massive Data Institute	306
Google	311
HUB Security Limited	321
IBM Research	331
IEEE Standards Association (IEEE-SA)	339
Immuta	342
Inpher	348
Intel Corporation	352
IQVIA Inc	362
Ixup Solutions	369
Juniper Networks	374
Karlsgate Inc	378
Knexus Research Corporation	386
Kotevska, Olivera	392
Kulshrestha, Anunay; Mayer, Jonathan; and Scheffler, Sarah	399
LeapYear Technologies	404
Mastercard	409
Meta	415
MITRE Corporation	426
MOSTLY AI Inc	439
Mozilla	445
NowVertical Group Inc	453
NTT Research	459
OpenMined	476
Palantir	484
Peisert, Sean	495
Prifina Inc	501
Privacy Analytics	512
PUBLIC Group International Ltd	525
R3	537
Restore the Fourth	547
Richardson, Douglas	552
Stealth Software Technologies Inc	563
Syntheticus Ltd.	574
The Ethical Tech Project	584

The Rise of Privacy Tech (TROPT)	590
Turner, Sean and Zucker-Scharff, Aram	693
University of Southern California Information Sciences Institute	696
Visa Inc	700
VMware	704
Wehbe, Joseph	715
Western Governors' Association	720
Work, Paul R.	739
World Data Exchange	745

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Accenture

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



July 9, 2022

Dr. Alondra Nelson
Acting Director
Office Science and Technology Policy
Re: Advancing Privacy-Enhancing Technologies (Docket Number: 2022-12432)

Dear Director Nelson,

As a leading global professional services company, Accenture provides a broad range of services and solutions in strategy and consulting, technology, interactive, and operations, that span all industries. We combine artificial intelligence (AI) with deep industry and analytics expertise to help our clients embrace these emerging, intelligent technologies confidently and responsibly and our 710,000 employees serve clients in more than 120 countries.

We are committed to create lasting change at the intersection of technology and human ingenuity for our business and communities. Our clients reach virtually every American consumer and rely on data and digital platforms to deliver products and services. Businesses that treat data collection and data strategy as part of the consumer experience will benefit from greater consumer willingness to share information.

We are committed to partnering with our clients to protect consumer privacy and support the establishment of a [national consumer privacy law](#) to protect consumers, promote innovation and advance U.S. competitiveness. Additionally, [Accenture Labs](#), which began conducting R&D on Privacy Enhancing Technology in 2015, are developing a new framework for [privacy-preserving data cooperatives](#). With a well-designed cooperative, companies can share data and collaborate without concerns around trust, compliance, privacy and data control or ownership. There are many technological techniques available to make this possible; confidential computing, homomorphic encryption and multiparty computation can all contribute. Another, often overlooked, benefit of privacy enhancing techniques is their potential for alleviating reliance on those energy-intensive technologies.

We appreciate the opportunity to submit comments to the Office of Science and Technology Policy 's Request for Information on "Advancing Privacy-Enhancing Technologies" and look forward to continuing to support your efforts.

Sincerely,

Paul Daugherty
Group Chief Executive – Technology & Chief Technology Officer
Accenture

Accenture
Comments to Office of Science and Technology Policy
Advancing Privacy-Enhancing Technologies
July 9, 2022

Question 1: *Specific research opportunities to advance Privacy Enhancing Technologies (PETs):*

While PETs are certainly beneficial in improving collaboration between entities and mitigating data privacy concerns, they are currently slow, expensive, and come at a large computational cost. Accenture strongly supports increased funding for research and development of classical PETs including privacy assessments and protection mechanisms for Differential Privacy, Federated Split Learning (FSL), Homomorphic Encryptions and Secure-Multi-Party Compute (MPC) solutions including cloud-based versions of such solutions to reduce computational costs, decrease risks of data breaches, and speed up data processing. In addition, Accenture supports the Department of Energy's (DoE) current research of a Quantum Internet, which could be accelerated to explore specific quantum-information protection techniques.

Questions 2 & 9: *Specific technical aspects or limitations of PETs: Existing barriers to PETs adoption:*

Homomorphic Encryption

The main limitation to mainstream adoption of Homomorphic Encryption is the computational intensity and cost of the processing. These factors limit the amount of data that can practically be used, which currently makes the process extremely slow, rendering it impractical and expensive for real-time or near real-time processing. Additionally, because the data remains encrypted throughout the process and there may be limits on the types or number of operations that can be performed, it is essential for the data processor and data owner to have agreements in place around the structure and content of the data as well as the processing that will take place so that the data processor cannot interrogate or experiment with the data.

Secure Multi-Party Communication

The main drawback to utilizing Secure Multi-Party Communication (MPC) technologies are the high computational costs needed to operate. MPC's require a lot of communication between parties, which can add further latencies during the computation process. Another factor with some MPC's is the complexity of representing a business problem as a logical circuit with a compliant structure, which can require some specialist skills. From a security perspective, one point to note is that MPC's don't protect against "poisoning" attacks, where one of the parties could attempt to maliciously influence the results of queries by another party by intentionally

using false or misleading data to intentionally lead to an answer which is not correct (i.e., exaggerating or understating a statistical result to drive another party to draw incorrect conclusions).

Quantum Information Based Privacy & Security Techniques

As we advance to a Quantum-Internet, meeting DiVincenzo's criteria for Quantum computing and communications, Quantum-Information security and privacy preservation techniques will be needed. Additionally, quantum security and privacy techniques could provide an alternative to classical PETs, which in their current versions are slow, vulnerable, and costly to operate.

Accenture has been ideating in quantum data privacy and security techniques. One of those techniques, US Patent Application #20220014364 - QUANTUM INFORMATION INTERCEPTION PREVENTION offers the strongest privacy, confidentiality and security currently conceivable. This invention promises to preserve security and privacy even against adversaries with unlimited resources, time and perfect knowledge (e.g., knowing protocols and cipher secrets). Technical aspects of this invention involve entangling a user's information qubit with two or more ancillary qubits and gate-cipher(s)—which together function as a quantum-key the user can control. The owner of this quantum-key would have the ability to control when, how and if the original information inside the protected qubit would be communicated or shared with another.

Additionally, this invention would enable the determination of whether the original information in the qubit was modified and provide the option to destroy the quantum-keys, thus effectively disabling the information in the protected qubit by making it incoherent. This information disablement happens even if the protected information qubit is no longer in control of the owner. This remote disablement aspect would be desirable if the information qubit is stolen or if the owner no longer consents to disclosing private information.

This invention also has the ability for private multi-party quantum computation. A quantum computer can perform non-measurement-based operations on the protected information qubit without access to the quantum-keys and return it to the owner. The owner can maintain physical control of the quantum-keys, guaranteeing that even if the other party attempted to decrypt or measure the qubit, they would only receive a random result which does not reveal any of the private information. If an adversary gained unauthorized access or control of a protected qubit and its quantum-keys, the adversary would only have one chance to reverse the quantum gate-ciphers on the entangled ancillary qubits.

We encourage the investment in quantum-based techniques as Quantum will enable a new paradigm of privacy and security techniques enabled by quantum computers and the quantum internet.

Question 3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:

Industry is increasingly leveraging data to improve analysis and decision-making—creating more personalized experiences for customers and greater operational efficiency. Currently, due to PET's limited capacity, the use cases of homomorphic encryption for multi-party ML are limited and only used in the inference phase. However, Federated Learning (FL) and Federated Split Learning (FSL) are better solutions for multi-party learning in the training phase. Initiatives and investments should include FL and FSL. Differential Privacy, Data Sanitization and Data Anonymization are also very impactful techniques that should be factored in to achieve or enhance privacy. As technology continues to improve, there is a wide variety of sectors that would benefit from PET's.

Healthcare:

The healthcare industry has some of the strictest regulations for data privacy and security. Homomorphic encryption could allow the healthcare industry to leverage machine learning services for crucial insights into areas such as medical diagnostics, leading to faster, more accurate diagnoses and more effective treatments.

A relevant healthcare use case to look to would be the MyHealthMyData (MHMD) program. MHMD, an EU-funded project, is looking at how to share anonymized data for medical care, research and development, while giving people ownership over their health data. The platform combines blockchain, smart contracts, dynamic consent and a comprehensive suite of data privacy and secure analytics tools including Homomorphic Encryption and MPC.

Financial Services:

Financial organizations are also under strict regulations for maintaining privacy of customer information. In decisions, such as loan or credit approvals that involve data from multiple owners/sources, Homomorphic Encryption can keep financial information secure while still allowing for the automation of predictive tasks such as loan approval and know your customer (KYC) services using machine learning.

Service Checks:

PETs will also be valuable in the aviation, railway, manufacturing, oil and gas industries, where we're seeing techniques like federated learning and SMPC utilized to predict the condition of in-service equipment or identify the root cause of an equipment failure. In these industries, an equipment vendor needs to collect data from many different customers and environments simultaneously to monitor and predict parts deterioration and optimal maintenance times. These vendors operate independently and by practice do not share performance data. Multi-party ML training, enabled by FL and SMPC, allows a model to both train and have inferences across these privately held datasets.

Another, often overlooked, benefit of privacy enhancing techniques is their potential for alleviating reliance on those energy-intensive technologies. Linking our long term privacy goals with our long term sustainability goals may bring additional attention, prioritization, and investment.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Access 4 Learning Community

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Larry L Fruth II, PhD
Executive Director/CEO
Access 4 Learning Community
Non-Profit Educational Technology Collaborative

Data access today is critical as major decisions on policy development, funding, program impact, and an array of other “data driven, time sensitive decisions”. This is all dependent on having access to the right data at the right time. Layer on top of this is the need for control by the appropriate data stewards in the sectors the data is held within – privacy and security layers!

The *Access 4 Learning Community (A4L)* is proud to respond to *Advancing Privacy-Enhancing Technologies* request for Information. We hope that this feedback and details of our successful data management and privacy work over the past 25 years will inform the agency’s understanding of how to accelerate the responsible development and adoption of PETs in a manner that promotes trust in data processing and information technologies.

The A4L Community, and its special interest group the Student Data Privacy Consortium (SDPC), is a unique, non-profit collaboration composed of schools, districts, local authorities, states, US and International Ministries of Education, software vendors and consultants. The Community is “Powered by SIF” as its major technical tool to help manage learning data simply, securely and in a scalable, standard way regardless of platform. The SDPC is designed to address the day-to-day, real-world multi- faceted issues faced when protecting learner information by setting common expectations between market providers and end users. The A4L Community has united these education technology end users and providers in an unprecedented effort to ‘*connect and secure effective learning ecosystems*’ to give teachers more time to do what they do best: teach.

The education sector is still struggling today with privacy enforcement in general. There is a great deal of data sharing that occurs in the education environment. This included not only internal data sharing but external data sharing between local education agencies and community partners, higher education agencies, edtech vendors and researchers. A4L & SDPC are on the verge of implementing privacy enforcement technologies through the Global Education Privacy Standard (GEPS) which includes Privacy Obligation Documents (PODs). PODs are the first iteration of what we call “privacy over the wire”.

A4L has had great success in assisting both schools and providers to navigate privacy issues while providing effective and secure online tools. One of SDPC’s cornerstone work has been the creation of the National Data Privacy Agreement (NDPA) that is in use now in 34 states. This NDPA has successfully bridged the gap between school and provider expectations around protecting student data. The SDPC Registry currently contains over 67,000 Data Privacy Agreements between schools and vendors for close to 8,000 applications impacting 34 million students. Through all of this work, the SDPC’s success can be attributed to the focus on bringing all sides together to find common ground and build from there.

The Access 4 Learning Community (A4L), linked to the Student Data Privacy Consortium (SDPC), has “moved the needle” on the marketplace addressing both interoperability and privacy controls of student data – the two must be taken collectively. These activities are global in nature but local in impact which has added to its success. Below are the areas of focus of A4L & SDPC related to data privacy;

Resource Registry

A set of “on the ground and real world” set of privacy tools allowing schools to manage and communicate on the software solutions impacting learning. The **SDPC Resource Registry** allows schools, districts, divisions, states, territories, and vendors to find resources, adapt them to their unique context and implement needed protections. This school-based tool allow schools to manage their applications and privacy “rules of the road” and currently is still growing with;

- 4 Countries collaborating
- 7,959 Resources in Registry Database
- 10,794 School Districts represented
- 34 States participating
- 86 Participating Vendors
- 70,000 Signed Data Privacy Agreements
- 34,000,000 Students supported by SDPC tools

National Data Privacy Agreement

The **National Data Privacy Agreement (NDPA)** has been developed with extensive review and comments from schools, districts, state organizations, marketplace providers and their legal representatives. It is designed to address common student data privacy concerns and streamline the educational application contracting processes for schools/districts who do not have the legal or fiscal resources and vendors who previously had to sign “one off” contracts with each of the over 13,000 US school districts. While the NDPA allows for any state specific legislative requirements, the majority of the privacy expectations are standardized and can be used by any entity as part of their Terms of Service Agreements.

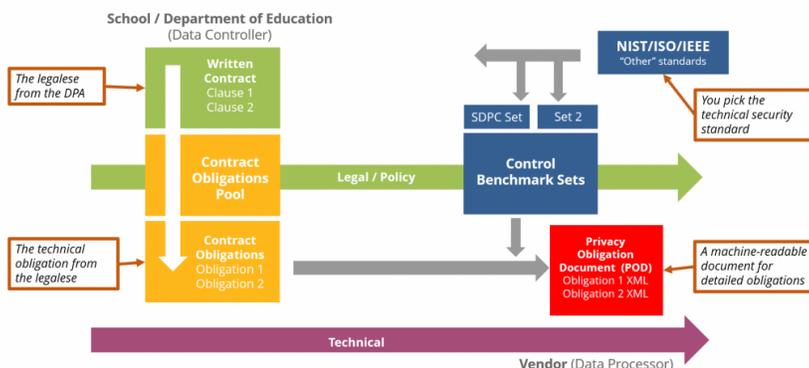
The NDPA is the first step in getting common expectations around safeguarding student privacy – but only the first step.

Global Education Privacy Standard (GEPS)

After the NDPA is established, the next step in getting common expectations around safeguarding student privacy established begs a question: What if you could automatically communicate these expectations between end users and vendors?

- Simply what you can and can't do with data
- Stipulates purpose - why you are being supplied with the data
- Legal obligations and technical benchmarks
- A green list of data elements you can access
- Data conditions (subsets of data e.g. senior years only)
- What you should do with the data if the data is no longer required
- Details of recipient of the data, who to contact if there are issues, who is handling the data, countries impacted

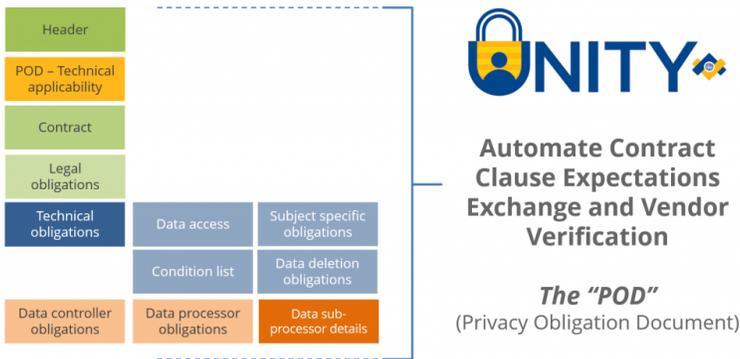
The **Global Education Privacy Standard (GEPS)** is designed to simplify, and in some cases automate, the software on-boarding process by aligning contract clauses to technical obligations to security standards. It is designed to set common expectations between vendors and their customers.



Privacy Obligation Document (POD)

A “POD” may sound foreign to you but this one addressing student privacy expectations is not a group of killer whales – on second thought that may not be a bad idea! The SDPC “**Privacy Obligation Document**” (**POD**) is a machine readable “meta-data” document that communicates the specifics of the privacy obligations each vendor is contractually bound to and ensures acknowledgement of those obligations by the vendor before they receive data. This split-second exchange and acknowledgement provides school districts, students, and parents with the assurance that their data privacy will be protected.

PODs contain all the required privacy metadata in a standard, industry accepted format. These obligations are driven by national and state laws as well as local requirements. The obligations are driven by Data Privacy Agreements (DPAs), including the National Data Privacy Agreement, executed by and between the LEA and Provider. The DPAs contain references to all applicable state and federal laws, technical obligations and security requirements where applicable. This will ensure that privacy *and* interoperability can be managed at the same time, bringing much-needed support to districts trying to keep up with the demands of exchanging data across their growing software infrastructure, while protecting the privacy of sensitive student, parent, and staff information. The POD is part of the *Global Education Privacy Standard (GEPS)* and can be “carried over the wire” utilizing the newest ‘Unity’ SIF Specification, developed by the A4L Community.

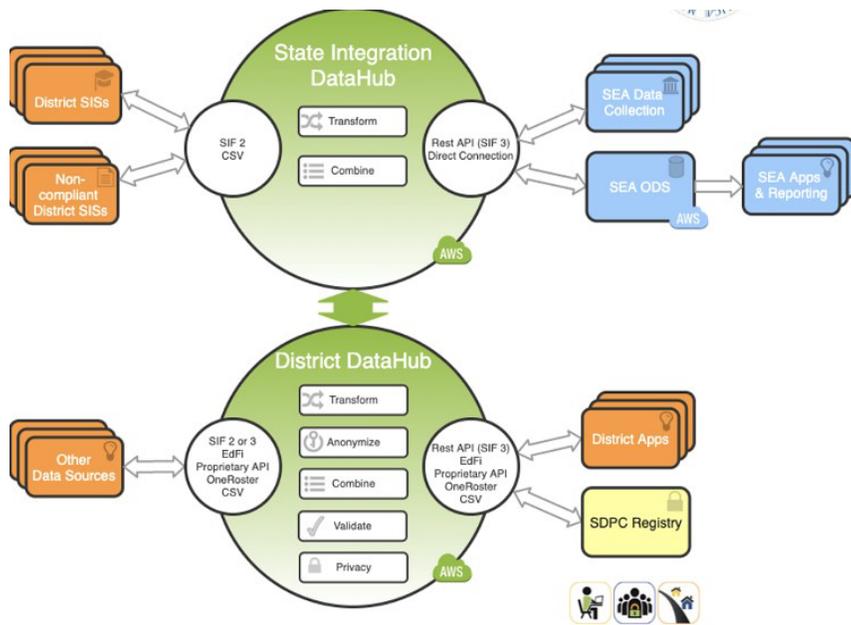


Data Hub Project

The Data Hub is a project to create a connected and secure ecosystem, balancing the needs of privacy and interoperability in one place. The main goals of the project are to:

- Align to open data standards
- Provide an easy-to-manage toolset for districts to handle connecting their many software systems
- Leverage existing state data connections
- Meet vendors “where they are” _to avoid external dependencies for project success
- Handle privacy and interoperability at the same time, in one system

The Data Hub project is also the first proof of concept for the use of PODs to enforce privacy obligations. No data is exchanged until there is an acknowledgement of the POD obligations by the receiver of the data. This ensures that the data is only being shared with entities that have acknowledged/certified their data privacy obligations.



Standardized Infrastructure

Almost all organizations that develop data models for their particular members/marketplace vertical do not provide an infrastructure for “data over the wire”. They allow for developer/user choice for moving the data between applications or in some cases provide suggested guidance. This does not allow for the full usage of PETs.

In 2014, the Access 4 Learning Community released an open standard infrastructure, SIF 3.0, leveraging a REST based approach to data exchange. The key contribution of this release was to define, coordinate and standardize the ways in which a RESTful educational service can be accessed securely, robustly, and in real time by multiple RESTful clients. This openly developed and freely accessible infrastructure blueprint is separate from any data model defining the payloads it carries, which means it can be used to support many different data models in many different locales such as being down currently with data models from Australia, New Zealand and North America and soon between other Standards Developing Organizations (SDOs) across various verticals. Designed to be separate, but support these data models, the standardized infrastructure can carry any data model and is unique in that it:

- Addresses privacy controls
- Is scalable for use
- Contains security controls
- Has been load tested
- Has an associated certification program

The benefits for a standardized infrastructure used between all data sectors in that it would:

- Pave the way for plug-and-play interoperability between sectors
- Empower meaningful validation and quality control
- Accelerates development.
- Allow for greater marketplace choice.
- Provides functionality clarity.
- Lowers barriers between sectors and geography

Summary

Access 4 learning and the Student Data Privacy Consortium have done more in the education sector to address operational privacy issues than any other organization. A4L has laid the foundation for secure data exchanges with over 20 years experience producing, publishing and implementing data interoperability standards. SDPC has laid the groundwork for addressing the legalities of data shAring across the entire educational ecosystem.

Now that A4L and SDPC have combined their expertise to address the issue of enforcing data privacy obligations from a technical perspective, we are well positioned to have an even larger impact on the education sector. The Global Education Privacy Standard and Privacy Obligation Documents are examples of Privacy Enforcing Technologies that will eventually become common practice in education. These PETs can be replicated across many sectors.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

ACT | The App Association

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

July 8, 2022

Office of Science and Technology Policy
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, District of Columbia 20504

RE: Comments of ACT | The App Association to the Office of Science and Technology Policy on its Request for Information on Advancing Privacy Enhancing Technologies

ACT | The App Association (App Association) appreciates the opportunity to submit views to the Office of Science and Technology Policy (OSTP) on its Request for Information on Advancing Privacy Enhancing Technologies (PETs). The App Association agrees with OSTP that PETs are an important tool for unlocking the full potential of the data economy and can help ensure that innovation in emerging technologies runs concurrently with a respect for basic human rights, promotes equity in data processing activities, and increases trust in the digital economy writ large.

The App Association represents thousands of small business software application development companies and technology firms that create the technologies that drive internet of things (IoT) use cases across consumer and enterprise contexts. Today, the ecosystem the App Association represents—which we call the app economy—is valued at approximately \$1.7 trillion and is responsible for 5.9 million American jobs.¹ Alongside the world’s rapid embrace of mobile technology, our members create the innovative solutions that power IoT across modalities and segments of the economy. App Association members exist at the cutting edge of the research, development, and implementation of PETs in their products and services.

Consumers who rely on our members’ products and services expect that our members will keep their valuable data safe and secure. The small business developer community the App Association represents practices responsible and efficient data usage to solve problems identified across consumer and enterprise use cases. Their customers have serious data security and privacy expectations, and as such, ensuring that the company’s business practices reflect those expectations by utilizing the most advanced technical protection mechanisms (e.g., end-to-end encryption) is a market-driven necessity. For this reason, we support the Administration’s goal of ensuring the United States leads the world in responsible data practices and technologies, including PETs, which are critical to our economic prosperity and national security, and to maintaining

¹ The App Association, State of the App Economy 2020, January 2021, <https://actonline.org/wp-content/uploads/2020-App-economy-Report.pdf>.

the core values behind America's scientific leadership, including openness, transparency, honesty, equity, fair competition, objectivity, and democratic values.

I. General Comments on Privacy Enhancing Technologies

The RFI notes that PETs encompass a “broad set of technologies that protect privacy,” listing examples such as “secure multiparty computation, homomorphic encryption, zero-knowledge proofs, federated learning, secure enclaves, differential privacy, and synthetic data generation tools.”² The RFI goes on to emphasize “privacy-preserving data sharing and analytics technologies,” a subset of PETs that help facilitate data sharing between entities and ensure those entities can perform advanced analytics without revealing the identities of the data subjects.

While we agree with the importance of advancing such PETs in the analytics space, we note that the constellation of PETs that app developers rely upon extends far beyond those few use cases. While there is no universally accepted taxonomy of PETs (or definition for the term, for that matter), existing efforts typically include categories of technologies that assist in the process of obtaining consent, data minimization, anti-tracking, encryption, anonymity, and control, among other categories, in addition to the technologies mentioned in the RFI.³ One strategy OSTP may consider when taking stock of the full spectrum of PETs for its analysis is to either bifurcate its research into business to business (B2B) and business to consumer (B2C) buckets, or to simply track the entire life-cycle of a given piece of data in various industry verticals, from collection, to processing by the first-party collector and subsequent processing by service providers or other third parties. This would help ensure that OSTP takes all possible PETs into account, including those utilized by B2B and B2C developers.

In general, we encourage OSTP to take as broad a view of PETs as feasible as it takes on the responsibility of coordinating the national strategy to ensure that these tools benefit individuals and society. This would track similar work carried out by allied governments and existing efforts at the congressional level. For example, the Privacy Commissioner of Canada took an inclusive view of PETs in its report, “A Review of Tools and Techniques,” saying, “PETs are intended to allow users to protect their (informational) privacy by allowing them to decide, amongst other things, what information they are willing to share with third parties such as online service providers,

² OSTP RFI on Advancing Privacy Enhancing Technologies, “Background”, June 9, 2022. <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>

³ See, e.g., Office of the Privacy Commissioner of Canada, “Privacy Enhancing Technologies -- A Review of Tools and Techniques,” November 2017, https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711/#fn9

ENISA, “PETs controls matrix - A systematic approach for assessing online and mobile privacy tools”, December 2016, <https://www.enisa.europa.eu/publications/pets-controls-matrix>

under what circumstances that information will be shared, and what the third parties can use that information for.”⁴ In the United States, the Promoting Digital Privacy Technologies Act, S.224, also takes a broad lens, defining PETs as “any software solution, technical processes, or other technological means of enhancing the privacy and confidentiality of an individual’s personal data in data or sets of data.”⁵

II. Specific Research Opportunities to Advance Privacy Enhancing Technologies

The App Association and the Innovators Network Foundation (INF) serve as principal resources in the privacy space for thought leadership, advocacy, and education for the global small business technology developer community.⁶ We regularly work to keep our members up to speed on the latest policy and legal developments and to translate those into practical and usable guidance to ease the burden of compliance.⁷ Furthermore, through our INF Privacy Fellowship, we support thought leadership that covers a wide range of privacy issues, including privacy enhancing technologies.⁸

We encourage OSTP to look to existing work from the privacy fellows and other leading academics on this topic as it conducts further research on PETs. For example, The Rise of Privacy Tech is an organization led by INF Privacy Fellow Lourdes Turrecha that serves as a leading conduit for startups in the privacy technology space to connect with funders, peers, and mentors in the industry and to catalyze privacy tech innovation. Recently, The Rise of Privacy Tech published its landscape analysis, “Defining the Privacy Tech Landscape,” which included a full cataloguing of the different technologies that encompass privacy tech, including PETs (noting that PETs are a subset under the

⁴ Office of the Privacy Commissioner of Canada, “Privacy Enhancing Technologies – A Review of Tools and Techniques”, November 2017, https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711/

⁵ U.S. House of Representatives, “Promoting Digital Privacy Technologies Act”, Sec. 2, February 4, 2021, <https://www.congress.gov/bill/117th-congress/senate-bill/224/text>.

⁶ ACT | The App Association, Innovators Network Foundation Announces Inaugural Privacy

Fellows (September 2019), available at: <https://actonline.org/2019/09/23/innovators-network-foundation-announces-inaugural-privacy-fellows/>; [Protecting America’s Consumers: Bipartisan Legislation to Strengthen Data Privacy and Security: Hearing before the Subcomm. on Consumer Protection and Commerce, 117 Cong. \(2022\) \(Statement of Graham Dufault, sr. dir. for public policy, ACT | The App Assoc.\), available at https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Testimony_Dufault_CPC_2022.06.14_0.pdf.](#)

⁷ See e.g., ACT | The App Association, General Data Protection Regulation Guide (May 2018), available at: https://actonline.org/wp-content/uploads/ACT_GDPR-Guide_interactive.pdf; What is the California Consumer Privacy Act (January 2020), available at: <https://actonline.org/wp-content/uploads/What-is-CCPA.pdf>.

⁸ ACT | The App Association, Innovators Network Foundation Announces Inaugural Privacy Fellows (September 2019), available at: <https://actonline.org/2019/09/23/innovators-network-foundation-announces-inaugural-privacy-fellows/>

larger privacy tech umbrella).⁹ Their investigation covered everything from key definitions and categorizing the different facets of the privacy tech stack, to a business analysis on the present and future of the privacy tech market.

Another privacy fellow, Dr. Lorrie Cranor, directs Carnegie Mellon's CyLab Security & Privacy Institute, which also conducts research at the cutting edge of certain PETs. For example, Dr. Cranor's scholarship on "privacy nutrition labels" has informed the rollout of similar labels on both of the major app platforms in recent months.¹⁰ Dr. Cranor's research team has also been at the forefront of developing internet of things security labels,¹¹ machine extractable opt-out choices,¹² and privacy enhancing plug-ins for app developers.¹³

III. Specific Sectors, Applications, or Types of Analysis That Would Particularly Benefit from the Adoption of PETs

App developers are already working to adopt and implement PETs in their products, services, and features in order to meet market demands. Here are a few examples of PETs that our members rely on every day:

- **On-device processing.** Apps utilize on-device processing for certain sensitive features to ensure that no external processing occurs and that the company cannot see or access the data. To share one key use case, our members currently use facial verification technologies embedded at the platform level, such as Apple's Face ID, to allow users to login to apps using a scan of their face from the camera app. An app developer can choose integrate Apple's Face ID as an option for users to select as one of the factors in a two-factor authentication scheme. For example, users often opt for two-factor authentication to improve device security in cases where an application stores sensitive personal information, such as bank account information. The mathematical representation of the individual's face (the gallery image) used to validate the comparison image

⁹ The Rise of Privacy Tech, "Defining the Privacy Tech Landscape, November 2021, <https://www.riseofprivacytech.com/wp-content/uploads/2021/11/TROPT-Defining-the-Privacy-Tech-Landscape-2021-v1.0-1.pdf>

¹⁰ Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "nutrition label" for privacy. In Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09. ACM Press. <https://doi.org/10.1145/1572532.1572538>

¹¹ P. Emami-Naeini, J. Dheenadhayalan, Y. Agarwal and L. F. Cranor, "Which Privacy and Security Attributes Most Impact Consumers' Risk Perception and Willingness to Purchase IoT Devices?," *2021 IEEE Symposium on Security and Privacy (SP)*, 2021, pp. 519-536, doi: 10.1109/SP40001.2021.00112.

¹² Kumar et al., "Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text". In WWW '20: The 2020 Web Conference, April 20–24, 2020, Taipei. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/1122445.1122456/>.

¹³ C Tianshi Li, Yuvraj Agarwal, and Jason I. Hong, Coconut: An IDE Plugin for Developing Privacy-Friendly Apps, Proc. ACM Interact Mob, Wearable Ubiquitous Technol, 2, 4, Article 178, December 2018 <https://doi.org/10.1145/3287056>

is stored within Apple's Secure Enclave on the device and is not available to the developer, Apple, or any other third party.¹⁴

- Encryption. The App Association supports fully leveraging technical measures including end-to-end encryption to protect data broadly, enabling key segments of the economy to function—from banking to national security to healthcare—by safeguarding access to, and the integrity, of data from unwanted interlopers. Encryption's role should not be understated – without encryption, entire economies and industries are put at a significantly heightened risk of their data being compromised. The importance of encryption to the app economy has only heightened during the COVID-19 pandemic and the increasing desire to perform traditionally offline functions in the digital space due to social distancing mandates. That's why we've been strong supporters of the National Institute of Standards and Technology's (NIST) efforts to support the development of encryption technologies, as well as their leadership in advancing risk-based scaled approaches to cybersecurity management in the NIST Cybersecurity Framework (which includes an emphasis on encryption as a technical protection mechanism), while opposing legislation seeking to undermine end-to-end encryption, such as the Lawful Access to Encrypted Data Act or the EARN IT Act.
- App Tracking Transparency. Even as federal lawmakers debate legislation that would put new guardrails around data sharing practices in the digital economy, app developers comply with a growing number of platform-level restrictions on certain types of data sharing with third parties. For example, Apple's App Tracking Transparency (ATT) tool creates a simple solution to the opt-in/opt-out binary by presenting users with a just-in-time push notification asking if they want to permit apps to track them across third-party tracking that follows them outside of the app onto the open web or even other third-party apps. This type of engineering solution has so far evaded an easy resolution in the policy world but has markedly improved user privacy outcomes along the way.¹⁵ We have also raised the concern that antitrust measures that prohibit restrictions by platforms on access to personal data would likely prohibit ATT and undermine privacy by outlawing key PET developments by the market.¹⁶ The development of PETs is too important to fall victim to proposals that impose blanket prohibitions on

¹⁴ Apple, "About Face ID advanced technology", September 14, 2021, <https://support.apple.com/en-us/HT208108>

¹⁵ Estelle Laziuk, "iOS 14.5 Opt-in Rate - Daily Updates Since Launch", Flurry (May 25, 2021), available at <https://www.flurry.com/blog/ios-14-5-opt-in-rate-att-restricted-app-tracking-transparency-worldwide-us-daily-latest-update/>.

¹⁶ Graham Dufault, "Antitrust and Privacy Part 1: The Market for Privacy on Mobile Platforms," ACT | THE APP ASSOCIATION BLOG (Apr. 26, 2022), available at <https://actonline.org/2022/04/26/antitrust-and-privacy-part-1-the-market-for-privacy-on-mobile-platforms/>; Graham Dufault, "Privacy and Antitrust Part 2: What's Really At Stake if Congress Prohibits App Store Management?" ACT | THE APP ASSOCIATION BLOG (Apr. 27, 2022), available at <https://actonline.org/2022/04/27/privacy-and-antitrust-part-2-whats-really-at-stake-if-congress-prohibits-app-store-management/>.

access restrictions, especially at the platform level, and we hope that OSTP takes this consideration into account when evaluating antitrust measures.

- Privacy Labeling. Over the past few years, the app marketplace has seen the gradual introduction of the “privacy nutrition label” concept. The contemporary version of these labels (drawing from more than a decade of scholarship with researchers proposing similar concepts in various forms)¹⁷ aims to perform a very simple function: make app developers’ privacy practices more understandable to the average consumer. Initial research demonstrates that many app developers welcome privacy nutrition labels as a convenient, efficient, and user-friendly way for them to demonstrate their privacy practices and see it as a major improvement from the previous practice of directing users to lengthy privacy policies for similar information.¹⁸ Though we believe the app platforms could do a better job of assisting developers in the creation and maintenance of the label, we believe the concept will help to maintain trust in the app ecosystem in the long run.

IV. Specific Laws that Could be Used, Modified, or Introduced to Advance PETs

The App Association has long-supported the passage of comprehensive federal privacy legislation that sets a strong baseline of consumer protection and creates legal certainty for American businesses. The App Association believes such legislation will naturally serve as a vehicle for incenting PETs as regulated entities look to comply with new requirements and ultimately compete on pro-privacy business practices. Recently, a bipartisan and bicameral group of lawmakers introduced a bill, the American Data Privacy and Protection Act (H.R. 8152) that we believe is the strongest effort at federal privacy legislation in years. Last month, the Energy & Commerce Committee’s Consumer Protection & Commerce Subcommittee invited the App Association to testify on the bill, which was subsequently marked up and approved by the Subcommittee.¹⁹

While seemingly neutral on precise categories or applications of PETs, the legislation would incent their advancement in several ways. First, the bill states that service providers and third parties “have the same responsibilities and obligations as a covered

¹⁷ Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A “nutrition label” for privacy. In Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09. ACM Press. <https://doi.org/10.1145/1572532.1572538>

¹⁸ Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. 2022. Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels. In CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 24 pages. <https://doi.org/10.1145/3491102.3502012>

¹⁹ [Protecting America’s Consumers: Bipartisan Legislation to Strengthen Data Privacy and Security: Hearing before the Subcomm. on Consumer Protection and Commerce, 117 Cong. \(2022\) \(Statement of Graham Dufault, sr. dir. for public policy, ACT | The App Assoc.\), available at https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Testimony_Dufault_CPC_2022.06.14_0.pdf.](https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Testimony_Dufault_CPC_2022.06.14_0.pdf)

entity with respect to [covered] data under all provisions of this Act.”²⁰ One PET relevant to this requirement is the practice of data tagging, where a data controller tags a consumer’s personally identifiable information with that user’s data processing preferences as indicated in responses to the company’s privacy policy or via specific data access requests. The tagged data elements can then be passed along with the appropriate instructions to service providers or other third parties in the data processing chain. Second, the bill’s provision that covered entities “shall not collect, process, or transfer covered data unless the collection, processing, or transfer is limited to what is reasonably necessary and proportionate” will also incent the development of PETs that facilitate data minimization.²¹ Examples of relevant PETs include differential privacy, secure multi-party computation, zero-knowledge proofs, edge computing, and local or on-device processing. Finally, Section 103, “Privacy by Design,” while not requiring covered entities to utilize specific PETs, will incent them to invest in PETs in order to satisfy their broad responsibility to “mitigate privacy risks, including substantial privacy risks, related to the products and services of the covered entity or the service provider, including their design, development, and implementation.”²² Moreover, under this section, the Federal Trade Commission would gain the authority to further opine on “what constitutes reasonable policies, practices, and procedures,” which could include guidance on the use of specific PETs.

V. **Conclusion**

The App Association appreciates OSTP’s consideration of the above views. We urge OSTP to contact the undersigned with any questions or ways that we can assist moving forward.

Sincerely,

Brian Scarpelli
Senior Global Policy Counsel

Matthew Schwartz
Public Policy Associate

²⁰ American Data Privacy and Protection Act (H.R. 8152), Sec. 302, <https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/BILLS-117hr8152ih.pdf>

²¹ Ibid. Sec. 101.

²² Ibid. Sec. 103 (a)(3).

ACT | The App Association
1401 K St NW (Ste 501)
Washington, DC 20005

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

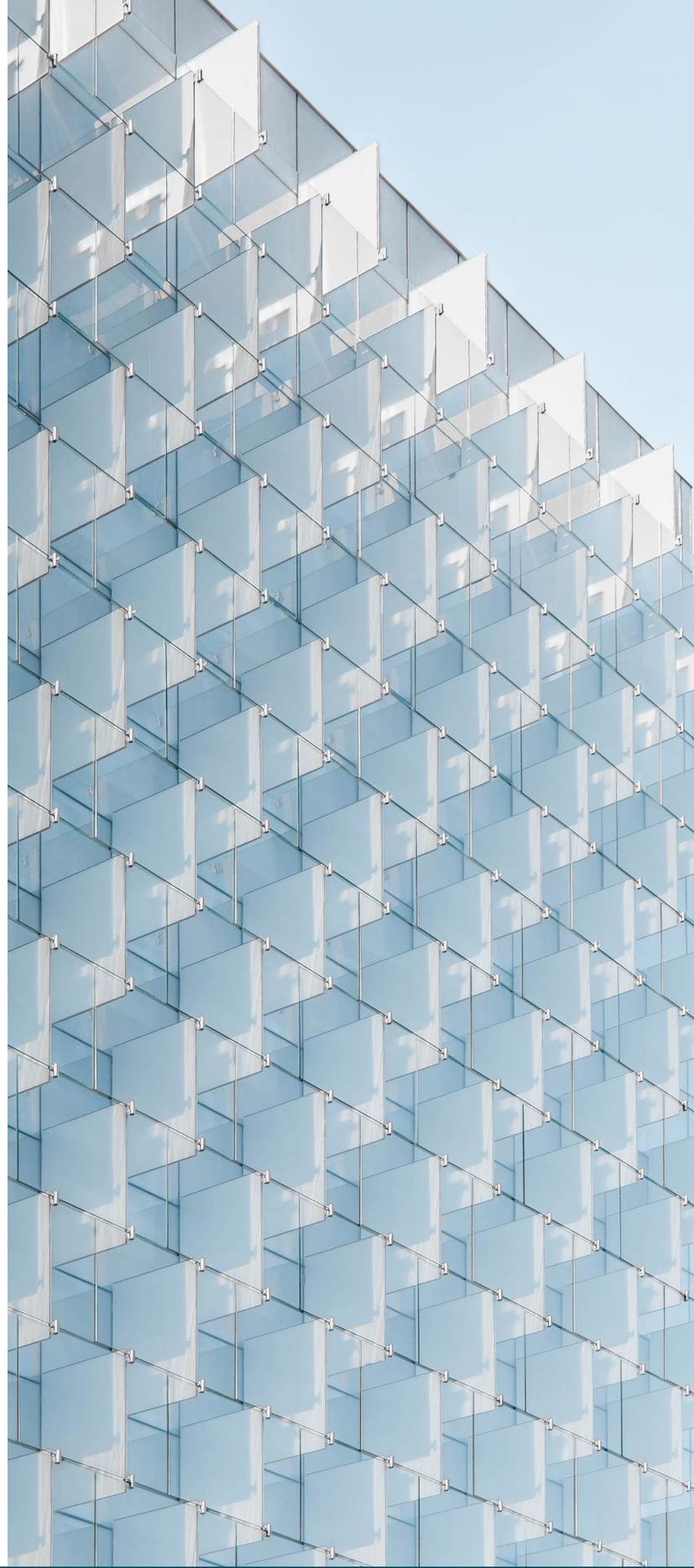
Agita Labs

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



AGITA
L A B S

REQUEST FOR INFORMATION RESPONSE



JULY 8, 2022

DRIVING DISCOVERY | PROTECTING PRIVACY



Table of Contents

- 01** COVER LETTER
- 02** EXECUTIVE SUMMARY
- 03** INFORMATION REQUESTED
- 08** SEQUESTERED ENCRYPTION:
UNDER THE HOOD



Cover Page



To whom this may concern:

At Agita Labs, we do not try to stop hacking.. Instead, our technology protects data *despite ongoing hacking*. This change in mindset is spearheaded by the work of Dr. Todd Austin and Dr. Valeria Bertacco, two computer science professors from the University of Michigan. In 2017, Austin and Bertacco built Morpheus, a secure computing platform that was designed to protect data from hacking. DARPA, the research arm of the US Department of Defense, put this system into a commercial red-teaming effort for three months, during which 500+ cybersecurity researchers were unable to successfully attack the system even a single time.

At the heart of Agita Labs' solution is a novel patented technology called *sequestered encryption* (or SE, for short). Sequestered encryption builds a hardware-based cryptographic wall between all software and sensitive data. As SE-protected software runs, all sensitive data is encrypted all the time, even when it is processed and analyzed. In fact, no one, not even programmers or superuser IT staff, can see data protected by sequestered encryption except the data owner who originally encrypted the data.

Agita Labs is excited to be considered for the development of the national strategy on privacy-preserving data sharing and analytics to better benefit individuals and the society as a whole. We know that secure computation holds great promise for advancing data breach protection, privacy enforcement, and zero-trust data sharing.

The opportunity to solve problems and create change around privacy for a range of institutions is the most attractive aspect of what we do. In this response, the objectives are to demonstrate what our technology can do, some barriers we have had with adoption when commercializing the technology, as well as some barriers around the technology itself.

Best regards,
Todd Austin, CEO
Valeria Bertacco, Chief Scientist
Sara McLean, Head of Business Development
Agita Labs, Inc., Ann Arbor, MI



Executive Summary

Secure computation holds great promise for advancing data breach protection, privacy enforcement, and zero-trust data sharing. This promise lies in the cryptographic-strength defenses that secure computation provides for data confidentiality - defenses that persist even when the system's software has been penetrated by attackers. These powerful defenses have emerged in several secure computation approaches: homomorphic encryption (HE) and multi-party computation (MPC). In addition to stopping data breaches, these novel secure computation technologies also enable new forms of zero-trust data sharing, making it possible to share sensitive data with untrusted parties while retaining control over who can view, process and learn from their data.

Existing secure computation technologies are not without their drawbacks, however. Today's secure computation frameworks possess many barriers to adoption: limited security defenses, performance overheads, and programming challenges. These constitute significant barriers to their adoption, limiting the organizations that will deploy these technologies and the applications for which these technologies can be employed. Additionally, commercial challenges exist in deploying secure computation technologies, including challenges surrounding consumer education, technology validation, market identification, standardization, and funding.

Agita Labs' has developed a novel form of hardware-based secure computation, called sequestered encryption (SE). Our SE technology addresses many of the technical concerns that have plagued other secure computation platforms. The capabilities of SE outmatch existing secure computation frameworks, providing confidentiality and integrity checking capabilities, plus safe data releases. The performance of SE-based secure computation is orders of magnitude faster than existing competing secure computation technologies, and with additional optimization and integration into the CPU processor, these overheads could be further reduced. Finally, the programmability of SE is much more approachable to developers, requiring only a few hours of training, rather than the highly idiosyncratic approaches embodied in competing approaches.

In this RFI response, we advocate for the use of sequestered encryption as an emerging technology that can address many of the technical barriers currently hampering the adoption of secure computation defenses. Moreover, we articulate with several recommendations for the US government to aid in overcoming both the technical and commercial barriers to the wide-spread adoption of secure computation technologies. These recommendations include support for additional funding for secure computation research and commercialization, exploration of a broader palette of secure computation technologies, and prioritizing government deployment of secure computation in its own sensitive IT operations.



The Promise of Secure Computation

Why Today's Data Security Is Not Very Secure

Data breaches are a regular--expensive and embarrassing--part of the current tech industry. IBM estimates the average data breach cost at more than \$4M, not including damage to public reputation and client relations. The root of this problem lies in popular approaches to data protection. Virtually all systems (even the most advanced systems that utilize trusted execution environments like Intel's advanced Software Guard Extension) protect data with software. As a result, to protect data we must prevent software hacking--a task that remains an eternally unsolved challenge in the security community.

Software hacking occurs when clever attackers identify vulnerabilities (typically in the form of bugs) in software that allow sinister exploits to step around the security defenses deployed on systems. Nearly all security breaches in the news are the result of software hacking, so stopping these exploits is a prime mission of the security industry. A closely related, but equally important, form of hacking is side channels. Side channels occur in the observable characteristics (e.g., run time) of software and hardware, revealing sensitive data being processed inside an application. While side-channel attacks are mostly academic today, there have been a few that have risen to the critical concerns, e.g., Kocher's attack that exposed RSA keys remotely, and most advanced data security technologies will work to address software hacking and side-channel attacks. In summary, if you cannot stop software hacking and side channels, you cannot protect data from attackers.

Why Not Put a Stop to Hacking?

A vast majority of the security industry is focused on stopping and detecting software hacking. To stop software hacking, security professionals attempt to identify and fix all the "bugs" in a program. Bugs are programmer errors, which often do not break an application, but attackers find clever ways to exploit these bugs. When bugs are found, they need to be fixed quickly before attackers can exploit them. This approach to security is often called "Patch and Pray" because finding *all* bugs is essentially impossible. If you ask a programmer if there exists any software without bugs, they will most likely say "No." Software is simply too complex, too rapidly evolving, and too intractable to lend itself to any form of high precision "bug hunting." Clearly, the odds of a software hack favor the attacker.

An important and powerful tool for bug hunting for formal verification. With formal verification, mathematical methods are used to prove that a program doesn't contain bugs that attackers could exploit. Formal verification is incredibly powerful because, when it works, it can find bugs before they can be exploited by attackers. The challenge with formal verification is two-fold; first, real programs are usually too complicated to fully analyze with formal verification. Formal security verification is a process of proving that something cannot happen, which is one of the most difficult computational challenges in computer science. Second, formal verification cannot identify the exploits that attackers will invent in the future. These exploits, which are often called zero-day exploits, are a primary concern of the security community, because these bugs represent those which the attacker community is aware of and the security defense community is not!

Further, even if organizations could find and fix all software bugs, they would still be susceptible to side channel attacks, which simply observe the operation of software and hardware to infer their secrets. Sophisticated and well-meaning developers can easily write completely bug-free code that is riddled with side channels, allowing any listening attacker to quickly understand the secrets held within the software. All considered, it is important to recognize that mainstream security defenses today cannot durably stop data breached. Vigilant and dedicated security teams can make breaches much harder, but these organizations are still at risk of getting breached.



Secure Computation Address Key Technical Challenges

Secure computation is a new form of computer computation that works directly on encrypted data without software needing to hold a data access key. Examples of systems that support this capability are those based on homomorphic encryption (HE) and multi-party computation (MPC). Systems that support secure computation build a cryptographic wall between all software and data, which immunizes these systems against disclosures of sensitive data via software hacking. While software hacking is effective on systems that perform computation in the clear, it has little agency over secure computation because once the hacker penetrates the system, all there is to steal is encrypted data. As illustrated in Figure 1, this desirable property of secure computation is referred to as *zero software trust*, because secure computation can maintain data security even if all the software in the system is hacked.

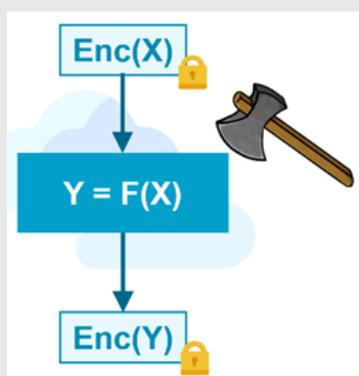


Figure 1: Secure computation runs directly on encrypted data.

Secure computation can also be a durable defense against side-channel attacks, if the underlying implementation of the secure computation capabilities doesn't leak information through observable channels, e.g., memory or operational timing. Most secure computation frameworks strive to eliminate vulnerabilities due to software hacking and side-channel attacks.

Secure Computation Addresses Privacy Challenges

Secure computation has significant potential future value to computing hardware that incorporates these capabilities, due to its ability to durably stop data breaches and create new forms of safe data sharing. The approach doesn't stop software hacking (which is a much more difficult and perhaps impossible task), but rather it renders the data breach inert since attackers can only steal encrypted sensitive data.

Zero software trust is more valuable than simply being a powerful security measure, it also creates many high-value opportunities for zero-trust data sharing. With zero-trust data sharing, it is possible to process sensitive encrypted third-party data without risk of attackers or programmers seeing, abusing, or stealing the data. For example, with secure computation, one could send their encrypted genome to an online (and potentially untrusted) server, which would then run a proprietary disease profiling algorithm on it using secure computation. The results of this computation would not be visible to anyone except the original data owner that encrypted the data, thus guaranteeing the privacy of the genome data. The prospect of zero-trust data sharing creates many business opportunities for secure computation in medical data sharing, fintech data sharing, genomics research, advertising analytics, private smart blockchains, secure health monitoring, secure avionics, privatized surveillance, etc.



Technical Barriers to Adoption of Secure Computation

While the promise of secure computation is great, the technical challenges in deploying it are also equally great. These challenges are both commercial and technical in nature, and in this section, we focus on the technical challenges, after which we illuminate the commercial challenges that we have experienced in our quest to deploy secure computation in customer operations. In this section, we limit our discussion to the technical barriers associated with homomorphic encryption (HE) and multiparty computation (MPC), the two dominant secure computation technologies in the marketplace today.

Performance Barriers: HE and MPC technologies are notoriously slow, with HE suffering from significant computation overheads and MPC suffering from heavy communication overheads. Compared to unprotected algorithms, these technologies can slow program performance by 100x - 1,000,000x slower. While there are certainly applications that can tolerate these slowdowns, there are many other important application areas that cannot, such as machine learning training, computer vision applications, and real-time applications.

Limited Defense Capabilities: HE and MPC technologies also suffer from limited security capabilities. Today, these technologies focus on the confidentiality of data and lack any support for ensuring the integrity or availability of computation resources, which are two other critical security concerns. For example, if a voting machine were built with HE technology, it could certainly keep votes secure from disclosure, but today's HE frameworks couldn't tell if an attacker manipulated the HE secure computation to add all the votes to their own preferred candidate. These types of attacks are called integrity attacks, because the attack compromises the integrity of the secure computation, making it possible for the attacker to create a new algorithm that likely benefits the attacker. The MPC community has begun to address this concern with zero-knowledge (ZK) proofs, but these proofs are *i)* specific to an algorithm and often require PhD-level expertise to craft, and *ii)* require significant additional computation resources to verify computation integrity. Finally, without strong integrity checking mechanisms existing secure computation frameworks cannot perform safe disclosures of encrypted data. A safe disclosure is a capability that allows the decryption of a specific program value that the data owners agree to release. Without strong integrity checking, attackers could abuse a decryption capability, and thus, today's secure computation frameworks require that the original data owner decrypt any value to be released. While a safe approach, this creates many application-specific deployment challenges. For example, regulatory access to secure computation isn't possible without the original data-owner performing the decryption. This would be a problematic situation especially if the data owner was committing infractions that the regulatory agency was trying to detect.

Significant Programming Challenges: The nature of HE and MPC defenses is that they don't protect software directly, instead, they protect linear integer arithmetic. As such, programmers must express their algorithms in the form of a linear integer mathematical expression before it can be protected by these frameworks. This requirement creates significant programming challenges that make it difficult or impossible to express certain algorithms. For example, the "less than" inequality test poses a significant challenge to HE, due to its non-linear nature. Thus, one doesn't see applications that require sorting being protected with HE defenses, since sorting algorithms rely heavily on relational tests. As such, mission critical applications such as databases cannot be readily protected with HE technology. In addition, the countless applications that rely on strings and floating-point numerics have significant challenges with HE adoption, since these fundamental data types are not well supported in today's secure computation frameworks.



Commercial Challenges to the Adoption of Secure Computation

With any startup comes multiple challenges in regards to adoption and commercialization. With a deep tech startup such as Agita Labs or other privacy enhancing technology companies, the challenges are often more complex. Being able to solve a problem for any opportunities that a privacy enhancing technology can present from a practical as well as economic standpoint can be the key challenge.

Education: Educating the market on how to use a PET and the benefits from it has been the single largest challenge in terms of adoption. Providing more training and development such as workshops, seminars, and education around what a privacy enhancing technology provides is needed for a successful commercialization. The need for more education around privacy enhancing technology is the largest challenge the PET community faces in terms of adoption.

Validation: Small or large enterprise businesses want to make sure that anyone they are working with has accreditation and third party validations or frameworks in place. Companies want to make sure when they are partnering with a new vendor that certain boxes are checked before moving forward. This holds true even more when you are working with data and protecting privacy. Privacy-enhancing technology is still new, therefore causing some hesitation on privacy and security conveners for their own organization. Being a small startup business and not being able to have specific proof points or third party validation can ultimately cause an enterprise business to not move forward.

Marketing Identification: Despite there being many challenges to overcome with new and innovative technologies such as a PET, narrowing down the correct market to target who understands the need for this technology has been another challenge. Once the defined markets have a better education and understanding on how a PET can be utilized in their environment, standardization will come soon after.

Standardization: Enterprise businesses do not always grasp the need for adopting a new technology within their organization. From the research and multiple conversions that have taken place, companies know very little around PETs. New technology is always a disrupter in the short run. The enterprise businesses have to upgrade their existing procedures and systems which can cause a major disruption. Large enterprise businesses have a lack of knowledge on how much time/resources a PET will need with their internal teams and until using a PET becomes a more standard practice, adoption to commercialize a PET will be at a slower pace.

Funding: Investors are always extremely excited and eager to hear about what Agita Labs and other privacy enhancing technology companies (and researchers) are working on, but shy away because they themselves are unclear as to the growth capacity of a PET, and what specific need or problem will be solved. Because a PET is still relatively new to the market, finding partners and channels to adopt the technology has been difficult.



Ways the Government Could Lessen Commercial Barriers

We see several opportunities where government involvement and resources could speed the adoption of secure computation, by lessening technical and commercial barriers.

Secure Computation Research Funding: Programs such as DARPA's SSITH and DPRIVE programs have been vital in advancing secure computation capabilities. DARPA SSITH's program led to the sequestered encryption technology detailed in this RFI. DARPA's DPRIVE program is developing significant improvements in the performance of homomorphic encryption, through the deployment of specialized hardware accelerators. Additional funding directed specifically toward secure computation (or privacy-enhanced technology) research would be a welcome opportunity, for both secure computation researchers and startups.

Broadening the Pallet of Secure Computation Capabilities: Today, much of the focus in the secure computation world is on homomorphic encryption. This is certainly warranted, since homomorphic encryption stands alone in its ability to minimize trust in the system software and hardware. But as the need to overcome the technical barriers to HE adoption persist, we feel that the government has become aware that additional more capable secure computation solutions do exist, such as Agita Labs' sequestered encryption.

Prioritize Government Deployment of Secure Computation Capabilities: Government entities see the value and problems that a privacy enhancing technology can provide. With that, if government entities start utilizing a PET this will help with adoption on the commercial side as well. Researchers and startups in the secure computation arena would welcome the US government in taking a leadership position in adopting secure computation technologies in its own IT operations, to advance its ability to stop data breaches, promote zero-trust data sharing capabilities, and provide additional commercialization opportunities for emerging secure computation products. On top of this, requiring new specific requirements or policies on data and privacy would help generate growth for privacy enhancing companies to work together to solve a problem that hasn't fully been solved as of yet.

Commercialization Grants: Government grants for small start ups would be extremely helpful for multiple reasons. A grant can help with the business expansion to develop new marketing strategies, improve business production and assist with overall business growth. With the help of grants from the government in terms of a privacy enhancing technology this can also help with technology adoption as well as training and development. Along with the obvious benefits a grant can provide when trying to commercialize privacy enhancing technologies, grants can improve project outcomes and expansion. These grants can accelerate a small start-ups business timeline which can lead to improved outcomes for the business.



Sequestered Encryption: Under the Hood

Sequestered encryption is enforced using a small hardware component called the SE enclave. The SE enclave is a trusted hardware processing element that can expose a secret key from the data owner (using industry-standard public-key cryptography), and then operate directly on encrypted program data using simple processing commands, such as ADD, MULTIPLY, and XOR. Since the results of any computation are encrypted, the inputs and outputs of any computation remain secret. The SE enclave is deployed into the cloud using existing FPGA-class nodes in Azure and AWS or on-premises with Intel-Altera CPUs, making this powerful hardware security technology ready to use in today's security and privacy-sensitive applications.

Sequestered encryption implements *encrypted computation*, which provides a highly durable defense against the disclosure of sensitive data because it sequesters (hides) all decrypted sensitive data and the data owner's keys used to access that data inside the hardware SE enclave. No software, including software from Agita Labs or the operating system, can access information inside the software-free SE enclave. Thus, even if the system's software gets hacked, not even the hacked software can get access to sensitive decrypted data or keys. In addition, SE enclave computation is free of any control, memory and timing side channels, and thus, you can rest assured that if attackers are observing the computation, there is nothing to be learned about your sensitive data.

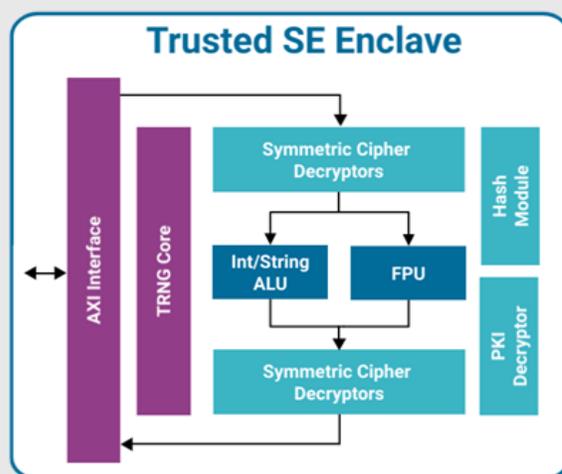


Figure 2: Agita Labs' sequestered encryption enclave.

The SE enclave enforces the integrity of any computation using a powerful patented *computational fingerprinting* technology. Computational fingerprints allow users to verify that *i)* the intended inputs to the computation were used in creating a result, and *ii)* the full unperturbed computation was run on the expected inputs. Any form of hacking to rearrange, replay, or otherwise manipulate SE computation is readily detected. This facility is invaluable in its ability to detect if the system software is being actively hacked or modified with malicious intent (e.g., supply chain attacks).

Added to this, our novel patented *safe datagrants* technology forms the basis for analysis of encrypted data, monetization of 3rd-party encrypted data, selective release of computed results, and 3rd-party auditing of encrypted data storage and computation. Safe datagrants allow a specific computation to decrypt a computed value (or transfer it to another security domain), as long as the computation that produced the value was not perturbed in any way.



Programming Interfaces

In addition to on-premises installation, Agita Labs has worked with Microsoft, Amazon, and Intel to deploy sequestered encryption into the Azure and AWS clouds. Deploying SE technology ensures that the computation is cryptographically secured to be confidential, undisturbed, and able to safely disclose data. Programmers utilize these capabilities in their business operations to:

- Protect sensitive data from data breaches
- Implement zero-trust data sharing of sensitive data with potentially untrustworthy parties
- Process and analyze sensitive 3rd-party data that is always encrypted and confidential

```
// calculate function F/DF value using Newton-Raphson method
enc_double rn_solver(enc_bool& converged, double maxerr,
                    unsigned maxiter, fn_type f, fn_type df)
{
    unsigned iter;
    enc_double val = 1.0;

    converged = false;
    for (iter = 0; iter < maxiter; iter++)
    {
        converged = enc_fabs(f(val)) <= maxerr;
        val = enc_cmov(converged, val, val - f(val)/df(val));
    }
    return val;
}
```

Figure 3: An example code protected with sequestered encryption, Newton-Raphson's Algorithm

Programmers access TrustForge capabilities using simple extensions to the C++, Python, or JavaScript programming languages. Programmers simply declare protected data types in their software (including integers, floating point, Booleans, and strings) and rebuild their programs to utilize SE computation; then computation on encrypted data is directed to the SE enclave where it is protected by cryptographic-strength defenses. Moreover, the SE programming interfaces enforce that programmers do not introduce vulnerabilities into their code, making SE-based programs *secure-by-construction*: if a computation compiles and runs with SE defenses, it is safe from all known forms of software hacking, data disclosure, and integrity attacks.



Competitive Analysis

The hardware integration and acceleration of the SE enclave give a significant advantage over HE and MPC frameworks, which require notably more computation and network communication. Experiments with the open-source VIP-Bench privacy benchmarks (co-developed with University of Michigan, NYU, and Addis Ababa Institute of Technology) have demonstrated that sequestered encryption technology can be many orders of magnitude faster than competing secure computation technologies. This advantage means that sequestered encryption can tackle performance-sensitive applications beyond the ability of other secure computation technologies (e.g., machine learning, recommendation, and computer vision applications).

Compared to other privacy-enhanced computation technology, Agita Labs' sequestered encryption delivers the most capable, performant, programmable, and secure computation on the market. Alternative secure computation technologies include homomorphic encryption (HE), or multi-party computation (MPC).

Significantly Better Performance: The hardware integration and acceleration of the SE enclave give a significant advantage over HE and MPC frameworks, which require notably more computation and network communication. Experiments with the open-source VIP-Bench privacy benchmarks (co-developed with University of Michigan, NYU, and Addis Ababa Institute of Technology) have demonstrated that sequestered encryption technology can be many orders of magnitude faster than competing secure computation technologies. This advantage means that sequestered encryption can tackle performance-sensitive applications beyond the ability of other secure computation technologies (e.g., machine learning, recommendation, and computer vision applications).

Enhanced Capabilities: Unlike mainstream security defenses, sequestered encryption defenses are not vulnerable to any known form of software hacking or side-channel attacks. In addition to containing software, virtually all TEEs today share microarchitectural resources with untrusted software making them susceptible to control, memory, and timing side channels. Unlike HE and MPC, the SE enclave can both protect the integrity of secure computation and safely release privacy-preserving information, if allowed by the data owner. In contrast, HE and MPC frameworks typically only release information by requesting the original data owner to decrypt those results. This approach becomes problematic in regulatory auditing applications, where the data owner may choose not to share sensitive data with a regulatory agency.

Straightforward to Program: While other secure computation frameworks protect mathematical expressions, sequestered encryption protects CPU processing commands. To protect computation with HE and MPC requires programmers to express their entire application as a mathematical expression. Moreover, HE often limits the depth of computation that is possible. These requirements negatively impede program development, requiring developers to perform potentially major surgery on an application (and brush up on their Taylor Series expansions). Sequestered encryption, in contrast, protects software directly, allowing programmers to readily port their existing unprotected applications to use SE defenses. Finally, SE-protected programs can utilize encrypted floating-point and string values, which typically pose a significant challenge to programmers in other secure computation solutions.

More Mature and Flexible Cryptography: SE technology is built on time-tested standard asymmetric and symmetric key ciphers. This ensures that the cryptography used in the system is mature and has received significant analysis by the cryptography community. In contrast, other secure computation frameworks, such as HE and MPC, often rely on application-specific cryptographic defenses that are less mature and have not yet received significant attention by the cryptography community. Additionally, as quantum computing inches closer to the mainstream, there is growing concern about the ciphers that will be appropriate to a post-quantum world. SE technology can readily incorporate post-quantum ciphers into its defenses once the cryptography community reaches consensus on what these ciphers should be.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

AHIP

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



601 Pennsylvania Avenue, NW T 202.778.3200
South Building, Suite 500 F 202.331.7487
Washington, D.C. 20004 ahip.org

July 8, 2022

The White House
Office of Science and Technology Policy (OSTP)
1650 Pennsylvania Avenue, NW
Washington, D.C. 20504

Sent Via Electronic Mail to: PETS-RFI@nitrd.gov

Re: Request for Information (RFI) Response: Privacy-Enhancing Technologies (PETs)

Dear White House Representative:

Everyone deserves the peace of mind of knowing that their personal health information is private and protected. With our long-standing commitment to protecting the health information of patients and consumers, AHIP appreciates this opportunity to submit comments on the Notice of Request for Information (RFI) on Advancing PETs.¹ **We support the development of a national strategy on privacy-preserving data sharing and analytics, along with associated policy initiatives.** Our comments and insights are based on health insurance providers' extensive experience in evaluating new technologies and best practices for consumer data privacy and security.

AHIP² members are accustomed to protecting the privacy and security of individuals' health information. While complying with the Health Insurance Portability and Accountability Act (HIPAA), the Health Information Technology for Economic and Clinical Health (HITECH) Act, State laws, and corresponding regulations, our members continually strive to institute best practices, add emerging technologies that meet or exceed the current legal and compliance expectations, and stay on the forefront of new developments and solutions to better protect their customers.

OSTP's evaluation can be important for privacy, security, and cybersecurity policies that serve as a significant "next step" in health care transformation by empowering data with technological protections that are based on our national values, individual rights, and the ongoing need to stay competitive with and ahead of advancements currently occurring in other countries and jurisdictions for economic and national security.

¹ 87 Fed. Reg. 35250.

² AHIP is the national association whose members provide health care coverage, services, and solutions to hundreds of millions of Americans every day. We are committed to market-based solutions and public-private partnerships that make health care better and coverage more affordable and accessible for everyone.

Given the importance of this initiative, as well as the breadth and scope of this work, we recommend the issuance of a proposed national strategy in the *Federal Register*, along with a focused outline of the components that are expected to be involved in the deployment and implementation. We believe that additional information from OSTP would allow for a more fulsome and substantive response addressing PETs based on OSTP's vision and plans.

A National Strategy for PETs

AHIP and our members are committed to supporting and advancing PETs. Yet, many stakeholders across the health care spectrum are in different places when it comes to their sophistication with these solutions and techniques. **The educating of stakeholders about a baseline set of terms and working definitions by OSTP would help further the discussions and focus on the issues to be encompassed in the national strategy.** We believe a subsequent opportunity to offer additional comments on these terms and working definitions would better inform the work and developments in this area for the national strategy.

There is also uncertainty about whether the RFI is intended to cover: (1) the ability of consumers to understand and manage their own choices based on privacy and control of individually identifiable and/or health data, and/or (2) the technological solutions and supporting policies that can foster anonymized data sharing, research, interoperability, and other needs utilizing enhanced privacy tools that inform data sets without compromising privacy and security on an individual basis. Clarifying the scope of the RFI in this regard can better inform future responses and allow for more relevant information to be offered.

Coordination of Efforts at the Federal Level

The Notice explains that the OSTP has been collaborating with several existing efforts, including the National Science and Technology Council's Fast Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics, Subcommittee on Networking and Information Technology Research and Development (NITRD), the National Artificial Intelligence (AI) Initiative Office, and the NITRD National Coordination Office. We appreciate the coordination to minimize duplicative or misaligned policies and believe the public would benefit from more information about these efforts.

For example, the National Institute of Standards and Technology (NIST) has been working on several projects, including the development of a Privacy Framework, a Cybersecurity Framework, a Risk Management Framework for AI, and other test projects and pilot programs to further evaluate the use of technology solutions and the potential risks and benefits of advancing such solutions in the health care and other sectors. In public forums, NIST has described working

with the OSTP, although it is currently unclear the extent to which NIST may be involved with PETs in this context.

In addition to NIST, other federal agencies are either charged with or are developing regulations and guidance to promote privacy, security, and cybersecurity. For example, the U.S. Department of Health and Human Services (HHS) Office for Civil Rights, the HHS Office of the Coordinator for Health Information Technology (ONC), the Centers for Medicare & Medicaid Services (CMS), the Federal Trade Commission (FTC), and many other federal agencies have to date issued some form of privacy, security and/or cybersecurity guidance. **We support a comprehensive and collaborative federal approach that will help consumers and other affected stakeholders better understand which agencies are engaged in this work, as well as the direction and focus of OSTP in terms of collaboration and priorities by industry sector (i.e., healthcare, retail, financial, etc.). This information can help commenters understand how the national strategy and related PET policies may align with, duplicate, or conflict with existing federal laws, regulations, and guidance.**

For the health care sector specifically, short-term as well as long-term goals should be developed. For other sectors, priority should be given to the national critical infrastructure sectors to help stage the development of PETs and application in these sectors.

Discussion of Emerging Technologies

Many recent technologies can advance business processes and data sharing techniques while also improving how we protect individually identifiable data. In healthcare, examples of these technologies include, but are not limited to:

- distributed ledger and blockchain-based technologies.³
- confidential computing examples such as the “Zero-trust Frameworks” for security.
- federated learning.
- “web3” which may further enhance consumers’ ability to control and restrict how their online data is used.
- “data sandboxes” which can offer a secure and governed environment in which certified de-identified health data is provisioned to vetted users.⁴

³ Blockchain use in healthcare can address needs for improved data movement, such as payer-to-payer exchange, which requires secure communication and the ability to share data between different entities. See also an article which illustrates one member’s efforts to test blockchain to improve efficiencies for Coordination of Benefits (COB) functions: <https://www.forbes.com/sites/michaeldelcastillo/2022/02/08/forbes-blockchain-50-2022/?sh=14c9409f31c6>.

⁴ One example of a “data sandbox” can be found at: <https://massdigitalhealth.org/mass-digital-health-programs/digital-health-sandbox-program/digital-health-sandbox-network/anthem>.

- synthetic data.⁵

These tools – when applied appropriately – can reduce privacy and security issues surrounding the release personal and sensitive information while advancing opportunities to use data in a multitude of contexts, such as validating and training AI algorithms on substantial amounts of data. In addition, PETs will enable advancements across research, grant, and start-up programs for companies as well as other specialized projects. **OSTP could encourage advancement of PETs across these programs by exploring federal support of safe harbors, “data sandboxes,” and demonstration projects.**

The health care industry can specifically benefit from understanding the goals and risks of PETs. Currently, much work is taking place to promote interoperability of health data on a national scale. This work is being conducted under the auspices of the HHS ONC, via a contract with The Sequoia Project, the Recognized Coordinating Entity for the Trusted Exchange Framework and Common Agreement (TEFCA). To the extent that PETs can further refine and protect data, this national initiative stands to benefit from such transformative work.

One result from fostering the greater use of PETs would be to allow healthcare consumers, providers and other entities to execute certain choices in an electronic environment based on consumer preferences and the ability of the provider or entity to share data. Preferences would need to be documented and executed, where feasible, but consideration should be given to those entities or individuals who are unable to execute such preferences perhaps because of patient safety or other concerns.

Furthermore, we strongly support use of PETs to advance research, particularly research that can benefit health outcomes, mitigate disease, help with early identification and treatment, and inform overall health and well-being. PETs that can help inform research and treatment outcomes are a key interest for consumers and our members. **The OSTP could encourage advancements specific to healthcare by supporting public-private collaborations and the use of PETs.**

It is also important to note that privacy cannot be discussed in a viable way without also considering security (as well as cybersecurity), as privacy and security go hand-in-hand to protect data and confidential information. Many PETs are built on the security infrastructure to protect data. **The focus of the RFI is currently on privacy but could be expanded to include security since PETs need the “what” of privacy controls and the “how” of security controls to be effective and work as intended.**

⁵ A recent article highlights the opportunities of synthetic data in healthcare: <https://www.wsj.com/articles/anthem-looks-to-fuel-ai-efforts-with-petabytes-of-synthetic-data-11652781602>.

Building public trust and acceptance centered around privacy, security, and cybersecurity will be essential components for moving forward in diverse settings and applications as PETs are utilized. Such confidence can help promote national acceptance and adoption.

Legal Considerations

In terms of laws, regulations, and guidance, a comprehensive and diverse series of general resources are available, but many do not specifically address PETs. At this stage, we caution against prescriptive regulations focused on PETs. We believe it would be prudent to allow PETs to be utilized and understood before constructing regulatory requirements. Creating regulatory requirements too early may hamper innovations in this area.

As public and private entities are learning more about PETs and deploying them in health and other sectors, flexibility and innovation will be key to learning what works and what may need improvements. Laws, regulations, and guidance should be developed at a future point when appropriate.

International Collaboration and Standards

Privacy considerations cannot be discussed in the United States without prioritizing the rights that are afforded to all Americans. The work that is taking place in international venues may prove to have some societal benefits, but in the United States any benefits must be balanced against the risks to individual rights. It is expected that PETs will protect individually identifiable data. However, as we have learned from cybersecurity campaigns and data breaches, no electronic system can provide absolute assurances that the systems and the data will be immune from intrusion or compromise. More understanding of the risks involved in potential cyber campaigns and breaches and how such incidents could be handled in the international context would be needed before American consumers could be asked to trust such technologies and processes.

In addition, varying international efforts can create confusion and result in inconsistent schemes. Many countries and jurisdictions have priorities and values that differ from U.S. values and laws. Some of our members comply with U.S. laws and regulations, as well as the European Union's Privacy Framework, the General Data Protection Regulation. Compliance with varying regulatory schemes is complex and can be costly to implement. **Within the confines of remaining consistent with U.S. rights and values, we encourage the OSTP to seek out ways to promote efficiency and reduce costs to benefit the health care consumer when data and PETs are used in international collaborations.**

July 8, 2022
Page 6

We appreciate the opportunity to comment on this important topic. Please contact me at -
with any questions.

Sincerely,

Danielle A. Lloyd
SVP, Private Market Innovations and Quality Initiatives

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Altman, Micah; Cohen, Aloni; and Vadhan, Salil

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

July 28, 2022
Via electronic filing

Jeri Hessman
Technical Coordinator
Fast-Track Action Committee (FTAC) on Advancing Privacy-Preserving Data Sharing
and Analytics.
NITRD/NCO
Email: <PETS-RFI@nitrd.gov>

Re: *RFI Response: Privacy-Enhancing Technologies*

Dear Members of the Committee,

This comment is informed by research with collaborators through the *Privacy Tools Project* at Harvard University.¹ In this broad, multidisciplinary project, we are exploring the privacy issues that arise when collecting, analyzing, and disseminating research datasets containing personal information. Our efforts are focused on translating the theoretical promise of new privacy protection and data utility measures into practical tools and approaches. In particular, our work aims to help realize the tremendous potential from social science research data by making it easier for researchers to share their data using privacy-protective tools.

Academic research in theoretical computer science, statistics and information science has demonstrated many challenges related to managing information privacy in the modern world.

In previous scholarly publications, we have offered several recommendations that we believe would help enable the wider sharing of research data while providing privacy protection for individuals.²

¹ See Privacy Tools for Sharing Research Data, <http://privacytools.seas.harvard.edu>. Also see prior Privacy Tools Group responses to related policy RFI's here: <https://privacytools.seas.harvard.edu/policy-engagement>.

² See Altman M, Wood A, O'Brien D, Vadhan S, Gasser U. Towards a Modern Approach to Privacy-Aware Government Data Releases. *Berkeley Technology Law Journal* 2015; 30(3):1967-2072; Vayena E, Gasser U, Wood A, O'Brien D, Altman M. Elements of a New Ethical and Regulatory Framework for Big Data Research. *Washington and Lee Law Review*. 2016;72(3):420-442.

Altman, Micah, Alexandra B. Wood, David O'Brien, and Urs Gasser. "Practical approaches to big data privacy over time." *International Data Privacy Law* 8(1):29-51. (2018).

Fluitt, J. Aaron and Cohen, Aloni and Altman, Micah and Nissim, Kobbi and Viljoen, Salome and Wood, Alexandra, Data Protection's Composition Problem (September 9, 2019). *European Data Protection Law Review (EDPL)*, Vol. 5, Iss. 3 (2019)

Altman, Micah and Cohen, Aloni and Falzon, Francesca and Markatou, Evangelia Anna (Lilika) and Nissim, Kobbi and Reymond, Michel Jose and Saraogi, Sidhant and Wood, Alexandra, A Principled Approach to Defining Anonymization As Applied to EU Data Protection Law (May 9, 2022). Available at SSRN: <https://ssrn.com/abstract=4104748> or <http://dx.doi.org/10.2139/ssrn.4104748>

Although our previous writings do not comment directly on all of the privacy-enhancing technologies under consideration by the committee, the writers judge that the risks discussed in these works apply to protected health information and that the broad findings and recommendations are readily applicable here. For these reasons, we recommend that the committee read and incorporate these recommendations, which are summarized below.

As a general framework, we have recommended the development of rules and guidance based on the following principles of a modern approach to privacy:

- Calibrating privacy and security controls to the intended uses and privacy risks associated with the data;
- When conceptualizing informational risks, considering not just reidentification risks but also inference risks, or the potential for others to learn about individuals from the inclusion of their information in the data;
- Addressing informational risks using a combination of privacy and security controls rather than relying on a single control such as consent or deidentification;
- Anticipating, regulating, monitoring, and reviewing interactions with data across all stages of the lifecycle (including the post-access stages), as risks and methods will evolve over time; and
- In efforts to harmonize approaches across regulations and institutional policies, emphasizing the need to provide similar levels of protection to research activities that pose similar risks.
- Recognizing that a single set of privacy and security controls is frequently not appropriate for all intended uses of the information. And designing for access using a tiered model is one in which data are made available to different categories of data users through different mechanisms.
- Recognizing that privacy risks constitute a specific form of informational harm. Privacy is not sufficient, even in theory, to ensure that decisions or algorithms based on personal data will be secure, non-discriminatory, explainable, reasonable, or immune to dangerous misuse.
- Recognizing that as the volume and complexity of data uses and publications grow rapidly across a broad range of contexts, it has become impossible to monitor all past data releases and anticipate all future attacks. Instead, PETS must address and control the cumulative information risks to participants.
- Rejecting security by obscurity, and instead recognizing that security requires protections to be based on public algorithms and protocols built and vetted by the greater security community.
- Recognizing that the risk of harm from the use and disclosure of information is not limited to data represented in any specific format. To be effective, definitions of use and disclosure in new or proposed legislation should be generally applicable for any type of data release—whether in the form of microdata, a summary table, an information visualization, statistical model coefficients, a trained model output by a machine-learning algorithm, a textual summary, or any other form.

We further note the importance of using protections that provide formal privacy guarantees where feasible. Many data-sharing models are compatible with a formal privacy guarantee called differential privacy. Differential privacy is a strong, quantitative notion of privacy that addresses both known and unforeseeable attacks, and is provably resilient to a very large class of potential misuse. In recently published work, we provide guidance on designing systems that integrate differential privacy protections throughout the information lifecycle and apply it in conjunction with a range of other complementary informational controls.³

In addition, the research cited above finds that addressing privacy risks requires a sophisticated approach, and the privacy protections currently employed in government releases of data do not take into account recent advances in data privacy research. We note that there is a wide range of technical, procedural, legal, educational, and economic controls available for managing privacy risks. However, most government data releases rely almost exclusively on a narrow set of interventions, namely redaction of identifiers and binary access control. This focus on a small set of controls likely fails to address the nuances of data privacy and utility, as well as the differences between data releases, which vary widely in terms of the intended uses of the data and the privacy risks involved.

This research also notes, as paraphrased, that advances in science and technology enable the increasingly sophisticated characterization of privacy risks and harms and offer new interventions for protecting data subjects. In our work,⁴ we describe a lifecycle approach that supports a systematic decomposition of the factors relevant to data management at each information stage, including the collection, transformation, retention, access or release, and post-access stages. Additionally, we propose a framework for developing guidance on selecting appropriate privacy and security measures that are calibrated to the context, intended uses, threats, harms, and vulnerabilities associated with specified research activities.

Figure 1 provides a partial conceptualization of this framework.⁵ In this diagram, the x-axis provides a scale for the level of expected harm from the uncontrolled use of the data, meaning the maximum harm the release could cause to some individual in the data based on the sensitivity of the information. This scale ranges from low to high levels of expected harm, with harm defined to capture the magnitude and duration of the impact a misuse of the data would have on an affected individual's life, and we have placed examples as reference points along this axis. The y-axis provides a scale for the

³ Wood, Alexandra, Micah Altman, Kobbi Nissim, and Salil Vadhan. "Designing Access with Differential Privacy." Handbook on Using Administrative Data for Research and Evidence-based Policy., Shawn Cole, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber (Eds.). Abdul Latif Jameel Poverty Action Lab, Cambridge, MA (2020).

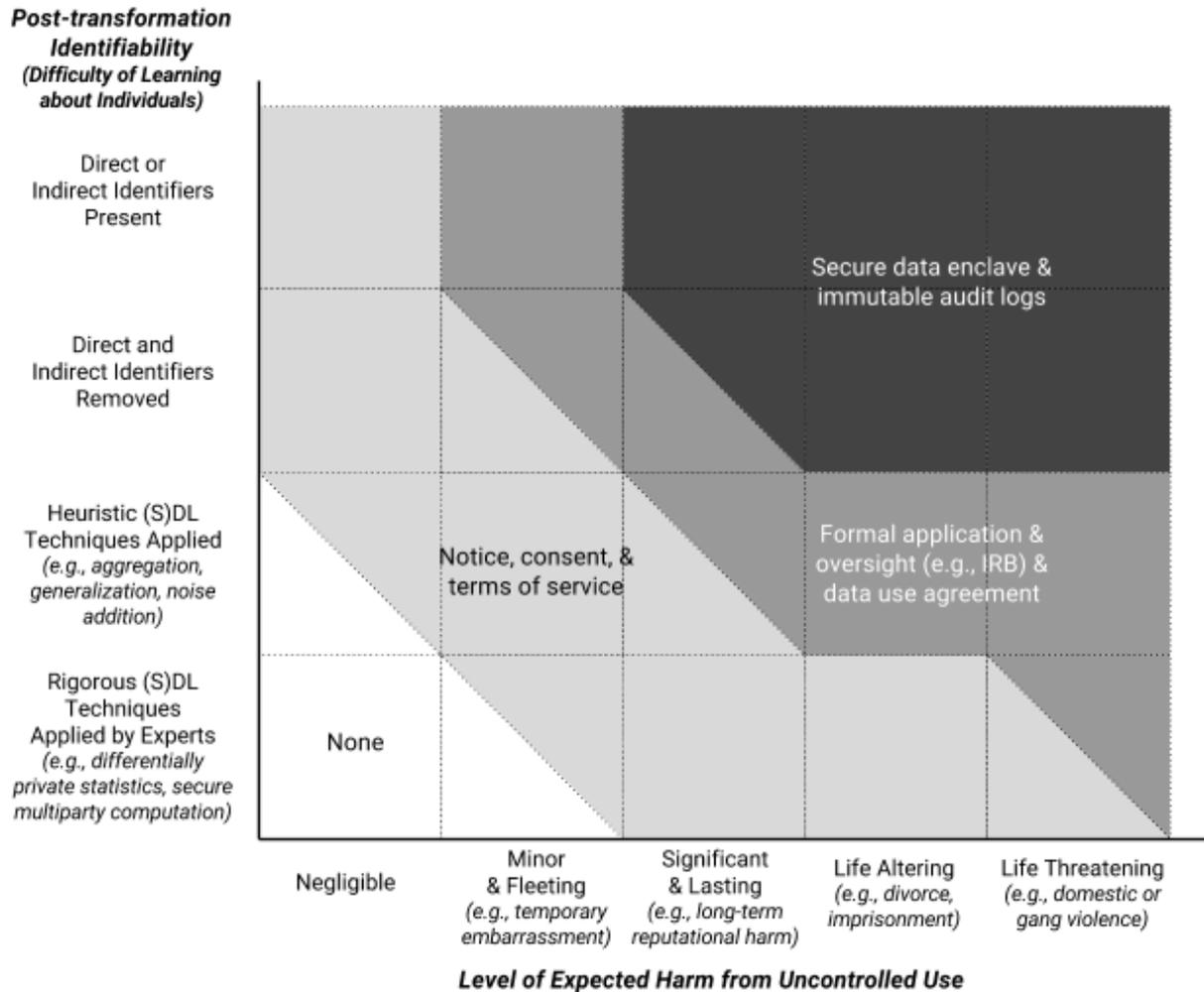
⁴ Altman M, Wood A, O'Brien D, Vadhan S, Gasser U. Towards a Modern Approach to Privacy-Aware Government Data Releases. Berkeley Technology Law Journal. 2015; 30(3): 1967-2072.

⁵ This diagram originally appeared in Altman M, Wood A, O'Brien D, Vadhan S, Gasser U. Towards a Modern Approach to Privacy-Aware Government Data Releases. Berkeley Technology Law Journal. 2015; 30(3): 1967-2072.

post-transformation identifiability or the potential for others to learn about individuals based on the inclusion of their information in the data. Several examples are provided as anchor points, ranging from data sets containing direct or indirect identifiers, to data shared using expertly applied rigorous disclosure limitation techniques backed by a formal mathematical proof of privacy.

The level of expected harm from uncontrolled use and the post-transformation identifiability of the data, taken together, point to minimum privacy and security controls that are appropriate in a given case, as shown by the shaded regions in the diagram. Regions divided by a diagonal line correspond to categories of information for which an actor could reach different conclusions based on the intended uses of the data or privacy standards that vary based on the applicability of regulation, contract, institutional policy, or best practice. The sets of controls within the shaded regions focus on a subset of controls from the more comprehensive set of procedural, economic, educational, legal, and technical controls we catalog in the work cited above. In practice, the design of a data management plan should draw from the wide range of available interventions and incorporate controls at each stage of the lifecycle, including the post-access stage. Also, note there are regions of this diagram that deviate from current practice in some domains. For example, we argue that data that have been de-identified using simple redaction or other heuristic techniques should in many cases be protected using additional controls, even though some existing standards do not expressly call for the use of additional controls when using such techniques.

Figure 1. Calibrating privacy and security controls.



For many activities, implementing a single set of privacy and security controls may not be appropriate for all intended uses of the information. For this reason, we generally recommend that regulators and data controllers implement a tiered access model. A tiered access model is one in which data are made available to different categories of data users through different mechanisms.

Figure 1 illustrates the relationship between transformation and release controls, and suggests how controls could be selected for different access tiers. For example, an investigator could provide public access to some data without restriction after robust disclosure limitation techniques have transformed the data into differentially private statistics. Data users who intend to perform analyses that require the full dataset, including direct and indirect identifiers, could be instructed to submit an application to an oversight body such as an institutional review board, and their use of the data would be restricted by the terms of a data use agreement. We argue that this framework, implemented through a data management plan and tiered access model, would help

data providers, data users, and oversight bodies calibrate the use of privacy and security controls to the contexts, threats, harms, and vulnerabilities associated with each specified research activity, as well as the purposes desired by different categories of data users.

In our prior work (cited above) we also call special attention to advanced data-sharing models and emerging formal approaches to privacy. We note that there are many privacy methods, PETs, and data-sharing models that can provide stronger privacy protection than traditional de-identification techniques that are in wide use today.

Although PETs have advanced rapidly, none offer a plug-and-play, high-performance, and high-quality solution over all common data protection use-cases: There remain many open questions that require technology-specific research and development: e.g. adapting different PET methods to specific computations and data types; performance tuning; and ease of use for developers, analysts, and the end-users of analytics -- including the public and policy-making communities.)

A broader and perhaps more important question is how to match and align PETs as appropriate to enable different intended uses, and protect against different harms and threats. The broad classes of harms that different PETs aim to mitigate, and the specific threats that they are effective in protecting against, vary widely across technologies. For example, secure multiparty computation aims to prevent harm to institutional data holders from other adversarial data holders while enabling computation on their joint data, by requiring collective approval for any computation on the joint data. In contrast to SMC, differential privacy aims to protect against privacy harms to individual data subjects by limiting the inferences that can be made from the informational outputs -- regardless of the specific type of computation; while personal data stores and data coops aim to reduce individual privacy harm and individual and group economic harm by limiting the domain of human activity in which computations and inferences are used.⁶ Further, most PETs aim to mitigate threats to individuals or specified institutional actors, and are not designed to provide wholesale protections to larger groups, such as marginalized communities.

Finally, we note that there is a paucity of research that examines the consequences of PET adoption or PET regulation on large interconnected social systems or society as a whole. It is well established that even privacy interventions that are provably effective and a Pareto-improvement in the short term can have a longer-term perverse effect on risk and global social welfare under many conditions (e.g. bounded rationality, externalities, adaptive system dynamics).⁷ And more and more frequently, the

⁶ For more details see Altman, Micah, Alexandra B. Wood, David O'Brien, and Urs Gasser. "Practical approaches to big data privacy over time." *International Data Privacy Law* 8(1):29-51. (2018).

⁷Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. "The economics of privacy." *Journal of economic Literature* 54, no. 2 (2016): 442-92.

introduction of PETs has broad consequences for data use and policy.⁸ Thus in the policy arena, it is important to consider not only the technical properties of the PET but to also consider the potential consequences of rules and policies requiring PETs and the way that both these rules and the widespread use of a PET may affect actor information and incentives.⁹

Data releases should incorporate more advanced data sharing models, including formal privacy models, where possible, as such techniques can enable wider access and use of data while providing robust privacy protection.

Thank you for your consideration of these comments.

Respectfully,

Micah Altman, Research Scientist, Center for Research in Equitable and Open Scholarship, MIT Libraries

Aloni Cohen, Assistant Professor of Computer Science and Data Science, The University of Chicago

Salil Vadhan, Vicky Joseph Professor of Computer Science and Applied Mathematics, School of Engineering and Applied Sciences, Harvard University

⁸ For a recent and particularly important case, see Boyd, Danah and Sarathy, Jayshree, Differential Perspectives: Epistemic Disconnects Surrounding the US Census Bureau's Use of Differential Privacy (March 15, 2022). *Harvard Data Science Review* (Forthcoming) , Available at SSRN: <https://ssrn.com/abstract=4077426>.

⁹ Altman, Micah, Alexandra Wood, and Effy Vayena. "A harm-reduction framework for algorithmic fairness." *IEEE Security & Privacy* 16, no. 3 (2018): 34-45.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Amazon Web Services (AWS)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



July 8, 2022

Re: Request for Information on Privacy Enhancing Technologies

Submitted To:

The White House
Office of Science and Technology Policy
Via Email

Submitted By:

Amazon Web Services, Inc.
12900 Worldgate Dr.
Herndon, VA 20170

Amazon Web Services (AWS) appreciates the opportunity to submit feedback to the Office of Science and Technology Policy (OSTP) in response to its Request for Information (RFI) on Advancing Privacy-Enhancing Technologies (PETs).¹

This RFI seeks input on technologies that advance mechanisms for privacy-preserving data sharing and analytics technologies to inform a national strategy that will put forth a vision for responsibly harnessing such technologies to benefit individuals and society. Some of the technologies and techniques referenced in the RFI include differential privacy, homomorphic encryption, and secure enclaves. AWS has invested considerable efforts into many of these technologies and techniques. We are grateful for this opportunity to provide feedback to OSTP on our efforts to help inform the national strategy that will help create an environment in which such technologies can be more broadly adopted and flourish to harness the power of data.

AWS is the world's most comprehensive and broadly adopted cloud platform, and serves millions of customers who are primarily businesses, non-profits, and government organizations. In this response we focus on efforts that AWS has undertaken to enable our customers with privacy-enhancing technologies. In addition to directly investing in PETs, AWS is also focused on related privacy topics including data governance, lineage, access controls, audits, and retention policies, among other areas. As a result of evolving privacy expectations of individuals, companies, and governments, the privacy needs of our customers are expanding rapidly, including the need for technologies that allow for disassociation and confidentiality. The way each company implements data privacy is dependent on their business goals; our goal is to provide customers a comprehensive set of technical capabilities that they can use to implement privacy into their architectures depending on their business needs and risk posture.

In our comments below, we describe specific PETs that we believe OSTP should focus on in its national strategy, as well as specific applications and research efforts that could help inform the national

¹ <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>

strategy, specific applications of these PETs in AWS products and services, as well as noting relevant limitations, tradeoffs, and barriers to adoption. Finally, we have set forth regulatory efforts that we believe can help in ensuring a thriving market for the creation, adoption, and use of PETs in the future.

Specific Privacy-Enhancing Technologies that Advance Greater Data Analysis and Sharing

Below we describe some PETs that AWS has focused on that can be leveraged to de-identify and disassociate data for the purpose of data analysis and sharing, and which we believe OSTP should also focus on in its national strategy. Increasingly, merging and leveraging different data sets can facilitate innovation and benefit the public. We have provided examples in each area of how such technologies have been utilized or could be utilized to enable data collaboration amongst multiple parties while disassociating the data from the data subject and preserving privacy.

Differential Privacy

We believe differential privacy (DP) is a powerful technique for protecting individuals' privacy when their data is used to derive insights about a data set, and OSTP should closely examine and invest further in this technology. Amazon Science has published a number of articles on differential privacy, which can be found here: <https://www.amazon.science/tag/differential-privacy>

The main idea behind DP is to add randomness — noise— to obscure the presence or absence of any single individual in the data set. Randomness can be integrated at different stages of an algorithm. For example, noise can be added to the data itself or to the output of the queries computed on the data. The more noise that is introduced, the greater the privacy protection, but more noise also degrades the utility of the algorithm outputs. This trade-off between protecting the privacy of users and providing useful insights about the population is controlled by a parameter known as *privacy budget*. The privacy budget is a finite resource that is consumed each time a query is run. Differential privacy is best suited for use cases where a small amount of error in the results of a query is acceptable, and does not destroy the utility of the query results. At AWS, we are actively conducting research and development into differential privacy technologies that make secure computation easier and more efficient. We believe this technique is an important building blocks for better data governance and privacy, and that there is significant privacy value to be gained from investments in this area.

Ongoing research on privacy-preserving machine learning in natural language models is summarized in this Amazon Science blog post: <https://www.amazon.science/blog/advances-in-trustworthy-machine-learning-at-alexai>. As noted in this blog post, differential privacy provides a rigorous way to quantify the privacy of machine learning models. The researchers investigated vulnerabilities presented in the differential-privacy literature and proposed computationally efficient mechanisms for protecting against them. Additionally, training machine learning models on synthetic, differentially private data can be a powerful way to make machine learning more privacy protective; these efforts are detailed further in this blog post: <https://www.amazon.science/blog/amazon-helps-launch-workshop-on-synthetic-data-generation>.

Differential privacy technology can be applied to enable multiple parties to engage in data collaborations. In data collaborations, one party (data consumer) can join their data with that of another party (data provider), and run aggregate queries (e.g. sum, count, average) against the joint dataset. These collaborations are commonly known as 'clean rooms' because neither party's raw data is revealed to each other in the process. AWS customers, particularly in the online advertising context, can

drive privacy protective multi-party data collaboration using data clean rooms:
<https://aws.amazon.com/blogs/industries/deploying-a-privacy-safe-data-clean-room-on-aws-with-snowflake/>.

Secure Enclaves

AWS Nitro Enclaves enables customers to create isolated compute environments to further protect and securely process highly sensitive data such as personally identifiable information (PII), healthcare, financial, and intellectual property data within their Amazon EC2 instances.

Nitro Enclaves helps customers reduce the attack surface area for their most sensitive data processing applications. Enclaves offers an isolated, hardened, and highly constrained environment to host security-critical applications. Nitro Enclaves enables a range of use cases that deal with the processing of highly sensitive data, such as securing private keys, tokenization, and multi-party collaboration. Nitro Enclaves' isolation, cryptographic attestation capabilities, and integration with AWS Key Management Service, are key features that provides customers with a practical approach to setting up multi-party collaboration. To illustrate how multiparty collaboration can work with Nitro Enclaves, we have recently published an example proof of concept (POC) on third-party bidding service for real estate transactions. In the POC, buyers will submit encrypted bids to the application. Once all the bids have been entered, the application will decrypt the bids, determine the highest bidder, and return a result without disclosing the actual bid amounts to any party. For more details, see:
<https://aws.amazon.com/blogs/compute/leveraging-aws-nitro-enclaves-to-perform-computation-of-multiple-sensitive-datasets/>.

Computing Over Encrypted Data

AWS Cryptography tools and services utilize a wide range of encryption and storage technologies that can help customers protect their data at rest and in transit. In some instances, customers also require protection of their data even while it is in use. To address this need, AWS is developing new techniques for cryptographic computing, an emerging technology that allows computations to be performed on encrypted data, so that sensitive data is never exposed. It can be the foundation used to help protect the privacy and intellectual property of data owners, data users, and other parties involved in machine learning activities. AWS presented on this class of techniques at our re:Invent conference in 2020, which can be accessed here, along with other cryptographic computing publications:
<https://aws.amazon.com/security/cryptographic-computing/>.

However, the computational resources and cost needed for this type of computing pose significant barriers to adoption. We believe this is an area of research that could benefit greatly from OSTP investment to materialize and operationalize.

Other cryptography efforts are also worth noting. Specifically, AWS has focused significant efforts on post-quantum cryptography for the cloud, and some of our efforts in this area can be found here:
https://d1.awsstatic.com/events/reinvent/2020/Building_PostQuantum_Cryptography_for_the_Cloud_SEC207.pdf. In 2020, AWS completed benchmarks of Round 2 Versions of the Bit Flipping Key Encapsulation (BIKE) and Supersingular Isogeny Key Encapsulation (SIKE) hybrid post-quantum Transport Layer Security (TLS) Algorithms, which were submitted to the National Institute of Standards and Technology (NIST) as part of NIST's Post-Quantum Cryptography standardization process:
<https://aws.amazon.com/blogs/security/round-2-hybrid-post-quantum-tls-benchmarks/>

Federated Learning

Federated learning allows data from multiple sources to be examined while limiting researcher access to each data source to help preserve privacy. For example, in some federated approaches, data contained in disparate sources can be queried for information on a certain question and only aggregated or de-identified data is returned to the researcher. Federation also enables the development of scalable architectures, where new data sources can be added to research models. As a result, use of federated systems can support initial proof-of-concepts that can then grow to include new data sources and further accelerate the time-to-science. Here is an example of the use of federated analysis in the European Healthcare system: <https://aws.amazon.com/blogs/industries/data-mesh-technology-can-enable-european-healthcare-system-collaborate/>

Regulatory Recommendations

PET adoption may be enhanced through meaningful measurement standards. We believe a risk-based approach which accounts for different ways to mitigate risks to acceptable levels, and different ways to measure that risk, would be a beneficial approach. It is important to address measurement standards from a flexible, risk-based approach (e.g. different models will reduce risk to different levels) to account for the vast number and type of contexts in which PETs would be deployed. Furthermore, data that has been de-identified should not be within the scope of privacy laws or regulations as disassociation would remove potential risks to individuals related to the data. This will also continue to promote the faster adoption of PETs as a way for organizations to reduce their privacy risks, and also retain the flexibility for the development of new and enhanced technologies.

We encourage OSTP to support efforts for sensible, comprehensive data privacy protections in the U.S. Such efforts should not deter innovation, and should encourage the development and adoption of techniques and controls that obstruct reidentification. Any such legislation should also clearly understand the different roles of data controllers and processor. Notably, efforts to ensure responsible and fair use of artificial intelligence, which is an application for many of the technologies above, should remain in a separate framework.

In addition, we encourage OSTP to work with each federal agency to accelerate adoption of PETs—particularly for those agencies that process more sensitive data. For example, OSTP should work with health agencies to accelerate use of PETs to spur biomedical innovation and provide patients with greater ability to direct the sharing of their data. OSTP should also work with the Department of Health and Human Services (HHS) to accelerate individual-level control and input on the sharing of their data. In this way, individuals could authorize the use of their data for different purposes, such as sharing their data with other care providers, discovering clinical trials that may be relevant for them, or volunteering data for research purposes. Cross-agency collaboration—including with the Centers for Medicare & Medicaid Services, Office of the National Coordinator for Health Information Technology, National Institutes of Health, and Office of Civil Rights—can identify opportunities to accelerate greater individual-level control. OSTP should work with HHS to examine the development of incentives (e.g., via electronic health record certification or the Promoting Interoperability program) and other policies to encourage healthcare providers to support individual-level control on the sharing of their data.

We thank you for the opportunity to respond and provide input into the development of the national strategy for advancing PETs.

Sincerely,

Shannon Kellogg

Vice President, AWS Public Policy --- Americas

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Anonos

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

To: PETS-RFI@nitrd.gov

Re: RFI Response: Privacy-Enhancing Technologies

From: Gary LaFever, CEO, Anonos, Inc.

Mark Little, Chief Data Strategist and Head of Engineering, Anonos,
Inc.

Type: Industry Response

1. Specific research opportunities to advance PETs: Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.

Encryption and access controls have become firmly established as standard practice for protecting data in transit and at rest. Processing of data with sensitive information, however, has suffered from a lack of PETs that are simultaneously practical to use, provide effective protection and preserve utility. Stated differently, organizations have not had access to PETs that provide efficient, scalable protection for data when in use.

Statutory Pseudonymisation, as first defined in Article 4(5) of the EU General Data Protection Regulation (GDPR), is rapidly becoming a de facto global standard for protection of data when in use and formally recognized by more than forty governments and non-governmental organizations (NGOs) around the world. Essentially identical statutory language is found in the EU GDPR, the UK GDPR, and the Data Protection regulations of Japan, South Korea, Brazil, and five US States (CA, VA, CO, UT, and CT), and formally acknowledged by the German Association For Data Protection And Data Security (GDD) and the World Economic Forum (WEF).

In each case, it has been embraced as a means for reconciling conflicts between maximizing data value and protection. Other countries and US states are looking to adopt similar provisions incorporating Statutory Pseudonymisation because of its unique ability to simultaneously maximize both data utility and data protection without being overly cumbersome, a significant advantage over other PETs (see submission to Item 2 below for further comments on this point).

The infographic is set against a dark blue background. At the top right is the ANONOS logo, which consists of two interlocking infinity symbols. Below the logo, there are several groups of flags and logos representing different regions and organizations. On the left is a grid of European Union member state flags, with the text 'EUROPEAN UNION' and the logo of the European Data Protection Board (EDPB) below it. In the center, there are flags for the UK, South Korea, Brazil, and Japan, each with its name written below. To the right of these are logos for GDD (German Association for Data Protection and Data Security) and WEF (World Economic Forum). Further right are flags for five US states: California, Virginia, Colorado, Utah, and Connecticut, with the text 'US STATES' below them. At the bottom, the title 'STATUTORY PSEUDONYMISATION' is written in large white capital letters. Below the title is a yellow banner with the text 'Recognized by 40+ Regulators and NGOs Around the Globe'. At the very bottom, the text 'Becoming New De-facto Standard' is written in large yellow capital letters.

To understand how Statutory Pseudonymisation delivers this advantage, it is necessary to look at the differences between past common use of the term pseudonymization and the statutory construction of the new definition, and the resulting implications.

Prior to the ratification of the GDPR, there were no statutes or laws defining the term pseudonymization, although the term has been in common use for many years. Most data protection practitioners would characterize it as a technique for obscuring personally identifying information (PII) that replaces direct identifiers with static tokens.

In contrast the EU GDPR defines Statutory Pseudonymisation as:

'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;[1]

The following graphic highlights how significantly different and more demanding the requirements are for Statutory Pseudonymisation than for the PET known variously as pseudonymization, hashing, tokenization, and key-coding.

Something New Under the Sun



Under the GDPR, the requirements of Article 4(5) fundamentally redefine Pseudonymisation to

- 1 Dramatically expand the scope to include all Personal Data, vastly more comprehensive than direct identifiers; and
- 2 Dramatically restrict the scope of additional information that is lawfully able to re-attribute personal data to individuals.

'pseudonymisation' means the processing of **personal data** in such a manner

- o that the **personal data can no longer be attributed**
- o to a **specific** data subject
- o **without** the use of **additional information**,





provided that **such additional information**

- o is **kept separately** and
- o is **subject to technical and organisational measures**
- o to ensure that the **personal data are not attributed** to an identified or identifiable natural person;

The first (**blue**) half of the Article 4(5) definition, by itself, means:

- o The **outcome must be for a dataset** and not just a technique applied to individual fields **because of the expansive definition of Personal Data** (all information that relates to an identified or identifiable individual) as compared to just direct identifiers;
- o Additional information could come from anywhere, **except the dataset itself**; and
- o Replacement of direct identifiers with **static tokens could suffice**.

However, when combined with the second (**purple**) half of the definition, the requirements regarding additional information mean that **any combination of additional information sufficient to re-attribute data to individuals must be under the control** of the data controller or an authorized party. To **achieve this level of protection**, it is necessary to:

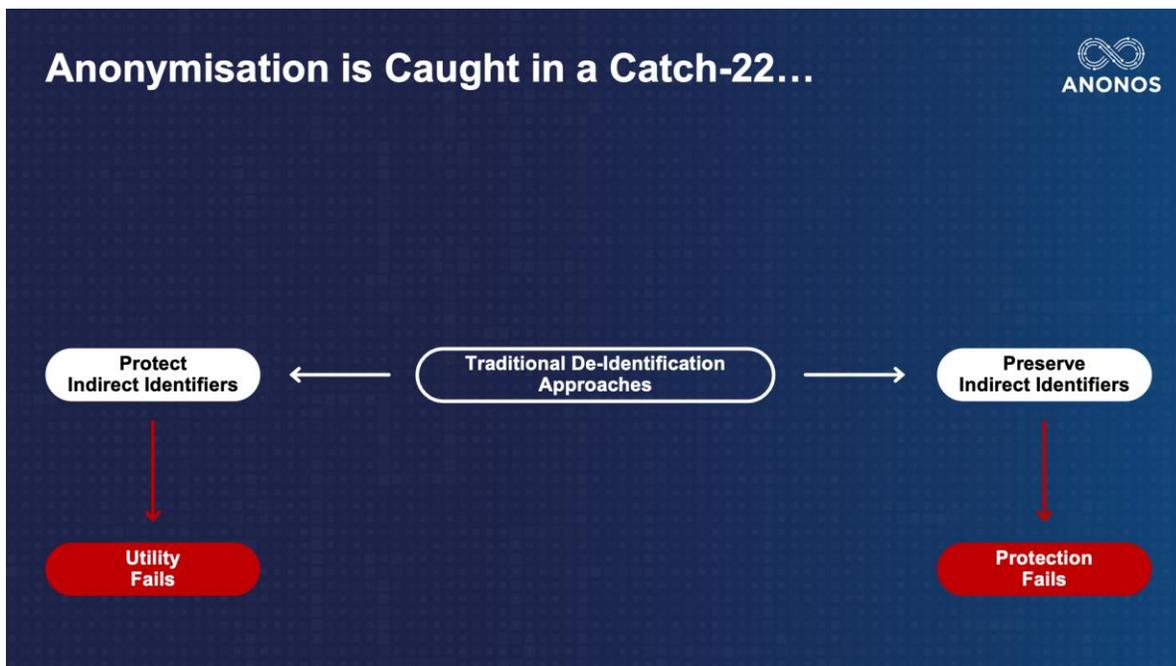
- o **Protect all indirect identifiers** as well as direct identifiers; and
- o Use dynamism by assigning different pseudonyms at **different times for different purposes** to avoid unauthorized re-linking via the Mosaic Effect (see <https://MosaicEffect.com/>).

This language fundamentally changes the meaning of the term in two ways. First, it dramatically expands the scope of applicability to Personal Data as defined under the EU GDPR (all information that relates to an identified or identifiable individual) which is much more comprehensive than direct identifiers that are PII. Second, the scope of additional information that can be used to

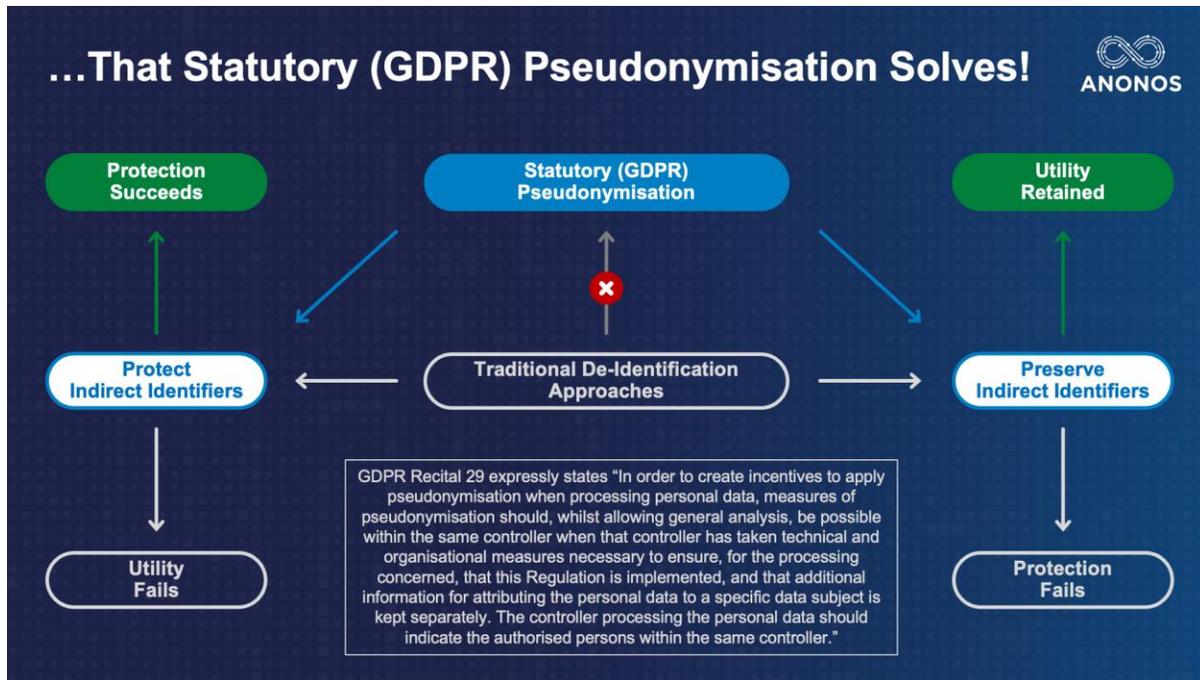
lawfully re-attribute personal information to an individual is dramatically narrowed so that all such information must be subject to technical and organizational controls limiting access to it.

Accordingly, it is clear that when defined this way, Statutory Pseudonymisation is no longer a description of a technique, but rather of an outcome describing the state of an entire data set, just as encryption or anonymisation transform cleartext into new states, viz., encrypted and anonymized. As such, transformation or replacement of individual fields must take into account the potential to reattribute data to an individual not only for direct identifiers, but also for quasi-identifiers. In some cases, it may be necessary or advantageous from a data protection standpoint to protect certain indirect-identifiers and attribute fields as well.

Under the EU GDPR, Anonymous [2] data is not considered Personal Data, and thus falls outside the scope of the regulation. In theory this makes anonymisation an attractive option for processing personal data. However, in a world awash in data, as a data protection technique, organizations attempting to rely on anonymisation find themselves in a catch-22[3]. On the one hand, if in attempting to anonymize data quasi-identifiers are (by definition) irreversibly protected, utility for analytics is largely destroyed, as the original values cannot be recovered. However, if the quasi-identifiers are not protected in an effort to retain analytic utility, there is little chance the data set will meet any meaningful standard of anonymity. Moreover, for data to qualify as anonymous under the GDPR, even the party creating an “anonymous” data set must not be able to reverse the protection. In practice, no one deletes source data after creating an “anonymous” version of the data, which means reversal of the protection is trivial (with the possible exception of aggregate data, which has limited utility in most instances).



What is not widely appreciated about Statutory Pseudonymisation as a PET is that it solves both of these limitations. Because, by definition, any protection of quasi-identifiers can be reversed when authorized, they can be aggressively protected with no loss in analytic utility. Statutory Pseudonymisation is often (incorrectly) characterized as weaker protection than anonymisation because it is still personal data and does not move processing outside the jurisdiction of the law, however, nothing could be further from the truth. Statutory Pseudonymisation actually provides superior protection against unauthorized reidentification with better utility than so-called anonymous data.



Extensive conversation with experts in EU data protection law – among them members of the European Data Protection Supervisor (EDPS) [4] and EU Member State Data Protection Authorities, including several involved in drafting the statutory language or subsequent regulatory guidance and recommendations (e.g., GDPR Article 4(5), EDPS recommendations for Schrems II compliant international data transfers) – confirms two things. First, that the above interpretation is exactly what was intended in the original construction of the GDPR statutory language and second, that the above analysis is not well understood outside the drafters of the language and a very small number of data protection experts.

In light of the foregoing, Federal research opportunities confirming the above, supporting software development, and providing education and training on the advantages and utility of Statutory Pseudonymisation as a PET should be a high priority.

2. Specific technical aspects or limitations of PETs: Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

The following chart evaluates the full range of data protection techniques, including both security-based approaches and traditional PETs against a series of criteria for evaluating the effectiveness of protection and the utility of the protected output. Rather than being a traditionally red/green or stoplight chart that evaluates all PETs against all criteria, this is a knockout chart. PETs are evaluated against the criteria sequentially from left to right, and once a PET fails to meet a criterion it is dropped from further consideration.

Limitations of PETs



Protections and Techniques	Type	Protects Data In use	Supports Protected Data Sharing and Multi-Cloud Processing	Supports AI and Machine Learning	Reconciles Conflicts Between Protection and Accuracy	Utility Comparable to Cleartext
Cleartext	None	NO				
Cleartext with Access Controls	Security	NO				
Trusted Execution Environment (TEE)	Privacy Enhancing Computation	YES	NO			
Multi-Party Computing (MPC)	Privacy Enhancing Computation	YES	YES	NO		
Homomorphic Encryption (HE)	Privacy Enhancing Computation	YES	YES	NO		
Differential Privacy	Privacy Enhancing Computation / Anonymisation	YES	YES	NO		
Cohorts/Clusters	Anonymisation	YES	YES	NO		
Masking	Anonymisation	YES	YES	YES	NO	
K-Anonymity	Anonymisation	YES	YES	YES	NO	
Tokenization	Anonymisation	YES	YES	YES	NO	
Generalization	Anonymisation	YES	YES	YES	NO	
Synthetic Data	Anonymisation / Privacy Enhancing Computation	YES	YES	YES	MIXED ¹	MIXED ¹
Statutory Pseudonymisation	Privacy Enhancing Computation	YES	YES	YES	YES	MIXED ²

¹Vendors claim and Buyers believe YES; informed commentary concludes NO.
²Buyers assume NO; informed commentary concludes YES.

Cleartext with Access Controls

Access controls are an essential component of data security. However, no matter how granular they are (e.g., attribute-based, tasked-based or even zero-trust) they are still binary; once granted, access is to clear-text. As a result, they do not provide protection for data in use.

Encryption

Encryption is the *sine qua non* for protecting data at rest and in transit. But to use data, for example in computation or analytics, it must be decrypted, at which point it is no longer protected at all.

Homomorphic Encryption (HE)

Research has been ongoing for many years in an effort to find a way to improve processing speed to a level even approaching commercial viability. Most information touting progress talks only of “improvements” and not actual processing throughput results, for good reason. Estimates suggest processing speeds that are 5 to 10 orders of magnitude slower than processing cleartext. That implies that computations that would take one millisecond in clear text would take anywhere from 1.5 minutes to nearly 4 months.

Multi-Party Computing (MPC)

A relatively new technique that is frequently (mis)represented as “encryption in use,” presumably for marketing purposes. The justification seems to be that more precisely, the encoding of data done to enable the shared computations is fairly characterized as a cryptographic technique, as is encryption. But as commonly used, encryption is not understood to be the encoding done in MPC, which results in cleartext values. In any case, MPC remains cumbersome, as it requires tremendous bandwidth for the communication and coordination required between the computing parties, which can be both expensive and results in processing speed penalties, limiting its use to niche applications.

Trusted Execution Environment (TEE) / Confidential Computing Environment (CEE):

Among the newest of new techniques, this approach sets up an on-processor enclave of a portion of system memory, and in some implementations, part of the CPU itself. Data is stored and moved around the processor in encrypted form until inside the enclave, where it is decrypted using a key only available within the enclave. Implementation is technically challenging, and often requires rewriting applications to work in the TEE. Additionally, the enclave is by definition a silo. Thus, this approach is not well-suited for data sharing and combining and multi-cloud or hybrid-cloud applications.

Differential Privacy and Cohorts/Clusters

By definition, these techniques provide results that are aggregated, and do not provide the record-level output necessary for most uses of data.

Anonymization

The following techniques, Masking, Generalization, Tokenization, K-Anonymity, Noise Introduction, and Synthetic Data all are used, typically by combining several together, in an effort to Anonymize data. However, in the effort to do so, they all fail to resolve the intractable trade-off between privacy

and utility that is inherent in anonymisation. In a big data world, they fail to deliver the privacy promised by anonymisation, and efforts to push them to their limits to do so ends up destroying the utility of the protected output.

Masking

This technique protects direct identifiers by masking or overwriting one or more characters. It requires the data, its use, and its users are all restricted/sequestered to prevent other unprotected fields in a record from being combined with the information in additional data sources to enable an individual to be distinguished from others or identified via linkage attacks (see <https://MosaicEffect.com>). This requirement to restrict access is inconsistent with the architectural requirements of increasingly prevalent use cases that require free flowing data and involve dynamically changing data sources, processes, and processors.

Generalization

This technique attempts to protect against reidentification by reducing the granularity of the original data. Classic examples include converting age to age ranges by range binning, or by rounding numerical values. Masking can also be used for generalization such as masking one or more trailing digits of a zip or postal code to create values that represent larger areas. By itself, this technique does little to protect identity as it is not useful for direct identifiers. It is often put into practice in an effort to achieve a specified level of k-anonymity (see below).

Tokenization

(Hashing/Key-Coding/Pre-GDPR Pseudonymization): These techniques: (i) only protect direct identifiers and (ii) protect those direct identifiers by replacing them with a recurring (persistent) token, making them effective only for limited, static use cases. They require that the data, its use, and its users are all restricted/sequestered to prevent other unprotected fields in a record from being combined with the information in additional data sources to enable an individual to be distinguished from others or identified via linkage attacks (see <https://MosaicEffect.com>). This requirement to restrict access is inconsistent with the architectural requirements of increasingly prevalent use cases that require free flowing data and involve dynamically changing data sources, processes, and processors.

K-Anonymity

K-anonymity techniques are intended to prevent a data subject from being singled out by grouping them with at least “k”-1 other individuals who share the same values for a specified subset of attributes in a data set. This subset of attributes, which are commonly referred to as quasi-identifiers because of their ability to, when used in combination, reveal identity. The quasi-identifiers are generalized as necessary (using techniques such as range binning, rounding, and masking) to ensure that all possible subgroups defined by the values of the quasi-identifiers have at least k individuals in them. In most cases, to achieve that status for all records in the data set, the required generalization severely degrades the utility of the data. To mitigate the degradation, a decision is made to be less aggressive in the generalization, and then suppress values or entire

records in those subgroups where k falls short of the specified level. Note however that this also results in degradation of data utility, as a result of distortion in the output dataset statistical properties relative to those in the original source data.

Noise Introduction

This technique involves intentionally changing values in a data set so that they are less likely to be useful in revealing identity while at the same time avoiding excessive degradation in data utility due to distortion of the statistical relationships among attributes. This technique explicitly trades off utility (i.e., accuracy) for privacy, and tends to fall short on both accounts.

Synthetic Data

The failure of synthetic data to adequately protect against identity disclosure is now well documented in academic papers. The current state of the art appears to be ~ 1% of data subjects at risk of identity disclosure, which is likely to be judged to be far short of the regulatory requirements for anonymous data. Efforts to reduce this risk inevitably come at the expense of accuracy, as maximizing accuracy leads to overfitting and duplicating unique records in the source data. Some organizations report accuracy rates of as low as 70%. An additional challenge relates to the incorporation of incremental records to an existing source data set, or the addition of additional tables. In order to properly preserve the statistical properties between records, fields and tables, these situations almost always will require regenerating the models used to create synthetic data.

Statutory Pseudonymisation

While it does address the forgoing criteria, there is no getting around the fact that a Statutorily Pseudonymised data set looks quite different from its cleartext source. For many aspects of analytics, particularly actual computation in algorithms this is not an issue at all, as the pseudonyms simply process as nominal or categorical strings. For analytics involving active participation by a person (e.g., exploratory data analysis, BI reports, feature engineering, results interpretation, etc.) this is clearly not the same as working with cleartext. That said, appropriate organizational controls used in conjunction with authorized reversals of pseudonyms to cleartext when necessary to advance processing means the issue is not insuperable, but more akin to a change in workflows.

[1] EU GDPR Article 4(5). Note that other jurisdictions have as a rule borrowed this construction verbatim (or nearly so).

[2] EU GDPR Recital (26).

[3] A problem in which the solution is denied by the problem itself. See <https://www.merriam-webster.com/dictionary/catch-22>.

[4] See attached report provided to the EDPS following a meeting with them and providing a mathematical proof of how data is protected using Statutory Pseudonymisation.

Attached Reference Document:

Anonos Data Embassy Overview Presented to European Data Protection Supervisor



ANONOS

Data Embassy and Variant Twins

This document describes the technical and mathematical underpinnings enabling Anonos software to overcome the well-established axiom:

“Data can be useful or perfectly anonymous, but never both.”

With Anonos software, people are no longer forced to choose whether they want data utility or protection.

Anonos enables them to have both.

Data Embassy and Variant Twins: Overview and Mathematical Underpinnings

****We recommend watching the following 9-minute video summary before reading this document <https://www.anonos.com/9minutes>**

Overview

Data Embassy Value Proposition

Anonos Data Embassy software uses technologically enforced protection to transform cleartext data into variable-resolution, use-case specific outputs called **Variant Twins**.¹ This fine-grained approach delivers:

- **Proactive Security/Privacy**, where data is fully protected **during use**, even in the event of a breach.
- **Compliance**: GDPR, CCPA, and more. Schrems II compliant surveillance-proof processing in US operated clouds, regardless of the location of servers.
- **Full Compatibility** with a wide range of both primary and secondary uses of data.
- **100% Accuracy**, [verified by external experts](#), relative to processing unprotected cleartext by enabling the relinking of the results of protected processing to source data under controlled conditions for authorised purposes only.

The technological controls embedded in Variant Twins travel with the data, enabling fully decentralised processing. **Anonos Data Embassy allows for maximum data utility, compliance with international privacy laws, and mitigated liability risk upon breach.**

Data Embassy Use Cases

Secondary Uses of Data

- Analytics, Business Intelligence and Reporting
- Machine Learning
- Artificial Intelligence
- Data Combing and Sharing

Primary Uses of Data

- Customer and Employee Support
- Transaction Processing
- Right to be Forgotten/Delete My Data

Location and Jurisdiction Independence

- On Premises
- Private Cloud
- Public Cloud and other International Transfers
- Multi-Cloud and Hybrid Cloud

Scalable Enterprise-Grade Protection

Data Embassy deploys with features and capabilities necessary for use at scale by global enterprises.

- **Data Protection Rules:** Configurable templates enable digitisation and technological enforcement of an organisation’s privacy policies so that they are no longer “just words” in a document. Configure and approve once per use case, automatically apply over and over.
- **Tagging:** Enables tags for field identifier types, statistical data types, use cases, jurisdictions, Variant Twin recipients, and more, enabling rapid and automated configuration of data transformers.
- **Group and Role-Based Permissions:** Establishes group and role-based permissions allowing users to authenticate and authorise controls for the necessary separation of responsibilities, segregation of duties, and “need-to-know” restrictions essential for demonstrating the technical and organisational controls over approvals for both protecting data and reversing protections.
- **Approvals:** Implements approvals for protection rules, policy deviation requests, data transformer configuration, Variant Twin creation and controlled relinking/reversal.
- **Auditability:** Maintains immutable records of all system activity through User IDs.
- **Design Studio with Privacy Engineering Tools and Aids:** Provides users tools for risk scoring, k-anonymity analysis, cardinality analysis, multiple preview modes, field/rule linking of data, to maintain referential integrity when required.
- **No Code Configuration:** Fast, easy, and scalable.
- **Fully Documented APIs:** Supports custom development via scripting, automation, and integration with existing data pipelines.
- **Modern Component Stack:** Kubernetes, Docker, Cassandra, PostgreSQL, Spark, Kafka, Kotlin, React, Vault, Keycloak; supports batch, streaming, high throughput, and high availability.
- **Improved Productivity:** Customers report a four-fold increase in the approval of projects, each in 25% of the time—achieving an overall 16x productivity gain in making high-value data available for use.

Compound Cryptographic Security

The use of multiple data transformation techniques, including cryptographic algorithms, enhances security and reduces the risk of access by an unauthorised party. Data Embassy integrates multiple transformation techniques, including:

- Omission of direct identifiers and replacement with random or deterministic pseudonyms (i.e., the same input value always results in the same output value) when it is not possible to omit such direct identifiers.
- Conversion of numerical fields to categorical whenever possible (e.g., age 25 to an age range of 20-29), while taking care to ensure that cardinality (number of unique values in a field) is neither too high (risking “fingerprinting”) nor too low (more easily guessed via brute force attacks).
- Replacement of the values in **all** categorical fields (direct identifiers, quasi-identifiers, indirect identifiers, and attributes) with deterministic pseudonyms.
- Scope delimiting the referential integrity of deterministic pseudonyms to the minimum needed for each given use case.
- Enforcement of patented dynamism by replacing data elements to be transformed with different pseudonyms at different times for different purposes (i.e., the same input value is assigned different pseudonyms for different projects, so outputs from one project cannot be used to “attack” outputs for another project). This is accomplished by using different initialisation vectors (commonly referred to as keys) and at least two cryptographic algorithms (see Appendix). This approach restricts the ability to combine or correlate otherwise seemingly unassociated information to defeat unauthorised reidentification via inference attacks and linkage attacks using the [Mosaic Effect](#).
- Application of k-anonymity to suppress records with uncommon combinations of quasi-identifier values (even when replaced with pseudonyms) to defeat singling out attacks.

Because of the integrated complexity of these multiple layers of protection, when properly implemented, it would not be possible to “attack” and reverse the protections to reattribute data to specific individuals without access to the keys and algorithms used to encode the data.

Despite the integrated complexity of this multi-layered protection, as with all processing before passing it to a processor for computation, the resulting output is ultimately converted into binary (0’s and 1’s), resulting in the same processing efficiency as unprotected cleartext – i.e., unlike homomorphic encryption or other methods of protecting the underlying data, the use of Anonos’ Variant Twins does not increase the need for computing resources.

The net effect of the above is protection for data **in use** that is analogous to SSL/TLS (Secure Sockets Layer/Transport Layer Security) protection for data **in transit**, since Data Embassy prevents unauthorised use of information, including potential personal details.

Data Embassy implements a Zero Trust Security Policy known as the principle of least privilege, or PoLP, that limits a user’s access rights to only what is strictly required for an assigned role, using centralised controls that technologically embed protections into data, **effectively restricting access rights to decentralised data flows to only the level of identifying data (including none) required for each authorised use case.**

Cleartext Utility

The layered approach to replacing cleartext source data with a cryptographically secure Variant Twin is fully compatible with most analytics, machine learning and AI processing, while delivering the accuracy of cleartext.

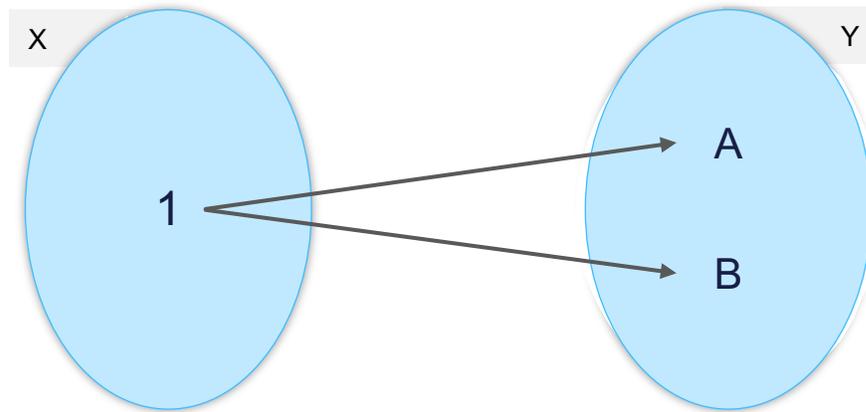
The reason for this surprising result is that replacing categorical values with deterministic pseudonyms has **zero** impact on the functionality of those data elements in the underlying algorithms. The algorithms only care that the values of the strings comprising categorical variables are (1) consistently used and (2) distinguishable. For example, in a model, Male/Female, M/F and a098dfae19 / ffud630rmf7 all work equally well, but only the last pair serves to prevent reidentification attacks.

However, the latter is not an option for a quasi-identifier like sex (nor for indirect identifiers nor attributes either) when attempting to anonymise data. The reason is that since anonymisation must be irreversible, much of the utility or insights that might be derived from the protected fields are irretrievably destroyed. However, with GDPR-compliant pseudonymisation, this type of protection is (1) necessary to meet the statutory requirements but (2) is definitionally allowed to be reversed under controlled conditions, ensuring the utility and insights are not lost, but are instead preserved.

Mathematical Underpinnings Of Variant Twin Cryptographic Security

Primer on Mathematical Functions

The concept of a function is very important in mathematics. You can think of a function as being a machine that takes in a certain set of inputs, and for each input item, generates one (and only one) output. Mathematically speaking, a function f is a mapping that takes elements from a set X and maps them to elements in a set Y . To emphasise that an element y in Y is the output of an element x in X , we commonly write $y = f(x)$.

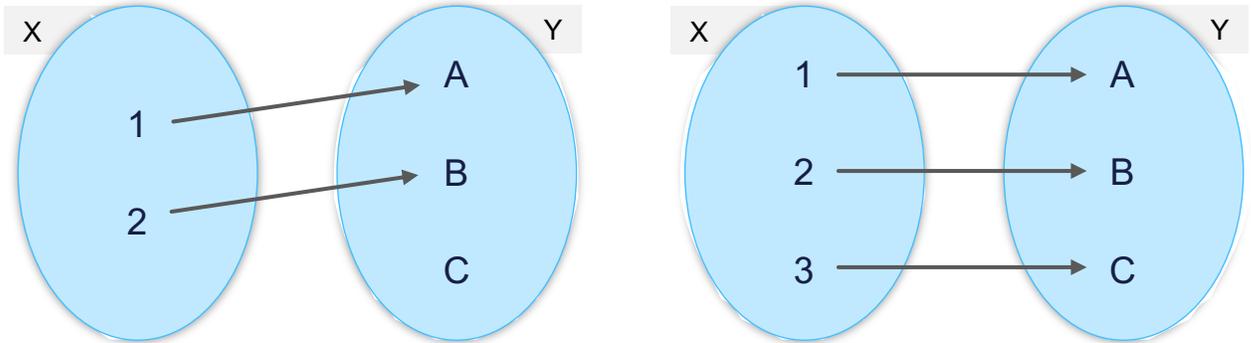


A one-to-many mapping, which is not a function.

Let f be a mapping from a set X to a set Y . If f maps an element of X to more than one element of Y , then f is said to be a *one-to-many*, which is not a function. An easy way to remember this is the phrase “he has had one too many, so he cannot function.” The below image shows an example of a one-to-many mapping, which is not a function.

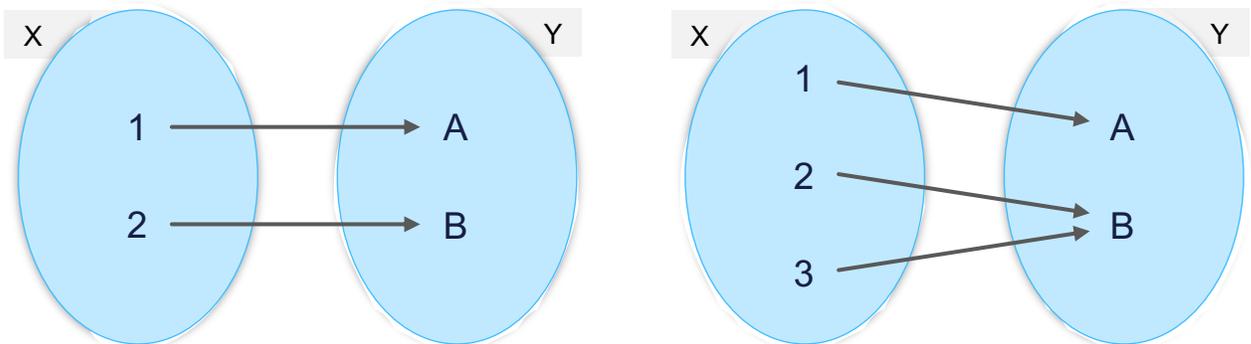
If f is a function from a set X to a set Y which maps each element of X to one (and only one) element in Y , then f is said to be *one-to-one* (or more mathematically sophisticated, *injective*).

An easy way to remember the term *injective* is to think of getting an injection or vaccination. The vaccination (injection) maps one (unvaccinated) person, to one (vaccinated) person. The below image demonstrates two injective functions.



An injective mapping, which is a function.

A function f from a set X to a set Y is said to be *onto* (or more mathematically sophisticated, *surjective*) if f maps an element of X to each element of Y . In other words, if for every y in Y , we can always find (at least one) element x in X such that $y = f(x)$, then f is surjective. The below images demonstrate two surjective functions.



A surjective mapping, which is a function.

We now introduce a special type of function, called a *bijection*. A function f which is both injective and surjective, is called a bijection. In other words, a bijection is a function between two sets X and Y where each element of both sets gets mapped directly with one and only one element of the other set. In other words, a bijection can be thought of as a relabeling of elements.

Bijections are very convenient functions to work with, in the sense that they have a particularly useful property: they are always invertible. What this means is that, if you have two sets X and Y and you know that the elements are connected via a bijective function f , if you have the element $y = f(x)$, and know the function f , you can always “work back” and find the element x which uniquely mapped to y . This is proven by the following theorem.

Theorem A: Let $f: X \rightarrow Y$ be a bijection. Then f is invertible.

Proof:

First, we must show that f^{-1} is the inverse of f :

Let $x \in X$ and $y = f(x)$
 By definition $f^{-1}(y) = x$
 By substitution $f^{-1}(f(x)) = x$
 By simplification $x = x$

Now,

Let $f: X \rightarrow Y$ be a bijection and let $y \in Y$
 Since f is surjective, there exists an element $x \in X$ such that $y = f(x)$
 Let $x = f^{-1}(y)$
 Since f is injective, we know that x is unique
 Let $y \in Y$ and $x = f^{-1}(y)$.
 By definition $f(x) = y$
 By substitution $f(f^{-1}(y)) = y$
 By simplification $y = y$

It also turns out that, if you have a function f between two sets X and Y which is invertible, then f is necessarily a bijection. This is proven by the following theorem.

Theorem B: Let $f: X \rightarrow Y$ have an inverse. Then f is a bijection.

Proof:

Let $f: X \rightarrow Y$ be invertible and have inverse $f^{-1}: Y \rightarrow X$.

We must show that f is surjective and injective.

Let $y \in Y$ and $x = f^{-1}(y)$.
 Then $f(x) = f(f^{-1}(y)) = y$
 So f is surjective

Next

Let $x_1, x_2 \in X$ such that $f(x_1) = f(x_2)$.

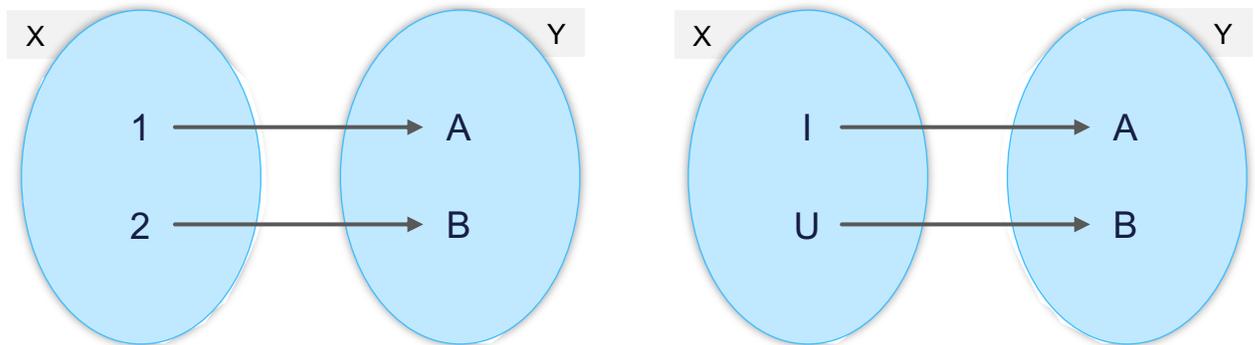
We must show that $x_1 = x_2$.

Let $y = f(x_1) = f(x_2)$
 Then $x_2 = f^{-1}(f(x_2)) = f^{-1}(y) = x$
 But also $x_1 = f^{-1}(f(x_1)) = f^{-1}(y) = x$.
 So $x_1 = x_2$ and f is injective.

The above two theorems, in combination, tell us that the concept of a function being invertible is the same concept as a function being a bijection. A function f is a bijection if and only if it is invertible. For this reason, the terms *bijective* and *invertible* for functions are, in fact, interchangeable.

However, an important thing to note is that one is only able to “reverse engineer” elements from a set Y back to the original input elements in the set X if we have (1) information on what the input set X originally is, as well as (2) information on the what the bijective function f looks like.

For example, consider the following two bijections:



In both scenarios, we have bijections from two different sets X_1 and X_2 to the set $Y = \{A, B\}$. If we only provide someone with the set Y , and no information on the bijection or X , then it is not possible to reverse engineer the elements of Y to determine which input elements were used to arrive at the elements of Y .

This is a particularly useful concept and is often used in cryptography, as one can encode elements of X into Y and sharing only the encoded elements of Y . Only those who have the necessary information, X and the bijection itself, are able to accurately reverse engineer the elements of Y to obtain the original elements of X .

Bijections have long been used in cryptography. In each case, the framework is as follows:

Suppose that person A wants to share the value x with person B confidentially. An external person C would like to discover what this value is.

A encodes x using a cryptographic algorithm that takes x and what is commonly called a key (or in some contexts an initialisation vector/value (IV)) to create y .

In some cases (symmetric encryption) the key must be treated as a secret and shared confidentially with B. In other cases (public-private key encryption) two keys are used, a public one for encrypting (that B gives to A) and a private one (that B alone holds) for decrypting. This equivalent to the operation $y = f(x)$ above.

B would use the appropriate key depending on the scenario above to obtain the original value of x . This is equivalent to the operation $x = f^{-1}(y)$ discussed earlier.

However, Person C is not able to recover x . Why?

Recall that in the earlier proofs, the function for recovering x is $x = f^{-1}(y)$. However, C is not in possession of source data, f , or of f^{-1} ("additional information" necessary to reverse protection). As a result, C is unable to compute x directly using f , nor to derive f from f^{-1} .

Note that even if they know the specifics of the cryptographic algorithm(s) used (which are often publicly published by an A to foster credibility in their data protection practices), that is at most partial information on f , as C is not in possession of the source data, keys/IVs or lookup tables necessary to have complete knowledge of f or f^{-1} .

Finally, the adoption of the various cryptographic algorithms used is grounded on security proofs predicated on validation of the absence of an efficient algorithm for determining x in the absence of knowledge necessary to reverse the protection. Here "no efficient algorithm" means that with current

or projected processing technology, it would take years, decades or even longer for a brute force attack to be successful.

Expressed another way, the only possible way for C, who is not in possession of the source data or the “additional information”, to reverse the protection is to try to try all possible values of x and all possible values for keys/IVs in the cryptographic algorithm, and for each guess, checking whether the result is y , which is not computationally feasible, by design, based on the specified length of keys/IVs. Note that this is true for any one field in the input/output dataset, and best practice is to use unique keys/IVs for each individual field.

Linking the Math to Data Embassy

Data Embassy uses these properties of functions, and the security of cryptographic algorithms in several ways.

First, for each record in a source data set, a single, random pseudonym is generated and attached to the source record. This is a pseudonym” for the record as a whole and serves as a pointer, or look-up value used when reversing the protections, and each of the elements in the record is thus mapped to the record-level pseudonym. In the language of functions, as to the individual elements in the record, this mapping is many-to-one, so it is surjective, but not injective. However, the mapping from the record as a whole to the record-level pseudonym is exactly one to exactly one and thus a bijection.

Because the record-level pseudonym is randomly generated, there is no formula or calculation available for returning from the pseudonym to the cleartext record. Instead, a Master Index (i.e., a look-up table) is created, with restricted access, which serves as the additional information (i.e., the inverse function) necessary to reverse that part of the process.

Then, a subset of the source record (“selected fields”), including the record-level pseudonym, and typically omitting at least any direct identifiers, as well as any other fields not needed for the intended use of the data in protected form (the Variant Twin). These selected fields are then mapped to corresponding fields in the output Variant Twin. This is an exactly one to exactly one mapping and thus the function that does so is a bijection.

Note, however, that this function is actually a composite of many different functions, as each field to be included in the Variant Twin has its own protection configuration. Some fields (e.g., the record level pseudonym, and numerical indirect-identifiers and attributes not suitable for conversion to categorical) are simply copied in cleartext form. In most cases, the remaining fields (quasi-identifiers, indirect identifiers and attributes that were natively or converted to categorical) are deterministically pseudonymised. Of importance, unless action is taken to expand the scope of the determinism beyond field-level, by default, every pseudonymised field will have its own unique, randomly generated key/IV.

By design, access to the source data/master index and keys/IVs is restricted via technical and organisational controls. As a result, unless authorised, recipients of a Variant Twin do not have access to them and thus have no access to, again using the mathematical vocabulary introduced above, X or f^{-1} and thus will be unable to reverse the protection directly. The only option left would be to attempt a brute force attack: for a given y , try all possible values of x , and for each x , all possible values for the key/IV and see if the result matches y . When properly implemented, in particular with regard to key/IV generation and length, (see Appendix), each of the chosen techniques has the property that

there is no “efficient algorithm” for doing so. Note that the techniques selected for use in Data Embassy are noted by [ENISA](#), and considered to be quantum computing secure.

However, an authorised user does have access to both X and f^{-1} , the additional information necessary to reverse the protection. For individual fields in the Variant Twin, including the record-level pseudonym f is a bijection, and thus always invertible to the original values in the source data set. Additionally, because of the inclusion of the record-level pseudonym, which attaches to all fields in the source, not just those selected for inclusion in the Variant Twin, access to omitted fields (in particular direct identifiers) is preserved. This extended relinking capability, which goes beyond mere reversal, is one of two key reasons why Data Embassy Variant Twins can ensure that 100% of the utility of cleartext is preserved. The second, of course, as noted earlier, is that the pseudonymised categorical values are fully compatible with advanced analytical techniques such as machine learning and AI, delivering 100% accuracy when compared to processing cleartext.

¹ Anonos systems, methods and devices are protected by a portfolio of granted international patents including, but not limited to: AUS 2018258656 (2021); US 11,030,341 (2021); CA 2,975,441 (2020); EU EP 3,063,691 (2020); US 10,572,684 (2020); CA 2,929,269 (2019); US 10,043,035 (2018); US 9,619,669 (2017); US 9,361,481 (2016); US 9,129,133 (2015); US 9,087,216 (2015); and US 9,087,215 (2015). See <https://www.anonos.com/patents> for more information.

APPENDIX - Data Embassy Pseudonymisation Algorithms

Anonos Data Embassy software pseudonymises data containing sensitive or regulated (e.g., personal) data. It enables users to process source data sets to generate two outputs:

1. A pseudonymised output (called a Variant Twin) in which case the data elements from one or more of the fields from the source data set are replaced with high entropy tokens.
2. A mapping output, called a Master Index, which contains the required information to recover the original data behind the high entropy tokens to reverse the pseudonymisation process.

Data elements can be transformed in a number of ways as an alternative to, or prior to, being replaced with pseudonyms:

1. Generalised using binning, masking, or rounding
2. Rescaling
3. Concatenation of two or more fields

Anonos Data Embassy can replace source data elements with four types of pseudonyms.

1. Reversible Deterministic
2. Reversible Non-Deterministic
3. Non-Reversible Deterministic
4. Non-Reversible Non-Deterministic

Pseudonym Type Definitions

Deterministic Pseudonyms

Deterministic pseudonyms replace recurring instances of the same data element value with the same pseudonym each time. Deterministic pseudonyms thus preserve referential integrity between pseudonyms over a defined scope. Possible examples of scope include within a single column, between two columns in a single table (e.g., country of origin and country of residence), across tables within a single database or between organisations, and across different databases.

When using deterministic pseudonyms with categorical fields, best practice is to use the narrowest scope consistent with the intended use case that preserves analytic utility, while still ensuring adequate protection against linkage and inference attacks. For example, by generalising age to age ranges using binning and then replacing each age-range value in an input data set with deterministic pseudonyms will result in a Variant Twin where it will be known that certain records have the same age-range, but not what that age-range is.

Random (Non-Deterministic) Pseudonyms

Pseudonyms that are random (non-deterministic) do not preserve consistency or referential integrity. Each recurrence of the same input value will be assigned a different unique pseudonym. The most common use is to create unique record-level pseudonyms that are used in controlled relinking. A second less common use is in the generation of test data, particularly if format preservation is used. This type of pseudonym is created using pseudorandom number generators that make use of the operating system kernel's entropy pool.

Reversibility

Pseudonyms can be reversible or non-reversible. These terms describe the options for recovering the original cleartext behind a pseudonym.

Non-reversible pseudonyms are generated using a technique called keyed hash message authentication code (Keyed HMAC). The algorithms are designed to be infeasible to reverse computationally. Instead, during the protection process, lookup tables called Master Indexes are created that map original cleartext to the generated pseudonym, which can be used to recover the cleartext when authorised.

Reversible pseudonyms are generated using symmetric encryption algorithms that take the source data element in cleartext and an encryption key as inputs to generate a ciphertext that is used as the pseudonym. The pseudonym can be directly reversed by using the key to decrypt the ciphertext.

Cryptographic Algorithms and Key Generation Details

The following table provides additional details regarding the specific algorithms used to generate each of the four types of pseudonyms.

Deterministic Reversible	Deterministic Non-Reversible	Non-Deterministic Reversible	Non-Deterministic Non-Reversible
Deterministic Authenticated Encryption using AES SIV IETF RFC 5297 with two 256-bit keys derived from PBKDF2 , one for the initialisation value and one as the encryption/decryption key.	Keyed HMAC using SHA256 with a 256-bit initialisation value derived from PBKDF2.	AES GCM using a 256-bit initialisation value derived from PBKDF2.	Java SecureRandom API in conjunction with the Operating System kernel's entropy pool.

SecureRandom Value Generation

Data Embassy uses the Java [SecureRandom](#) API coupled with the OS kernel's entropy pool in order to convert source data elements into non-deterministic, non-reversible pseudonyms that are completely unrelated to the underlying data.

The actual code that generates these values is:

```
byte[] bytes = new byte[DEFAULT_SIZE]; PRNG.nextBytes(bytes);
return Hex.encodeHexString(bytes);
```

The DEFAULT_SIZE is the size of a pseudonymous token in bytes (16) and PRNG is an instance of SecureRandom.

The Java SecureRandom class provides a cryptographically strong random number generator (RNG) which complies with the specification FIPS 140-2, as mentioned in the [official Java documentation](#).

Data Embassy uses this same strategy to generate the row-level pseudonyms for each record in the source data. Through Anonos' patented Controlled Relinking approach, authorised users can use these row-level pseudonyms to get back to the cleartext values for any field in the source data or produce a newly protected Variant Twin with new configuration values.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

**Archer, David W.; Varia, Mayank; Smart, Nigel; Malozemoff, Alex;
Darais, David; Baum, Carsten; Rosulek, Mike; Tromer, Eran; and Near,
Joe**

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

A Response to the
Request for Information on Advancing
Privacy-Enhancing Technologies
by the
Office of Science and Technology Policy

David W. Archer, PhD - Principal Scientist, Galois, Inc.
Mayank Varia - Associate Professor, Boston University
Nigel Smart - Professor, KU Leuven; Chief Academic Officer at Zama
Alex Malozemoff, PhD - Principal Researcher, Galois, Inc.
David Darais, PhD - Principal Scientist, Galois, Inc.
Carsten Baum - Assistant Professor, Aarhus University
Mike Rosulek - Associate Professor, Oregon State University
Eran Tromer - Associate Research Scientist, Columbia University
Joe Near - Assistant Professor, University of Vermont

Respondent type: Industry and Academic Collaboration

08 July 2022

Preface

In this response, we address selected questions from the OSTP RFI where we have deep expertise and long experience. Our response team regularly collaborates on PET technology research and development, addressing social and legal implications of PETs, and providing practical proofs of PET usefulness in multiple disciplines. Our viewpoints combine both industry and academia across the USA and northern Europe.

In addition to our direct responses below, we encourage OSTP to also carefully study the work of the United Nations Privacy Preserving Techniques Task Team, accessible at this website: <https://unstats.un.org/bigdata/task-teams/privacy/>. Members of our response team, particularly Dr. Archer and Prof. Varia, led development of the original UN Handbook of Privacy-Preserving Computation Techniques, initially published in 2018, and are contributing significantly to its second version due later this year. Prof. Varia is also writing a companion guide for the law profession to assess the extent to which PETs satisfy and open new affordances under privacy and data protection regulations.

1. Specific research opportunities to advance PETs:

While PETs have come a long way in the past 10 years, some significant advances must be made to bring them into practical mainstream use, for example in the Government sphere. The remaining areas that we believe need the most attention are:

- **Programmability.** Today, most PET applications are hand-coded and optimized by a small handful of expert cryptographers. This approach is analogous to the machine-code level programming used in the early days of computers, prior to the advent of high-level programming languages (HLLs) and the abstractions they provide. The few attempts at such HLLs for PETs to date have been proofs-of-concept, unsuitable to address practical problems. HLLs and their compiler toolchains for PET programming that are easy to use, yet provide abstractions useful across multiple PET families, are an area of research critical to transitioning PETs into everyday use. We note that the IARPA-funded HECTOR program, unfortunately canceled after its first year, was aimed in significant part at this concern.
- **Debuggability.** Identifying and mitigating defects during the software development process remains a critical productivity drain for all programmers. This problem is dramatically amplified for PET programming, but remains largely unsurfaced because most PET programs so far remain very simple and thus easy to reason about. While typical debugging of PET programs may fall into the realm of “just an engineering problem”, there is a deeper problem: growing insistence by organizations (especially national security activities) that formal methods in computer science be used to mathematically verify programs before adoption. Thus a salient, unaddressed area of research in PETs is the development of formal verification strategies and technical approaches for their verification.

- **Compilation tools.** Today, we rely on purpose-built compilers to transform application source code into executable code, with a distinct compiler that provides specific treatment for each PET framework. For example, compilers for linear secret sharing (LSSS) frameworks aim to minimize the number of communication rounds, while FHE compilers aim to minimize (sometimes similar, but often different) costs such as computation depth and number of bootstrappings. Compilers that successfully accommodate diverse frameworks, yet provide commonality of programmer support, error reporting, and other features is an important area for development.
- **Certification.** New cryptographic approaches are subject to certification by diverse federal authorities (for example, NSA's CyberSecurity Directorate). Approval by those authorities is often much more rapid and successful when the candidate protocols are based on well-understood, already-approved cryptographic primitives. However, nascent protocols for PETs are often developed in academia using novel primitives that do not meet that standard. Thus research in PETs that rely on standard cryptographic methods would benefit the nation in terms of getting PETs to a place where they can be readily certified or authorized to operate in mission settings.
- **PET hardware acceleration in low-power regimes.** PETs may find significant adoption at the network edge, even though current commercial focus often seems to be "in the cloud". Examples of edge processing include tactical computers at forward operating bases (FOBs) processing diverse signal sources that must be kept private; or satellite or ground-station based computation on imagery data that is sensitive, as a precursor to transmitting analytic results to mission users. Because PETs are still generally very resource-intensive, and because computational power at the network edge is always at a premium, special-purpose accelerator hardware specifically designed for low-power operation will be a critical technology to drive adoption. Research is needed on how to design such accelerators to be most efficient in both power and speed, as well as on how to divide the workload of secure computation between software on a typical host CPU and hardware acceleration. We point to the DARPA DPRIVE program as a first effort in this direction.

Today, certain PETs are showing more promise than others. We believe that the most promising PET technologies today are the following:

- Private set intersection (PSI) with (generally statistics-based) computation on the resulting intersection. Statistical analysis across sensitive data held by multiple Government activities or agencies is a growing need, called for in multiple legislative initiatives as well as the Federal Data Strategy. However, the conundrum of how to share such data while respecting the need for its confidentiality continues to confound agencies. PSI offers the most efficient sensitive data linkage approach available, while providing cryptographic protections over the data being shared. We point for example to a recent pilot project at the National Center for Educational Statistics, conducted as a collaboration between Galois, Inc. and Georgetown University, and using PSI technology from Galois to operate at full scale for nationwide data.
- Use of multi-party computation, and in particular linear secret sharing, for example to secure sensitive cryptographic artifacts such as signing keys for digital signature protocol

(including post-quantum signature schemes) is also a growing need with promising PET support. At present, the NIST candidate post-quantum signature schemes such as Crystals/Dilithium do not have such *threshold signing* capability. However, promising prototypes are under construction now, and early indications suggest that performance may be on par with mission-scale throughput.

- Interactive zero knowledge proof protocols for complex proofs such as software cybersecurity look particularly promising. Current leading edge work demonstrates the ability to prove the existence of vulnerabilities such as the OpenSSL Heartbleed weakness, and upcoming enhancements will offer proof of key properties of software such as *memory safety*. Non-interactive ZK proofs that verify correct computation while keeping the inputs to and results of the computation private is another area where ZK PETs are promising, with some relatively-simple instances already deployed in financial-technology applications.
- Fully homomorphic encryption has entered early deployment in a few places, for well-structured applications. As called out above, programmability and debuggability still must be addressed, but FHE (especially with hardware acceleration as needed) appears to be on track to address well-chosen applications.

2. Specific technical aspects or limitations of PETs:

We highlight two technical limitations of PETs:

- First, PETs necessarily and purposefully inhibit the ability to perform manual cleaning and vetting of sensitive input data, because those inputs are by nature encrypted, and because secure computation algorithms for cleaning data are too complex to be computationally efficient. This problem necessitates changes in the data cycle workflow when using PETs. That said, we emphasize that two common techniques remain viable: performing data cleaning at the source *prior* to encryption and subsequent inclusion within a PET-enabled workflow; and performing automated techniques such as outlier detection and data reliability testing with PET protections applied to these techniques along with the rest of the analysis. The latter leverages the expressive power and programmability of these PETs.
- Second, PET statistical disclosure techniques such as *differential privacy* provide strong privacy benefits by preventing computation outputs from revealing input data, at the expense of adding distortion to those outputs. Here, we wish to emphasize that the relaxation of perfect accuracy is inherent to the task of providing output privacy, rather than being a specific flaw of differential privacy. Other attempts at disclosure limitation based on suppression or the use of quasi-identifiers have consistently been shown to be re-identifiable; recent examples include <https://arxiv.org/pdf/2202.13470.pdf> and <https://queue.acm.org/detail.cfm?ref=rss&id=3295691>. Additionally, privacy budgets for differential privacy techniques that might appear to be excessive due to their worst-case nature have increasingly been found to have matching attack bounds that are viable to execute in practice.

3. Specific sectors, applications, or types of analysis that would particularly benefit from using PETs

The classes of applications that particularly benefit from PETs are significantly influenced by the performance limitations of PETs today. In particular, (1) the analysis used in the application should occur on a recurring but infrequent basis; (2) The computation itself should be rather straight-forward (e.g., simple statistics vs. training a neural net), if all the data were actually present & centralized. The challenge should not be the computation but rather the inability to collect all relevant data centrally; and (3) participants should be willing to accept some delay in receiving a response. While there do exist high performance and real-time applications for PETs, the reality is that most potential PET applications today would over-stress current data processing systems. However, we point to several applications that meet the criteria above and are thus practical today:

- **Government statistics reporting across multiple organizations.** As mentioned above, OSTP might refer to a report on using PETs in a prototype for the US Department of Education here: <https://mccourt.georgetown.edu/news/a-federal-government-privacy-preserving-technology-demonstration/>
- **De-confliction of computer network resource use in law enforcement operations.** As described above, *private set intersection* is a particularly promising PET. Securely de-conflicting the use of network vantage points (that is, computers) from which diverse law enforcement agencies monitor and mitigate cyber-crime is a natural fit for such technology.
- **High-security digital signatures,** particularly for national security use. Highly sensitive digital signature keys or other encryption keys should never be materialized in one place. Secure multi-party computation is currently being explored as a practical way to materialize such keys in a secure form that is immune to exfiltration.
- **Cross-border financial crime detection.** The finance industry is often an early adopter of new technology, and emerging statutes such as GDPR make cross-border data sharing impossible. PETs appear to be a natural fit for solving this conundrum. Significant work is already being done in this space.
- **Distinguishing perpetrators from informants.** At the recent Theory and Practice of Multi-Party Computation conference, the firm Roseman Labs provided a nice example: Law enforcement personnel are interested in finding perpetrators of human trafficking, but distinguishing them from known trafficking victims who are in contact with an NGO. A form of private set intersection was used to combine lists of suspects and such victims, across multiple organizations.
- **Enabling pathways to justice for survivors of sexual assault and harrassment.** Survivors of sexual assault and harrassment are often very hesitant to come forward and report such incidents. At the same time, perpetrators are well known for exploiting legal channels to suppress claims of assault and abuse. Research has found that equities are dramatically improved, and survivors are more willing to come forward and take action if they know that others have been victimized by the same perpetrator. Keeping survivor

data private and immune to inappropriate perpetrator suppression is a natural fit for PETs. We encourage OSTP to refer to the work at <https://www.mycallisto.org/> . In particular, we refer the reader to this paper:

<https://www.projectcallisto.org/callisto-cryptographic-approach.pdf>.

Dr. Archer and Prof. Varia worked on development of the PETs used there, and served on the cryptography advisory board for that organization.

- **Privacy-preserving digital assets.** Financial privacy is crucial to emerging digital finance systems, such as Central Bank Digital Currencies (CBCD). Indeed, the Federal Reserve opined that “Protecting consumer privacy is critical. Any CBDC would need to strike an appropriate balance, however, between safeguarding the privacy rights of consumers and affording the transparency necessary to deter criminal activity.”

<https://www.federalreserve.gov/publications/files/money-and-payments-20220120.pdf>

Zero-knowledge proof systems offer a powerful approach for achieving these aims in blockchain-based systems, as explored by Zerocash (<http://zerocash-project.org>) and subsequent operational deployments and enhancements.

- **Robustness to supply-chain threats.** Modern IT systems rely on the integrity of their computational components, which may be compromised by supply chain compromise, intrusion, or faults. Zero-knowledge proofs offer a way to assure the integrity of the final output, even if underlying computer or network components are compromised — while protecting the confidentiality of sensitive information implicated in the derivation of that data. Current realizations still entail a high overhead, but show great promise for use in high-assurance critical applications.
- **Data integrity.** A related goal is assuring the integrity of *data* which may have been manipulated in undesired ways. For example, ensuring the truthfulness of images is a major challenge, especially given the evolution of Deep Fake capabilities. Academic technology-demonstrators show that zero-knowledge proofs can assist in vetting authenticity of image files (<https://www.cs.tau.ac.il/~tromer/photoproof>), and are ripe for extension to other media types and concrete applications.

We also encourage OSTP to engage with the Department of Homeland Security and obtain the presentations given at the recent PETS4HSE (PETs for the Homeland Security Enterprise) workshop there. Dr. Archer’s presentation provided several examples of practical PET use suitable for DHS.

6. Specific mechanisms, not covered above, that could be used to advance PETs:

We emphasize strongly that the way to put PETs into practice is to demonstrate what they can do – what new capabilities are achievable with PETs that were not achievable before. History in the realm of cyber security shows clearly that bringing security techniques to bear on current solutions - “sprinkling security over the top” - does not lead to new adoption. Indeed, if a capability is in use today, even insecurely, users are far more likely to accept the current risks and continue, rather than change to a secure mechanism. (This is an unfortunate state of

affairs, because attack patterns always improve over time, making current mechanisms progressively less secure unless protected with cryptographically provable security. We refer for example to the Census Bureau's own example attack on their own previous de-identification technology). Thus what's important for PET adoption is to *prove out new ideas*, focusing on those that are firmly grounded in the needs of practical, real-world use. To that end, we recommend the use of grants to the States to develop PET-based novel solutions that benefit their citizens and that come with firm demonstration metrics for privacy. Statistical applications, such as for example analysis of how life-long educational outcomes correlate to economic wellbeing, may be a good place to start.

9. Barriers, not covered above, to PET adoption:

As with any technology that depends on data sharing, a key barrier to adoption is the *willingness* to share data, whether secure or not. That is, PETs can come into their own only when two or more parties agree to pool data to do something new. We recommend that OSTP review the Lunar Ventures report on the potential business upside of such sharing. The problem is that such coming together to even discuss sharing is hard for companies, or due to anti-trust regulations may be deemed illegal, and in any case struggles against the corporate mindset of protecting intellectual property. Getting over this initial resistance for enough commercial use cases is a salient barrier to wide-spread adoption of PETs commercially. Similar barriers appear in the Government sphere.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Argonne National Laboratory (ANL)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to RFI on Privacy-Enhancing Technologies

Prepared by

Kibaek Kim (Computational Mathematician)

Alec Poczatek (Cybersecurity Analyst)

Dan Harkness (Interim Group Leader)

Minseouk Ryu (Postdoctoral Researcher)

Ravi Madduri (Computer Scientist)

Submitted on July 8, 2022



Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL,
60439

Introduction

About Argonne National Laboratory

Argonne National Laboratory (ANL), one of the U.S. Department of Energy (DOE) national laboratories for science and engineering research, is managed for DOE by UChicago Argonne, LLC. Argonne's mission is to apply a unique mix of world-class science, engineering, and user facilities to deliver innovative research and technologies. Research at Argonne includes energy, biological and environmental systems, advanced computing, and national security. In the following paragraphs, we describe some of the unique capabilities and facilities at Argonne that would help advance the research and development of PETs.

In 2019, Argonne National Laboratory began establishing an AI testbed with following goals: (1) to provide an open and unbiased environment for the evaluation of emerging AI accelerator technologies designed to accelerate training and inference for deep learning models; (2) to document and make available to others information about use cases, software compatibility, software integration requirements, and realized performance on a variety of test problems; and (3) to support collaborations with AI technology developers, academic computer science and data science departments, commercial sector players in AI, and the DOE laboratories. The testbed will work with AI technology companies to make their systems available to the academic, laboratory, and commercial AI developer communities. The specific systems to be deployed in the testbed are not yet known but are expected to include systems such as the Graphcore IPU, Wave Computing DPU, Google Edge TPU, and Loihi neuromorphic chip.

In 2021, Argonne implemented the necessary policies, procedures, controls, and systems to support the analysis of human subject data, called the Argonne Biomedical Learning Environment (ABLE). ABLE is built on Argonne's moderate enclave that implements NIST 800-53 controls with a separation of duties between administration and research activities. When completed, the enclave will have the necessary privacy controls implemented into all processes providing compliance with HIPAA and FISMA requirements for handling protected health information and personally identifiable information. The ABLE environment at Argonne consists of several high-level resources including cutting-edge hardware for deep learning and machine learning.

In 2022, installation of Aurora, a new Intel-Cray exascale computer, began at the Argonne Leadership Computing Facility (ALCF). When fully deployed, Aurora is projected to have a peak performance of more than 2 exaflops, with over 10,000 nodes and over 10 petabytes of aggregate system memory. The revolutionary architecture will support large-scale simulation (PDEs, traditional HPC), data-intensive applications (scalable science pipelines), and deep learning and emerging science AI (training and inferencing

Argonne Responses

In this section we provide responses to specific topics in the RFI. We use the topic description followed by Argonne's response.

1. **Specific research opportunities to advance PETs: Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.**

Privacy-preserving federated learning (PPFL)

Federated learning (FL) enables training a machine learning model from distributed data sources without collecting the data to a central location. Specifically, FL is capable of training a model by sharing not the raw data but the processed results (e.g., gradients of the loss function, local model parameters, or hidden representation of the neural network models) between a server and clients (i.e., data owners). This capability is favorable not only to data owners with privacy concerns but also to data analysts who need more data for better learning. It is beneficial especially when the data owners cannot transfer data to the central server because of privacy concerns of citizens or legal frameworks (e.g., medical data) or in areas with national security interests. FL itself, however, cannot guarantee data privacy because the intermediate results communicated during the FL process can be utilized to estimate the local data (e.g., inference attack [1]). See Figure 1 generated by the authors for the report to the DOE Advanced Scientific Computing Advisory Committee Meeting in March 2022. This situation thus calls for the development of privacy-preserving FL (PPFL) that integrates privacy-preserving techniques in FL, thereby opening new research directions in the FL community.

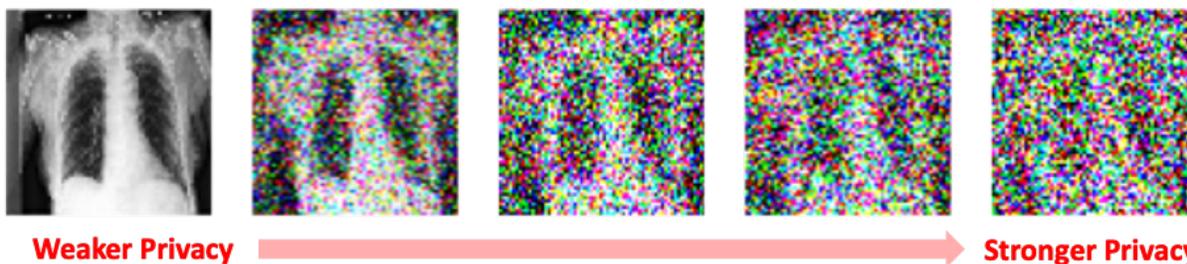


Figure 1. Reconstruction of chest X-ray images from model weights by leveraging the communication during FL with weaker privacy (left) and stronger privacy (right). [23]

PPFL is a recent advance in FL that aims to ensure data privacy by introducing privacy-preservation techniques such as differential privacy, homomorphic encryption, and secure computations. Differential privacy, the state-of-the-art privacy technique for quantifying and limiting information disclosure by random perturbation that is supported by theoretical justifications, has been widely adopted for PPFL mainly because of its computational efficiency compared with other privacy-preserving techniques. For example in [12], the computational efficiency of differential privacy is compared with homomorphic encryption. Differentially private FL, however, encounters inevitable trade-offs between data privacy and learning performance (i.e., more noise added for ensuring stronger data privacy downgrades the learning performance). Needed, therefore, are novel PPFL algorithms that provide higher accuracy under strong data privacy. For example, the inexact alternating direction method of multipliers algorithm developed in our recent work [17] provides the same level of differential privacy as the state-of-the-art PPFL algorithm but outperforms it by exploiting the objective perturbation for differential privacy and multiple local update techniques. *Among the research directions in PPFL, critical are the development of advanced training algorithms and communications with differential privacy and other privacy-preserving techniques.*

More challenges and opportunities exist and arise from different settings of FL, each of which further complicates privacy-preserving procedures. First, data distributed across the clients can be unbalanced and not identically and independently distributed (IID). Non-IID data can make the model training challenging. For example, in the cross-device setting (e.g., millions of sensors and edge devices), a subset of clients may be sampled at every iteration of a global update, resulting in the underlying distributions changing at every step of the algorithm. Another challenge is the communication required for training a model across the clients. On the one hand, FL with data distributed in a few large silos (i.e., cross-silo FL) may experience computational load imbalance across the clients, causing some local model trains to take more time than others. On the other hand, in the case of cross-device FL, a large number of devices may suffer from communication bottlenecks with limited network capacities. Moreover, in both cross-silo and cross-device settings, some clients may be temporarily unavailable or fail during the training. Efficient communication between a server and clients is also a key challenge in PPFL as more communication rounds could increase the risk of data leakage because more intermediate results can be utilized for the reverse-engineering process for estimating the locally stored data. Therefore, advanced communication strategy (e.g., asynchronous update, network topology, and compression) needs to be explored.

PPFL on multimodal datasets

Data modality will also lead to different settings of FL. Depending on the distributed data modality, FL can also be categorized as (i) horizontal FL (HFL) [2,3,4,5,6,7] or (ii) vertical FL (VFL) [8,9,10,11,12]. In the HFL setting, every client shares the same data features but different data samples, which can be considered as a data matrix (i.e., data samples and features are

represented by rows and columns, respectively) split horizontally and distributed over multiple clients. In the VFL setting, every client shares the same data samples but different data features, which can be considered as the data matrix split vertically and distributed over multiple clients. The key difference between HFL and VFL is that HFL is limited to unimodal data, whereas VFL is applicable to multimodal data. We observe that VFL has been underexplored, whereas HFL has been actively studied in both academics and practice. Moreover, from the privacy-preserving perspective, we expect that achieving data privacy will be more challenging with VFL than with HFL, mainly because of the expensive and complicated communication required for multimodal data in VFL. Moreover, most FL has been developed for supervised learning. In many applications (e.g., anomaly detection using satellite data [13,14] and feature training from multimodal data [15,16]), data labels are often missing, calling for developing FL algorithms for unsupervised HFL and VFL.

Trusted AI using PETs

Trusted AI models are resilient to model shift and preserve the privacy of data used in training. To enable the creation of trusted AI in PETs, advances are needed in building trusted data repositories and cyber infrastructure practices that enable continuous training of models resilient to bias and ensure fairness to improve the overall confidence of the general public. Achieving these advances requires the development of well-labeled (also known as AI-ready) data, managed by using FAIR (Findability, Accessibility, Interoperability, and Reproducibility) principles along with well-documented APIs to securely access the data. Publicly available AI-ready datasets are important in developing baseline models that can later be used in implementing continuous training and validation of AI models. These, coupled with the implementation of a “trusted model repository,” can provide AI practitioners details on provenance and training data distributions, along with model parameters that are useful in assessing the quality and maturity of the model.

Data and model shifts in AI models dent the applicability of AI in real-world settings. Models are better performing when they are trained from the real world and are continuously updated to accurately capture changes in data distribution. However, real-world data in many domains is often hard to access because of privacy concerns and reidentification attacks. PPFL offers a framework to address the data and model shifts in AI models, especially when trained on real-world data with sufficient privacy guarantees. ***We envision the creation of a continuously updated, trusted AI model repository trained by using disparate private data, leveraging PPFL technologies to mitigate bias and improve confidence in the applicability of AI models to the real world.***

Continuous training and validation of AI models using PPFL technologies require a fair amount of computational power depending on the privacy-budget allocated to the differential privacy algorithm. Stronger privacy protection means a larger privacy budget and hence more computational power. ***We envision leveraging DOE’s supercomputing facilities to provide the computing required to create and maintain the trusted model repository.*** Frameworks such as Parsl [24], FuncX [25], and Flux [26] that support for distributed computational

workloads exist but need to be integrated with PPFL frameworks to fulfill the vision. PPFL frameworks should be integrated in IAM (identity and access management) systems to leverage secure data transfer technologies and ensure that large-sized models and datasets can be transferred reliably.

DOE runs the world's fastest supercomputers and more recently has set up private enclaves that are certified to host private data securely. Leveraging PETs to build trustworthy AI models that are robust and up to date requires a comprehensive redesign of data management and computational strategies to create an ecosystem of capabilities. **We envision the creation of a comprehensive framework that evaluates models trained using PPFL technologies against various threats and attack models.**

Confidential computing on leadership-class systems

Confidential computing uses hardware-based techniques to isolate data, specific functions, or an entire application from the operating system, hypervisor or virtual machine manager, and other privileged processes. Data is stored in the trusted execution environment (TEE), where it is impossible to view the data or operations performed on it from outside, even with a debugger. The TEE ensures that only authorized code can access the data. If the code is altered or tampered with, the TEE denies the operation. Technologies such as HPCrypt [27] allow creation of secure private enclaves on supercomputing resources using cryptographic technologies that are integrated into policies that govern execution on supercomputers. To enable confidential computing on leadership-class computing facilities, **we envision integration of differential privacy with privacy-enhancing cryptography techniques** such as secure multiparty computation, private set intersection, private information retrieval, zero-knowledge proofs, and fully homomorphic encryption with TEEs.

2. Specific technical aspects or limitations of PETs: Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

Evaluation of existing PPFL packages

A few open-source FL frameworks exist. These include Argonne Privacy-Preserving Federated Learning (APPFL) [18], Open Federated Learning (OpenFL) [19], Federated Machine Learning (FedML) [20], TensorFlow Federated (TFF) [21], and PySyft [22]. The existing software packages have been developed to implement various FL algorithms with different

communication protocols while leveraging the existing ML tools such as PyTorch and TensorFlow. Some packages also implement privacy-preserving techniques such as differential privacy. Here we briefly discuss the capabilities of each framework.

APPFL: This package has been developed by Argonne National Laboratory via multiple DOE-ASCR projects. It provides PPFL tools for users in practice while allowing research communities to implement, test, and validate various ideas for PPFL. The current implementation includes a number of training algorithms with differential privacy. The package supports the PyTorch data and model objects and the communication via MPI or gRPC. The package was highlighted at the DOE Advanced Scientific Computing Advisory Committee Meeting in March 2022.

OpenFL: This framework has been developed by Intel. It was initially developed as part of a research project on FL for healthcare and designed for a multi-institutional setting. In OpenFL, an environment is constructed based on collaborator and aggregator nodes that form a star topology; in other words, all collaborator nodes are connected to an aggregator node. Communication between nodes is through gRPC via a mutually authenticated transport layer security network connection. However, OpenFL does not support different communication protocols other than gRPC which limits adoption in heterogenous computing environments.

FedML: This is an open research library to facilitate FL algorithm development and fair performance comparison. It supports on-device training for edge devices, distributed computing, and single-machine simulation. It utilizes gRPC and MQTT for device communication to simulate cross-device FL on real-world hardware platforms. Also, it utilizes MPI for simulating FL in a distributed-computing setting. It implements weak differential privacy that aims to prevent a backdoor attack, which requires less noise in training data compared with what is required for ensuring data privacy. However, FedML does not support adding custom differential privacy algorithms that limits its applicability to track and adopt the latest developments in the space.

TFF: This framework is available from Google for machine learning and other computations on decentralized data. In TFF, an FL environment is constructed by using multiple GPUs that are used as clients. Also, TFF can be simulated on a Google Cloud platform. Currently, TFF supports FedAvg and differential privacy for private federated learning. TFF only supports gRPC-based communication protocols that limit the applicability in heterogenous computing environments.

PySyft: This framework is available from OpenMined, an open-source community. In PySyft, an FL environment is constructed by Virtual Workers, WebSocket Workers, or GridNodes. While Virtual Workers live on the same machine and do not communicate over the network, the others leverage WebSocket as a communication medium to ensure that a broad range of devices can participate in a PySyft network. Currently, PySyft supports FedAvg and differential privacy for private federated learning.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs: Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of Executive Order 14058: Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government.

Sectors and application domains

Biomedical applications. In biomedicine, data fusion and centralized AI analyses are difficult to realize because of privacy concerns and reidentification attacks. This situation has resulted in multimodal biomedical datasets governed by federal regulations and consortium-specific data usage agreements that have made centralized data collection and analysis difficult. PPFL provides a great opportunity to unlock biomedical insights from multimodal data. Applications include building better epidemiological models by combining multimodal private data from different healthcare providers across the international borders, for example for early detection of pandemics and better risk prediction models for various diseases across different populations. The applications of PETs, especially PPFL, have the potential to create robust AI models [28].

DOE user facilities. DOE's Basic Energy Sciences (BES) program has identified its first priority research opportunity as follows: "Efficiently extract critical and strategic information from large, complex datasets at BES's scientific user facilities." User facilities such as Argonne's Advanced Photon Source (APS) are valuable resources that generate multimodal data for numerous different types of use cases. Some of these cases involve Controlled Unclassified Information acquired in support of projects related to NNSA, DOD, and other agencies. However, sharing raw data across geographically separated instruments, especially across institutions, is challenging; and there are concerns (e.g., citizen privacy, legal frameworks for medical data) that those experiments could not make full use of advances in AI/ML or in areas with national security interests. If raw data can stay in place by sharing only the processed results, it can make more information beneficial to larger communities without explicitly sharing any sensitive data.

Critical Infrastructure. The electric power grid is at the heart of a complex system of interdependent critical infrastructures. It represents many of the common characteristics and challenges of complex engineering systems. Any disruption to the system can lead to dire consequences for the functioning of society and the economy. The electric power grid is undergoing a modernization currently pushed by 100% clean energy targets and enhanced connectivity through distributed sensing and control devices (e.g., PMUs, smart meters). Future

grid operation is anticipated to be distributed/decentralized, as opposed to the existing centralized operations, by utilizing a large amount of distributed data available at the edges of the network. However, such data collected by edge devices may not be available to the central operator in real time because of the large volume of the datasets and data privacy. At the same time, the electric grid is facing complex operational challenges due to increasing threats from natural disasters (e.g., hurricanes, wildfires) and human-initiated cyber incidents. In this context, an important challenge for complex engineering systems is data-driven decision making for resilient operation by accounting for the distributed/decentralized monitoring and control. While potentially reducing the impacts of single-point failures, distributed or decentralized operations introduce new challenges associated with secure data integration from different sources. These challenges need to be addressed by an AI-driven/assisted monitoring and control system with the overarching constraints of data privacy and security.

Cybersecurity federated models. Cybersecurity is often seen as a single-organization-centric model. However, adversaries often target multiple organizations using the same tactics, techniques, and procedures. Cybersecurity federated models (e.g., Argonne's CFM) are community-based security models to promote a global defense against common threats within large heterogeneous and distributed organizations. Many of these models have already approached privacy-enhancement through anonymization and obfuscation (removing identifying information) or through encryption (limiting who can view the information). Anonymization and obfuscation techniques tend to overscrub information to minimize the risk of missing identifiable elements. Encryption does not actually fix the problem but instead just limits the exposure to those trusted to decrypt it. Cybersecurity federated models hold much promise for moving beyond simple information sharing to instead empowering collaborative analysis. Such capabilities, however, will require better privacy-enhancing techniques. With the right assurances of privacy, the future of cybersecurity federated models could include distributed analytics at the edge; distributed query and response capabilities, allowing trusted analysts to locate relevant data across their peers; and federated learning to improve artificial intelligence for cybersecurity.

Cybersecurity improvements with user-behavior. A significant effort within the U.S. government to push toward tighter security models such as zero trust. This direction requires that security architectures have sufficient data to determine whether any action is a malicious one in order to prevent exploits from occurring in near-real time. Preservation of privacy is imperative in this regard, making data difficult to obtain as de-anonymization becomes easier with more available data points to analyze.

Many cybersecurity challenges also stem from the need to be able to de-anonymize data in the event of legal requirements where information must be used in investigations or other scenarios in which the digital information is relevant. This data could be the only evidence of data breaches, data manipulation, or other malicious activity, making privacy protections that much harder to balance.

Data collection such as packet captures of internet traffic in an organization also stand to be a valuable resource in detecting malicious behavior. But because such packet captures contain all

data sent by a given user as well as the data received by that user, they pose a significant potential for privacy issues. Indeed, the captured data can include passwords, personally identifiable information, email contents, schedules, and myriad other personal data. It also may include company intellectual property and any other type of data that gets sent across a network. This type of data is one of the hardest types of datasets for cybersecurity researchers to obtain, since very few datasets have been curated enough to ensure that none of this privacy infringing information is contained within it.

Fingerprinting, or tracking users based upon correlating factors on how they interact with computer systems, is also a common technique utilized by cybersecurity operations teams to identify whether the user is a known threat. Implementing user fingerprinting is utilized heavily in advertising agencies as well to try to sell people products as more information is gathered about their personal interests and spending habits. The same technology utilized by these advertising agencies is used to determine whether users are attempting to access data that they shouldn't be. The aggregated data gathered about these users can allow for the identification of a specific person. Here PETs can assist with ensuring that user privacy is maintained while still ensuring that threats are identified.

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs: This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

The Privacy Group under the Cybersecurity Priority area of the National Institute for Standards and Technology has developed comprehensive guidelines and evaluated tools developed by academia and leading internet companies that can be incorporated in various application domains.

References

- [1] Shokri, Reza, et al. "Membership inference attacks against machine learning models." *2017 IEEE symposium on security and privacy (SP)*. IEEE, 2017.
- [2] Konečný, Jakub, et al. "Federated learning: Strategies for improving communication efficiency." *arXiv preprint arXiv:1610.05492* (2016).
- [3] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." *Artificial intelligence and statistics*. PMLR, 2017.
- [4] Wei, Kang, et al. "Federated learning with differential privacy: Algorithms and performance analysis." *IEEE Transactions on Information Forensics and Security* 15 (2020): 3454-3469.
- [5] Truex, Stacey, et al. "LDP-Fed: Federated learning with local differential privacy." *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*. 2020.
- [6] Huang, Zonghao, et al. "DP-ADMM: ADMM-based distributed learning with differential privacy." *IEEE Transactions on Information Forensics and Security* 15 (2019): 1002-1012.
- [7] Agarwal, Naman, et al. "cpSGD: Communication-efficient and differentially-private distributed SGD." *Advances in Neural Information Processing Systems* 31 (2018).
- [8] Wei, Kang, et al. "Vertical Federated Learning: Challenges, Methodologies and Experiments." *arXiv preprint arXiv:2202.04309* (2022).
- [9] Chen, Tianyi, et al. "Vafl: a method of vertical asynchronous federated learning." *arXiv preprint arXiv:2007.06081* (2020).
- [10] Romanini, Daniele, et al. "PyVertical: A vertical federated learning framework for multi-headed SplitNN." *arXiv preprint arXiv:2104.00489* (2021).
- [11] Sun, Jiankai, et al. "Vertical federated learning without revealing intersection membership." *arXiv preprint arXiv:2106.05508* (2021).
- [12] Wang, Chang, et al. "Hybrid differentially private federated learning on vertically partitioned data." *arXiv preprint arXiv:2009.02763* (2020).
- [13] Liu, Yi, et al. "Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach." *IEEE Internet of Things Journal* 8.8 (2020): 6348-6358.
- [14] Zhang, Kai, et al. "Federated variational learning for anomaly detection in multivariate time series." *2021 IEEE International Performance, Computing, and Communications Conference (IPCCC)*. IEEE, 2021.
- [15] Ngiam, Jiquan, et al. "Multimodal deep learning." *ICML*. 2011.
- [16] Zhao, Yuchen, Payam Barnaghi, and Hamed Haddadi. "Multimodal Federated Learning on IoT Data." *2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 2022.
- [17] Ryu, Minseok, and Kibaek Kim. "Differentially private federated learning via inexact ADMM with multiple local updates." *arXiv preprint arXiv:2202.09409* (2022).
- [18] Ryu, Minseok, et al. "APPFL: Open-Source Software Framework for Privacy-Preserving Federated Learning." *In: Proceedings of the 36th IPDPS 2022 Workshops (to appear)*. 2022.
- [19] Reina, G. Anthony, et al. "OpenFL: An open-source framework for Federated Learning." *arXiv preprint arXiv:2105.06413*(2021).

- [20] He, Chaoyang, et al. "FedML: A research library and benchmark for federated machine learning." *arXiv preprint arXiv:2007.13518* (2020).
- [21] Bonawitz, K., H. Eichner, and W. Grieskamp. "TensorFlow federated: machine learning on decentralized data.(2020)."
- [22] Ziller, Alexander, et al. "Pysyft: A library for easy federated learning." *Federated Learning Systems*. Springer, Cham, 2021. 111-139.
- [23] Helland, Barbara. "View from Germantown: Advanced Scientific Computing Research Update". U.S. Department of Energy, Office of Science.
https://science.osti.gov/-/media/ascr/ascac/pdf/meetings/202203/ASCAC_202203_Presentation_Helland.pdf Accessed: 2022-07-07
- [24] Y. Babuji, et al. 2019. Parsl: Pervasive Parallel Programming in Python. In 28th International Symposium on High-Performance Parallel and Distributed Computing (HPDC'19). ACM, 25--36.
- [25] Ryan Chard, et al. 2020. FuncX: A Federated Function Serving Fabric for Science. HPDC '20. Association for Computing Machinery, New York, NY, USA, 65--76.
- [26] Dong H. Ahn, et al, "Flux: Overcoming Scheduling Challenges for Exascale Workflows", Future Generation Computer Systems, Volume 110, 2020, Pages 202-213
- [27] HPCrypt Data Protection System, developed by the Lawrence Livermore National Laboratory, <https://ipo.llnl.gov/technologies/hpcrypt>
- [28] The clinician and dataset shift in artificial intelligence. The New England Journal of Medicine, 2021

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Arm

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



Office of Science and Technology Policy

“Notice of request for information on Advancing Privacy-Enhancing Technologies”

Document Number 2022-12432

8 July 2022

On behalf of Arm¹, please find comments to the consultation on advancing privacy-enhancing technologies, or “PETs”. As the leading global provider of semiconductor intellectual property (“IP”) and supporting technology, Arm is at the forefront of developing the types of technologies discussed in the consultation and fully supports the U.S. Government and the Office of Science and Technology Policy promoting and supporting accelerated development, deployment and adoption of PETs.

Over the years, Arm has developed and licensed a number of technologies that have been used to provide protection of data in use from even privileged software. TrustZone™² has been in the market in the form of central processing unit (CPU) IP since 2004 and can protect a set of fixed software payloads from the operating systems and hypervisors. TrustZone is used on billions of devices today to provide use cases such as mobile device management, payment, and user authentication through biometrics. TrustZone is designed to allow device vendors to implement security use cases associated with the platform, these are limited in number and so in general this is also true for the amount of memory that can be used by TrustZone workloads. Arm has also provided support for Virtualization in CPU IP since 2011. Virtualization is often used in deployments to protect security use cases from a primary operating system (OS) kernel. With virtualization there is no specific limit on the amount of memory resource that can be used by a secure workload, that is shielded from a primary OS kernel, but both the hypervisor and code in TrustZone need to be trusted by such a secure workload. This might not be practical in deployments where the provider does not have full control of the stack or where those pieces of software are large and complex.

In June 2021, Arm made a significant announcement for PETs with the launch of technical specifications for Arm Confidential Compute Architecture (Arm CCA). This combines the properties of TrustZone and virtualization, and provides trusted execution environments (TEE), called Realms, which are protected from supervisory code such as hypervisors or kernels, as well as from TrustZone code, or other Realms.³ As is the case with virtual machines, Realms are not limited on the amount of memory they can use.

To further contribute to the development and adoption of PETs, Arm is also a Premier Member of the Confidential Computing Consortium, under the Linux Foundation, which is the leading group developing open source, and open standards and specifications focused on Confidential Computing and specifically the area of “data in use”. This was the key missing piece to

¹ See [Defining the Future of Computing – Arm®](#)

² See, for instance [TrustZone for Cortex-A – Arm®](#)

³ For a more complete discussion on CCA, see [Unlocking the power of data with Arm CCA - Architectures and Processors blog - Arm Community blogs - Arm Community](#)

significantly enhancing security and privacy, as protection methods have existed for “data at rest” and “data in transit” and work continues to strengthen those two distinct areas.

Protecting data while in use was difficult as it required silicon, software, and other technology providers to coalesce around significant changes to underlying computing architecture, but the results will create a more secure digital world. As Arm’s Senior Vice President and Chief Architect, Richard Grisenthwaite, described in a blog on the announcement of Arm CCA:

Currently, applications and virtual machines place huge amounts of trust in the supervisor software (kernels or hypervisors) that manage them. Supervisors can access the resources used by applications for their program code and data. Exploits against supervisors can therefore leak confidential data or algorithms held in the applications. Confidential computing changes the traditional trust relationship between applications and supervisors by removing the supervisor’s right to access the resources used by the application, while retaining the right to manage them. Removing that right of access is critical because the devices we use today handle large quantities of confidential data. Cloud systems can be running payloads from many different customers, while mobile devices can contain both personal and business information, from medical data to company emails. Confidential computing reduces the need to trust unseen technology within any compute environment. Arm CCA extends workload isolation to enable a provider to shift from a position where service providers **will not** access customer data, to one where they **cannot** access customer data - thereby reducing the volume of software that must be trusted, the attack surface for hackers, and the potential for customer data or algorithm breaches. Arm CCA introduces a new kind of confidential computing environment called a Realm, which protects the data and code, even in use.⁴

And as Richard goes on to conclude: “Our vision for Arm CCA is to protect all data and code wherever computing happens, while empowering developers to implement strong privacy controls...”

⁴ See [Arm CCA will put confidential compute in the hands of every developer – Arm®](#)

In response to the specific questions from the consultation, please find Arm's responses below:

1. *Specific research opportunities to advance PETs: Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.*

Arm response

The Department of Defense (DoD), Defense Advanced Research Project Agency (DARPA) has funded work into a technology areas known as “fully homomorphic encryption” or “FHE” through a project named Data Protection in Virtual Environments (DPRIVE).⁵ This technology area is one of the most promising ways of doing privacy-enhancing analytics on extremely sensitive data sets, and will have significant application across DoD and commercial applications. Four teams of researchers have been selected to perform this work and Arm are acting as a subcontractor to the SRI International team.⁶ DPRIVE and similar projects should be a priority for the USG to continue funding, particularly if additional appropriations are made for technology research through the U.S. Innovation and Competitiveness Act (USICA) or through the potential increased funding allocations Congress is considering in the Fiscal Year 2023 National Defense Authorization Act.

3. *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs: Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of [Executive Order 14058: Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government](#).*

Arm response

Cloud will be the first area of computing where PETs will be deployed in scale. This has the potential to drive significant new use cases from sensitive workloads such as medical and

⁵ See [Data Protection in Virtual Environments \(darpa.mil\)](#)

⁶ See [DARPA Selects Researchers to Accelerate Use of Fully Homomorphic Encryption](#)

financial services which are currently done on premises out of concern for potential compromise in cloud servers. Research suggests that a significant proportion of server workloads (65% generally rising to above 90% in financial services) are currently hosted in private data centers. Data residency is ranked as a top reason for maintaining private infrastructure. Additionally, certain government use cases could transition to the cloud with cloud deployments of Arm CCA and other comparable technologies. A gold standard for cloud providers is to achieve a state where they are unable to access customer data.

Veracruz is an open-source software project⁷, adopted by the CCC, which is developing a framework for privacy-preserving collaborative computing that may be useful for Federal Agency sharing and collaboration, as posed in the question. The aim of Veracruz is to explore and create a demonstrator for how secure distributed systems can be built using strong containerization technology and remote attestation protocols. As the Wiki discusses, possible use cases include privacy-preserving machine learning, delegated computation, and commitments.⁸

Secure transport protocols like TLS will be updated to support mutual attestation, enabling workloads everywhere to establish the trustworthiness of cloud services before data is exchanged or trusted. When PETs that support platform and workload attestation are deployed across a broader range of form-factors, cloud services will also be able to assess the trustworthiness of devices like edge servers, laptops, and phones. It is expected that the PETs will be deployed on microprocessors before they're deployed with microcontrollers.

4. *Specific regulations or authorities that could be used, modified, or introduced to advance PETs: Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking authorities under the Office of Management and Budget, the Federal Trade Commission, and financial regulatory bodies, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (e.g., Other Transaction Authority) to procure PETs for development.*

Arm response

The U.S. approach to privacy and security has historically been based on a sector-specific, sensitivity-based approach. Health information, financial information, and children's and minors' online activity, for instance, all have been required to be handled in specific manners due to unique sensitivities. While the U.S. Congress is currently debating and attempting to legislate a more consistent and comprehensive approach to general data and privacy protections, the agencies regulating these more sensitive forms of data which already require

⁷ See [Veracruz Project – Just another Linux Foundation Projects 2 site \(veracruz-project.com\)](https://veracruz-project.com)

⁸ See [What are some use cases for Veracruz? · veracruz-project/veracruz Wiki · GitHub](https://veracruz-project.com/wiki/veracruz)

special treatment should encourage, incentivize and in some instances mandate use of PETs. This should not be read as support for mandating a specific technology; as is the case in cybersecurity which the U.S. government generally addresses appropriately, technologies evolve, improve, and become outdated more quickly than government rulemaking. Rather, having already determined certain data types are more sensitive and therefore deserve special protections, appropriate agencies should consider whether commercially available technologies and industry best practices for protecting and preserving privacy of those types of data require more regulation or could be incentivized in other ways, as has been done in the cybersecurity space.

Further, the US government should use its procurement and purchasing power to drive adoption and broader take up of PETs. This could be done by moving certain government workloads currently done on-premises to the cloud in instances where cloud provider use or PETs that can offer comparable levels of data privacy protections to on-premises.

Lastly, NIST standards such as FIPS 140-2 have been instrumental in helping buyers procure secure products and services. The US government may like to consider supporting or developing a similar standard to define security requirements for cloud instances that support PETs such as Confidential Computing. This would help government and private sector organizations choose between products. In the absence of standardization, organizations will need to undertake their own security assessment, an activity that requires specialized resources and access to information pertaining to each vendor's platform architecture; this would result in either applications staying on-premises, or data being placed at risk.

8. *Existing best practices that are helpful for PETs adoption: Information about U.S. policies that are currently helping facilitate adoption as well as best practices that facilitate responsible adoption. This includes existing policies that support adoption, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This also includes information about where and when PETs can be situated within tiered access frameworks for accessing restricted data, ranging from publicly accessible to fully restricted data.*

Arm response

USG should include discussion of PETs in all relevant government and industry guidance documents, and should keep them updated on a regular basis as the technology develops. These include, but are not limited to, the NIST Cybersecurity Framework,⁹ the NIST Privacy Framework,¹⁰ CISA's Privacy Impact Assessments,¹¹ and all relevant FTC privacy and security guidance.¹²

⁹ [Cybersecurity Framework | NIST](#)

¹⁰ [Privacy Framework | NIST](#)

¹¹ [Privacy Impact Assessments | Homeland Security \(dhs.gov\)](#)

¹² [Privacy and Security | Federal Trade Commission \(ftc.gov\)](#)

Further, an expert agency could be tasked with drafting a guidance document that would enable migration to, and greater uptake of, current and emerging PETs. NIST have done something similar to this with its work to prepare for and facilitate the migration and adoption of post-quantum cryptography when that technology is ready for wide deployment.¹³ While not an exact parallel, something similar could be done to prepare for and facilitate greater adoption of PETs among government agencies.

Conclusion

Arm appreciates the opportunity to contribute to the Office of Science and Technology Policy's information gathering on PETs, and share perspectives on how to advance their development and adoption. Please be in touch if we can provide additional information on our responses or other questions the Office may have.

Respectfully Submitted,

Vince Jesaitis¹⁴
Hugo Vincent
Charles Garcia-Tobin
Mark Knight

¹³ See [Migration to Post-Quantum Cryptography | NCCoE \(nist.gov\)](#) and <https://www.nccoe.nist.gov/sites/default/files/2022-06/Migration-to-PQC-05-16.pdf>

¹⁴ Contact for additional questions:

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Asemio

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Asemio, LLC
Managing Director
(Industry Respondent)

White House Request for Information:
Advancing Privacy-Enhancing Technologies
July 8, 2022

The following Request for Information response has been prepared for the United States Office of Science and Technology by Asemio, a technology firm that uses privacy-enhancing technologies to help organizations share sensitive information securely.

1. Specific Research Opportunities to Advance PETs:

Research Topics

The federal government may find it promising to invest in research that clarifies national data ecosystem advancement measurements. Privacy-enhancing technologies exist within a complex and sprawling web of dynamic systems, including cultural phenomena (e.g., political, economic, and social agendas), governance frameworks, and technical infrastructure. This dynamic super-system could benefit from being mapped and from an attempt to quantify and qualify relevant components. By implementing PETs situated inside a measurable system, we can better understand the efficacy of the sum of the elements in an individual data ecosystem.

One of the lacking aspects in modern data ecosystem research is the dearth of ‘research on research.’ There is limited information publicly available on the ‘time tax’ and the investment required to produce insights from various public good data systems. This lack of knowledge leaves technologists and those engaging in data system modernization efforts in the dark about how to measure the efficacy of these improvements. If the federal government were to invest in measuring and quantifying our data system improvement efforts, we surmise that the clarity offered by those efforts would encourage and accelerate the adoption of PETs. That adoption would present a clear opportunity to cut costs and maximize resources in many different domains.

An additional area for possible research is in increased interoperability between PET systems. PETs are a diverse and wide ranging set of technologies, and as PETs continue to be adopted, there is a need for research into effective PET-switching infrastructure. For example, Chicago has embraced innovation using PETs and has multiple PET initiatives spanning the public health, healthcare, and education domains through the work being done by the CAPriCORN network, Cook County Health department, and School Health Access Collaborative efforts in coordination with Public Health Institute of Metropolitan Chicago.¹ This makes sharing information across domains difficult, despite improved sharing of data being a highly valuable goal of modern technology. Improved access to data would mean that clinical health leaders in the city would benefit from greater access to SDoH data to get a full picture of the community members they are caring for and public health leaders would benefit from more information on what is happening at the clinical level, among other benefits. The federal government would provide significant value by advancing our understanding of protocols or standards that could enable PET interoperability.

¹ Kho, Abel N, et. al. CAPriCORN: Chicago Area Patient-Centered Outcomes Research Network. J Am Med Inform Assoc. 2014 Jul-Aug;21(4):607-11. <https://doi.org/10.1136/amiajnl-2014-002827>; “Open Source PPRL (Privacy Preserving Record Linkage),” Linkja, accessed July 8, 2022, <https://linkja.github.io/>; Tamar Westphal, “School Health Access Collaborative,” Public Health Institute of Metropolitan Chicago, March 22, 2022, <https://phimc.org/initiatives/shac/>

2. Specific Technical Aspects or Limitations of PETs:

Technical Aspects

A clear understanding of the intended use of data (e.g., to access or share raw data; to perform analysis with other partners; to publish the insights from the data collected, etc.) is essential before employing any PETs. PETs are often not technically interchangeable and are restricted to certain use cases. Therefore it is important to deploy the correct PETs as per their intended use to optimize efficiency. The intended use of data also significantly impacts the type of agreements and governance necessary, the lift required for participating agencies or organizations, and community buy-in needed for projects.²

It is imperative to understand that PETs are not the ultimate solution for privacy concerns and should not be treated like so. These are mere technologies that can aid in minimizing the negative impacts of collecting and sharing sensitive data. These technologies can have accountability issues and therefore should be used as a tool in the larger organizational safety kit for safe and secure data sharing.

Limitations

There are a number of perceived, and actual, limitations that have slowed progress on integrating PETs into modern integrated data systems. Specifically, one perceived limitation of secure hash encoding (SHE), a type of PET, indicates that the deterministic nature of SHE record linking is too strict and results in an unacceptable number of false negatives because the technique is especially sensitive to common data input errors (e.g., misspelling of names). In our experience, proper data cleaning and normalization techniques ameliorate this perceived limitation and F-scores reach upwards of 95% and higher—an efficacy that seems sufficient for general statistical insight.

PET is a large umbrella term that encompasses many different technologies and techniques. The diversity of techniques means limitations are present across a spectrum and should be approached individually to be understood accurately. The complexity of this landscape and implications of the disparate approaches can be overwhelming for organizations and institutions to fully understand. Although over the long term PETs have the potential to lower costs in data ecosystem advancements by reducing compliance and security risks, the up-front costs associated with employing a new system using PETs can be a limiting factor. Many PETs are relatively new, and the obscure nature of the technology can be especially daunting to those outside of the cryptography and analytics domains. The lack of familiarity, paired with the lack

² “PETs Adoption Guide,” PETs Adoption Guide, accessed July 8, 2022, <https://cdeiuk.github.io/pets-adoption-guide/adoption-guide/>.

of guidance and regulation surrounding their use, increases the uncertainty of adopting a new technology and slows progress towards making these tools more accessible.

Many PETs require a level of technical acumen to implement and adopt. The most beneficial measures that could mitigate risk in this process would be architectural guidance from the federal government to outline a standardized and secure way to build systems using PETs. Implemented incorrectly or at sub-par quality, PETs run a risk of creating a false sense of security that is built on software that is vulnerable to attacks.

3. Specific Sectors, Applications, or Types of Analysis that Would Particularly Benefit from the Adoption of PETs:

High potential for the adoption of PETs

Sectors dealing with high political, social, and economic risk due to the loss or exposure of sensitive information are among those that would benefit most from the widespread implementation of PETs. This includes sectors and applications where data are exceptionally decentralized, use cases where PETs can reduce the risk of unintentional disclosures, and use cases where PETs might assist in data portability and interoperability.

One of the most promising approaches to privacy-enhanced data systems that invites further attention is the use of secure hash encoding (SHE).³ SHE is currently being used by various municipalities such as Chicago, Illinois and Tulsa, Oklahoma, as well as by leading technology companies and federal partners such as Datavant.

One of the most promising applications of SHE is the use of micro-level data in community data systems. The value here is in unlocking administrative data for use in natural world experimentation. The ‘wicked problems’ that our communities across the United States face include complex domain-specific dynamics (e.g., social determinants of health role in whole person health).⁴ Attempts to tame these problems could benefit greatly from safe, secure, easy, and ethical access to these community-level microdata. Much of the data that could inform decision making around these complex issues is sensitive, personally identifiable data used for record linking and lies in “decentralized” data stores that exist in disparate organizations. It is difficult to share these data across organizational governance boundaries due to their sensitive nature.⁵ In general, much of the information that could be used to advance public good is decentralized, or federated. Our country’s support networks are made up of local CBOs, state

³ D. Vatsalan, et al., A taxonomy of privacy-preserving record linkage techniques, *Information Systems* (2013), p. 9, <http://dx.doi.org/10.1016/j.is.2012.11.005>

⁴ Horst W. Rittel and Melvin M. Webber, “Dilemmas in a General Theory of Planning,” *Classic Readings in Urban Planning*, June 2018, pp. 52-63, <https://doi.org/10.4324/9781351179522-6>.

⁵ Asemio, “Architecting Resilient and Adaptive Communities through Technological Innovation,” Asemio, Accessed on July 8, 2022, <https://www.asemio.com/architecting-resilient-and-adaptive-communities-through-technological-innovation/>.

agencies, federal entities, tribal governments, community stakeholders including activists, philanthropists, other public and private entities, and many more.

Currently, there are a number of prototypical organizations and government systems that are employing PETs and benefiting from the improved insights and decrease in associated governance costs. In Chicago and Tulsa, PETs integrate clinical and SDoH data to better understand the underlying clinical, education, and public health issues that are contributing to poor economic and social outcomes. Using these technologies has been beneficial for the creation of a rapid analytics infrastructure. The benefits can be seen in increased compliance, reduction in security risks, improved control, and more comprehensive integration of community organizations—both vertically (e.g., state to community to region to cross-region) and horizontally (e.g., health, human services, justice involvement, education).

4. Specific Regulations or Authorities that Could be Used, Modified, or Introduced to Advance PETs:

Updates to Existing Regulations

Clarifying the use of PETs in relation to existing regulatory controls (e.g., HIPAA and FERPA) would greatly benefit our national ability to share and link data across institutions. Current standards are often ambiguous and subjective, especially when it comes to the issue of de-identification. Updating regulatory and technical standards (e.g., from the National Institute on Standards of Technology) as well would help guide organizations looking to implement PETs on best practices, expectations, and acceptable use.

5. Specific Laws that Could be Used, Modified, or Introduced to Advance PETs:

State & Federal Law

Existing regulatory and policy frameworks, in addition to underinvestment in antiquated infrastructure, has greatly challenged public interest data and technology innovation. We must address architecture that enables alignment to the complex and changing public health, political, and social landscapes. Speed of adaptability aligned to rate of change is critical as there are likely to be no “silver bullet” problem-solving innovations.

Ways to advance PETs include introducing new or modifying existing provisions in state or federal law. These changes could include creating safe harbors or defining how use, disclosure, safeguards, and breaches are handled. In addition, legislation could be created or modified to incentivize the development and adoption of PETs. This legislation could look like tax breaks or subsidies for companies that develop or use PETs. It could also include requirements that certain

entities use PETs in order to protect user data. There is a need to take into account all of the different stakeholders involved in the development and adoption of PETs, including not only the federal government, but state and local governments as well as private sector entities.

New Legislation

In addition, a comprehensive response should also consider how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. For example, such legislation could include provisions that would provide incentives for entities to develop and adopt PETs. Also, a comprehensive response should take into account international law as it applies to privacy and data sharing among international entities. Finally, a comprehensive response should also address how to ensure that PETs are used in a way that protects the privacy of individuals. Any strategy that focuses on the development of PETs without considering how they will be adopted or used is incomplete.

6. Specific Mechanisms, Not Covered Above, that Could be Used, Modified, or Introduced to Advance PETs:

Other Mechanisms Within the Federal Government

The government could also adopt policies and procedures that encourage the use of PETs within the federal government itself, such as interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, and data use or sharing agreements. One example of this would be to require the use of PETs for all data sharing projects between the federal government and state, local, tribal, and territorial governments. This could be an effective way to promote the use of PETs by government entities at all levels due to the increased demand that would result from such a requirement.

Involvement of Non-Governmental Organizations

Non-governmental organizations hold large amounts of social data that could be used without compromising the privacy of their service populations using PETs. There are a number of ways that their involvement in federal adoption of these technologies could have a positive impact, such as:

- Involving educational institutions to lead research, create awareness, and catalyze community engagement to promote responsible use of PETs;
- Engaging institutions in cross-sectoral research to develop and deploy PETs responsibly, and;

- Promote adoption of PETs in both Government and Non-Governmental Organizations and support NGOs with the expertise to implement PETs in performing due diligence and maintaining standards across the board.⁶

7. Risks Related to PETs Adoption:

Although privacy is often presented as a binary state (i.e., something is private or not), in actuality, it exists more on a spectrum. This spectrum of “how private” data is exists in tension with “how useful” data is. This tension between data utility and data privacy is at the heart of the problem that PETs are trying to solve. In essence, PET is attempting to increase the utility of data while at the same time increasing (or at least keeping constant) the privacy protections of sensitive data.

The level of privacy guarantee when using privacy-enhancing technology can often be misconstrued or oversimplified or over trusted. There is a difference between mathematical/theoretical privacy and the application of realistic privacy and security measures. For example, reidentification can occur as a result of recombination attacks. Therefore, PETs warrant robust mechanisms that provide a higher level of security against information technology attacks and also statistical attacks.

Various cryptographic attacks (e.g., frequency, rainbow table, and dictionary) still present a security issue for secure hash encoding, however proper use of salts and concatenating linking fields will limit the effectiveness of these kinds of attacks on personally identifiable data. This approach is recommended for system architecture. More concerning is the use of statistical hacks and/or data recombination attacks, which pose a greater risk for re-identification.

⁶ “Privacy Enhancing Technologies,” Royal Society, pp. 54-57, accessed July 8, 2022, <https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/>.

Reference List

1. Kho, Abel N, et. al. CAPriCORN: Chicago Area Patient-Centered Outcomes Research Network. *J Am Med Inform Assoc.* 2014 Jul-Aug;21(4):607-11. <https://doi.org/10.1136/amiainl-2014-002827> ; “Open Source PPRL (Privacy Preserving Record Linkage),” Linkja, accessed July 8, 2022, <https://linkja.github.io/> ; Tamar Westphal, “School Health Access Collaborative,” Public Health Institute of Metropolitan Chicago, March 22, 2022, <https://phimc.org/initiatives/shac/>.
2. D. Vatsalan, et al., A taxonomy of privacy-preserving record linkage techniques, *Information Systems* (2013), <http://dx.doi.org/10.1016/j.is.2012.11.005>
3. Ibid.
4. “PETs Adoption Guide,” PETs Adoption Guide, accessed July 8, 2022, <https://cdeiuk.github.io/pets-adoption-guide/adoption-guide/>.
5. Horst W. Rittel and Melvin M. Webber, “Dilemmas in a General Theory of Planning,” *Classic Readings in Urban Planning*, June 2018, pp. 52-63, <https://doi.org/10.4324/9781351179522-6>.
6. Asemio, “Architecting Resilient and Adaptive Communities through Technological Innovation,” Asemio, March 30, 2022, <https://www.asemio.com/architecting-resilient-and-adaptive-communities-through-technological-innovation/>.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Austin, Lisa; Lie, David; Nikolov, Aleksandar; and Papernot, Nicolas

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Name of the persons filing the comment:

- Dr. Lisa Austin, University of Toronto,
- Dr. David Lie, University of Toronto,
- Dr. Aleksandar Nikolov, University of Toronto, s
- Dr. Nicolas Papernot, University of Toronto,

Respondent type: Academic

We respond inline to a subset of the topics included in the request for information published in the federal register. The topics are indicated in bold below, along with the corresponding topic number included in the original request for information.

1. *Specific research opportunities to advance PETs:* Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.

Several techniques have been proposed to train on distributed data with the aim of protecting sensitive data. Most prominently, Federated Learning (FL) allows individual models to be trained on distributed data, where the model gradients, parameters, or other forms of model updates are shared and then aggregated by a central party. This framework achieves data minimization because it does not directly centralize data at the server orchestrating the protocol.

However, in its vanilla form, FL does not provide privacy (in the sense of differential privacy, see response to topic 2 below) since gradients or parameters can still reveal information about the training set of each participating party. The crux of why FL is inherently vulnerable to data reconstruction attacks is that it is designed to provide confidentiality (data does not leave user devices) rather than privacy (output of the computation does not leak sensitive attributes from the users' input). Without additional privacy measures, FL cannot protect users from the central party reconstructing their data.

One key takeaway from this line of work is that privacy-preserving FL, in the presence of an untrusted central party, is not yet practical. Hence, without trust in the central party, FL cannot provide any privacy guarantees to its users. It is therefore important for users to either understand the trust they are placing in the central party when participating in a FL protocol--or to require that new FL protocols be put forward which do not make these implicit assumptions about trust.

We believe that the latter line of research deserves much more attention and support from funding agencies. A promising research direction is to develop new protective measures that can be integrated in the design and deployment of FL to reduce the trust required in the central party. One particular aspect is the addition of noise inside a cryptographic protocol, such as secure multiparty computation, to jointly obtain differential privacy. As of yet, however, known constructions' communication costs are prohibitive so we will investigate new ones.

2. *Specific technical aspects or limitations of PETs: Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.*

We argue that defining privacy in terms of de-identification and deanonymization techniques is inadequate because it focuses on modifying personal information rather than the method of analysis of this information. We start by illustrating the challenges inherent in de-identification through examples of how naive approaches can fail. We argue that such failures are inherent when de-identification focuses on the data rather than the analysis procedures. We also give a high-level, non-technical overview of recent frameworks of privacy protection developed by researchers in computer science, AI, and statistics, which, unlike the definition above, are centered around algorithms rather than the data themselves.

Because of the subtleties in defining de-identification, it is often instructive to start instead with failures to de-identify data. One naive attempt at de-identifying data is to simply strip the data from personally identifying information, i.e., to remove common identifiers like names, addresses, social insurance numbers, etc. There are many examples of how this approach usually fails, because seemingly non-identifying pieces of information can often be linked together to uniquely identify people. This observation goes back to pioneering work by Latanya Sweeney, who famously identified former Massachusetts Governor William Weld's health records from his date of birth, gender, and zip code.¹ This is known as a "reconstruction attack" (when the attacker reconstructs a data set containing personal information) or a "re-identification attack" (when the attacker can also link the personal information to identifiable individuals). To

¹ Latanya Sweeney. Simple demographics often identify people uniquely. *Health* 671, 1–34 (2000).

illustrate how it works, we will take a lesser-known and more recent example: the re-identification of the Myki data set by Culnane, Rubinstein, and Teague.²

Myki is a smart card-based public transit ticketing system used in the state of Victoria in Australia: travellers tap on when boarding a bus or train, tap off when exiting, and the system registers their trip and charges them accordingly. As part of a datathon in 2018, the Victoria government released a large data set of Myki usage information, which contestants could use to analyze the state's public transit system. The data set contained information about trips: when and where a card was tapped on, when and where it was tapped off, and which bus or train line was used. By way of de-identification, card ID numbers were substituted with random strings, and, of course, no names or addresses of registered users were included. However, if the same card was used for several trips, then the same random string was listed for each trip.

The first step researchers took in breaking the anonymity of this data set feels surprisingly innocuous: they found their own cards in the data set. This is simple: you can look up several times and places where you took a bus. For example, the times you took a bus to work on two consecutive work days, and one weekend trip. If there is only one card in the data set that was used for all of these trips, then this must be your card. This works most of the time because it turns out that only two trips are usually enough to identify a card uniquely.

While identifying your own card scarcely feels like a privacy violation, it enables a much more damaging second step: identifying the card of someone you know. For example, knowing which card in the data set is yours, you can easily identify the card of a coworker. Perhaps the coworker takes the bus home at the same time as you, and maybe you had that one work dinner together. You can check which cards were used with yours at these times, and, again, it turns out that, more often than not, there is a unique match in the data set. Having identified the card of your coworker, you can find out any other trip they have taken: weekend trips, doctor visits, or other excursions that they probably expect to be private.

The researchers also showed that such privacy breaks can be extended further by linking the data set with publicly available information. They cross-referenced the tweets of Anthony Carbines, a member of the Victorian Legislative Assembly, with the data set, and, since Carbines often proudly tweets about the Victoria public transit system when he's boarding a train, they could identify his card as well.

Let us draw several lessons from this attack:

- Anything is potentially personally identifying information, because combining a few innocuous pieces of information often identifies a person uniquely.
- Who does the identification matters. It is likely that whoever is curious about your private information already knows a lot about you. This knowledge can be leveraged to find out

² Chris Culnane, Benjamin Rubinstein, Vanessa Teague. Stop the Open Data Bus, We Want to Get Off

even more.

- De-identification is hard in a connected world. Twitter, LinkedIn, and other social networks and websites provide easily accessible information that an attacker can leverage to facilitate re-identifying people in a supposedly de-identified data set.

In many practical scenarios, an adversary will not have direct access to the data but instead to a by-product of the data. This could be, for instance, the output of a database query, or the prediction of a machine learning model trained on the data. One may hope that, because such by-products only contain aggregate statistical information, they are inherently privacy-preserving. This hope, unfortunately, does not bear out in reality.

As a first example, let us take reconstruction attacks from counts. Researchers and official statistics agencies often publish data summaries in the form of tables which contain counts of how many people satisfy some property: these can be summaries of surveys, or voting polls, or census data published by the US Census Bureau. Such counts feel safer, from a privacy perspective, than fine-grained data sets such as the Myki data set. They can, however, still enable an attacker to identify individuals. Imagine, for example, that someone poses two counting queries to the data set: “How many computer science professors at the University of Toronto smoke?” and “How many computer science professors at the University of Toronto who were not born in Bulgaria smoke?” Each count is (probably) a relatively large number, so would not seem like a threat in isolation, but subtracting the second number from the first identifies immediately whether one of the authors of this blog post smokes.

More sophisticated versions of this kind of attack were introduced by Dinur and Nissim,³ who showed that, even if some noise is added to the counts, they can still be used to reconstruct private information about most individuals in a data set. Recently, researchers showed these attacks are much more than a theoretical threat, and can be successfully carried out against Diffix, a system specifically designed to answer counting queries while protecting privacy [CN20]. Moreover, the US Census Bureau has conducted its own reconstruction experiments and concluded that reconstructing sensitive microdata from the 2010 Decennial Census is possible, at least theoretically.⁴ On a high-level, reconstruction attacks are possible whenever the attacker can get accurate enough answers to many uncorrelated counting queries, i.e., counting questions that ask about properties that do not overlap much. (A precise mathematical condition was identified by Muthukrishnan and one of the authors of this response.⁵)

What this points to is that, rather than focusing on making a particular set of data private (e.g., through de-identification by removing personally identifying information), the scientific community has discovered that making an analysis technique (or algorithm) private provides more meaningful guarantees. Seeing a data set in isolation makes it hard, if not impossible, to

³ Irit Dinur, Kobbi Nissim. Revealing information while preserving privacy. PODS 2003.

⁴ Simson Garfinkel, John Abowd, Christian Martindale. Understanding Database Reconstruction Attacks on Public Data. ACMQueue, Vol. 16, No. 5 (September/October 2018): 28-53.

⁵ Aleksandar Nikolov, S. Muthukrishnan, Optimal Private Halfspace Counting via Discrepancy. STOC 2012.

decide whether the data are successfully de-identified. As we observed already, this depends on what an attacker trying to re-identify individuals knows, and what additional information may be available from other sources. Without knowing who may try to do the re-identification, and what information they possess, or which individuals or what new information they are interested in, we cannot decide if a data set is safe for publication. If we know, however, the method (i.e., the algorithm) through which the data was analyzed to produce some by-product of it, such as a table of counts or a machine learning model, then we can actually make guarantees that hold against any possible attacker, with any kind of side information. This insight was developed initially in work by Dwork, McSherry, Nissim, and Smith,⁶ who introduced a seminal framework for private data analysis known as differential privacy. Differential privacy has seen an increasing number of recent adoptions, both in industry (by Google, Apple, Facebook, among others) and by official statistics agencies, most notably the US Census Bureau starting with the 2020 Decennial Census. This shift away from de-identification and the associated focus on the dataset also requires a shift away from transferring the modified dataset: the privacy guarantee now comes from the method of analysis so the data must be kept secure and access to it restricted. To obtain access to the dataset, one would have to specify (a) what analysis they are interested in running on the data and (b) how they will ensure that this analysis is privacy-preserving. For example, (b) could take the form of a proof of differential privacy for the algorithm that is used to analyze the data.

Indeed, as we mentioned, differential privacy defines privacy with respect to an algorithm used to analyze data, rather than with respect to the data themselves.

Informally, the definition can be described using a thought experiment. Imagine that we execute the algorithm on a data set, and we also execute it on the data set with the data of one individual modified. To be differentially private, the algorithm must behave almost identically in these two scenarios, and this must happen for all data sets and all modifications to the data of any one individual. For example, a differentially private algorithm must have the property that if it produces a particular machine learning model from a data set that includes your data, then it is almost as likely to produce the same model if your data were removed or changed completely. This means that an adversary observing or using the model output by this algorithm is unable to tell whether your data (or any other person's data) were used to train the model at all. Then, whatever information an attacker may learn about you from observing the algorithm's output could have also been learned even if your data were never seen by the algorithm.

Differentially private algorithms have the property that they reveal statistical information about the data, such as, for example, correlations between genetic factors and medical conditions, but do not reveal information about particular individuals, such as whether they have some medical condition. This is because statistical information does not depend crucially on any one person's data, while private personal information does. We should note that the actual technical definition is quantitative, and the level of privacy protection is tunable, allowing us to trade privacy off for the machine learning model's accuracy, or to query answers produced by the algorithm.

⁶ Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. TCC 2006

4. *Specific regulations or authorities that could be used, modified, or introduced to advance PETs:* Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking authorities under the Office of Management and Budget, the Federal Trade Commission, and financial regulatory bodies, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (e.g., Other Transaction Authority) to procure PETs for development.

In general, regulations and authorities should contribute or encourage exploratory demonstrations of the benefits of PETs. It can be difficult for researchers to demonstrate the benefits of a technology because the current regulations do not allow them to deploy a prototype of it on relevant data—while at the same time said regulations will not be updated unless there is a demonstration of the benefits of said technology. For instance, it is difficult to demonstrate the benefit of differential privacy for certain applications if current regulations mandate that data which is handled in these applications be de-identified. This is because differential privacy promises that an analysis can be made privacy-preserving without any form of de-identification. Hence, by forcing researchers to evaluate differential privacy on de-identified data, we prevent them from demonstrating the full potential of differential privacy. Put another way, we should ask: what new forms of data analysis can differential privacy enable? Rather than asking: how does differential privacy compare to de-identification. Regulations and authorities here could help in several ways, including by providing access to “sandboxes” in which PETs can be experimented with to motivate more significant changes to the law.

5. *Specific laws that could be used, modified, or introduced to advance PETs:* Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.

In general, the regulatory environment needs to make three changes: 1) a shift in its approach to risk, 2) the development of technical standards, and 3) the creation of regulatory frameworks

for data intermediaries. We outline these three shifts below, but do not comment on the specifics of how these might be implemented within the US legal landscape.

1) Shifting the Approach to Risk

While the question of whether a person is “identifiable” from some mass of information is central to many privacy concerns – and often whether data processing is regulated or not – it does not catch the many ways in which contemporary data processing creates privacy vulnerabilities. For example, an individual might not be identifiable in any *single* database, but the accumulation of information about this specific individual across [multiple databases](#) potentially increases what can be known about this individual. Waiting for that risk to materialize into a “reasonably foreseeable” risk of identifiability (or whatever is the applicable legal standard in a particular jurisdiction) is to address responsible data analysis too late in the data pipeline.

This is why we disagree with [risk-based approaches](#) that focus *only* on identifiability. Consider the analogy of speeding on a highway: I might not be wronging any specific individual when I speed, but my behaviour is risky and we regulate it through imposing speed limits in order to reduce this risk. Similarly, my data processing might not identify any specific individual, but my methods might still be risky and we need to regulate to reduce this broad risk—rather than focus on identifiability only.

The way to do this is to ensure that all organizations that process “information about persons” utilize reasonable practices to minimize privacy risks. By privacy risks, we mean here the processing of information about persons in a manner that increases the risk of identifying an individual or inferring information about a specific individual. As we have already outlined above in our response to question 2, mitigating privacy risks should not focus on manipulating the data (the de-identification approach), but should instead involve focusing on the algorithms and the computing environment more generally.

Although this shift in approach to privacy risks would require regulating the processing of “information about persons” rather than regulating the processing of PII, it would not necessarily have to dramatically shift other aspects of privacy regulation. For example, Canada has proposed [new privacy legislation](#) that would regulate de-identified, but not anonymous, information but would exempt de-identified information from some of the obligations that pertain to PII (including consent in a number of contexts). We agree with this approach although we are disappointed that the proposed legislation continues to focus on de-identification and fails to adopt the broad privacy risks minimization principle we advocate for here.

2) Development of Technical Standards

There is a strong need for the development of international standards regarding re-identification risks. We outline two components of this: regulatory and technical.

On the regulatory side, there is an increasingly international push to embrace “privacy by design” (PbD). Generally speaking, PETs are considered too narrow and technical to be a full solution to the complexities of PbD. Some of those complexities include the need to comply with

multiple laws, the often open-ended nature of some legal obligations, the need to account for how general legal norms get interpreted and re-interpreted.⁷ Advocates of PbD often point to the need for a broad approach that embraces technical measures but also work processes, management structures, etc.⁸ That said, we believe that technical standards that address one aspect of privacy – the risk of identifying an individual or inferring information about a specific individual – are possible. Managing this risk is common to most privacy laws across different jurisdictions, even if those laws may address this risk differently. For example, jurisdiction A might require a low level of risk when dealing with “sensitive” information whereas jurisdiction B might not require that. Having technical standards for measuring and addressing this risk is important even if other aspects of PbD might remain jurisdiction-specific and require more holistic approaches.

On the technical side, the creation of strong standards requires a shift away from de-identification as the approach to managing privacy risks. As we outlined in our answer to question 2, it is difficult to measure the risk of re-identification when focusing on data in isolation, as it depends upon knowing what the attacker might know as well as what other information is available (which also changes with time). However, approaches like differential privacy can give more accurate guarantees.

3) Data Intermediaries

Ensuring that privacy risks – understood here as the risk of identifying an individual or inferring information about a specific individual – are mitigated when analyzing information about persons requires a high level of technical expertise. Organizations may wish to analyze information in a privacy-protective way but see the costs of this as prohibitive. We think there is great scope for the development of data intermediaries that could manage this for organizations. The EU has recently signaled a move in this direction with the introduction of the [Data Governance Act](#). Some of the interest in Europe lies in creating intermediaries that could help individual data subjects better manage their right to control their PII.⁹ However, other models focus on reducing the legal and technical complexities involved in data-sharing, particularly in relation to managing re-identification risks.¹⁰ Creating a regulatory environment that provides clarity regarding issues of liability and oversight is important to create the right incentives for the successful creation of these intermediaries.¹¹

Data intermediaries could also provide additional options for managing some of the risks involved in the use of some PETs. For example, as we outlined in our answer to question 1,

⁷ Bert-Jaap Koops & Ronald Leenes, “Privacy Regulation Cannot Be Hardcoded. A Critical Comment on the ‘Privacy by Design’ Provision in Data-Protection Law” (2014) 28:2 Intl Rev L Comp & Tech 159.

⁸ Anver Levin, “Privacy by Design by Regulation: The Case Study of Ontario” (2018) 4 Can J Comp & Contemp L 115 at 118.

⁹ Sylvie Delacroix and Neil D. Lawrence, “Bottom-up data Trusts: disturbing the ‘one size fits all’ approach to data governance” *International Data Privacy Law*, Volume 9, Issue 4, November 2019, Pages 236–252.

¹⁰ Lisa M. Austin and David Lie, “Safe Sharing Sites” (2020) 94 NYU L Rev 581.

¹¹ Lisa M. Austin and David Lie, “Data Trusts and the Governance of Smart Environments: Lessons from the Failure of Sidewalk Labs’ Urban Data Trust” (2021) 19 *Surveillance and Society* 255.

privacy-preserving FL with an untrusted central party is currently infeasible. Technical solutions to this can be developed but, in addition, the creation of trusted intermediaries that could manage this risk is an alternative pathway.

7. Risks related to PETs adoption: Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

We highlight two risks associated with the adoption of PETs. First, PETs remain highly technical and evaluating the guarantees they provide often requires an understanding of statistical analysis. For instance, there is a lot of confusion between the technical notions of confidentiality and of differential privacy (see above our response to topic 1). This in turn means that end users need to place a lot of trust in companies handling their data because (a) they are not certain which guarantee is needed to protect their data against the risks they are concerned about and (b) even if they know which guarantee is needed they may be unable to assess how strong the guarantee is and/or should be. There is room here to develop technical standards that are endorsed by legislators and regulators to provide more transparency to end users and make enforcing privacy more actionable for both regulators and companies handling data.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Bitfount

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

RFI Response: Privacy-Enhancing Technologies

Bitfount Ltd.
June 27, 2022

[Bitfount Ltd.](#) is a private company working to safely unlock the value of data for the benefit of humankind. Bitfount makes previously intractable data interactable. We do this through the development of a platform, based on the paradigm of federated data science, which removes barriers to data collaboration without compromising on data privacy or quality. As an interested party in OSTP's efforts to advance PETs, Bitfount is delighted to provide the following comments regarding OSTP's recent RFI.

Contact: Blaise Thomson, CEO;

Topic 1.

While there are a wide range of ongoing research activities in the PET field, most of those with which we've come into contact deal with the generation and development of the techniques themselves rather than how to develop standards such that they could be more easily adopted. The various players in the space will not be able to come to agreement on the development of standards for techniques which dictate what 'level' of privacy protection (in the form of epsilon for differential privacy, for example) or deployment requirements without a more overarching set of standards. We recommend investing both in the development of a set of standards around differential privacy, synthetic data, and secure multi-party computation with which all parties can live.

As an extension of this work, we recommend performing research into and developing easy on-ramps for businesses with PET-appropriate use cases. A large barrier to adoption is simply knowledge that PETs are an acceptable solution to data management and privacy challenges from the point of view of regulators. Research into how best to educate businesses and create incentives for adoption could go a long way.

Topic 2.

Current limitations of PETs include the sheer number of definitions of privacy, impact on accuracy of outputs, insufficient compute power, and lack of benchmarks in the field.

- **Definitions of privacy:** There are several definitions of privacy which have been developed overtime. We view this as a limitation of the PET field given the wide range of accepted definitions, even within a given PET technique, is likely to cause confusion among those who might adopt PETs. For a technology that is meant to reduce uncertainty, having so many complex options may lead to reluctance to adopt PETs by less-informed entities.
- **Impact on accuracy:** While PETs provide a useful mechanism for analyzing otherwise sensitive data, they do affect the accuracy of the output a data scientist receives from performing a query or running a model. This is a limitation of the technology in cases where algorithms are making decisions which could have real-life consequences for citizens. The accuracy loss can typically be quantified, but there is a tradeoff between accuracy and privacy leakage.

- **Insufficient compute power:** For the average organization, executing day-to-day business processes integrated with PETs would be too costly given they often require significantly more computational power than running data operations with no privacy-preserving techniques applied. Note, this includes time cost, as data science teams tend to prioritise speed and will not want to adopt a slower technology if it is not required.
- **Lack of benchmarks:** For techniques such as differential privacy, we as a collective industry do not have easily digestible benchmarks describing what level of privacy preservation is appropriate given the type of data you are using and the level of risk you can tolerate in association with the data. A lack of benchmarks for epsilon in various situations means we struggle to make it ‘turnkey’ for businesses to adopt PETs and understand what they need to implement.

Topic 3.

On a high-level, we believe techniques such as federated data science in combination with differential privacy can provide a viable solution to challenges such as enabling cross-border data transfer under a new version of Privacy Shield.

On a more tactical level, we see tremendous potential for the use of PETs in industries holding data typically considered sensitive or protected, but that could be ‘unlocked’ in the public interest if analysis were properly de-risked. These include healthcare and life sciences, defense, energy, and international trade and commerce.

- **Healthcare:** PETs can be leveraged in healthcare settings, especially during clinical trial design, execution, and post-trial analysis, to ensure HIPAA compliance while reducing data overhead. Additionally, we believe PETs have a place in enabling researchers to develop the next generation of digital biomarkers and diagnostic models without putting patient data at risk.
- **Defense:** PETs can be leveraged to improve data collaboration between allied countries, while minimizing risk to citizenry. PETs can add another layer of protection in data operations for government agencies or their contractors when performing machine learning or general data analysis. PETs can also

be leveraged to enforce data access and usage controls across various database systems within the defense sector.

- **Energy:** PETs can be leveraged to enable multi-party data consortia between energy companies and local authorities to monitor ESG goals without creating competitive risk.
- **International trade & commerce:** With PETs, we believe several multi-national companies would be able to better operate, while protecting the data sovereignty of a given nation's citizens. However, to unlock this use case, companies would need to better understand from regulators where the use of PETs to access data in a federated manner falls on the spectrum of x-border data transfer.

Topic 4.

We would value clearer guidance in relation to regulation or authorities on the following:

- A formal definition of anonymity with the use of PETs. At what point do regulatory bodies consider data activity to render individuals anonymous if PETs are in use?
- The introduction of federated data science as a mechanism to enable a version of Privacy Shield to protect x-border data transfer between the US and EU countries.

Topic 6.

There is concern in the field that the institutions investing most in the advancement, adoption, and development of PETs, namely Meta, Microsoft, Google, and IBM, are setting the agenda for how they will be used and in what contexts. We recommend developing an independent set of standards dictating when it is appropriate to leverage which kind of PETs and what thresholds are for different types of data, use cases, and levels of risk tolerance in a given enterprise. These would not be legal mandates, but rather would allow businesses to make informed decisions about how the data they've collected is leveraged as opposed to accepting the parameters laid out by these companies.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Brave Software

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

RFI Response: Privacy-Enhancing Technologies

Organization: Brave Software

Organization Type: Industry

Authors: Shivan Kaul Sahib (Privacy Engineer and PM), Pete Snyder (Senior Privacy Researcher and Director of Privacy)

Date: July 8, 2022

Brave Software welcomes the opportunity to make brief comments in response to the Office of Science and Technology Policy's request for information on Advancing Privacy-Enhancing Technologies to help inform development of a national strategy on privacy-preserving data sharing and analytics.

Comments

Risks with PETs that rely on centralization (topics 2 and 7)

Some privacy-enhancing schemes rely on centralization of user data and resources to operate. For example, Google's [Signed Exchanges](#) [1] and [Private Prefetch Proxy](#) [2] proposals improve privacy for users by making Google's cache servers sit in-between the user and the website; the website owner does not come to know the user's IP address but the cache server learns all websites a user would potentially connect to. Privacy-preserving protocols like these centralize the collection of sensitive user data by one entity which then hides it from other entities. These kinds of privacy-enhancing technologies are problematic because they only improve privacy if you absolutely trust the centralizing party. If not, they are strictly worse for privacy. In addition, such schemes create a single point of privacy failure.

Even when the sensitive user data being collected is encrypted and the centralizing parties cannot view the raw data, which is the case with several new [privacy-preserving data analytics schemes](#) [3] that rely on multi-party computation, there are still concerns around cost of operation. Schemes that rely on multi-party computation tend to be extremely expensive and complex to operate, effectively ruling out deployment by smaller organizations. In effect, this means that centralization increases and economic competition decreases when it comes to deploying PETs. **To alleviate these concerns, there should be a focus on researching and developing PETs that can enhance privacy while not being prohibitively expensive or complex to deploy.**

Client telemetry (topics 3 and 6)

Browsers and other user clients rely on sending telemetry signals to a backend server in order to identify problems, detect breakages and inform product decisions. However, there is a risk of sending potentially-sensitive information about a user in these signals. There is active work being done in the [Internet Engineering Task Force \(IETF\)](#) [4] to standardize protocols that enable privacy-preserving data analytics, particularly in the [Privacy Preserving Measurement \(PPM\)](#) [5] and [Oblivious HTTP Application Intermediation \(OHAI\)](#) [6] working groups. Also note that some browsers like Brave have already deployed [privacy-preserving analytics systems](#) [7].

One of the protocols being discussed in the Privacy Preserving Measurement (PPM) IETF working group is Distributed Secret Sharing for Private Threshold Aggregation Reporting (STAR) ([standardization draft](#) [8], [research paper](#) [9]), which aims to provide k-anonymity privacy guarantees for client-submitted data using a combination of secret sharing and Oblivious Pseudorandom Functions (OPRFs). The focus of STAR is privacy that is cost-effective and easy to deploy.

Privacy-preserving client telemetry is an important area for research and development that often relies on multiple parties, and it's important that this work happens in open standardization bodies. Open source development of privacy-respectful protocols that prevent centralization should be encouraged.

Global Privacy Control (topics 4 and 5)

[Global Privacy Control \(GPC\)](#) [10] is a signal that allows users to notify websites and businesses of their privacy preferences when it comes to selling or sharing data. It is implemented in the Brave browser as well as the DuckDuckGo Privacy Browser, Mozilla Firefox, EFF Privacy Badger and others. GPC is recognized as a legitimate Do Not Sell request on behalf of a user according to the [California Consumer Privacy Act \(CCPA\)](#) [11]. **Privacy regulations should make it easy for users to indicate preferences similar to how CCPA explicitly supports the GPC signal. In general, legislation should allow privacy-focused client software to make privacy-relevant assertions on behalf of their users. User signals like Global Privacy Control should be supported on a federal level.**

References

1. “Signed Exchanges (SXGs).” *web.dev*, <https://web.dev/signed-exchanges/>. Accessed 7 July 2022.
2. “Private prefetch proxy in Chrome.” *Chrome Developers*, 11 May 2022, <https://developer.chrome.com/blog/private-prefetch-proxy/>. Accessed 7 July 2022.
3. Thomson, Martin. “Privacy Preserving Attribution for Advertising.” *The Mozilla Blog*, 8 February 2022, <https://blog.mozilla.org/mozilla/privacy-preserving-attribution-for-advertising/>. Accessed 7 July 2022.
4. *IETF | Internet Engineering Task Force*, <https://www.ietf.org/>. Accessed 7 July 2022.
5. “Privacy Preserving Measurement (ppm).” *IETF Datatracker*, <https://datatracker.ietf.org/wg/ppm/about/>. Accessed 7 July 2022.
6. “Oblivious HTTP Application Intermediation (ohai).” *IETF Datatracker*, <https://datatracker.ietf.org/wg/ohai/about/>. Accessed 7 July 2022.
7. “Privacy-Preserving Product Analytics (P3A).” *Brave Browser*, 22 August 2019, <https://brave.com/privacy-preserving-product-analytics-p3a/>. Accessed 7 July 2022.
8. “draft-dss-star-00 - STAR: Distributed Secret Sharing for Private Threshold Aggregation Reporting.” *IETF Datatracker*, 7 March 2022, <https://datatracker.ietf.org/doc/draft-dss-star/>. Accessed 7 July 2022.
9. “[2109.10074] STAR: Secret Sharing for Private Threshold Aggregation Reporting.” *arXiv*, 21 September 2021, <https://arxiv.org/abs/2109.10074>. Accessed 7 July 2022.
10. *Global Privacy Control — Take Control Of Your Privacy*, <https://globalprivacycontrol.org/>. Accessed 7 July 2022.

11. "California Consumer Privacy Act (CCPA) | State of California - Department of Justice - Office of the Attorney General." *California Department of Justice*, <https://oag.ca.gov/privacy/ccpa/#heading7b>. Accessed 7 July 2022.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

BurstIQ

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

RFI: Request for Information on Advancing Privacy-Enhancing Technologies

Response by:

Frank Ricotta, Chief Executive Officer
Wendy Charles, PhD, Chief Scientific Officer
Robert Lubeck, Chief Growth Officer

BurstIQ, Inc.
9635 Maroon Circle, #310
Englewood, CO 80112

QUESTIONS:

1. Specific research opportunities to advance PETs:

Response:

We believe that the current generation of PET is fast approaching its event horizon. Some technical advances may extend the horizon based on the current approach that requires further research and new methods that will change the fundamental paradigm.

Specific topics include:

Privacy Enhanced Data: There exists a fundamental flaw in the underlying premise of PET solutions. Many proposals focus on systems and not on the data. These systems erroneously assume that data does not inherently contain any intelligence. Therefore, organizations take a traditional systems approach to data protection by controlling access to the data, where/how data may be processed, and/or techniques that protect sensitive information through anonymization. However, we believe that a significant untapped area for further research should focus on a data-centric approach with Privacy-Enhanced Data (PED). At BurstIQ, we describe this as “smart data” or “trusted data.” The foundation of PED is based on the concept that data can become “self-aware data objects.” In short, PED is a new data construct that fuses context

and trust attributes with the data itself into a new data object instead of managing these attributes in the surrounding systems. With data attributes such as metadata, rights, source, origin, chain of custody, time, place, and use permissions, PEDs can then provide key parameters to PET solutions to drive their behavior. Several methods can maintain the integrity of a PED and verify key attributes. Specifically, blockchain methodology offers a solid foundation for signing/attesting PEDs and tracking ownership, use, and authentication. Please review our Smart Data White Paper at <https://burstiq.com/smart-data-white-paper/> or by requesting the PDF from info@burstiq.com to learn more about PED capabilities.

Generative Adversarial Networks (GANs): GANs are a form of artificial intelligence (AI) where at least one AI network focuses on learning while at least one other AI network focuses on judging the behavior of the first. While not necessarily designed for privacy protections per se, GANs support how data is used—or not misused—in intelligent systems.

Extended trust to the edge: We believe there is a lot of promise around W3C digital identity frameworks with trust anchors, W3C Verifiable Credentials, and non-fungible tokens (NFTs) related to individual privacy and data ownership.

Scalability: Additional research is needed to improve the scalability of many PETs. For example, while zero-knowledge proofs and homomorphic encryption have the potential to extend the current approach event horizon, the technologies are currently too slow and immature to gain wide-scale adoption (Sharma et al., 2020; Tomaz et al., 2020).

2. Specific technical aspects or limitations of PETs:

Response:

We advise against overly optimistic utilization of any technology as a panacea for privacy-related issues. Instead, advocate for a layered privacy approach in the manner organizations should approach cybersecurity.

Concerning de-identification, synthetic data transformation offers an excellent alternative to identifiable data for hypothesis generation or feasibility analyses; however, synthetic data is not typically sufficiently accurate to generate diagnostic/treatment algorithms (Raghunathan, 2021).

Further, tokenizing attribution fields offer another potential option. While this approach can maintain statistical significance, current techniques require the creation of duplicate data sets.

Last, these methods' next evolution should be real-time based on a trusted presentation driven by access, ownership, and permissions, which requires PED as a foundational capability.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:

Response:

PETs are particularly beneficial for providing individuals' health-related data for the healthcare and life sciences industries and protecting intellectual property in any industry.

Elaborating on the *health-related research industry*, researchers frequently obtain de-identified health information from electronic health record systems, data repositories, or data marketplaces. However, de-identified data create challenges for data linking and (by nature of being de-identified using the HIPAA Safe Harbor provision in 45 CFR § 164.514(b)(2)) lack dates of service necessary for assessing longitudinal trends for an individual or population. For these circumstances, we recommend tokenizing attribution fields.

Federate learning is also promising PET for any collaboration involving the need to share data or algorithms among collaborators without revealing intellectual property. As an example of these solutions, the MELLODDY Project is a collaboration of 10 pharmaceutical companies that share algorithms to improve the predictive capacity of molecular compounds for drug discovery (Burki, 2019). The partnership uses the Owkin blockchain to maintain molecular compounds in each pharmaceutical company's own servers and refine the algorithms using each's private data without sharing proprietary molecular compounds with each other. Only the algorithm is shared using a blockchain-based federated learning protocol.

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs:

Response:

First, we'd like to caution legislators that privacy-preserving technologies alone are not sufficient to protect individuals' sensitive information. Specifically, the concept of de-identification has become largely an illusion, with consideration that data scraping and matching require little more than basic programming and analytics skills to link data sets (Chiauzzi & Wicks, 2019; Narayan & Felten, 2014). While PETs can reduce exposure to individually-identifiable information, we advocate for modifications of the current federal HIPAA Privacy Rule to include more comprehensive privacy protections than the outdated list of HIPAA identifiers at 45 CFR § 164.514(b)(2)(i) to factor capabilities of data linking with modern computing.

Next, as a health information technology provider, we request more protections for the privacy of 365,000 health-related apps not subject to the HIPAA Privacy and Security Rules (Help Net Security, 2021). We applaud the progress of the American College of Physicians and the American Telemedicine Association in creating the *Digital Health Assessment Framework (2022)* and the Federal Trade Commission's (FTC) *Mobile Health App Best Practices (2016)*.

While these are valuable guidelines, there is no requirement for software developers to use these guidelines, resulting in significant vulnerabilities to the privacy of health information for an estimated 23 million U.S. users (Help Net Security, 2021). Unfortunately, there are few consequences for breaches, and there has been no FTC enforcement of the 2009 Health Breach Notification Rule to date (FTC, 2021). We urge the Federal Trade Commission to implement and enforce more stringent privacy protections for health information not subject to HIPAA regulations.

5. Specific laws that could be used, modified, or introduced to advance PETs:

Response:

First, we would like to use this section to dispel the falsehood that blockchain-based privacy-preserving strategies (as a blanket statement) are incompatible with the General Data Protection Regulation (GDPR) or U.S. privacy requirements. Blockchain compatibility with laws or statutes that mandate the “right to be forgotten” or the “right to be deleted” depends entirely on the platform and programming utilized. Several published methods provide direction for programming methods for actual data or obfuscation that meet the letter and spirit of these regulations (e.g., Herian, 2020; Yang et al., 2020). Further, a European Commission report—written in part by one of us (WC) through the European Observatory Blockchain Observatory and Forum—determined that health-oriented blockchains can meet GDPR requirements if designed appropriately (Livitckaia et al., 2022). For an extensive overview of blockchain compliance with U.S. regulations for health-oriented research, please read a journal article authored by one of us (Charles et al., 2019, <https://doi.org/10.3389/fbloc.2019.00018>).

Second, there is a growing patchwork of privacy statutes in U.S. states (Lively, 2022). While there is considerable overlap in the nature of privacy protections, the unique nature of some states’ privacy requirements involves extensive software customization to meet these unique requirements. We advocate for a federal privacy act, similar to GDPR, that unifies the state-level statutes to allow for more straightforward programming and consistent approaches to designing PETs.

6. Specific mechanisms not covered above that could be used, modified, or introduced to advance PETs:

Response:

We advocate for creating PET technology-specific standards that would allow organizations to design and implement PETs responsibly and with more confidence. The National Institutes of Standards and Technology has created a Privacy-Enhancing Cryptography Project (2017) and published an excellent overview of cryptography-based PET (Brandão & Peralta, 2021) as a

starting point. We encourage the development of specific standards that software developers could utilize for best practices.

With gradual advances made in quantum computing—that will render some passwords and hashing algorithms obsolete in 5-10 years (Fedorov et al., 2018), we also advocate for standards that will reduce the likelihood that quantum computing could defeat some of the current PET approaches.

Finally, we would like to see significant advancements in digital identity standards to include technical constructs extending to data “ownership” and revocation methods. Also, we advocate for regulatory frameworks that clearly state that individuals have the right to “own” and manage their digital identities and data—especially their health data.

7. Risks related to PETs adoption:

Response:

We view centralized governance as the primary risk for PETs. Specifically, data privacy remains at risk when organizations insist on owning and controlling individuals’ data with minimal oversight or transparency.

With the emergence of Web3 technologies built on distributed governance, there is a risk that regulatory agencies not infringe on this market innovation and inadvertently encourage centralized data control. For more information on trusted governance structures, please review the section on governance at <https://burstiq.com/trust-as-a-differentiator-protecting-human-data-in-your-products-and-services-burstiq/>.

Government agencies have a role in creating frameworks for proper data rights and appropriate uses. For example, when individuals are presented with options for conducting business with an organization or using a website, we advocate for “opt-in” data permissions instead of the typical default of “opt-out.”

8. Existing best practices that are helpful for PETs adoption:

Response:

We advocate for thorough risk assessments and staff training, regardless of the technologies used for managing sensitive information. Organizations are also encouraged to perform ethical risk assessments (e.g., Allen et al., 2020; LaPointe & Fishbane, 2019) to identify vulnerabilities in particular populations, including the prospect that certain age groups, individuals who lack

access to broadband technology or sufficient computing devices, or individuals with physical or cognitive disabilities may lack access to PETs.

9. Existing barriers, not covered above, to PETs adoption:

Response:

There must be a clear delineation between uses of PETs and uses of private data. We recognize that government agencies cannot oversee all data utilization. However, agencies should provide standards and frameworks that drive interoperability, shift data ownership from Big Tech to individuals, and improve cybercrime enforcement.

10. Other information that is relevant to the adoption of PETs:

We advocate for a clear and consistent definition of “privacy-enhancing technology,” and what types of technology would be included in a solution framework.

Further, data rights and ownership should be foundational to the entire suite of PET solutions. Presently, there is a near-complete lack of clarity on this topic.

Last, AI should also be a key focus going forward with frameworks about AI ethics, bias, and digital twin civil rights as an extension of a person.

REFERENCES:

Allen, M., Vasiliu-Feltes, I., Ingraham, A., Mysore, I., Vaughn, M., & Charles, W. (2020). *Blockchain ethical design framework for healthcare*. Government Blockchain Association. <https://www.gbaglobal.org/blockchain-ethical-design-framework-for-healthcare/>

Brandão, L. T. A., & Peralta, R. (2021, November 3). *Privacy-enhancing cryptography to complement differential privacy*. <https://www.nist.gov/publications/privacy-enhancing-cryptography-complement-differential-privacy>

Burki, T. (2019). Pharma blockchains AI for drug development. *The Lancet*, 393(10189), 2382. [https://doi.org/10.1016/S0140-6736\(19\)31401-1](https://doi.org/10.1016/S0140-6736(19)31401-1)

Charles, W., Marler, N., Long, L., & Manion, S. (2019). Blockchain Compliance by Design: Regulatory Considerations for Blockchain in Clinical Research. In *Frontiers in Blockchain* (Vol. 2). <https://doi.org/10.3389/fbloc.2019.00018>

Chiauzzi, E., & Wicks, P. (2019). Digital trespass: Ethical and terms-of-use violations by researchers accessing data from an online patient community. *Journal of Medical Internet Research*, 21(2), e11985. <https://doi.org/10.2196/11985>

Digital Health Assessment Framework, V1. (2022, June 24). <https://confluence.external-share.com/content/94649754-10cd-4803-9b67-91c784673d70>

Federal Trade Commission. (2016, April 4). *Mobile health app developers: FTC best practices*. <http://www.ftc.gov/business-guidance/resources/mobile-health-app-developers-ftc-best-practices>

Federal Trade Commission. (2021, September 15). *Statement of the Commission on breaches by health apps and other connected devices*. https://www.ftc.gov/system/files/documents/rules/health-breach-notification-rule/statement_of_the_commission_on_breaches_by_health_apps_and_other_connected_devices.pdf

Fedorov, A. K., Kiktenko, E. O., & Lvovsky, A. I. (2018). Quantum computers put blockchain security at risk. *Nature*, 563(7732), 465–467. <https://doi.org/10.1038/d41586-018-07449-z>

Help Net Security. (2021, February 12). *mHealth apps consistently expose PII and PHI through APIs*. <https://www.helpnetsecurity.com/2021/02/12/mhealth-apps-expose-pii-phi-through-apis/>

Herian, R. (2020). Blockchain, GDPR, and fantasies of data sovereignty. *Law, Innovation and Technology*, 12(1), 156–174. <https://doi.org/10.1080/17579961.2020.1727094>

LaPointe, C., & Fishbane, L. (2019). *The blockchain ethical design framework*. Georgetown University. <https://beeckcenter.georgetown.edu/wp-content/uploads/2018/06/The-Blockchain-Ethical-Design-Framework.pdf>

Lively, T. K. (2022, June 9). *US state privacy legislation tracker*. <https://iapp.org/resources/article/us-state-privacy-legislation-tracker/>

Livitckaia, K., Charles, W., Larrañaga Piedra, U., Niemerg, M., Hasselgren, A., & Papadopoulou, E. (2022). *Blockchain applications in the healthcare sector*. European Union Blockchain Observatory & Forum: An initiative of the European Commission. https://www.eublockchainforum.eu/sites/default/files/reports/eubof_healthcare_2022_FINAL_pdf.pdf

Narayan, A., & Felten, E. W. (2014). *No silver bullet: De-identification still doesn't work*. Princeton University. <https://www.cs.princeton.edu/~arvindn/publications/no-silver-bullet-de-identification.pdf>

National Institute of Standards and Technology | Computer Security Resource Center. (2017, January 3). *Privacy-enhancing cryptography*. CSRC | NIST. <https://csrc.nist.gov/projects/pec>

Raghunathan, T. E. (2021). Synthetic data. *Annual Review of Statistics and Its Application*, 8(1), 129–140. <https://doi.org/10.1146/annurev-statistics-040720-031848>

Sharma, B., Halder, R., & Singh, J. (2020, January). Blockchain-based interoperable healthcare using zero-knowledge proofs and proxy re-encryption. *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), Bengaluru, India.
<https://doi.org/10.1109/comsnets48256.2020.9027413>

Tomaz, A. E. B., Nascimento, J. C. D., Hafid, A. S., & De Souza, J. N. (2020). Preserving privacy in mobile health systems using non-interactive zero-knowledge proof and blockchain. *IEEE Access: Practical Innovations, Open Solutions*, 8, 204441–204458.
<https://doi.org/10.1109/access.2020.3036811>

Yang, J., Onik, M. M. H., & Kim, C.-S. (2020). Blockchain technology for protecting personal information privacy. In M. Ahmed (Ed.), *Blockchain in data analytics* (pp. 122–144). Cambridge Scholars Publisher.
https://books.google.com/books/about/Blockchain_in_Data_Analytics.html?id=z_zLDwAAQBAJ

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

**Canetti, Ran; Kaptchuk, Gabe; Reyzin, Leonid; Smith, Adam; and Varia,
Mayank**

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to the RFI on Advancing Privacy-Enhancing Technologies

Ran Canetti, Gabe Kaptchuk, Leonid Reyzin, Adam Smith, and Mayank Varia
Boston University (academic institution)

July 8, 2022

PETs provide an extremely powerful and versatile toolbox for controlled processing of data, while keeping the data itself hidden. Consequently, PETs allow for a radical rethinking of the possible uses of data, providing fine-grained control over which functions of data are to be exposed (and to whom) and which are to be kept hidden. When properly and responsibly used, this power of PETs can greatly advance democracy for the benefit of all individuals and communities. However, when used carelessly and irresponsibly, PETs can also cause great harm. Therefore, the adoption and development of PETs should be done in a careful and principled way, while making sure to adequately educate the relevant constituents in the power and dangers of using PETs, and while taking care to provide adequate checks and balances that are appropriate to each and every use case.

More specifically, used properly, PETs can cut the Gordian knot that pits transparency and accountability against privacy and secrecy in a seemingly inherent way: It provides ways to obtain information that is inherently derived from private data but is crucial for good governance, while keeping the data private and at the same time providing public assurance to the correctness and relevance of the obtained information. But when deployed improperly or recklessly, PETs risk exacerbating inequities, reducing privacy, diminishing transparency and accountability, and undermining public confidence in government agencies and their data products. Furthermore, the power of PETs inevitably leads to a series of policy questions about *when* and *how* PETs should be deployed.

The goal of this report is to advise government agencies regarding the processes and regulations that need to be put in place in order to incentivize responsible development and adoption of PETs, while minimizing the risk of harmful deployments of PETs. For that purpose, we first elaborate on the potential benefits, as well as on the dangers and pitfalls. Then we proceed with our recommendations (which are derived from said benefits and dangers). Specifically, we answer four of the questions listed in the Government's Request for Information:

- The democratizing promise of public sector use of PETs to achieve socially beneficial outcomes. (Addressing Q3 about benefits from the adoption of PETs.)
- The individual, group, and societal harms that can arise from misuse of PETs. (Addressing Q7 about risks related to PETs adoption.)
- The benefit of separating the policy questions about a system's goals and values from the technical questions about how to combine PETs to realize it. (Addressing Q8 about existing best practices that are helpful for PETs adoption.)
- A path forward toward developing and deploying PETs that serve the public good. (Addressing Q9 about overcoming sociotechnical and trust barriers, and improving equity.)

About the Authors

We are a group of cryptography and privacy researchers at Boston University who develop and analyze privacy enhancing technologies (PETs). Our areas of expertise cover the mathematical and cryptographic foundations of PETs (including [1–16]), the practical development and deployment of systems involving PETs [17–25], the social and policy contexts in which these technologies are deployed [26–32], and serving on committees that advise national statistics organizations [33, 34]. Members of our group have been at the forefront of the envisioning of PETs, as well as the main algorithmic and analytical tools that build PETs, over the past 30 years.

We are encouraged to see the Office of Science and Technology Policy, the National Artificial Intelligence Initiative Office, and Networking and Information Technology Research and Development take an active interest in furthering the adoption of PETs. We appreciate and welcome the opportunity to provide important information and context about PETs based on our decades of experience in cutting-edge cryptographic research from the early 1990s through today.

Q3. The Promise of Public Sector PETs

In the modern information era, data is powerful. At its best, the explosion in data availability and the massive cost reduction for data analysis promise to bring about a more efficient and democratic world. For governments, data-driven insights can lead to evidence-based policymaking [35] that improves the government’s ability to efficiently allocate resources and respond to the needs of its citizens. In the private sector, services offered by companies to people can be effortlessly customized based on individual traits and habits. However, this idealistic promise is undercut by several real-world risks like data breaches, bulk surveillance, and a massive network of data brokers [36] that collectively solidify existing power and exacerbate inequities.

As a consequence, we describe in this section **the democratizing power in *not* having data**. PETs allow for the possibility of having our cake and eating it too: they enable the social benefits of data-driven policymaking while mitigating the harms of social control, censorship, and authoritarianism. These technologies are coming to maturity just as they are most needed and are now fast, simple, and expressive enough to use in many public sector and civil society applications.

To showcase the democratizing and empowering promise of PETs in government and civil society applications, we briefly describe a few of the applications of PETs that we have designed, developed, and deployed [37] together with domain experts and impacted communities. We focus on the human rights and civic values that were addressed using PETs.

- **Transparently describe government actions** [19]: We participated in the design of PETs for the 2020 Decennial Census. The implementation of differential privacy provides strong privacy guarantees, as opposed to heuristic ones. Another major advantage relative to prior disclosure limitation techniques is the ability to *publish the specific methodology performed by the government*, so that researchers can account for this distortion; this has brought the complex debate about statistical disclosure limitation into the public discourse [38].
- **Promote grassroots initiatives and volunteer activism** [20–22]: Jointly with the Boston Women’s Workforce Council, we deployed PETs to measure the gender and racial wage gap in the greater Boston area based on real salary data of about 1/6 of the Boston workforce. A key challenge was *designing an easy to understand and easy to use* system so organizations were comfortable contributing their payroll data toward a measurement that they all agreed would have social benefit, but whose value relied on a high rate of voluntary participation.

- **Provide a voice to all** [23]: Inspired by the #MeToo movement, we designed a system together with the non-profit organization Callisto that allows multiple survivors of a repeat sexual offender to be securely matched together so they can act with a stronger voice. Our biggest focus was to design a privacy-respecting system that *protects survivors’ assault reports from litigation or investigation discovery requests*, including a subpoena to Callisto itself.
- **Stop bulk surveillance** [25]: PETs allow financial institutions to better detect fraud and money laundering [39], which promotes public safety. Our fraud detection tool developed with Fiverity focuses on the ethical and legal obligations to *safeguard personal information of real people* (i.e., non-fraudulent accounts) even while searching for instances of fraud.
- **Promote constitutional rights** [26, 27]: PETs allow the court system to protect law enforcement’s sources and methods while simultaneously providing criminal defendants with their due process rights to inspect and challenge the resulting evidence gathered. A key challenge here was *integrating the use of PETs into the legal process* by which prosecutors and defense attorneys converge on a property that the defense wants the prosecution to prove.

In summary, PETs expand the Pareto frontier of data analyses that can be performed: they enable us to design new methods to conduct statistical analyses that simultaneously *enhance accuracy and quality, respect personal autonomy and consent choices, and enforce data minimization and purpose restrictions*. As a result, PETs offer the promise of conducting socially beneficial and privacy-respecting statistical analyses that would have been difficult—if not impossible—to compute otherwise. But we must also be prudent about the fact that PETs could be used to conduct socially detrimental and privacy-invasive calculations as well; we discuss these risks next.

Q7. The Risk of PETs: *When* Should PETs Be Deployed?

As we have explored above, there are an immense number of opportunities to deploy PETs that will serve the social good, making a clearly positive impact of society. As such, it might be tempting to deploy PETs wherever possible. However, **deploying PETs without careful consideration upfront**, but instead mimicking Silicon Valley’s “move fast and break things” mentality, will not only fail to realize the potential of PETs, but **could create tangible negative impacts for society**. In particular, we are concerned about the following negative impacts that can arise if PETs are deployed inappropriately or inconsiderately:

- **PETs Deployments that Reduce Privacy.** The title “privacy enhancing technology” is somewhat a misnomer, in that a typical understanding of the word “privacy” might lead the casual observer to assume the wrong properties of a PET. *It is very possible for a PET to reduce the privacy afforded to individuals and organizations when deployed—and, indeed, this will likely be the case if not deployed with care.* The reason for this is threefold:
 - (1) A single PET on its own may only address a subset of privacy considerations, and may in fact make some participants worse off than if the statistical analysis never occurred. For instance, a PET that provides only input privacy might calculate an accurate statistic but fail to provide disclosure limitation, leading to reconstruction attacks that reveal specific individuals’ data.
 - (2) The baseline privacy expectations of an individual, impacted group, or organization might be very high, such that *any* change to the system might actually reduce their privacy. For example, creating a differentially private release of a database will reduce

the privacy afforded to individuals with personal information in that database if the prior baseline is that no statistical releases from that database occurred.

- (3) PETs provide *fine-grained* notions of security. That is, privacy can be defined virtually arbitrarily, within the context of the application, and a PET meeting that definition of privacy can be constructed. As an illustrative example: several proposals for election security include a zero knowledge proof that each voter is included in the final vote tally, but by contrast it would be terrible to produce a zero knowledge proof that a particular citizen voted for a specific candidate.
- **Privacy Washing.** Using a PET to perform a harmful computation does nothing to mitigate the harm. Indeed, using a PET often serves to distract or obfuscate to the casual observer the harm that deploying the computation would cause them. Examples of this concern include proposals to deploy PETs to allow for bulk surveillance of encrypted data [40, 41], scan all cloud uploads against abusive content [42], and perform targeted advertising with private real-time auctions [43–46]. Large parts of the cybersecurity community and civil liberties groups have decried these proposals [47–52]—not because the cryptographic techniques were flawed, but because these specific proposals *obscure from the broader debate over whether these tasks should occur at all, raise questions about whose privacy is being protected and prioritized, and failed to provide public accountability and auditing comparable to existing non-PET systems.*
 - **Amplifying Market Concentration.** Many PET deployments (especially secure multi-party computation and differential privacy) allow organizations to learn insights from data to which they would not otherwise have access. While this facilitates some of the best applications of PETs such as data-driven policymaking, sharing insights in this way is not without risks. Industrial organizations can leverage access to these insights to solidify competitive advantages (*e.g.* training better machine learning models using private data); the cost of implementing PETs also adds barriers to entry for small companies. Government bodies can further solidify their control over the population (*e.g.* using private geolocation data to target police deployments in marginalized communities). Such deployments of PETs tend to *further consolidate power in monopolistic businesses or autocratic governments*, rather than uplift democratic principles.
 - **Losing the Public Trust.** Trust is built slowly and shattered quickly. Missteps in deploying PETs—either by choosing applications poorly, faulty implementations, or misunderstanding threat models—will likely prevent the carefully considered privacy-enhancing systems that would have the most positive impact. *People and organizations might conflate previous privacy-reducing PETs deployments with future privacy-enhancing PETs deployments*, simply because some of the underlying technical tools are shared and they are worried about rehashing previous debacles. In the worst case, a government misstep in deploying PETs could trigger a backlash that would prevent statistical agencies from access data that they currently have, as the misstep might be seen as a sign of ineptitude.

Q8. Specifying Policy Objectives and Assessing Democratic Values

In this section, we advocate for **regulation that establishes a process for vetting PETs used within the U.S. federal government that transparently specifies its intended functionality, and allows for democratic debate about potential risks and possible mitigations.**

With the excitement surrounding a number of successful deployments of PETs [53], it is sometimes easy to forget that PETs are *tools*—means to an end. *Asking if a PET is appropriate in a specific scenario is like asking if a new kind of asphalt might be helpful for building a highway.* While asphalt is an essential tool, the discussion hides the more important questions such as: who will be harmed and who will be helped by the road’s construction? These are questions of *power*, that must be answered in rigorous consultation with impacted communities [54, 55]. Only once these questions have been adequately settled can detailed engineering discussions start.

Put another way, cautious deployment of PETs requires asking the question “when should a PET be deployed.” Engaging with this question requires *specifying concretely* what information the envisioned system should release and what it should keep private, and evaluating the merits of such a specification. Importantly and perhaps counterintuitively, *this discussion can happen entirely independently of the actual PET techniques that would be used and combined to realize this specification.* Instead, the discussion should focus on the “input-output” behavior of the system, *i.e.* the information possessed by each party and what each party should learn after interacting with the system. Fortunately, researchers in cryptography and programming languages have already designed the principles necessary to follow this approach.

Existing practice: Specifying ideal functionalities. While discussing the potential deployment of a PET without ever mentioning the PET might seem impossible or too abstract, this is in fact a common practice in cryptography—albeit in a highly formalized context. To capture the wide range of security properties that a system under consideration should achieve (and also the properties that it doesn’t claim to have), cryptographers write an *ideal functionality, which is a formal specification of the behavior of a system* that abstracts away all low-level implementation details including which PET, if any, should be used. The ideal functionality can describe the system as though it were purely magic and can leverage fictional, perfectly trustworthy parties. A cryptographer will then prove that some particular set of PETs, properly interleaved, successfully produces a perfect imitation of the ideal functionality, *i.e.* creates something that is “as good as” the idealized specification and moreover is insatiable in the real world.

For example, an (informal) ideal functionality describing an end-to-end encrypted messaging system that scans for abusive content might specify that (1) users can send a messages to a service provider, along with an intended recipient of that message; (2) if the user’s message matches a list of prohibited and abusive content, the service provider learns the contents of the message. Otherwise, the service provider learns only the length and timing of the messages, but nothing about the contents; and (3) the service provider can choose to deliver the message to the intended recipient at a time of its choosing. Notice that this specification need never mention the word encryption. Indeed, encryption might be one tool used to realize this functionality, but other approaches could, in principle, exist. Moreover, it is possible to have a policy discussion on the merits of such a system purely by considering the ideal functionality. And by removing any mention of cryptography, the ideal functionality is easier for the public to understand and debate.

A path forward toward the design of PETs. The practice of writing these ideal specifications has technical value [2–4]. Additionally, we believe that it has *tremendous untapped social value.* In some sense, the practice of writing and scrutinizing these specifications independently of the underlying implementation modularizes the process of analyzing deployments of PETs. In particular, it *cleanly separates the social question of what goals a PET deployment should accomplish*—which society should collectively decide through the democratic process—*from the technological question of how these goals should be met,* which can be safely relegated to cryptographers and information

security experts once the values questions have been debated and decided.

There are a number of reasons to promote this two-step process toward the design of PET-enabled systems, including:

- *Receiving feedback and constructive criticism from everyone in a democratic society, including marginalized voices.* History has shown time and again that systems (computational or otherwise) that may have good outcomes for some people can also yield disparate impacts on other groups in the community [54, 55].
- *Inserting useful “friction” into the design process* [56, 57]. As mentioned above in our discussion of risks, making everything computable isn’t always a good thing. Whereas people could point in the past toward the “impossibility” of a computation in order to stop it, the existence of PETs forces a more careful introspection upfront about which data analyses will promote rather than hinder human rights.
- *Providing transparency and legitimacy of government actions.* It is particularly important to describe upfront how a PET-enabled system will work (for instance, through a System of Records Notice or similar) because by its very nature the system cannot be radically transparent during use. In fact, auditing the design of the system is just one part of transparency; we must also allow for auditing the execution of the system as we describe next.

Q9. Doing it Right: *How* Should PETs Be Deployed?

If PETs are to serve democratic values, the deployment of PETs must follow democratic principles. Making it easy for people to trust the fidelity of a PET deployment is critical to making the deployment successful. Unfortunately, there are numerous potential pitfalls into which the designer of PET can fall that will quickly shatter any trust that has been developed. To that end, we **suggest a series of guiding principles that can mitigate the potential risks of PET deployments and increase the chances that they will reach their full democratic promise.**

Public Development Process. *Choices about the particular techniques used in a PETs deployment encode the values of its designers.* A particular ideal functionality can be realized with many different combinations of individual privacy enhancing technologies. Choosing one PET over another—or even using different mathematical assumptions or parameters to realize a single PET—can have concrete and disparate consequences for the level of privacy that the PET deployment provides. For example, it is possible to realize secure multiparty computation using either an honest-majority assumption, in which only a colluding majority of the parties participating in the protocol could compromise the privacy of the computation, or a dishonest-majority assumption, in which the data will remain private if even a single party behaves honestly. Similarly, differential privacy puts an upper bound on the amount of information that can leak about an individual during a statistical disclosure, but the parameter choices meaningfully change the qualitative impact of this bound on individuals and communities.

When developing a new PET deployment, it is critical that the choice of techniques and parameters be made *publicly*. NIST has long followed this process with its standardization efforts [58–61], and successfully garnered the trust of both security experts and the general public. At a minimum, having public process ensures that experts can identify vulnerabilities before it is too late. More importantly, disclosing the techniques in use *allows the public to debate if the values encoded in those techniques are appropriate for the application.*

Public Verifiability. The validity of PET deployments should be verifiable—even by an adversarial party who didn’t participate in any aspect of the original design or execution, and who has vested interest in demonstrating the opposite. Because PETs are unfamiliar, mathematically complex, and intrinsically *hide something*, a careless deployment of PETs may be used to cast doubt on the results of the system. But, with careful construction, it is possible to *deploy PETs that affirmatively offer public verification of their proper execution and that can provide evidence to counteract any accusation that might discredit the system.*

We believe that there are two crucial types of verifiability that are *necessary* in order to make PET deployments trustworthy:

- **Results Verifiability.** *Privacy must not come at the expense of public oversight and accountability.* Systems that do not leverage PETs can be carefully checked by data scientists, and the results that they produce can be verified by simply looking at the underlying data. In principle, any member of the public who doubts a result could get access to the data and check on their own. Many PET deployments, on the other hand, make this kind of radical transparency impossible for experts and the general public alike—indeed, this is a core motivation for using PETs. As such, PET deployments should use techniques to ensure the results are convincing to *everyone*, like verifiable computation and zero knowledge proofs.

We remark that this challenge is not new; indeed, election officials have grappled for decades with the challenge of providing public verifiability of the election tallying process, while also guaranteeing voter privacy and protecting against voter coercion. Privacy and anti-coercion threats necessitate that the election system make it impossible to check how each individual person voted, thereby necessitating more onerous methods to provide public verifiability such as risk-limiting audits, cameras to observe the tallying process, and so on. Systems employing PETs must similarly include methods to demonstrate correct behavior and obtain public legitimacy, and new research may be required to achieve this goal.

- **Code Auditability.** It is also necessary to follow good software development practices when developing and deploying PETs: source code should be publicly available and written in a way that meaningful auditing is feasible. The best way to ensure that a software package does not contain security-critical vulnerabilities is to *allow experts ample opportunity to attempt to break it.* Additionally, elegance and simplicity should be the watchwords of PETs software development. If it is too difficult or painful to engage with the software, none of the necessary audits will happen and the organization deploying the PET may have a false sense of security.

Plan for the Long Term. Because PETs limit the ways in which data can be used, it is critical to think through the full lifetime of the data before committing to a particular PET deployment. The use of PETs can be *irrevocable*, so making decisions based only on short-term needs may make it impossible to accomplish critical tasks in the long term. There are three ways in which we anticipate that foresight is necessary before making an initial PET deployment or modifying a system in operation:

- **Understanding Irrevocable Decisions.** *PET systems require careful consideration up-front about which types of extensibility are explicitly desired versus explicitly forbidden.* For example, once data is encrypted, then losing the decryption key is tantamount to deleting the data. This can be a blessing if you *want* to delete the data, for instance when decommissioning a system at its end-of-life, but it can be a substantial issue otherwise. More sophisticated PETs can also embed irreversible decisions. For example, if a survey is conducted using privacy preserving techniques, and a desired correlation is not computed, it may be infeasible to

go back and collect the data again. . . though once again, we stress that *this may be a desirable feature in some contexts to limit mission and scope creep*. Alternatively, if data is collected using local differential privacy with a particular privacy budget, then later decreasing the amount of noise in the data set (in an effort to get more accurate statistics) would be a complicated process that may require new data collection.

- **System Changes Impacting Public Trust.** When a deployment of PETs does permit changes after initial deployment, any use of this power to change how the system works may *shatter the public’s trust in the PET—and in the organization running the privacy-enhanced system*. For example, a government agency might announce that it plans to collect data and only release the results using (central) differential privacy under some published privacy parameter. If the agency then reduces the privacy in the release after data collection has occurred, this would constitute a breach of the norms surrounding the use of the data. As such, *retaining the technical flexibility to change the system later is not always preferable to making irrevocable decisions upfront*, as it brings about its own problems. Recent work has explored how to limit access to this power altogether and to ensure public verifiability for any modifications made to a PET system in use [28,62].
- **Composing with Future PET Deployments.** As PETs deployments proliferate, *there will be a desire to have deployments interoperate*, which may be impossible without prior planning. For example, the 2020 Census uses differential privacy in its statistical disclosures. The current deployment does not support verifying that the randomness injected into the results was selected honestly—it is technically possible that a malicious data operator manipulated the results by carefully selecting the “random” values given to the disclosure avoidance algorithm. While it is in principle possible to prove that the randomness used was selected independently of the census data (using zero-knowledge to prevent any further privacy leaks), the current system design does not support the composition of the Census differential privacy release and zero-knowledge proofs—and there may be no easy way to retrofit the system later.

The challenges described above are generally not technically complicated; the necessary crypto and privacy tools to avoid these bad outcomes are generally well-understood and sufficiently adaptable to address most realistic deployment scenarios. For instance, a simple countermeasure to the Census issue described above is to post cryptographic commitments to the sensitive data in order to maintain the ability to prove properties about it in the future. Instead, *the rate-limiting resource is human foresight and patience*.

Explainability. Especially for public facing PET deployments, ensuring that the deployment is explained clearly—both to technical and non-technical audiences [63]—is critical to engendering trust. Several recent works in the domain of usable security by us [29,30,64] and others (*e.g.*, [65–68]) have highlighted the need for clearer communication about PETs. Without a clear explanation, people might not trust the system with their data when appropriate, or they might expect more privacy guarantees that the system will actually provide. *Misalignment of expectations and reality will cause distress and mistrust*. As such, developing clear communication about a technological deployment is critical to its long-term success.

Prioritizing Privacy & Functionality over Efficiency. It is prudent to take a realistic evaluation of the efficiency needs of PET deployments. We acknowledge that PETs are not at the point where they can be used for high-performance, low-latency, real-time systems. But those statistical calculations are rare. Particularly for government deployments of PETs, having protocols run for

hours or days may be *completely practical*. And when compared to the legal burdens or outright impossibility of performing a valuable analysis otherwise, the cost of the computing power is often downright cheap.

Research literature into PETs offer a variety of different options for balancing the privacy afforded to the data, the functionality the PET provides, and the concrete efficiency of the resulting system. It is important to be realistic in PET deployments about the level of efficiency that is actually tolerable, as opposed to what would be ideally desired. Prioritizing the efficiency of the system often results in *worse* privacy guarantees and poorer functionality—recall that the most efficient system is one that preserves no privacy at all. Being pragmatic about the acceptable performance in the context of a particular application radically changes the space of PETs that are practical and ready for deployment *today*.

Conclusion

In conclusion, we summarize the most important takeaways from this report:

- Although data is powerful, it can also be toxic. There is thus great value in *not* having access to data so that it cannot be altered, breached, censored, surreptitiously surveilled, or otherwise abused.
- At its best, PETs offer the promise to improve government transparency and oversight, promote grassroots initiatives, provide a voice to all communities, mitigate bulk surveillance, and promote human rights.
- Deploying PETs carelessly can lead to tangible harms for society, such as reducing the privacy of some people or entrenching existing power structures. We offer several principles to mitigate these risks and allow PET deployments to reach their full promise.
- We recommend distinguishing the objectives of a PET system from its cryptographic implementation. Providing ideal specifications allows everyone to contribute toward the democratic discourse about the goals and values that a PET system *should* imbue.
- All aspects of PET deployments should be verifiable. Furthermore, the verification process should be public and simple-minded, so that even those parties who have a vested interest in demonstrating the opposite will have no choice but to be convinced in the correctness of the results. The public must be able to view the initial PET design and implementation, and separately also verify that it is producing correct results in use.
- When designing PET systems, one must decide upfront which aspects of the system are immutable (and therefore limits scope creep) versus extensible (and therefore adaptable as organizational boundaries and trust decisions change). A “middle ground” here is to allow for changes to be made, but in a manner that enforces public oversight and auditability.
- The ability to explain to the public what a PET algorithm does is crucial, even if it requires sacrificing some accuracy or efficiency. Misalignment of expectations and reality will lead to distress and mistrust.
- For many government applications considering PETs, efficiency is not the main bottleneck to adoption today. Instead the main challenge is human planning and patience throughout the process of PET design, evaluation, deployment, and oversight.

References

- [1] Mihir Bellare, Ran Canetti, and Hugo Krawczyk. Keying hash functions for message authentication. In Neal Koblitz, editor, *Advances in Cryptology – CRYPTO’96*, volume 1109 of *Lecture Notes in Computer Science*, pages 1–15, Santa Barbara, CA, USA, August 18–22, 1996. Springer, Heidelberg, Germany.
- [2] Ran Canetti. Universally composable security: A new paradigm for cryptographic protocols. In *42nd Annual Symposium on Foundations of Computer Science*, pages 136–145, Las Vegas, NV, USA, October 14–17, 2001. IEEE Computer Society Press.
- [3] Ran Canetti. Universally composable security. *J. ACM*, 67(5):28:1–28:94, 2020.
- [4] Ran Canetti, Alley Stoughton, and Mayank Varia. EasyUC: Using EasyCrypt to mechanize proofs of universally composable security. In Stephanie Delaune and Limin Jia, editors, *CSF 2019: IEEE 32st Computer Security Foundations Symposium*, pages 167–183, Hoboken, NJ, USA, jun 25–28 2019. IEEE Computer Society Press.
- [5] Ran Canetti and Hugo Krawczyk. Analysis of key-exchange protocols and their use for building secure channels. In Birgit Pfizmann, editor, *Advances in Cryptology – EUROCRYPT 2001*, volume 2045 of *Lecture Notes in Computer Science*, pages 453–474, Innsbruck, Austria, May 6–10, 2001. Springer, Heidelberg, Germany.
- [6] Yevgeniy Dodis, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In Christian Cachin and Jan Camenisch, editors, *Advances in Cryptology – EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 523–540, Interlaken, Switzerland, May 2–6, 2004. Springer, Heidelberg, Germany.
- [7] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *TCC 2006: 3rd Theory of Cryptography Conference*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284, New York, NY, USA, March 4–7, 2006. Springer, Heidelberg, Germany.
- [8] Christina Garman, Matthew Green, Gabriel Kaptchuk, Ian Miers, and Michael Rushanan. Dancing on the lip of the volcano: Chosen ciphertext attacks on apple iMessage. In Thorsten Holz and Stefan Savage, editors, *USENIX Security 2016: 25th USENIX Security Symposium*, pages 655–672, Austin, TX, USA, August 10–12, 2016. USENIX Association.
- [9] Benjamin Fuller, Mayank Varia, Arkady Yerukhimovich, Emily Shen, Ariel Hamlin, Vijay Gadepally, Richard Shay, John Darby Mitchell, and Robert K. Cunningham. SoK: Cryptographically protected database search. In *2017 IEEE Symposium on Security and Privacy*, pages 172–191, San Jose, CA, USA, May 22–26, 2017. IEEE Computer Society Press.
- [10] Joël Alwen, Binyi Chen, Krzysztof Pietrzak, Leonid Reyzin, and Stefano Tessaro. Scrypt is maximally memory-hard. In Jean-Sébastien Coron and Jesper Buus Nielsen, editors, *Advances in Cryptology – EUROCRYPT 2017, Part III*, volume 10212 of *Lecture Notes in Computer Science*, pages 33–62, Paris, France, April 30 – May 4, 2017. Springer, Heidelberg, Germany.
- [11] Arka Rai Choudhuri, Matthew Green, Abhishek Jain, Gabriel Kaptchuk, and Ian Miers. Fairness in an unfair world: Fair multiparty computation from public bulletin boards. In Bhavani M. Thuraisingham, David Evans, Tal Malkin, and Dongyan Xu, editors, *ACM CCS 2017: 24th Conference on Computer and Communications Security*, pages 719–728, Dallas, TX, USA, October 31 – November 2, 2017. ACM Press.
- [12] Ran Canetti, Yilei Chen, Leonid Reyzin, and Ron D. Rothblum. Fiat-Shamir and correlation intractability from strong KDM-secure encryption. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2018, Part I*, volume 10820 of *Lecture Notes in Computer Science*, pages 91–122, Tel Aviv, Israel, April 29 – May 3, 2018. Springer, Heidelberg, Germany.
- [13] Albert Cheu, Adam D. Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2019, Part I*, volume 11476 of *Lecture Notes in Computer Science*, pages 375–403, Darmstadt, Germany, May 19–23, 2019. Springer, Heidelberg, Germany.
- [14] Ran Canetti, Aloni Cohen, Nishanth Dikkala, Govind Ramnarayan, Sarah Scheffler, and Adam D. Smith. From soft classifiers to hard decisions: How fair can we be? In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 309–318. ACM, 2019.
- [15] Ran Canetti, Yael Tauman Kalai, Anna Lysyanskaya, Ronald L. Rivest, Adi Shamir, Emily Shen, Ari Trachtenberg, Mayank Varia, and Daniel J. Weitzner. Privacy-preserving automated exposure notification. Cryptology ePrint Archive, Report 2020/863, 2020. <https://eprint.iacr.org/2020/863>.
- [16] Ran Canetti, Benjamin Fuller, Omer Paneth, Leonid Reyzin, and Adam D. Smith. Reusable fuzzy extractors for low-entropy distributions. *Journal of Cryptology*, 34(1):2, January 2021.

- [17] Nikolaj Volgushev, Malte Schwarzkopf, Ben Getchell, Mayank Varia, Andrei Lapets, and Azer Bestavros. Conclave: Secure multi-party computation on big data. In *Proceedings of the Fourteenth EuroSys Conference*, pages 3:1–3:18. ACM, 2019. <https://arxiv.org/pdf/1902.06288.pdf>.
- [18] John Liagouris, Vasiliki Kalavri, Muhammad Faisal, and Mayank Varia. Secrecy: Secure collaborative analytics on secret-shared data. *arXiv*, abs/2102.01048, 2021. <https://arxiv.org/abs/2102.01048>.
- [19] Daniel Alabi, Audra McMillan, Jayshree Sarathy, Adam Smith, and Salil Vadhan. Differentially private simple linear regression. *Proceedings on Privacy Enhancing Technologies*, 2022(2):184–204, 2022.
- [20] Boston Women’s Workforce Council. Data privacy. <https://thebwwc.org/mpc>.
- [21] Museums Moving Forward. Data study. <https://museumsmovingforward.com/data-study>.
- [22] Greater Boston Chamber of Commerce. Pacesetters. <https://bostonchamber.com/networks/pacesetters/>.
- [23] Anjana Rajan, Lucy Qin, David W. Archer, Dan Boneh, Tancrede Lepoint, and Mayank Varia. Callisto: A cryptographic approach to detecting serial perpetrators of sexual misconduct. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 49:1–49:4. ACM, 2018. <https://www.projectcallisto.org/callisto-cryptographic-approach.pdf>.
- [24] Fireblocks. About fireblocks. <https://www.fireblocks.com/about>.
- [25] Fiverity. A holistic approach to digital fraud detection. <https://www.fiverity.com>.
- [26] Kenneth A. Bamberger, Ran Canetti, Shafi Goldwasser, Rebecca Wexler, and Evan J. Zimmerman. Verification dilemmas in law and the promise of zero-knowledge proofs. *Berkeley Technology Law Journal*, 37(1), 2022. Available at SSRN: <https://ssrn.com/abstract=3781082>.
- [27] Dor Bitan, Ran Canetti, Shafi Goldwasser, and Rebecca Wexler. Using zero-knowledge to reconcile law enforcement secrecy and fair trial rights in criminal cases. In *2nd ACM Symposium on Computer Science and Law (to appear)*, 2022. Available at SSRN: <https://ssrn.com/abstract=4074315>.
- [28] Matthew Green, Gabriel Kaptchuk, and Gijs Van Laer. Abuse resistant law enforcement access systems. In Anne Canteaut and François-Xavier Standaert, editors, *Advances in Cryptology – EUROCRYPT 2021, Part III*, volume 12698 of *Lecture Notes in Computer Science*, pages 553–583, Zagreb, Croatia, October 17–21, 2021. Springer, Heidelberg, Germany.
- [29] Rachel Cummings, Gabriel Kaptchuk, and Elissa M. Redmiles. “I need a better description”: An investigation into user expectations for differential privacy. In Giovanni Vigna and Elaine Shi, editors, *ACM CCS 2021: 28th Conference on Computer and Communications Security*, pages 3037–3052, Virtual Event, Republic of Korea, November 15–19, 2021. ACM Press.
- [30] Gabriel Kaptchuk, Daniel G. Goldstein, Eszter Hargittai, Jake M. Hofman, and Elissa M. Redmiles. How good is good enough for COVID19 apps? The influence of benefits, accuracy, and privacy on willingness to adopt. *Digital Threats: Research and Practice*, 2022. Available at <https://arxiv.org/pdf/2005.04343.pdf>.
- [31] Sarah Scheffler and Mayank Varia. Protecting cryptography against compelled self-incrimination. In Michael Bailey and Rachel Greenstadt, editors, *USENIX Security 2021: 30th USENIX Security Symposium*, pages 591–608. USENIX Association, August 11–13, 2021.
- [32] Aloni Cohen, Sarah Scheffler, and Mayank Varia. Can the government compel decryption? Don’t trust – verify. In *2nd ACM Symposium on Computer Science and Law (to appear)*, 2022.
- [33] U.S. Bureau of Economic Analysis. Advisory committee on data for evidence building. <https://www.bea.gov/evidence>.
- [34] UN Committee of Experts on Big Data and Data Science for Official Statistics. Privacy preserving techniques task team. <https://unstats.un.org/bigdata/task-teams/privacy/index.cshhtml>.
- [35] 115th Congress. Foundations for evidence-based policymaking act of 2018. <https://www.congress.gov/bill/115th-congress/house-bill/4174>.
- [36] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books, 1st edition, 2018.
- [37] Boston University. Open-source libraries, tools, and platforms that enable the implementation and deployment of applications that employ secure multi-party computation. <https://github.com/multiparty>.
- [38] danah boyd and Jayshree Sarathy. Differential perspectives: Epistemic disconnects surrounding the us census bureau’s use of differential privacy. In *Harvard Data Science Review (Forthcoming)*, 2022.
- [39] U.K. Financial Conduct Authority. 2019 Global AML and Financial Crime TechSprint. <https://www.fca.org.uk/events/techsprints/2019-global-aml-and-financial-crime-techsprint>.

- [40] Dorothy E. Denning and Dennis K. Branstad. A taxonomy for key escrow encryption systems. *Commun. ACM*, 39(3):34–40, 1996.
- [41] David Chaum. Privategrid: online communication with strong privacy. In *Real World Cryptography*, 2016.
- [42] Abhishek Bhowmick, Dan Boneh, Steve Myers, Kunal Talwar, and Karl Tarbe. The Apple PSI system. https://www.apple.com/child-safety/pdf/Apple_PSI_System_Security_Protocol_and_Analysis.pdf, 2021.
- [43] Google. Federated learning of cohorts (floc). <https://github.com/WICG/floc>.
- [44] Google. Turtledove. <https://github.com/WICG/turtledove>.
- [45] Microsoft Corporation. Parakeet. <https://github.com/WICG/privacy-preserving-ads/blob/main/Parakeet.md>.
- [46] Microsoft Corporation. Multi-party computation of ads on the web (MaCAW). <https://github.com/WICG/privacy-preserving-ads/blob/main/MACAW.md>.
- [47] Harold Abelson, Ross J. Anderson, Steven M. Bellovin, Josh Benaloh, Matt Blaze, Whitfield Diffie, John Gilmore, Matthew Green, Susan Landau, Peter G. Neumann, Ronald L. Rivest, Jeffrey I. Schiller, Bruce Schneier, Michael A. Specter, and Daniel J. Weitzner. Keys under doormats. *Commun. ACM*, 58(10):24–26, 2015.
- [48] Susan Landau. Punching the wrong bag: The deputy AG enters the crypto wars, October 2017. <https://www.lawfareblog.com/punching-wrong-bag-deputy-ag-enters-crypto-wars>.
- [49] Hal Abelson, Ross J. Anderson, Steven M. Bellovin, Josh Benaloh, Matt Blaze, Jon Callas, Whitfield Diffie, Susan Landau, Peter G. Neumann, Ronald L. Rivest, Jeffrey I. Schiller, Bruce Schneier, Vanessa Teague, and Carmela Troncoso. Bugs in our pockets: The risks of client-side scanning. *arXiv*, abs/2110.07450, 2021. <https://arxiv.org/abs/2110.07450>.
- [50] Ran Canetti and Gabriel Kaptchuk. The broken promise of Apple’s announced forbidden-photo reporting system – and how to fix it. <https://www.bu.edu/riscs/2021/08/10/apple-csam>, 2021.
- [51] Bennett Cyphers. Don’t play in google’s privacy sandbox. <https://www.eff.org/deeplinks/2019/08/dont-play-googles-privacy-sandbox-1>, 2019.
- [52] Bennett Cyphers. Google’s floc is a terrible idea. <https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea>, 2021.
- [53] Mayank Varia. A survey of MPC offerings. https://drive.google.com/file/d/1NT_vdxRC8YEPlkQa2KHw22ai9IshyU73/view, 2018.
- [54] Catherine D’Ignazio and Lauren F. Klein. *Data Feminism*. MIT Press, 2020.
- [55] Sasha Costanza-Chock. *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press, 2020.
- [56] Paul Ohm and Jonathan Frankle. Desirable inefficiency. *Fla. L. Rev.*, 70:777, 2018.
- [57] Paul Ohm. Statement of Paul Ohm, Professor, Georgetown University Law Center and Member, Commission on Evidence-Based Policymaking. In *Testimony for the Hearing on Protecting Privacy, Promoting Policy: Evidence-Based Policymaking and the Future of Education, Before the Committee on Education and the Workforce, U.S. House of Representatives*, 2018. Available at <https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2019/03/CEP-Comissioner-Paul-Ohm-Testimony-from-1.30.18.pdf>.
- [58] National Institute for Science and Technology. Cryptographic standards and guidelines. <https://csrc.nist.gov/projects/cryptographic-standards-and-guidelines/archived-crypto-projects/aes-development>, 2016.
- [59] National Institute for Science and Technology. Hash functions. <https://csrc.nist.gov/projects/hash-functions/sha-3-project>, 2017.
- [60] National Institute for Science and Technology. Post-quantum cryptography. <https://csrc.nist.gov/Projects/post-quantum-cryptography>, 2017.
- [61] National Institute for Science and Technology. Lightweight cryptography. <https://csrc.nist.gov/Projects/lightweight-cryptography>, 2017.
- [62] Prabhanjan Ananth, Abhishek Jain, Zhengzhong Jin, and Giulio Malavolta. Pre-constrained encryption. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

- [63] Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke, and Salil Vadhan. Differential privacy: A primer for a non-technical audience. *Vanderbilt Journal of Entertainment & Technology Law*, 21(1):209–275, 2018. <http://www.jetlaw.org/journal-archives/volume-21/volume-21-issue-1/differential-privacy-a-primer-for-a-non-technical-audience/>.
- [64] Lucy Qin, Andrei Lapets, Frederick Jansen, Peter Flockhart, Kinan Dak Albab, Ira Globus-Harris, Shannon Roberts, and Mayank Varia. From usability to secure computing and back again. Cryptology ePrint Archive, Report 2019/734, 2019. <https://eprint.iacr.org/2019/734>.
- [65] Ruba Abu-Salma, Elissa M. Redmiles, Blase Ur, and Miranda Wei. Exploring user mental models of End-to-End encrypted communication tools. In *8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*, Baltimore, MD, August 2018. USENIX Association.
- [66] Omer Akgul, Ruba Abu-Salma, Wei Bai, Elissa M. Redmiles, Michelle L. Mazurek, and Blase Ur. *From Secure to Military-Grade: Exploring the Effect of App Descriptions on User Perceptions of Secure Messaging*, page 119–135. Association for Computing Machinery, New York, NY, USA, 2021.
- [67] Christian Stransky, Dominik Wermke, Johanna Schrader, Nicolas Huaman, Yasemin Acar, Anna Lena Fehlhaber, Miranda Wei, Blase Ur, and Sascha Fahl. On the limited impact of visualizing encryption: Perceptions of E2E messaging security. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 437–454. USENIX Association, August 2021.
- [68] Dominik Wermke, Nicolas Huaman, Christian Stransky, Niklas Busch, Yasemin Acar, and Sascha Fahl. Cloudy with a chance of misconceptions: Exploring users' perceptions and expectations of security and privacy in cloud office suites. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, pages 359–377. USENIX Association, August 2020.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Carnegie Mellon University

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Carnegie Mellon University Response to Office of Science and Technology Policy (OSTP) RFI on Advancing Privacy-Enhancing Technologies

Alessandro Acquisti

Professor, Heinz College of Information Systems and Public Policy

Travis Breaux

Associate Professor, Institute for Software Research, School of Computer Science

Jason Hong

Professor, Human-Computer Interaction Institute, School of Computer Science

Norman Sadeh

Professor/Co-Director, CMU Privacy Engineering Program

Institute for Software Research, School of Computer Science

Steven Wu

Assistant Professor, Institute for Software Research, School of Computer Science

Introduction

This document is a response to OSTP's Request for Information on Advancing Privacy-Enhancing Technologies (87 FR 35250), by several faculty at Carnegie Mellon University's CyLab Security and Privacy Institute. The authors' areas of expertise include AI, HCI, privacy, cybersecurity, software engineering, fairness, economics, and more.

We have organized our response below by the questions posed in the RFI.

1. Specific research opportunities to advance PETs:

Studies to Understand What Influences End-Users to Adopt PETs

Support more studies as to what technical, usability, social, and other factors lead end-users to adopt (or not) different kinds of PETs, as well as what kinds of interventions are useful. There are already a vast number of PETs, but there also seems to be limited adoption of them by end-users. For example, in 2018, Google reported that less than 10% of their users used two-factor authentication, despite the practice being widely advocated by security experts. A deeper understanding of why people do or don't adopt PETs could help lead to better designs and interventions.

For example, Das et al [DKDH14] found that about half of changes in security behaviors were due to a social influence. In follow-up work, Das et al [DKDH14] found that simple social interventions (e.g. "108 of your friends use extra security settings") increased people's likelihood of adopting features. In related work, Krsek et al [KWDH+22] found that seeing what strangers recommended for Facebook settings as well as self-reflection were also effective in getting people to change their privacy settings.

In another line of work, Das et al [DKDH15] analyzed large-scale data about use of security features from Facebook users and found that social influence affects one's likelihood to adopt a security feature, varying based on the observability of the feature, its adoption rate among one's friends, and the number of distinct social circles from which those feature-adopting friends originate.

Researchers have also examined how people naturally come up with surprising ways of managing their own privacy. For example, Jin et al [JGRY+22] found that, in the context of their own smart homes, people would unplug devices or cover up cameras so that they would not be monitored while at home. Having a better understanding of these kinds of naturalistic approaches could help in terms of understanding people's mental models, as well as why people choose these relatively simple techniques over more sophisticated ones.

One potentially promising avenue of research here is to develop stage models of adoption for privacy. For example, in health care, some people need more general awareness of their health problem while others are already active in preventative behavior and only need support in maintaining that behavior. By having a better model of behavior change, researchers and practitioners can develop better kinds of systems, user interfaces, and interventions to help foster adoption. Other relevant examples of the type of research that could help promote the adoption of PETs include the development and evaluation of different types of nudges. For instance, Story et al. have recently reported on their exploration of Protection Motivation Theory (PMT), Action Planning (AP), and Coping Planning (CP) bases for designing interventions designed to promote adoption of the Tor browser [SSC+22].

Studies to Understand What Influences Developers to Adopt PETs

Support more studies as to what technical, usability, social, and other factors lead developers to adopt (or not) different kinds of PETs in their products, as well as what kinds of interventions are useful. Developers are one of the greatest points of leverage for improving privacy, and so it's worth investigating more as to how to best educate, influence, and support this group.

For example, Witschey et al [WZWM+15] interviewed 42 developers and quantified why they did or did not adopt different security tools, using Diffusion of Innovations as a framework for organizing the results. Both Li et al [LLDH21] and Tahaei et al [TLV22] studied how developers talk about privacy, analyzing the /r/AndroidDev subreddit and StackOverflow respectively, generally finding that developers have low understanding of privacy and that app store requirements have a much greater influence than privacy laws.

Tahaei et al [TRLH+22] also examined the developer documentation for managing privacy on popular ad networks, and conducted user studies with developers to understand their perceptions. Tahaei et al found that documentation about privacy regulations was scattered across several pages, and used terms and language developers found hard to understand. There was also a mismatch in expectations, with ad networks putting the burden of complying with regulations on developers, while developers felt ad networks should be responsible for ensuring compliance with regulations.

Privacy Annotations to Improve the Privacy Ecosystem

Many programming languages support annotations, which are metadata about a program in the source code added by a developer. This metadata can be used by the compiler, the IDE, and other tools to manage workflow, documentation, and so on. Li et al first proposed *privacy annotations* in 2018 [LAH18], where developers specify the purpose of data collection, data retention, where the data will be sent, and so on. These annotations can be used by the IDE to offer hints about alternative APIs to use and to help with auditing of Android apps. In 2021, Li et al [LRAC+22] extended this work so that privacy annotations could be used to auto-generate useful user interfaces for informing end-users of data access by apps and offering user interfaces for configuring privacy settings.

More broadly, we believe privacy annotations can greatly improve the entire privacy ecosystem. Developers can use them at programming time, to help with many basic privacy-related development tasks, such as choosing appropriate APIs, adding privacy user interfaces, and filling out privacy nutrition labels. These same annotations can also be used by other members of a development team to audit the data collection behaviors of apps. These annotations can also be embedded into compiled apps, making it easier for app stores and other third parties to audit the behavior of apps and ensure that they are doing what they say they will do. One long term and ambitious idea would be to have annotations across all components of a system, including front-end as well as back-end cloud computing, enabling full end-to-end support for developers and for auditors.

App Manifests for Declaring and Enforcing App Behaviors

A common problem with almost all apps is *over access*. A sleep monitor app might use microphone data, but in reality it only needs loudness. A weather app might use GPS data, but only really needs city granularity. A smart TV app might access the raw viewing logs, but only needs a summary of most viewed channels and duration rounded to the nearest hour. Overaccess happens because there is a mismatch between what APIs there are for accessing data and what data the app actually needs.

Jin et al [JLHK+22] proposed three key ideas to address overaccess. First, all apps (smartphone, IoT, browser plugins, etc) are required to declare their data collection behaviors in a short and human-readable app manifest. Second, behaviors are specified using a small and pre-defined set of operators for gathering and transforming the data. Third, the manifest is enforced by a trusted hub which also runs the operators. Together, these form a trusted data flow that can help with minimizing the granularity of data shared with others.

An example flow for a sleep monitor might be “get access to microphone every minute, transform it to loudness, send to sleep.com”. While an auditor might not fully know how the data will be used, they can easily understand that only loudness granularity will be sent out. This approach also makes it possible to auto-generate a privacy nutrition label from a manifest, track who knows what about you, modify the data flow based on any kinds of rules (e.g. no face data should be sent out), and auto-generate consistent user interfaces across apps.

Jin et al investigated this approach with manifests in the context of IoT. One promising angle for research here is to extend the approach to other platforms, e.g. browser plugins, smartphones, cloud data, etc. Another angle is to investigate how to combine this approach with privacy annotations, making it so that there is a single unified approach that offers the best of both worlds. This combined approach could also make it possible to have a full end-to-end approach that works across distributed components, e.g. front-end and cloud backend.

Improving Auditing of Apps and Devices by Developers and Third Parties

Many PETs focus on preventing unauthorized access or disclosure of sensitive data. However, relatively little work has focused on making it easier to detect the same, which would help with ensuring that apps and devices comply with any stated behaviors and privacy regulations.

For example, an auditor internal to the company might want to do white box testing, tracing data from collection, transmission to cloud storage, processing, and reporting. Right now, this is a labor intensive process, with many disparate tools currently only usable by people with a high level of technical expertise.

Similarly, third parties such as researchers, consumer advocates, and journalists might also be interested in auditing apps and devices, to gather data and develop new techniques for checking for compliance at scale. However, there are a lack of standards and tools, making it hard even for experts to understand what data is being collected in the first place. Furthermore,

encryption techniques like TLS, while a best practice for privacy and security, unintentionally make it hard to conduct legitimate research.

PETs to Support Developers, Ecosystem Operators and Regulators and Assist with Compliance Analysis

Research has shown that software developers, both large and small, struggle to ensure that their products and services comply with applicable regulations. For instance, in their analysis of over a million Android Apps in the Google Play Store, researchers found that apps had an average of over 3 potential compliance issues [ZSS+19] (see also [ZWZ+17]). Research combining code analysis, privacy policy text analysis, as well as wizards designed to prompt developers to carefully reflect on their data practices can go a long way in helping developers ensure that their products and services are compliant (e.g., [ZSS+19,GFR+22]). We have only started to scratch the surface of what is possible in this space. Research has consistently shown that developers generally lack adequate expertise and support to properly disclose their data practices and ensure that their code is compliant. Similar technologies can also be used to help ecosystem operators (e.g., app stores, browser extension stores, IoT ecosystem operators) ensure that technologies developed by third parties and made available by the ecosystems to their user communities comply with relevant regulations. Those same technologies can also be used to conduct at-scale compliance analysis, something that regulators badly need if they are to keep pace with the emergence of ecosystems such as app stores that feature several million apps (e.g. [ZWZ+17,ZSS+19]).

PETs for the Enterprise to Study Data Flows Within and Across an Organization

Research on “PETs for the Enterprise,” which moves beyond web and mobile applications to consider how data is processed within and across an organization’s enterprise architecture. This focus includes how services built on Kubernetes, and Apache NiFi, Spark, and Hadoop for data processing, integrate with technical privacy controls for revocable authorization, confidentiality and disassociation. PETs for the enterprise can aid decision makers in planning privacy-at-scale by rationalizing how data flows within their service infrastructure, and verifying that against their corporate privacy policy on data protection. Service-oriented architectures that allow companies to decentralize data management are frequently used to scale performance to meet demand across government services. This places agencies at risk of over-collection and repurposing data [BSH15], particularly, when the volume of data represents hundreds of thousands of data subjects with thousands of data types.

Enhancements to formal methods for modeling data flow could allow enterprise planners to pinpoint where within a large system, and how to apply PETs as agencies increasingly look to share information across services. The ability to design systems to grant and revoke consent as individual privacy preferences and data use policies change over time, without degrading ongoing data processing and analytics is an open area of research [RBP+23].

Empirical Social Science Research on Impact of PETs

Support more social science research on the impact of PETs, and in particular empirical studies, including natural and field experiments, on the downstream implications and outcomes of the

deployment of PETs, including 1) the analysis of the economic trade-offs associated with the adoption of PETs in industry, government, or by consumers; 2) the investigation of how the potential burdens of adoption or potential losses from decreased granularity of data are allocated among different stakeholders, and how those burdens can be mitigated and minimized; 3) the study of how the usage of PETs affects or alters the allocation, across different stakeholders, of the value accrued from data analytics.

2. Specific technical aspects or limitations of PETs:

Generalizable techniques for tracing information flows

Restrictions on data sharing, including access control and de-identification techniques, rely on developer claims and assertions made at design time that data can only be accessed through specific endpoints, otherwise data leaks can arise that circumvent these restrictions. The efficacy of these restrictions is no greater than the reliability of these claims and assertions. In source code analysis, tools have been developed to trace information flow from collection points to sharing and use points to detect leaks and privacy policy violations [SWH+16]. Due to the wide variety of programming languages and software development frameworks or Application Programmer Interfaces (APIs), technical challenges exist in tracing information flow throughout filesystems, graphical user interfaces (GUIs) [WQH+18, QNG+19], and networked services [ZWS+20]. Solutions have largely been one-off examples tied to specific operating systems and APIs. Due to the availability of application binaries, advances in information flow tracing have largely accrued on mobile applications. Techniques that are generalizable beyond specific operating systems, programming languages and frameworks, and that can be passively integrated into developer toolchains, are needed to realize end-to-end information flow tracing to detect privacy leaks at scale.

Risk metrics for sensitive data

Confidentiality and disassociation techniques introduce cost to software development both in terms of software design, implementation and maintenance effort, as well as a perceived loss of data utility in certain applications. Techniques are needed to assess the privacy risk associated with sharing data to enable the application of PETs to the most sensitive data. Evidence exists that data subjects perceive privacy risk differently based on the ambiguity of how data is used [BBR+16], and their experience and social and physical distance to the privacy harm [BB18]. African Americans and Hispanics, for example, perceive greater risk from data sharing by law enforcement [BRA+19]. Bhatia and Breaux have designed and evaluated reliable privacy risk metrics that can be used to measure the perceived risk by data subjects to inform design decisions [BB18]. However, their work discovered two unaddressed technical challenges in privacy risk measurement: (1) risk changes as new datasets are composed from other datasets, yet the change in risk due to composition is unknown, including both the direction and magnitude of such risks; and (2) risk is subject to environmental effects, including the recency and severity of known privacy harms. Addressing these limitations of risk metrics would allow developers to combine PETs in ways that result in predictable, measurable impacts to data subject privacy as risks evolve.

Development of Human-AI Interaction Technologies for PETS

While new regulations such as CCPA/CPRA are requiring more detailed data practice disclosures and the availability of privacy choices (e.g., opt-ins and opt-outs) that were not previously required, these new regulations also have highly unrealistic expectations when it comes to what lay users can realistically be able to do (e.g., [LLS14,LLS+14]). No one reads privacy policies and no one has the time to analyze all the privacy choices and configure all the privacy settings now available to users. What is required is the development of intelligent interfaces that can help people take advantage of these regulatory advances and effectively regain control of their data (e.g. [LAS+16,RBW+19]). Recent advances in the development of Natural Language Processing and Machine Learning techniques to automatically analyze the text of privacy policies and help answer people's privacy questions offer the promise of enabling people to take advantage of more detailed data practice disclosures without requiring people to actually read the text of privacy policies (e.g., see CMU's Usable Privacy Policy Project [UPPP22,WSL+18]). Similarly machine learning techniques that can model people's privacy policies and help provide recommendations to users and reduce the burden required to configure an otherwise unmanageable number of privacy decisions could also go a long way in helping users (e.g, see work on personalized privacy assistants [PPA22,LAS+16, DDS+18]). Yet significantly more research is required to configure these technologies and ensure that users retain effective control over their privacy decisions while benefiting from the predictive power of these technologies.

Development of APIs and Infrastructure for PETS

In addition to developing intelligent interfaces to assist users with the management of their privacy, it is important to recognize that new technologies such as the Internet of Things also lack basic standards and APIs to communicate privacy policies and expose privacy settings to users. It is therefore critical to also promote the development of infrastructures, APIs and standards to support such communication. A particularly prominent effort in this regard is the development at Carnegie Mellon University of a Privacy Infrastructure for the Internet of Things (e.g. see [IoTPI22a, IoTPI22b,DDS+19,FYS21]). Additional research is required to promote the development and adoption of such infrastructure and promote the adoption of standards necessary for the adoption of such infrastructures.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETS:

Emergency Response

Both immediate emergencies (e.g. earthquakes, flash floods, or terrorist attacks) as well as longer running ones (e.g. fires, hurricanes) could benefit greatly from more real-time data, which is possible from the smartphones that people carry as well as from the large number of cameras and other sensors in neighborhoods and cities. Some examples might include using location data to estimate how many people are still in a building for recovery, running private queries on people's smartphone photos to look for a specific individual's face (think Boston Marathon bombing), and using cell phone call data records to infer a major spike in activity in a localized area (possible accident or terrorist attack).

Urban Analytics

It is possible to use large-scale location traces to understand people's behaviors in cities (see for example [CSHS12]). Using geotagged social media, cell phone call data records, or smartphone tracking libraries embedded in popular apps, it is possible to acquire large-scale traces of where people go in a city. This kind of data can be helpful for urban planning, traffic analysis, understanding economic impacts of road closures, how people use parks, the walkability of different neighborhoods, and more. However, the sensitive nature of this kind of data leads to a large number of privacy concerns.

Public Health

Access to public health suffers from disparities in economic status that model disparities in trust in government data collection, processing and surveillance, including public health surveillance.

Telemedicine is one area where PETs could improve adoption. Telemedicine appointments provide unprecedented convenience to patients with inflexible work and childcare schedules, as well as unprecedented access to a patient's physical space outside the clinic or hospital, whether at home or at work. Privacy and security risks in telehealth are well known [HK14], but research into this vulnerable population and how PETs can improve access is lagging, and health apps and connected devices continue to risk privacy harm due to data breach [FTC21].

Contact tracing is another area that could benefit from PETs. Contact tracing seeks to understand who an infected individual may have been in contact with, and can help with early isolation measures. Despite its potential, and despite the fact that there are a number of tracking mechanisms (e.g. Bluetooth, GPS on smartphones, QR codes), electronic-based contact tracing measures have seen only limited adoption. Privacy is one of several barriers to widespread adoption of electronic contact tracing mechanisms [LCYB+21].

Distributed community sensing

Many people are already starting to deploy sensors in their homes, but these sensors tend to be isolated from one another. Having a programmable substrate that links many sensors in a neighborhood or across an entire city together would enable the creation of many new kinds of apps. For example, by linking together many smart speakers, one could create a noise map of the entire city, which could be used by public health researchers as well as individuals wanting to purchase a home. By linking together smart doorbells, one could create a smarter Amber Alert spotter, a lost dog spotter, a neighborhood parking space estimator, a pedestrian counter, a car traffic counter, a snow plow spotter, and more. However, privacy is a major barrier here, as such a system could be easily abused to stalk individuals or to infer if someone is home or not.

Covert Spying in Everyday Situations

Here, we mention two cases where PETs could help. The first is mitigating intimate partner violence. Freed et al [FPML+18] showed how malicious individuals use known spyware as well as rather innocuous technologies to track and monitor intimate partners. Better PETs could help victims detect misuse of these technologies or hide their activities.

The second is with temporary rentals, such as Airbnb and VRBO. A number of media articles have reported on how people have found undisclosed surveillance devices in these rentals, cameras in particular. Song et al [SHCH20] presented the results of a survey of 192 participants who have stayed in an Airbnb before, with 50% stating that they searched for devices, and 8 participants actually finding undisclosed devices. Assisting people in finding what sensing devices are nearby, where exactly those devices are, whether they are active or not, and how the data will be used would greatly help in this rental scenario, as well as many other smart home and smart building scenarios.

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs:

The role of the FTC, FDA and HHS in regulating health apps can be improved to cover services at the “margin” of medical devices and electronic health records. The FTC released a recent 2021 press release highlighting the privacy risks of health apps and connected devices [FTC21], whereas the FDA has moved to establish rules for Software as Medical Device. While many health apps do not meet the eligibility criteria to be classified as medical devices, nor are they included under the umbrella of the HIPAA Privacy Rule, they are collecting sensitive health data, which has become a gap in regulation and oversight. Regulators should work with industry to deploy lightweight PETs that can increase privacy assurances while supporting innovation in this area.

Regulations Requiring Standards and APIs to Make Privacy Usable

While new regulations such as CCPA/CPRA are requiring more detailed data practice disclosures and the availability of privacy choices (e.g., opt-ins and opt-outs) that were not previously required, these new regulations also have highly unrealistic expectations when it comes to what lay users can realistically be able to do (e.g., [LLS14,LLS+14]). No one reads privacy policies and no one has the time to analyze all the privacy choices and configure all the privacy settings now available to users. An example is the cookie management prompts that pop up on nearly every website one visits, forcing users to repeatedly enter their choices over and over again as they move from one website to the next. If users could specify their choices once and for all in their browsers and let their browser communicate these choices to the websites they visit, user burden could significantly be reduced (e.g., see Smullen et al. for a discussion of this and related issues [SYF+21]). Regulation that would require websites to support simple APIs that allow browsers to communicate these preferences would go a long way to reduce unnecessary burden and allow people to actually exercise choices that are otherwise only theoretical. The same is true for the communication of privacy policies and the presentation of privacy choices in the context of the Internet of Things (e.g., see Zhang et al. for a discussion of this issue in the context of video analytics scenarios [ZFB+21]). In short, APIs and standards to support these scenarios would go a long way towards helping people truly benefit from new regulations such as CCPA/CPRA (as well as similar regulations introduced in a number of other states).

Moving from Opt-Out to Opt-Ins

Beyond the above, US consumers would significantly benefit from regulations that move away from opt-out choices and make opt-in the default approach to exercising one's privacy choices [CS19]. Opt-out choices have been shown to be consistently gamed by industry, with these choices often buried deep in the text of privacy policies and/or websites and users being required to jump through an unrealistic number of hoops [HPW+19]. Opt-in choices provide a level playing field where consumers have to first provide consent before companies can engage into potentially invasive data practices and tend to much better reflect people's privacy preferences (e.g., [SYZ+21]).

5. Specific laws that could be used, modified, or introduced to advance PETs:

See comments provided under 4.

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:

Data Sets for Evaluating Different PETs

Having access to large-scale data sets could help with evaluating the effectiveness of different PETs and the tradeoffs involved. For example, location data of many individuals could help in developing and comparing new kinds of algorithms. Note that large-scale location data sets do exist, but tend to be proprietary and/or expensive, making them out of reach of the vast majority of researchers. Given the sensitive nature of location data, there would also need to be other kinds of safeguards in place to prevent accidental release as well as misuse of the data.

See also comments already provided under 4 on regulations to require the adoption of standards and APIs to make privacy usable.

7. Risks related to PETs adoption:

No comment from our team.

8. Existing best practices that are helpful for PETs adoption:

No comment from our team.

9. Existing barriers, not covered above, to PETs adoption:

A High Burden on Developers for Compliance

In an analysis of a popular discussion board for Android developers, Li et al [LLDH21] found that developers often viewed privacy as a burden with little benefit for themselves. For example, developers felt that they had to learn new kinds of APIs, keep up to date with vague documentation and changes to the Android operating system, and comply with new policies on Google Play app store, all without any clear benefit to their app. Using a colloquialism, for developers, privacy is currently no carrots and all sticks. Li et al recommended looking for ways to reward developers for their effort, for instance prioritizing their app on search results or highlighting apps that were especially good with respect to privacy.

This same paper also found that discussions of privacy amongst Android developers rarely happened due to new privacy laws or regulations, and that developers rarely talked about privacy issues when proposing new apps or getting feedback from other devs. Instead, privacy was brought up primarily because of external events, such as changes to the app store or the operating system.

Encryption Makes it Hard for Researchers and Third-Party Auditors to Evaluate Privacy

Encryption techniques like TLS are a best practice for privacy and security, making it hard for attackers to eavesdrop on potentially sensitive communications. However, these same techniques make it hard for legitimate researchers and third parties to assess and to audit what data an app or device is sending out. Whether it is supported by an operating system, library, device, or app, having some kind of explicit setting that lets legitimate researchers and auditors examine data would be incredibly helpful in improving transparency and might also facilitate adoption of many new kinds of PETs.

10. Other information that is relevant to the adoption of PETs:

Division of Labor for Privacy Across the Entire Ecosystem

The division of labor as to who is responsible for what aspects of managing privacy is currently unclear. Hong [Hong17] argued that the burden of privacy is currently too heavy for end-users, and that instead we should consider how other parts of the ecosystem can help. Using an analogy with spam email, it used to be the case that end-users had to manually delete all spam themselves. Over time, however, email service providers started using machine learning to filter emails, network providers started deploying protocols for blocking bad emails, and law enforcement took down egregious offenders, with the end result being vastly fewer spam emails.

However, it is unclear how best to achieve a similar result for privacy, partly because there are vastly more stakeholders involved. Furthermore, there are many misconceptions and disagreements as to who should be responsible for what aspects of privacy. For example, as noted earlier in [TRLH+22], ad networks placed the responsibility of complying with privacy on developers, but developers felt that ad networks should be responsible instead.

In all cases, we should aim to avoid policies and regulations that place undue burden on end-users, as that has repeatedly been shown to be ineffective in practice.

References

- [ABL20] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. "Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age." *Journal of Consumer Psychology* 30.4. (2020): 736-758.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3688497
- [ATW16] Alessandro Acquisti, Curtis Taylor, and Liad Wagman. "The economics of privacy." *Journal of Economic Literature* 54.2. (2016): 442-92.
<https://www.aeaweb.org/articles?id=10.1257/jel.54.2.442>
- [BB18] Jaspreet Bhatia and Travis D. Breaux. "Empirical Measurement of Perceived Privacy Risk." *ACM Transactions on Computer-Human Interaction*. 25, 6, Article 34. (December 2018) 47 pages. <https://doi.org/10.1145/3267808>
- [BBR+16] Jaspreet Bhatia, Travis D. Breaux, Joel R. Reidenberg and Thomas B. Norton. "A Theory of Vagueness and Privacy Risk Perception." In Proceedings of the *IEEE 24th International Requirements Engineering Conference (RE)*. (2016) pp. 26-35, doi: 10.1109/RE.2016.20.
<https://ieeexplore.ieee.org/abstract/document/7765508>
- [BRA+19] Brooke Auxier et al. "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information." PEW Research Center. (November 15, 2019)
<https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>
- [BSH15] Travis D. Breaux, Daniel Smullen, and Hanan Hibshi. "Detecting repurposing and over-collection in multi-party privacy requirements specifications." *2015 IEEE 23rd International Requirements Engineering Conference (RE)*. (2015) pp. 166-175, doi: 10.1109/RE.2015.7320419. <https://ieeexplore.ieee.org/document/7320419>
- [CS19] L Cranor and N Sadeh, "Congress, Make Privacy the Rule - Not the Exception." Op-Ed in the Hill. (February 1, 2019)
<https://thehill.com/opinion/cybersecurity/428121-congress-make-privacy-the-rule-not-the-exception/>
- [CSHS12] J Cranshaw, R Schwartz, JI Hong, N Sadeh. "The livelihoods project: Utilizing social media to understand the dynamics of a city." ICWSM 2012. (2012)
<https://ojs.aaai.org/index.php/ICWSM/article/view/14278>
- [DDS+18] A Das, M Degeling, D Smullen, and N Sadeh. "Personalized Privacy Assistants for the Internet of Things," 2018 IEEE Pervasive Computing: Special Issue - Securing the IoT. (April 2018) https://www.privacyassistant.org/media/publications/IEEE_magazine_2018.pdf

- [DKDH14] S Das, ADI Kramer, LA Dabbish, JI Hong. "Increasing security sensitivity with social proof: A large-scale experimental confirmation." CCS 2014. (2014) <https://dl.acm.org/doi/10.1145/2660267.2660271>
- [DKDH14] S Das, THJ Kim, LA Dabbish, JI Hong. "The Effect of Social Influence on Security Sensitivity." SOUPS 2014. (2014) <https://dl.acm.org/doi/10.5555/3235838.3235851>
- [DKDH15] S Das, ADI Kramer, LA Dabbish, JI Hong. "The role of social influence in security feature adoption." CSCW 2015. (2015) <https://dl.acm.org/doi/10.1145/2675133.2675225>
- [FPML+18] D Freed, J Palmer, D Minchala, K Levy, T Ristenpart, N Dell. "'A Stalker's Paradise': How Intimate Partner Abusers Exploit Technology." CHI 2018. (2018) <https://dl.acm.org/doi/10.1145/3173574.3174241>
- [FTC21] FTC Warns Health Apps and Connected Device Companies to Comply With Health Breach Notification Rule. (September 2021) <https://www.ftc.gov/news-events/news/press-releases/2021/09/ftc-warns-health-apps-connected-device-companies-comply-health-breach-notification-rule>
- [FYS21] Yuanyuan Feng, Yaxing Yao, Norman Sadeh, "A Design Space for Privacy Choices: Towards Meaningful Privacy Control in the Internet of Things." Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. (May 2021) <https://www.privacyassistant.org/media/publications/chi21-design-space.pdf>
- [GFR+22] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain and Norman Sadeh, "Helping Mobile Application Developers Create Accurate Privacy Labels." IWPE'22. (May 2022) https://usableprivacy.org/static/files/Gardner_IWPE_2022.pdf
- [HK14] Timothy M. Hale, Joseph C. Kvedar. "Privacy and Security Concerns in Telehealth." AMA Journal of Ethics, 16(12): 981-985. (2014) doi://10.1001/virtualmentor.2014.16.12.jdsc1-1412 <https://journalofethics.ama-assn.org/article/privacy-and-security-concerns-telehealth/2014-12>
- [Hong17] "The privacy landscape of pervasive computing." IEEE Pervasive Computing 16 (3), 40-48. (2017) <https://ieeexplore.ieee.org/document/7994573>
- [HPW+19] H Habib, S Pearman, J Wang, Y Zou, A Acquisti, LF Cranor, N Sadeh, "" It's a scavenger hunt": Usability of Websites' Opt-Out and Data Deletion Choices." Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. (April 2020) https://usableprivacy.org/static/files/habib_chi_2020.pdf
- [IoTPI22a] Carnegie Mellon University's "Privacy Infrastructure for the Internet of Things" portal. (2022) <https://www.iotprivacy.io/login>

[IoTPI22b] “Personalized Privacy Assistants for the Internet of Things” project website. (2022)
<https://privacyassistant.org/iot/>

[JGRY+22] H Jin, B Guo, R Roychoudhury, Y Yao, S Kumar, Y Agarwal, JI Hong. “Exploring the Needs of Users for Supporting Privacy-Protective Behaviors in Smart Homes.” CHI 2022. (2022)
<https://dl.acm.org/doi/abs/10.1145/3491102.3517602>

[JLHK+22] H Jin, G Liu, D Hwang, S Kumar, Y Agarwal, J Hong. “Peekaboo: A Hub-Based Approach to Enable Transparency in Data Processing within Smart Homes.” IEEE Security and Privacy 2022. (2022)
<https://www.computer.org/csdl/proceedings-article/sp/2022/131600b571/1C1O8pmx6jm>

[KWDH+22] I Krsek, K Wenzel, S Das, JI Hong, L Dabbish. “To Self-Persuade or be Persuaded: Examining Interventions for Users’ Privacy Setting Selection.” CHI 2022. (2022)
<https://sauvikdas.com/uploads/paper/pdf/38/file.pdf>

[LAH18] T Li, Y Agarwal, JI Hong. “Coconut: An IDE plugin for developing privacy-friendly apps.” Ubicomp 2018. (2018) <https://dl.acm.org/doi/10.1145/3287056>

[LAS+16] B Liu, MS Andersen, F Schaub, H Almuhiemedi, S Zhang, N Sadeh, A Acquisti, and Y Agarwal, “Follow My Recommendations: A Personalized Assistant for Mobile App Permissions”, Symposium on Usable Privacy and Security (SOUPS '16). (June 2016)
<https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-liu.pdf>

[LCYB+21] T Li, C Cobb, JJ Yang, S Baviskar, Y Agarwal, B Li, L Bauer, JI Hong. “What makes people install a COVID-19 contact-tracing app? Understanding the influence of app design and individual difference on contact-tracing app adoption intention.” Pervasive and Mobile Computing. (2021) <https://www.sciencedirect.com/science/article/pii/S1574119221000833>

[LLDH21] T Li, E Louie, L Dabbish, JI Hong. “How Developers Talk About Personal Data and What It Means for User Privacy: A Case Study of a Developer Forum on Reddit.” CSCW 2021. (2021) <https://dl.acm.org/doi/10.1145/3432919>

[LLS14] B Liu, J Lin, N Sadeh, “Reconciling Mobile App Privacy and Usability on Smartphones: Could User Privacy Profiles Help?” 23rd International Conference on the World Wide Web (WWW 2014). (July 2014) https://www.cs.cmu.edu/~bliu1/Bin_Liu_WWW2014_Reconciling.pdf

[LLS+14] J Lin, B Liu, N Sadeh, and JI Hong, “Modeling Users’ Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings.” 2014 ACM Symposium on Usable Security and Privacy (SOUPS 2014). (July 2014)
<https://www.usenix.org/system/files/conference/soups2014/soups14-paper-lin.pdf>

- [LNAH21] T Li, EB Neundorfer, Y Agarwal, JI Hong. "Honeysuckle: Annotation-guided code generation of in-app privacy notices." Ubicomp 2021. (2021)
<https://dl.acm.org/doi/10.1145/3478097>
- [LRAC+22] T Li, K Reiman, Y Agarwal, LF Cranor, JI Hong. "Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels." CHI 2022. (2022)
<https://dl.acm.org/doi/10.1145/3491102.3502012>
- [QNG+19] Xue Qin, Robert Neuhaus, Diego Gonzales, Xiaoyin Wang, Travis Breaux, and Jianwei Niu. "Taming web views in the detection of Android privacy leaks." In Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security (HotSoS '19). Association for Computing Machinery, New York, NY, USA, Article 23, 1–2. (2019)
<https://doi.org/10.1145/3314058.3317732>
- [RBW+19] Abhilasha Ravichander, Alan W Black, Shomir Wilson, Thomas Norton and Norman Sadeh, "Question Answering for Privacy Policies: Combining Computational and Legal Perspectives." 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP 2019), Hong Kong, China. (November 2019)
https://usableprivacy.org/static/files/ravichander_emnlp_2019.pdf
- [RBP+23] Marco Robol, Travis D. Breaux, Elda Paja, Paolo Giorgini. "Consent verification monitoring." *ACM Transactions on Software Engineering Methodology (TOSEM)*. (January 2023) <https://arxiv.org/pdf/2206.06406>
- [SHCH20] Y Song, Y Huang, Z Cai, JI Hong. "I'm All Eyes and Ears: Exploring Effective Locators for Privacy Awareness in IoT Scenarios." CHI 2020. (2020)
<https://dl.acm.org/doi/abs/10.1145/3313831.3376585>
- [SSC+22] Peter Story, Daniel Smullen, Rex Chen, Yaxing Yao, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, Florian Schaub. "Increasing Adoption of Tor Browser Using Informational and Planning Nudges." Proceedings on Privacy Enhancing Technologies 2022 no. 2, pp. 1–32. (2022) https://peterstory.me/publications/story_popets_2022.pdf
- [SWH+16] Rocky Slavin, Xiaoyin Wang, Mitra Bokaei Hosseini, James Hester, Ram Krishnan, Jaspreet Bhatia, Travis D. Breaux, and Jianwei Niu. "Toward a framework for detecting privacy policy violations in android application code." In Proceedings of the *38th International Conference on Software Engineering (ICSE '16)*. Association for Computing Machinery, New York, NY, USA, 25–36. (2016) <https://doi.org/10.1145/2884781.2884855>
- [SYF+21] D Smullen, Y Yao, Y Feng, N Sadeh, A Edelstein, and R Weiss. "Managing Potentially Intrusive Practices in the Browser: A User-Centered Perspective." Proceedings on Privacy Enhancing Technologies 2021 (4), 500-527. (October 2021)
<https://sciendo.com/it/article/10.2478/popets-2021-0082>

[TLV22] M Tahaei, T Li, K Vaniea. "Understanding Privacy-Related Advice on Stack Overflow." PETS 2022. (2022) <https://petsymposium.org/2022/files/papers/issue2/popets-2022-0038.pdf>

[TRLH+22] M Tahaei, KM Ramokapane, T Li, Ji Hong, A Rashid. "Charting App Developers' Journey Through Privacy Regulation Features in Ad Networks." PETS 2022. (2022) <https://mohammad.tahaei.com/publication/pets-2022-charting-regulations-adnetworks/pets-2022-charting-regulations-adnetworks.pdf>

[UPPP22] Usable Privacy Policy Project website.(2022) <https://usableprivacy.org/>

[WQH+18] Xiaoyin Wang, Xue Qin, Mitra Bokaei Hosseini, Rocky Slavin, Travis D. Breaux, and Jianwei Niu. "GUILeak: tracing privacy policy claims on user input data for Android applications." In Proceedings of the *40th International Conference on Software Engineering (ICSE '18)*. Association for Computing Machinery, New York, NY, USA, 37–47. (2018) <https://doi.org/10.1145/3180155.3180196>

[WSL+18] Shomir Wilson, Florian Schaub, Frederick Liu, Kanthashree Mysore Sathyendra, Daniel Smullen, Sebastian Zimmeck, Rohan Ramanath, Peter Story, Fei Liu, Norman Sadeh, Noah A. Smith, "Analyzing Privacy Policies at Scale: From Crowdsourcing to Automated Annotations." *ACM Transactions on the Web*, 13, 1. (December 2018) https://usableprivacy.org/static/files/swilson_tweb_2018.pdf

[WZWM+15] J Witschey, O Zielinska, A Welk, E Murphy-Hill, C Mayhorn, T Zimmerman. "Quantifying developers' adoption of security tools." 2015 Joint Meeting on Foundations of Software Engineering. (2015) <https://dl.acm.org/doi/10.1145/2786805.2786816>

[ZFB+21] Shikun Zhang, Yuanyuan Feng, Lujo Bauer, Lorrie Faith Cranor, Anupam Das, and Norman Sadeh, "'Did you know this camera tracks your mood?': Understanding Privacy Expectations and Preferences in the Age of Video Analytics." *Proceedings on Privacy Enhancing Technologies*, 2021, 1. (April 2021) <https://www.petsymposium.org/2021/files/papers/issue2/popets-2021-0028.pdf>

[ZFD+20] Shikun Zhang, Yuanyuan Feng, Anupam Das, Lujo Bauer, Lorrie Faith Cranor, Norman Sadeh, "Understanding People's Privacy Attitudes Towards Video Analytics Technologies." School of Computer Science Technical Report CMU-ISR-20-114 Carnegie Mellon University. (December 2020) <http://reports-archive.adm.cs.cmu.edu/anon/isr2020/CMU-ISR-20-114.pdf>

[ZSS+19] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel Reidenberg, N. Cameron Russell, and Norman Sadeh, "MAPS: Scaling Privacy Compliance Analysis to a Million Apps." *Privacy Enhancing Technologies Symposium (PETS 2019)*, 3. (July 2019) <https://usableprivacy.org/static/files/popets-2019-maps.pdf>

[ZWS+20] Xueling Zhang, Xiaoyin Wang, Rocky Slavin, Travis Breaux, and Jianwei Niu. “How does misconfiguration of analytic services compromise mobile privacy?” In Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering (ICSE '20). Association for Computing Machinery, New York, NY, USA, 1572–1583. (2020)
<https://doi.org/10.1145/3377811.3380401>

[ZWZ+17] S Zimmeck, Z Wang, L Zou, R Iyengar, B Liu, F Schaub, S Wilson, N Sadeh, SM Bellovin, JR Reidenberg, "Automated Analysis of Privacy Requirements for Mobile Apps." NDSS'17: Network and Distributed System Security Symposium. (February 2017)
<https://usableprivacy.org/files/news/NDSS17.pdf>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Center for AI and Digital Policy (CAIDP)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



Comments of the

THE CENTER FOR AI AND DIGITAL POLICY (CAIDP)
to the
OFFICE OF SCIENCE AND TECHNOLOGY POLICY (OSTP)
on the
THE PROMOTION OF PRIVACY-ENHANCING TECHNOLOGIES (PETs)

On behalf of the Center for AI and Digital Policy (CAIDP), we write in response to the Request for Information (RFI) on Advancing Privacy-Enhancing Technologies.¹ The CAIDP is an independent non-profit organization that advises national governments and international organizations on artificial intelligence (AI) and digital policy. The CAIDP currently serves as an advisor on AI policy to the OECD, the Global Partnership on AI, the Council of Europe, the European Union, and other international and national organizations. We work with more than 200 AI policy experts in over 50 countries.

The CAIDP has previously expressed strong support for AI policies that advance democratic values and promote broad social inclusion based on fundamental rights, democratic institutions, and the rule of law.² In our report *Artificial Intelligence and Democratic Values*, we set out several recommendations for national governments.³ In the U.S. country report, we noted favorably that the “U.S. and UK announced plans to promote Privacy Enhancing Technologies (PETs), including low-data AI, the deletion of unnecessary data, and techniques for robust anonymity.”⁴

The Office of Science and Technology Policy (OSTP) has now issued a “Request for Information on Advancing Privacy-Enhancing Technologies.”⁵ We support the OSTP initiative.⁶

¹ “Request for Information on Advancing Privacy-Enhancing Technologies.” *Federal Register: The Daily Journal of the United States Government*, The Office of Science and Technology Policy, June 6, 2022. <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>.

² CAIDP Statements, <https://www.caidp.org/statements/>

³ CAIDP, *Artificial Intelligence and Democratic Values* (2022), <https://www.caidp.org/reports/aidv-2021/>

⁴ *Id.* at 472.

⁵ *Id.*, *supra* 1.

⁶ CAIDP wrote earlier to the OSTP in support of PETs. “New technologies such as AI pose new challenges for privacy, dignity, autonomy, and equality. Metrics for explainability, interpretability, and transparency should be established to protect fundamental rights, human

We agree that PETs can “promote continued innovation in emerging technologies in a manner that supports human rights and shared values of democratic nations.”⁷ We further support the goal “to accelerate the responsible development and adoption of PETs in a manner that maximizes the benefit to individuals and society, including increasing equity for underserved or marginalized groups and promoting trust in data processing and information technologies.”⁸ We believe this is one of the important challenges facing the United States in the realm of AI and digital policy.⁹

In these comments, we clarify the meaning of the critical term “Privacy Enhancing Techniques,” place the current OSTP initiative in the context of other government efforts to promote PETs, identify examples of poorly conceived Privacy Enhancing Techniques, and warn that without adequate guidance many more unsafe systems will be deployed, placing users at risk and diminishing public confidence in this initiative. We make three specific recommendations:

- 1) Conduct independent evaluation of PETs prior to deployment
- 2) Promote expiration dates for commercial PETs
- 3) Incorporate PETs in the AI Bill of Rights

We need to underline at the outset that we use the term “Privacy Enhancing Technologies” precisely, as it was originally conceived, to describe **techniques that “minimize or eliminate the collection of personal data.”**¹⁰ This is also the definition adopted by the National Academies study on the HIPAA Privacy Rule which recommended that “the federal government should support the development and use of Genuine privacy-enhancing techniques

well-being, and to increase public trust. These metrics alongside Privacy Enhancing Technologies would help protect privacy.” *Comments of CAIDP to OSTP on National Artificial Intelligence Research and Development Strategic Plan* at 4, Mar. 4, 2022 <https://www.caidp.org/statements/> .

⁷ Id., supra note 1.

⁸ Id.

⁹ In a statement to the US Congress, we set out a broad range of recommendations for the United States, including implementing the OECD AI Principles, establishing a process for meaningful public participation in the development of national AI policy, establishing an independent agency for AI oversight, establishing a right to algorithmic transparency, and supporting the Universal Guidelines for AI. *CAIDP Statement to House Armed Services Committee regarding US AI Policy* (Mar. 25, 2021), <https://www.caidp.org/app/download/8305652763/CAIDP-HASC-03252021.pdf>

¹⁰ Marc Rotenberg, *Eurocrats Do Good Privacy: The contrast between a decorated cryptographer in Europe and one trying to avoid prosecution in the United States is more than curious*, *Wired*, May 1, 1996, (describing early government efforts to promote “Privacy Enhancing Technologies”), <https://www.wired.com/1996/05/eurocrats/>. See also Herbert Burkert, *Privacy-Enhancing Technologies: Typology, Critique, Vision* in *Technology and Privacy: the New Landscape* 143–67 (eds., Philip E. Agre & Marc Rotenberg 1997);

that *minimize or eliminate the collection of personally identifiable data.*¹¹ As the OSTP itself has explained, examples of PETs include utilizing low-data artificial intelligence, deleting unnecessary data, and creating techniques for robust anonymity.¹² From this perspective, **the aim of PETs is not to enable further transfers of personal data but rather to limit the collection of personal data in the first instance.** There are many reasons to favor this approach

- Data breaches, criminal hacking, and espionage remain a primary concern for all organizations that choose to collect personal data.¹³
- Genuine PETs reduce privacy and security risks as data that is not collected cannot be misused by the data collector or be subject to data breach.¹⁴
- Genuine PETs protect vulnerable groups, particularly children. For example, President Biden recently called on Congress to “strengthen privacy protections, ban targeted advertising to children, and demand tech companies stop collecting personal data on our children.”¹⁵
- Genuine PETs are aligned with well-established privacy norms, including the GDPR and many US privacy laws.¹⁶
- PETs typically seek to implement Fair Information Practices, and where possible, to minimize or eliminate the collection of personally identifiable information.¹⁷

¹¹ The Institute of Medicine, *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*, Sharyl J. Nass, Laura A. Levit, and Lawrence O. Gostin, Editors 55 (2009) (Recommendation IIIa) (emphasis added)

¹² “U.S. and U.K. Governments Collaborate on Prize Challenges to Accelerate Development and Adoption of Privacy-Enhancing Technologies.” *The White House*, The United States Government, 14 June 2022, <https://www.whitehouse.gov/ostp/news-updates/2022/06/13/u-s-and-uk-governments-collaborate-on-prize-challenges-to-accelerate-development-and-adoption-of-privacy-enhancing-technologies/>.

¹³ IBM, *Cost of a Data Breach 2021* (“2021 had the highest average cost in 17 years: Data breach costs rose from USD 3.86 million to USD 4.24 million, the highest average total cost in the 17-year history of this report.”), <https://www.ibm.com/security/data-breach>

¹⁴ Testimony of Marc Rotenberg, *Privacy in the Commercial World*, U.S. House of Representatives, March 1, 2001 https://archive.epic.org/privacy/testimony_0301.html.

¹⁵ The White House, *Remarks of President Joe Biden, State of the Union Address*, March 1, 2022, <https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/03/01/remarks-of-president-joe-biden-state-of-the-union-address-as-delivered/>

¹⁶ General Data Protection Regulation (GDPR), Art. 5(1)(c) (“Personal data shall be: . . . adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data minimisation’) (Principles relating to processing of personal data.)

¹⁷ Marc Rotenberg, *Fair Information Practices and the Architecture of Privacy, (What Larry Doesn't Get)*, 2001 Stan. Tech. L. Rev. 1 (2001). Marc Rotenberg, *Protecting Human Dignity in*

- Genuine PETs encourage the development of innovative techniques that are less dependent on the collection of personal data.¹⁸
- Genuine PETs minimize processing and are therefore aligned with emerging norms for AI policy that consider the environmental impact of big data models.¹⁹
- Genuine PETs are aligned with democratic values as they reduce the risk of mass surveillance.

We recognize that there are other techniques, including Privacy by Design and Privacy-Preserving Technologies (such as Differential Privacy) that incorporate techniques to enable the transfer of personal data. Where it is necessary to transfer personal data, the most robust methods should be adopted. That explains, for example, the central requirement that communications networks are built on end-to-end encryption.

We also respect the desire to enable data analysis for medical research and other fields of social benefit, but we caution that these same techniques for data aggregation can easily be used to enable mass surveillance and target vulnerable communities.²⁰ **The OSTP must be extremely cautious, in its efforts to promote data transfers, that it does not enable methods that could easily be turned against democratic values and marginalized communities.** As the OSTP itself has recently stated:

[T]here are also risks that PETs could provide a false veneer of privacy, misleading people into believing that a data sharing arrangement is more private than it really is. Furthermore, in some cases, PETs could exacerbate existing problems with certain types of data analysis, such as discriminatory analysis resulting from biased data.²¹

the Digital Age, UNESCO (November 2000),
<https://unesdoc.unesco.org/ark:/48223/pf0000121984>

¹⁸ See, e.g., David Chaum, *Achieving Electronic Privacy*, *Scientific American* (August 1992) (Chaum developed techniques to provide authentication without identification, a cornerstone of the PETs paradigm as such techniques enable transactions without requiring the disclosure of personal data), <https://www.scientificamerican.com/article/achieving-electronic-privacy/>

¹⁹ “Recommendation on the Ethics of Artificial Intelligence.” *Unesdoc.unesco.org*, Nov. 23, 2021, <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.

²⁰ Chris Buckley and Paul Mozur, *How China Uses High-Tech Surveillance to Subdue Minorities*, *NY Times*, May 22, 2019, <https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html>

²¹ The White House, *Advancing a Vision for Privacy-Enhancing Technologies*, June 28, 2022 <https://www.whitehouse.gov/ostp/news-updates/2022/06/28/advancing-a-vision-for-privacy-enhancing-technologies/>

This concern should guide the OSTP’s work in this field.

Subsequent Developments with PETS

The Madrid Privacy Declaration

Technical experts and civil society organizations have carried forward the effort to promote genuine Privacy Enhancing Technologies, based on the definition set out above. The Madrid Privacy Declaration of 2009, undertaken at the annual meeting of the Data Protection commissioners, reaffirmed international instruments for privacy protection, identified new challenges, and called for concrete actions.²² The Madrid Declaration was endorsed by over 100 organizations and 200 experts. Among other recommendations, the Declaration “Reaffirm[ed] support for *genuine Privacy Enhancing Techniques that minimize or eliminate the collection of personally identifiable information* and for meaningful Privacy Impact Assessments that require compliance with privacy standards.” The Declaration also “Recommend[ed] comprehensive research into the adequacy of techniques that deidentify data to determine whether in practice such methods safeguard privacy and anonymity.”

These two recommendations – support for genuine PETS and research to evaluate such techniques – could be the cornerstone of the OSTP’s work going forward.

The G20

Privacy Enhancing Technologies also arise in the context of global efforts to promote Data Free Flows with Trust (DFFT). The concept was developed by the recently deceased, former Japanese Prime Minister Shinzo Abe.²³ Prime Minister Abe underscored the importance of privacy protection, explaining that the DFFT regime should be built on “non-personal data.”²⁴ The G20 nations, of which the US is a member, have endorsed this concept of PETS. At the 2020 G20 meeting the Digital Ministers stated:

The cross-border flow of data, information, ideas and knowledge generates higher productivity, greater innovation, and improved sustainable development. At the same time, we recognize that the free flow of data raises certain challenges, such

²² “Madrid Declaration.” *The Public Voice*, Nov. 3, 2009, <https://thepublicvoice.org/madrid-declaration>.

²³ *Prime Minister Abe’s AI Policy and Data Governance Legacy*, CAIDP Update 1.7 (Sept. 1, 2020), <https://dukakis.org/center-for-ai-and-digital-policy/caidp-update-prime-minister-abes-ai-and-data-governance-legacy/>

²⁴ *Id.*

as the protection of privacy and personal data. G20 members recognize the need to address these challenges, in accordance with relevant applicable legal frameworks, which can further facilitate data free flow and strengthen consumer and business trust, without prejudice to legitimate public policy objectives, including by: . . . exploring and better understanding technologies such as privacy enhancing technologies (PETs).²⁵

The G7

The G7 is another global forum that has promoted Privacy Enhancing Technologies.²⁶ The Data Protection and Privacy officials of the G7 nations also issued a statement on Data Free Flows with Trust which said, “human dignity, must be central to AI design; AI must be transparent, comprehensible, and explainable; and the data protection principles of purpose limitation and data minimization must apply to AI.” They said that “‘red lines’ are needed for AI systems that are not compatible with our values and fundamental rights.”²⁷

1) The Need for Independent Evaluation of PET

PETs offer great promise. However, it is necessary to ensure that there is independent evaluation of these techniques prior to deployment. Companies and government agencies should not be allowed to represent that they have established Privacy Enhancing Techniques without independent evaluation. Several consumer privacy cases, as well as government surveillance programs, have demonstrated the shortcomings of that approach.

For example, in 2008, the FTC sued Ask.com for misrepresenting the privacy technique for the search engine AskEraser, after a group of consumer privacy organization identified flaws in the privacy technology.²⁸ As the organizations explained:

The company purports to provide an Internet search engine that provides privacy protection by limiting the collection and use of Internet search histories. In fact,

²⁵ *Ministerial Declaration*, G20 Digital Economy Ministers Meeting, July 22, 2020, <http://www.g20.utoronto.ca/2020/2020-g20-digital-0722.html>.]

²⁶ *G7 Digital and Technology Track – Annex 2: G7 Roadmap for Cooperation on Data Free Flow with Trust*, Apr 28, 2021, http://www.g8.utoronto.ca/ict/2021-annex_2-roadmap.html

²⁷ G7 United Kingdom 2021, *Data Free Flows with Trust*, Sept. 8, 2021, <https://www.caiddp.org/app/download/8342900463/g7-attachment-202109.pdf>

²⁸ EPIC, Center for Digital Democracy, Consumer Action, Fairfax County Privacy Council, Patient Privacy Rights, U.S. Bill of Rights Foundation, *In the Matter of Ask.Com, Complaint and Request for Injunction, Request for Investigation and for Other Relief*, Jan. 19, 2008, https://epic.org/wp-content/uploads/privacy/ask/epic_askeraser_011908.pdf

the product does not work as advertised: Internet search histories will be retained without notice to Internet users. Moreover, AskEraser requires Internet users to disable genuine privacy features, and it exposes Internet users to additional tracking, monitoring, and profiling by means of a Persistent Identifier.

Ask.com claimed that AskEraser, would delete search histories “within hours.” The company advertised that the new search tool “will offer its searchers unmatched control over their privacy.” However, Ask.com placed a persistent unique identifier, also known as a “cookie,” on the user’s device that would be stored for two years. With the persistent identifier, the company gathered sensitive personal data such as IP addresses, the address of the last URL visited before arriving at Ask.com. And the company actually prevented users from deleting the persistent identifier if they were to use the service. As a consequence, Ask.com “privacy technique” allowed the company to track and monitor the user for as long as the user continued to use the service. The FTC determined that this was an unfair and deceptive trade practice.²⁹

A similar problem arose with Snapchat, a social media app which claimed that users could make photos ‘vanish.’³⁰ In fact, an investigation revealed that the photos were retained. Snapchat also transmitted users’ location data and collected their address books without consent. According to the Electronic Privacy Information Center (EPIC), Snapchat made multiple misrepresentations to consumers about its product that stood in stark contrast to how the app actually worked. The Federal Trade Commission agreed with EPIC, pursued an investigation, and obtained a settlement.³¹

²⁹ *In the Matter of Snapchat, Inc.*, No. C-4501, before the Federal Trade Commission, Dec. 23, 2014, <https://www.ftc.gov/system/files/documents/cases/141231snapchatcmpt.pdf>

³⁰ *Privacy watchdog EPIC files complaint against Snapchat with FTC*, Los Angeles Times, May 17, 2013 (“Snapchat is the app that promises to delete photos but it doesn’t,” said Marc Rotenberg, EPIC’s executive director. “We have no problem with apps that make photos vanish. But they should work as promised, and if they don’t the Federal Trade Commission should investigate.”), <https://www.latimes.com/business/la-xpm-2013-may-17-la-fi-tn-privacy-watchdog-epic-files-complaint-against-snapchat-with-ftc-20130517-story.html>; EPIC, *In the Matter of Snapchat: Complaint, Request for Investigation, Injunction, and Other Relief*, May 16, 2013, (“Despite promising to its users that photos and videos sent via Snapchat will ‘disappear forever,’ Snapchat photos and videos remain available to others even after users are informed that the photos and videos have been deleted.”) <https://epic.org/wp-content/uploads/privacy/ftc/EPIC-Snapchat-Complaint.pdf>

³¹ Federal Trade Commission, *Snapchat Settles FTC Charges That Promises of Disappearing Messages Were False: Snapchat Also Transmitted Users’ Location and Collected Their Address Books Without Notice Or Consent*, May 8, 2014, <https://www.ftc.gov/news-events/news/press-releases/2014/05/snapchat-settles-ftc-charges-promises-disappearing-messages-were-false>. See also Andrea Peterson, *Snapchat agrees to settle FTC charges that it deceived users*, The

The examples of AskEraser and Snapchat demonstrate that there must be independent evaluation of Privacy Enhancing Techniques. Any strategy for PETs that fails to incorporate such safeguards will almost certainly encourage faulty technology, place users at risk, and diminish public support for the initiative.

A related problem concerns the definition of PETs. For example, the recent US-UK initiative relies on a loose definition of “PETs” to justify what will likely be massive collections of personal data by law enforcement agencies. UK Minister Julia Lopez, said that she was “delighted that the U.K. and U.S. are working with regulators on both sides of the Atlantic to help realize the potential of novel privacy-enhancing technologies (PETs) to tackle financial crime.” She went on to describe how the U.K.’s National Data Strategy outlines “the promise of trustworthy data access.” According to the UK official, “PETs have the potential to facilitate new forms of data collaboration to tackle the harms of money laundering, while protecting citizens’ privacy.”³² But the obvious question that needs to be considered is what mechanisms of oversight will be established to ensure that citizens privacy will be protected when law enforcement agencies are in control of the privacy technologies. In most simple terms, “who will watch the watchers?” If that question cannot be answered at the outset, such programs should not go forward.

There are many examples where such representations of privacy preserving techniques by government agencies turned out, on closer inspection, to be false. For example, the US federal agencies responsible for the development of wiretapping techniques in the 1990s, known as “Carnivore,” claimed that the only data that would be accessed was the data lawfully accessible under a judicial warrant. But independent investigation revealed that the data could be obtained by government officials outside the scope of the warrant.³³ Similarly, the developers of the Total Information Awareness program claimed that they had established privacy safeguards but that was only with regard to data access by low-level government employees and did nothing to limit

Washington Post, May 8, 2014 (“Julia Horwitz, Consumer Protection Counsel at the Electronic Privacy Information Center which originally complained to the FTC about Snapchat, told the Post it was happy with the resolution. ‘We’re extremely pleased that the FTC is taking its data privacy protection seriously and is recognizing behaviors by companies like Snapchat that breach promises to consumers,’ she said. ‘This was a real success. But this consent order’s true effectiveness depends upon the agency’s consistent enforcement over the next 20 years,’ she cautioned.”)

³² “U.S. and U.K. Governments Collaborate on Prize Challenges to Accelerate Development and Adoption of Privacy-Enhancing Technologies.” *The White House*, The United States Government, 14 June 2022, <https://www.whitehouse.gov/ostp/news-updates/2022/06/13/u-s-and-uk-governments-collaborate-on-prize-challenges-to-accelerate-development-and-adoption-of-privacy-enhancing-technologies/>.

³³ IIT Research Institute, *Independent Review of the Carnivore System*, Dec. 8, 2000, https://www.justice.gov/archive/jmd/carniv_final.pdf

the ability of department heads to repurpose the use of the data collected. More recently implementation of the Cyber Information Sharing Act has been subject to scrutiny precisely because it enables the transfer of personal data to government agencies outside the judicial process, relying on novel techniques for privacy protection.

Another project of concern is currently underway at the Global Partnership on AI (GPAI). The GPAI is investigating the use of AI-powered cameras for surveillance and monitoring of outdoor and working environments.³⁴ The intention is that the cameras will be programmed to intervene if specified events are detected, such as a fire or medical emergency. But of course, this is also a system of mass surveillance that will capture the images and conversations of identifiable individuals in real time. The technical challenge will be to eliminate the massive amount of personal data that will be routinely gathered with the goal of identifying the events of interest without compromising privacy. But the organizational challenges, rarely considered at the outset, will be to prevent the reuse of the data gathered for other unrelated purposes.

A simpler and more effective solution may be simply to avoid systems that involve the massive collection of unnecessary personal data and deploy instead techniques that are specifically designed to identify the risk of fire or to alert personnel in the case of medical emergencies. These techniques are likely to be more reliable, more effective, less complex, and less expensive. A device designed solely to detect fire, as compared with a general-purpose system that gathers massive amounts of personal data, is an excellent example of a PET as it would accomplish its task without collecting unnecessary personal data.

2) The Need for Expiration Dates for PETs

In addition to independent evaluation for PETs, we also strongly recommend a certification program that would indicate a time period during which the security of the PET would be assured. This is necessary because rapid advances in cryptography and data analytics have made clear that popular techniques will over time no longer be secure. For example, MD5 a popular cryptographic hashing function, developed by Ron Rivest in 1991, was later found to have extensive vulnerabilities.³⁵

A National Academies of Sciences study that seeks to promote Privacy Enhancing and Privacy-Preserving Techniques recommended the use of expiration dates to provide legal certainty for those who offer and deploy Privacy Enhancing Techniques.³⁶ We support this

³⁴ “AI at Work Observation Platform.” *GPAI*, Nov. 2021, <https://gpai.ai/projects/future-of-work/ai-at-work-observation-platform/>.

³⁵ Wikipedia, MD5, <https://en.wikipedia.org/wiki/MD5>

³⁶ Robert M Groves, Michael E Chernew, Piet Daas, Cynthia Dwork, Ophir Frieder, Hosagrahar V Jagadish, Frauke Kreuter, Sharon Lohr, James P Lynch, Colm O’Muircheartaigh, Trivellore

proposal and recommend that it be incorporated in the OSTP initiative to promote Privacy Enhancing Technologies.

3) Advancing PETs in the Context of the AI Bill of Rights

Finally, we recommend that the OSTP advance Privacy Enhancing Technologies in the context of the AI Bill of Rights much as civil society organizations and technology experts promoted Privacy Enhancing Technologies and related research in the context of the Madrid Privacy Declaration. *The AI Bill of Rights should incorporate an affirmative obligation to adopt Privacy Enhancing Techniques that have been subject to independent review and a privacy impact assessment.*

Conclusion

We support the OSTP initiative to promote Privacy Enhancing Technologies. Such techniques are particularly important as more services are digitized and more data is gathered. But we caution that genuine PETs will minimize or eliminate the collection of personal data. It is a fundamental conceptual mistake to assume that PETs are intended to facilitate the transfer of personal data.

Many of the great challenges facing our nation, including measuring the consequences of climate change, require virtually no collection of personally identifiable information.³⁷ But the collection of personal data poses specific challenges that cannot be ignored as the OSTP itself has acknowledged. The precise definition of Privacy Enhancing Techniques will help ensure that the social benefits are maximized, the risks to democratic values and marginalized groups are minimized, and true innovation occurs.

Thank you for your consideration of our views. We welcome to opportunity to discuss further.

Raghunathan, Roberto Rigobon, and Marc Rotenberg. 2017. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*, National Academies of Sciences, Engineering, and Medicine paper, <https://privacytools.seas.harvard.edu/publications/innovations-federal-statistics-combining-data-sources-while-protecting>

³⁷ Marc Rotenberg, *Let's Use Government Data to Make Better Policy: It's a no-brainer, as long as privacy concerns are taken seriously*, Scientific American (Oct. 4, 2017), <https://blogs.scientificamerican.com/observations/lets-use-government-data-to-make-better-policy/>



Sincerely,

Marc Rotenberg
CAIDP President

Merve Hickok
CAIDP Chair

Karine Caunes
CAIDP Global Program Director

/S/

Dr. Lorraine Kisselburgh
CAIDP Board Members

Maison Bergeron
CAIDP Research Assistant

Sandra Lattner
CAIDP Research Assistant

Center for AI and Digital Policy

1100 13th St., NW, Suite 800

Washington, DC 20005

caidp.org

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Chu, Bessie

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

From:
To:
Subject: Fwd: PET Recommendations
Date: Friday, July 8, 2022 2:19:13 PM

Get [Outlook for iOS](#)

From: Bessie Chu <>
Sent: Friday, July 8, 2022 2:08 PM
To: PETS-RFI <PETS-RFI@nitr.gov>
Subject: PET Recommendations

Hello,

I am the Product Manager at Cape Privacy, a company that specializes in PETs. A few thoughts below.

1. Investing in and educating the public about PETs is crucial. China is going to be able to leapfrog the US in areas such as healthcare and AI/ML research generally due to unfettered access to their citizens' data. The US can be an AI/ML powerhouse, but it requires the consent and therefore understanding of its citizens of the benefits and how risk to their personal data can be reduced.
2. Investing in multiple areas of PETs. The US should place multiple bets on confidential computing (hardware based frameworks) as well as software-based frameworks such as secure multiparty computation and homomorphic encryption to process data. In addition to the data processing layer, investments should be done in areas such as differential privacy to democratize the use of more data. The census is a good example of this, but a lot more data could be unlocked if there is some standardization.
3. Supporting a broad range of institutions doing this work in the United States and allied nations - tertiary institutions, large companies, and start-ups like mine are doing this work at multiple angles of attack. Other states are doing so, and thus the United States government should be making investments and playing an active role in coordinating research in a way that also benefits national security.

Best,

Bessie Chu

All e-mails to and from this account are for NITRD official use only and subject to certain disclosure requirements.
If you have received this e-mail in error, we ask that you notify the sender and delete it immediately.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Cloudflare

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



VIA EMAIL

Dr. Alondra Nelson, Director
Office of Science and Technology Policy
Eisenhower Executive Office Building
725 17th Street NW
Washington, D.C.

Re: RFI response: Privacy-Enhancing Technologies, FR Doc. 2022-12432

Dear Dr. Nelson,

We write in response to the Request for Information (RFI) published by the Office of Science and Technology Policy seeking comment on Privacy-Enhancing Technologies (PETs). Cloudflare appreciates the opportunity to comment, and we hope the exercise will result in a useful sharing of information amongst stakeholders.

Cloudflare has faced a variety of challenges in seeking to deploy PETs, ranging from industry resistance to restricting data collection to differing views among governments about new technology. We believe the U.S. government can play an important role in helping both to encourage adoption of new PETs and to build consensus between governments about how to address new PETs.

Cloudflare submits the following comments, which will address our own experience in PETs, including collecting data in privacy-enhancing ways; a brief discussion of research opportunities, risks, and limitations of PETs; our views on the regulation of PETs; and our recommendations for further work to advance and support the development of PETs.

Introduction and Cloudflare Background

Cloudflare is an Internet performance and security company that is on a mission to help build a better Internet. Privacy-enhancing technologies are critical components to achieve that goal. We operate a global network of points of presence in 270+ cities, within 50ms of 95% of the world's population. We use that network to provide services that help people increase the security and performance of their web sites and services.

Cloudflare's Work on PETs

Since its inception in 2010, Cloudflare has invested in the development and deployment of cutting-edge cryptography and Internet standards to help people improve the security and performance of their websites and services. Our company's goal is to enable a research

environment that generates both new knowledge and technology that can lead to innovative products. Cloudflare's research team¹ works hand-in-hand with both product and engineering to help drive long-term positive outcomes for both Cloudflare's customers and the Internet at large. Our researchers are active contributors to the Internet Engineering Task Force² (IETF) and other standards organizations, and our extensive network allows us to bring privacy-enhancing technologies from the experimental stage through deployment in our extensive network.

We've released a number of services that deploy state-of-the-art, privacy-enhancing technologies for DNS and other communications to help individuals, businesses, and governments alike, and we've made substantial contributions to the development of privacy-enhancing Internet protocols. Here are a few notable examples of our work in the area of PETs:

- **DNS Privacy:** Since 2018, Cloudflare has maintained a free DNS resolver (1.1.1.1) that powers approximately 16.1 million DNS queries per second. 1.1.1.1 supports private and encrypted requests from clients via DNS over HTTPs (DoH) and DNS over TLS (DoT), providing confidentiality for DNS traffic in transit to Cloudflare. DNS requests can contain some sensitive data, such as your location, the domains and subdomains you have visited, the time of day requests were submitted, and how long you stayed on certain sites. The 1.1.1.1 resolver never writes the querying IP addresses together with the DNS query to disk, and unless someone is using 1.1.1.1 for Families, we don't receive the EDNS Client Subnet Header (which might reveal location information).

1.1.1.1 also supports other privacy-enhancing mechanisms for DNS, including query name (QNAME) minimization and omission of EDNS Client Subnet extensions in DNS messages, both of which limit information about what and who is requesting resolution of a particular name to authoritative servers.

Even with these standard mechanisms deployed, 1.1.1.1 is operated by a single entity that can still technically link all queries to client IP addresses. As a result, Cloudflare went beyond the basics to develop and deploy Oblivious DNS over HTTPS (ODOH),³ which separates IP addresses from queries, so that no single entity can see both at the same time.⁴

- **Transport Layer Security (TLS)/QUIC:** Cloudflare has played an active role in developing transport security protocols in wide use on the Internet, from our initial deployment of TLS 1.3 in 2014,⁵ to 2018 when we made it the default setting on all of

¹ For background on Cloudflare's research team, including research areas, staff profiles, and publications, see <https://research.cloudflare.com/>.

² See The Internet Engineering Task Force (hereinafter, IETF) at <https://www.ietf.org/>.

³ See IETF, Oblivious DNS over HTTPS, at <https://datatracker.ietf.org/doc/rfc9230/>.

⁴ See E. Kinnear, P. McManus, T. Pauly, T. Verma & C.A. Wood, Oblivious DNS over HTTPS, IETF Independent Submission, June 2022, available at <https://www.rfc-editor.org/rfc/rfc9230.pdf>.

⁵ See Matthew Prince, Introducing Universal SSL, The Cloudflare Blog, Sept. 29, 2014, at <https://blog.cloudflare.com/introducing-universal-ssl/>.

our customer sites.⁶ Cloudflare was also one of the early adopters of QUIC and is today one of the largest deployments.

- **Encrypted Client Hello (ECH):** ECH is a new extension for TLS that will significantly improve its privacy. Today, a number of privacy-sensitive parameters of the TLS connection are negotiated in the clear. This leaves a trove of metadata available to network observers, including the endpoints' identities, how they use the connection, and so on. ECH encrypts the full handshake so that this metadata is kept secret. Together with major web browsers and other stakeholders, Cloudflare has helped develop the standard and support initial experiments.
- **Privacy Pass:** Cloudflare developed and released the privacy-enhancing technology called Privacy Pass in 2017,⁷ and it has since evolved into a standard used by many website operators (it's on track⁸ to be adopted as a standard by the IETF). Privacy Pass lets users prove their identity across multiple sites anonymously without enabling tracking. When people use anonymity services or shared IPs, it makes it more difficult for website protection services like Cloudflare to identify their requests as coming from legitimate users and not bots. To help reduce the friction for these users — which include some of the most vulnerable users online — Privacy Pass provides them with a way to prove they are legitimate across multiple sites on the Cloudflare network. This is done without revealing their identity, and without exposing Cloudflare customers to additional threats from malicious bots.
- **Distributed Aggregation Protocol (DAP)⁹:** Cloudflare researchers have been instrumental in developing DAP, a new protocol to enable privacy-preserving measurement. The protocol involves a large set of clients and a small set of servers and is used to compute aggregate statistics over the clients' inputs without learning the inputs themselves. DAP enables the use of recent advances in multi-party computation in order to ensure that no input is ever seen in the clear by any server. Several organizations are actively working on DAP. For example, the Internet Security Research Group¹⁰ (ISRG) recently raised one million dollars to fund Divvi Up, a system for privacy-respecting aggregate statistics.¹¹ Apple and Google released a white paper last

⁶ See Alessandro Ghedini, You get TLS 1.3! You get TLS 1.3! Everyone gets TLS 1.3!, The Cloudflare Blog, May 16, 2018, at <https://blog.cloudflare.com/you-get-tls-1-3-you-get-tls-1-3-everyone-gets-tls-1-3/>.

⁷ See Nick Sullivan, Cloudflare Supports Privacy Pass, The Cloudflare Blog, Nov. 9, 2017, at <https://blog.cloudflare.com/cloudflare-supports-privacy-pass/>.

⁸ See Tommy Pauly, Steven Valdez, & Christopher A. Wood, The Privacy Pass HTTP Authentication Scheme, IETF, July 6, 2022, at <https://datatracker.ietf.org/doc/draft-ietf-privacypass-auth-scheme/>.

⁹ See Tim Geoghegan, Christopher Patton, Eric Rescorla, & Christopher A. Wood, Distributed Aggregation Protocol for Privacy Preserving Measurement, IETF, May 4, 2022, at <https://datatracker.ietf.org/doc/draft-ietf-ppm-dap/>.

¹⁰ See Internet Society Research Group, at <https://www.abetterinternet.org/>.

¹¹ See Dan Fernelius, ISRG Raises More Than \$1M For Advancing Divvi Up, Divvi Up, June 28, 2022, at <https://divviup.org/blog/2022-06-28-announcing-divviup-funding/>.

year explaining how their COVID-19 exposure notifications system was based on a predecessor of the new DAP standard.¹²

Deep Dive: Standardizing Multi-Party Computation for Privacy-Preserving Measurement (RFI topic #1)

A type of privacy-enhancing technology that OSTP identified as an area of particular interest is secure multi-party computation. Within that field, privacy-preserving measurement (PPM) is a critical operation that refers to the measurement of digital data in ways that are privacy-preserving for common tasks such as sharing, collaboration, and analysis. In its consideration of specific research opportunities to advance PETs (topic #1 of the RFI), PPM should be one of the areas that OSTI includes as part of its national strategy. Organizations and researchers often want to measure certain operations in which sensitive or personal data are present. For example, an Internet infrastructure company such as Cloudflare might want to measure how long it takes for clients to render a web page served by Cloudflare. Or a medical research institution or public health agency might want to measure the spread of infection in a community. In both of these situations, the measuring entity's interest lies in aggregated data, not in individual data points or identities themselves. By using PPM techniques, the measuring entity can calculate the aggregated data and associated values in a way that avoids any exposure of the personal or sensitive data.

In early 2022, IETF established the PPM working group.¹³ The objective of the working group is to investigate and develop protocol standards for aggregating user measurements in privacy-preserving ways, including the deployment of new cryptographic techniques. The goal is to ensure that the server (or multiple, non-colluding servers) can compute the aggregated value without learning the value of individual measurements. To do so, the IETF PPM working group will standardize protocols for deployment of these techniques on the Internet.

Cloudflare researchers played an integral role in forming the IETF PPM working group. Cloudflare continues to help guide the standardization process by co-editing documents, guiding the agenda in meetings, and developing or helping to maintain open-source implementations.

The direction of the work of the IETF's PPM working group is determined primarily by use cases brought to the working group. So far, this includes applications like web browser telemetry, web analytics, telemetry for COVID-19 exposure notification platforms,¹⁴ and various machine learning tasks. Any organization, be it a private company or a public institution, is free to bring use cases to the group and participate in the protocol design process.

¹² See Apple & Google, Exposure Notification Privacy-Preserving Analytics (ENPA) White Paper, Apr. 2021, available at https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ENPA_White_Paper.pdf.

¹³ See Privacy Preserving Measurement (PPM), IETF, at <https://datatracker.ietf.org/wg/ppm/about/>.

¹⁴ See Apple & Google, supra note 12.

Like many IETF working groups, the IETF PPM working group is made possible by engagement with the academic community. Through journal articles, conference papers, and other artifacts, researchers provide critical insights into the (in)security of proposed protocols and, especially in the case of PPM, devise solutions for specific problems. We have seen a number of papers in recent years proposing multi-party computation (MPC) schemes for specific (classes of) secure data aggregation tasks. These works draw from and build on a wide variety of sub-disciplines of cryptography, including zero-knowledge proof systems, (partially) homomorphic encryption, oblivious transfer, and others.

In many of these systems,¹⁵ the computation of the aggregate is distributed amongst multiple servers as follows: Clients “split” their measurements into multiple “secret shares” and send one share to each server. Once all the shares have been uploaded, the servers use them to reconstruct the desired aggregate result. Splitting the measurements into secret shares ensures they are kept secret: As long as at least one server executes the protocol honestly, no individual’s measurement is ever observed in the clear. At the same time, many of these systems are designed to ensure that the result is always computed correctly, even in the presence of malicious (or merely misconfigured) clients that would otherwise provide invalid inputs to the computation.

This paradigm improves significantly upon the status quo. While a number of solutions exist for improving privacy of the measurement aggregation process, none other than MPC allows the aggregate result to be computed both precisely and without revealing measurements to the data collector.¹⁶ MPC also provides an opportunity for a third-party audit of the collection process, thereby ensuring much stronger security performance.

The “cost” of stronger security is additional operational complexity, since the reliability of the data collection system now depends on multiple servers instead of just one. Addressing this complexity is one of the principal tasks of the IETF PPM working group. The DAP standard mentioned above is the first draft taken up by the IETF PPM working group.¹⁷ The objective of the DAP specification is to spell out precisely how to execute MPC protocols for secure aggregation over HTTPS. The specification targets a particular class of constructions, called

¹⁵ See, e.g., Ulfar Erlingsson, Vasyl Pihur, & Alesandra Korolova, RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response, Aug. 25, 2014, available at <https://arxiv.org/pdf/1407.6981.pdf>; Henry Corrigan-Gibbs & Dan Boneh, Prio: Private, Robust, and Scalable Computation of Aggregate Statistics, Proceedings of the 14th USENIX Symposium on Networked Systems Design and Implementation (NSDI '17), Mar. 27-29, 2017, Boston, MA, available at <https://www.usenix.org/system/files/conference/nsdi17/nsdi17-corrigan-gibbs.pdf>; Dan Boneh, Elette Hoyle, et al., Lightweight Techniques for Private Heavy Hitters, 2021 IEEE Symposium on Security and Privacy (SP), 2021, pp. 762-776, available at <https://eprint.iacr.org/2021/017.pdf>; and Joseph J. Pfeiffer, Denis Charles, Davis Gilton, et al., Masked LARK: Masked Learning, Aggregation and Reporting workflow, Oct. 27, 2021, available at <https://arxiv.org/pdf/2110.14794.pdf>.

¹⁶ Differentially private systems such as RAPPOR [EPK14] in which clients randomize their uploaded measurements, provide some degree of privacy protection, but the aggregate result is not computed precisely.

¹⁷ See Tim Geoghegan, Christopher Patton, Eric Rescorla, & Christopher A. Wood, Distributed Aggregation Protocol for Privacy Preserving Measurement, IETF, May 5, 2022, at <https://datatracker.ietf.org/doc/draft-ietf-ppm-dap/>.

Verifiable Distributed Aggregation Functions (VDAFs).¹⁸ A specification of VDAFs, including a few concrete instantiations from the literature, is under development in the CFRG¹⁹ research group of the IRTF.²⁰ The goal of this work is to provide cryptography researchers with an explicit set of design criteria for developing new constructions. In turn, the DAP specification provides a clear roadmap to deployment of these constructions.

Our hope is that the work developed in the IETF PPM working group will enable organizations of all sizes to adopt a private-by-default posture for their data collection and aggregation processes. Today the primary challenge for adopting something like DAP for a given use case is that there may not be a VDAF that is suitable for that use case. Thus, one way OSTP can support this work is to describe to the research community the government's current set of use cases:

- Which agencies, state or federal, benefit from collecting information about visitors to their websites or other Internet properties? How is this information used? Which statistics or other aggregate artifacts are most useful?
- Are there situations in which information is not collected because it is deemed too sensitive to collect?

Research Opportunities, Risks, and Limitations of PETs (RFI topics #2 and #7)

The RFI's description of privacy-enhancing technologies presenting numerous likely benefits stemming from secure and privacy-preserving data analysis techniques was spot-on. As a technology company with a substantial commitment to furthering research that helps build a better Internet, Cloudflare looks forward to increased collaboration and innovation across industry and academia as a result of the deployment of PETs.

However, Cloudflare also recognizes that deploying PETs is an exercise not without risk. Deploying PETs means less surface area for surveillance by third parties, including governments. Cloudflare has faced resistance in deploying PETs like DOH that limit access to data from industry players that seek to use that data, as well as from governments—including European governments—that are concerned that it will limit the availability of web browsing data. Some governments have shown active opposition to PETs by blocking Internet access or traffic.²¹ Such reactions to PETs make the Internet less open, less secure, and less reliable; they

¹⁸ See Richard Barnes, Christopher Patton, & Phillip Schoppmann, Verifiable Distributed Aggregation Functions, IETF, May 26, 2022, at <https://datatracker.ietf.org/doc/draft-irtf-cfrg-vdaf/>.

¹⁹ See Internet Research Task Force, Crypto Forum Research Group, at <https://irtf.org/cfrg>.

²⁰ See Internet Research Task Force, <https://irtf.org/>.

²¹ Examples abound. China, Russia, and South Korea have all been reported to block websites that deploy ESNI, a type of PET that encrypts Internet metadata. See Kevin Bock, David Fifield, Amir Houmansadr, Dave Levin, et al., Exposing and Circumventing China's Censorship of ESNI, *Censorship.ai*, Aug. 7, 2020, at <https://geneva.cs.umd.edu/posts/china-censors-esni/esni/> (China); Nguyen Phong Hoang, Michalis Polychronakis, & Phillipa Gill, Measuring the Accessibility of Domain

could also have the unintended effect of splintering the global Internet.²² A national strategy that seeks to encourage the deployment of PETs has to integrate international efforts, building consensus around the need to deploy new protocols and approaches. OSTP should also, as part of its national strategy, explicitly enlist the U.S. CTO Team to encourage the U.S. government to leverage those PETs that will help the U.S. government effectively deliver services in a privacy-preserving manner.

Another perceived risk in some quarters of the use of PETs is that they might make systems less reliable. OSTP could help mitigate this perceived risk by supporting research that examines the reliability of PETs, and in working with industry and academic partners to further education on the appropriate use of PETs.

Privacy-enhancing technologies show great promise, but they also have limitations. One limitation of some PETs is that they (intentionally) make it harder for the data collector to “explore” the data. In this case, technologies like OHTTP²³ (the generalization of ODoH) may be worthy of consideration. More generally, it is likely that different use cases will call for the deployment of different PETs. One size will not fit all.

Regulation of PETs (RFI topics #4 and #5)

“Privacy-enhancing technologies” is a new name for a set of tools, techniques, and practices that have been a focus of research and development for many years. At Cloudflare, we take a proactive approach towards privacy, and we believe privacy is not only about responding to different regulations. Privacy is about building technology that helps customers do a better job protecting their users. It is about minimizing the exposure of sensitive data and protecting user privacy and personal information. It is about helping to build a better Internet. In the regulatory arena, Cloudflare has actively supported efforts to develop a framework for U.S. federal privacy standards, urging policymakers to adopt technology-neutral approaches that allow standards to change and improve as technology does. In the context of privacy-enhancing technologies, we encourage the same approach.

In its work across federal agencies and departments, OSTP plays an important role in ensuring that the executive branch understands the effects of regulation, funding, and use of privacy-enhancing technologies. Any regulation of this area should strive to be responsive to the

Name Encryption and Its Impact on Internet Filtering, forthcoming Passive and Active Measurement Conference 2022, Feb. 1, 2022, available at <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/d1cc0cc9b507a8d9ef71c90c35caa9919b95a077.pdf> (Russia); and Caleb Chen, South Korea Expands Internet Censorship to HTTPS With First Countrywide Use of SNI, PIABLOG, Feb. 19, 2019, at filtering <https://www.privateinternetaccess.com/blog/south-korea-expands-internet-censorship-to-https-with-first-countrywide-use-of-sni-filtering/> (South Korea).

²² See A Declaration for the Future of the Internet, The White House, Apr. 28, 2002, available at https://www.whitehouse.gov/wp-content/uploads/2022/04/Declaration-for-the-Future-for-the-Internet_Launch-Event-Signing-Version_FINAL.pdf.

²³ See Martin Thomson & Christopher A. Wood, Oblivious HTTP, IETF, Feb. 15, 2022, at <https://datatracker.ietf.org/doc/draft-ietf-ohai-ohttp/>.

dynamic nature of the technology sector, and aim for non-prescriptive and flexible guidance. Any regulation would need to be compatible with other privacy-protecting regulations around the globe, including GDPR, CCPA, and other industry-specific privacy laws in the United States, such as HIPAA. Many of these regulations predate the development of some PETs, and as in any dynamic area of research and development, it is certain that the technology will outpace any regulation.

OSTP, especially its CTO Team, could contribute to the adoption of PETs by incentivizing federal agencies and departments to consider and deploy PETs when reasonable and appropriate to do so. Research into incentive structures or safe harbors for the private sector (including critical infrastructure) around the use of PETs could be informative for regulators worldwide.

Lastly, we believe that strong encryption is key to privacy and security on the Internet, and to the successful development of PETs. OSTP's continued advocacy for strong encryption (and opposition to efforts to weaken encryption), including its use by publicly funded organizations, will go far in improving the privacy of sensitive data.

Recommendations for Advancing and Supporting PETs (RFI topics #1 and #6)

Throughout this response, we have made a number of recommendations for ways that OSTP could advance and support privacy-enhancing technologies. In this section we consolidate them into a single list:

- Across federal agencies and departments, push forward standards-based approaches towards privacy-preserving data sharing and analytics (perhaps, for example, in cooperation with NIST and its Privacy Framework²⁴);
- Conduct a baseline assessment of federally funded research on PETs to date and make that baseline assessment publicly available;
- Encourage federal funding of research in PETs (including research that examines the reliability of PETs);
- Work to ensure that any regulation of this area is responsive to the dynamic nature of the technology sector, and aim for non-prescriptive and flexible guidance that is compatible with other privacy-protecting regulations around the globe;
- As part of its national strategy, explicitly enlist the U.S. CTO Team in working to ensure that the U.S. government leverages those PETs that will help the U.S. government effectively deliver services in a privacy-preserving manner;
- Describe to the research community the government's current set of use cases:

²⁴ See NIST Privacy Framework, at <https://www.nist.gov/privacy-framework>.

- Which agencies, state or federal, benefit from collecting information about visitors to their websites? How is this information used? Which statistics or other aggregate artifacts are most useful?
- Are there situations in which information is not collected because it is deemed too sensitive to collect?
- Integrate international efforts, building consensus around the need to deploy new protocols and approaches to encourage the deployment of PETs;
- Work with industry and academic partners to further education on the appropriate use of PETs; and
- Continue to support the adoption of encryption to protect privacy and oppose efforts to weaken encryption.

Conclusion

Over the next few years, the global tension between the need to keep data as private as possible when using digital services and the perceived need by governments to exert more control over content flowing over the Internet is likely to continue to increase as data becomes increasingly private and encrypted through the development and deployment of privacy-enhancing technologies. Keeping this tension in mind (and supporting research specifically on exploring this tension) will help OSTP formulate a national strategy on privacy-preserving data sharing and analytics that will be more resilient, and resulting policy initiatives that will be productively geared toward responsible use of PETs to benefit individuals and society.

At Cloudflare, we are excited by the promise of privacy-enhancing technologies to protect sensitive information while enabling more innovative and collaborative research. Cloudflare appreciates OSTP's specific and comprehensive questions, and the breadth of expert opinion they will generate for the field. We look forward to continuing to engage with OSTP on privacy-enhancing technologies as this process of developing a national strategy on PETs moves forward.

Sincerely,

/s/ Alissa Starzak

Vice President, Global Head of Public Policy

Attachment: References

References

Selected publications with Cloudflare authors relevant to privacy-enhancing technologies

RFC 9230: Oblivious DNS over HTTPS, Internet Engineering Task Force (IETF). 2022. Eric Kinnear, Patrick McManus, Tommy Pauly, Tanya Verma, Christopher A. Wood.

Might I Get Pwned: A Second Generation Compromised Credential Checking Service, 31th USENIX Security Symposium (USENIX Security 22). Bijeeta Pal, Mazharul Islam, Marina Sanusi, Nick Sullivan, Luke Valenta, Tara Whalen, Christopher A. Wood, Thomas Ristenpart, Rahul Chattejee.

Oblivious DNS over HTTPS (ODOH): A Practical Privacy Enhancement to DNS, Proceedings on Privacy Enhancing Technologies 2021, Volume 4, pp. 575–592. 2021. Sudheesh Singanamalla, Pop Chunhapanaya, Jonathan Hoyland, Marek Vavruša, Tanya Verma, Peter Wu, Marwan Fayed, Kurtis Heimerl, Nick Sullivan, Christopher A. Wood.

Privacy Pass: Bypassing Internet Challenges Anonymously, Proceedings on Privacy Enhancing Technologies, no. 3 (2018), pp. 164-180. 2018. Alex Davidson, Ian Goldberg, Nick Sullivan, George Tankersley, Filippo Valsorda.

Selected Cloudflare blog posts relevant to privacy-enhancing technologies

“Empowering Your Privacy,” Emily Hancock, Chief Privacy Officer, 28 January 2020, <https://blog.cloudflare.com/empowering-your-privacy/>.

“Supporting the latest version of the Privacy Pass Protocol,” Alex Davidson, 28 October 2019, <http://blog.cloudflare.com/supporting-the-latest-version-of-the-privacy-pass-protocol/>.

“Certifying our Commitment to Your Right to Information Privacy,” Emily Hancock and Rory Malone, 29 July 2021, <http://blog.cloudflare.com/certifying-our-commitment-to-your-right-to-information-privacy/>.

“Improving DNS Privacy with Oblivious DoH in 1.1.1.1,” Tanya Verma and Sudheesh Singanamalla, 8 December 2020, <http://blog.cloudflare.com/oblivious-dns/>.

“Cloudflare supports Privacy Pass,” Nick Sullivan, 9 November 2017, <http://blog.cloudflare.com/cloudflare-supports-privacy-pass/>.

“Cloudflare Zaraz launches new privacy features in response to French CNIL standards,” Yair Dovrat and Yo'av Moshe, 15 June 2022, <https://blog.cloudflare.com/zaraz-privacy-features-in-response-to-cnil/>.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Computer & Communications Industry Association (CCIA)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Before the
Office and Science and Technology Policy
Washington, D.C.

In re

Request for Information on Advancing
Privacy-Enhancing Technologies

Document Number 2022-12432

COMMENTS OF
THE COMPUTER & COMMUNICATIONS INDUSTRY ASSOCIATION (CCIA)

In response to the Request for Information (“RFI”) published in the Federal Register at 87 Fed. Reg. 35250 (June 9, 2022), the Computer & Communications Industry Association (“CCIA”)¹ submits the following comments to the Office of Science and Technology Policy (“OSTP”):

I. Introduction

CCIA is pleased to provide comment on how the federal government can not only facilitate but encourage implementation of Privacy-Enhancing Technologies (“PETs”) for employing and transferring data, particularly the data of consumers and Internet users.

Privacy is best secured through anonymization, a goal that might never fully be reached, or, if reached, would strip data of all utility. PETs, most notably differential privacy, are the best alternative to anonymization. Rather than attempt data de-identification, which is increasingly difficult and susceptible to reverse-engineering, PETs enable data minimization. But, as the RFI notes, PETs are not yet widely adopted.

Though the United States presently lacks a federal privacy statute of general application, many of our federal agencies presently are authorized to adopt consumer-protection measures focused on data privacy and security. In addition, within any federal privacy legislation that it takes up in the future, Congress could consider creating safe harbors, and possibly exemptions,

¹ CCIA is a nonpartisan, not-for-profit trade association. For 50 years, CCIA has promoted open markets, open systems, and open networks. CCIA members employ more than 1.6 million workers, invest more than \$100 billion in research and development, and contribute trillions of dollars in productivity to the global economy. A complete list of CCIA members is available at <http://www.cciagnet.org/members>.

when PETs are in use. The federal government might be best suited to creating incentives that will make PETs commonplace tools for protecting consumer privacy.

II. How PETs Maximize the Benefits of Data While Minimizing Risk

The central question in privacy analysis is how much specificity of information the data user is willing to forego versus how much security the data owner one wants to preserve. Historically the risk-reward tradeoff in data privacy could be plotted as a straight line between no access/zero risk and full access/unlimited risk. Where a particular data policy fell between those points was a function of operations that were largely manual – truncating Social Security numbers and data deletion, for instance. Experience has shown that such measures are neither scalable nor reliable.

Previously the strategy for securing data was anonymization through “injecting noises into the original data,” like blurring a photograph, which obscures much of the original data in order to render the data set de-identified.² But, like a blurred photograph, “the more protection there is, the less useful the released data are.”³ PETs have shifted the paradigm of choosing between data utility and data security. They have significantly lowered the baseline of risk associated with handing over one’s data and made the incremental increases of risk much smaller. PETs have introduced machine learning (artificial intelligence) into data security, which “has the power to reveal information that would not be obvious to a human evaluating a dataset unassisted.”⁴

A. Forms of PET.

Forms of PET include differential privacy and federated learning. Differential privacy, or “DP”, defines privacy as a matter of mathematical probability rather than an absolutist question whether data does or does not correspond to a particular individual. DP can be described as “a privacy-protecting layer between raw data and a user of the data.”⁵ It thus provides responses to queries without “enough information to identify any individual included in the dataset.”⁶ DP is

² Fang Liu, Ph.D., A Statistical Overview on Data Privacy, 34 Notre Dame J.L. Ethics & Pub. Pol’y 477, 478 (2020).

³ Fang Liu, *supra* n.2, at 478.

⁴ Andrea Scipa Els, Artificial Intelligence as a Digital Privacy Protection, 31 Harvard J.L. & Tech. 217, 218 (2017).

⁵ Anna Myers & Grant Nelson, Differential Privacy: Raising the Bar, 1 Geo. L. Tech. Rev. 135, 137 (2016).

⁶ Andrea Scipa Els, *supra* n.4, at 220.

considered “state of the art” in privacy protection, because it cannot be reverse-engineered – DP does not reveal data, but rather it reveals the result of machine-learning decisions about data.⁷

Federated learning is a process that uses AI to import small sets of data from users into the cloud, then immediately aggregate and average those sets with other users’ sets in a centralized learning model. No one user’s data is stored uniquely, but rather it uses a “collaborative learning method” to “train” an AI model on answering questions via supplied data.⁸ These “meta-updates,” so-called,⁹ are the data sets against which a specific algorithm is applied to answer one particular question, which further ensures that identifiable consumer data is not exposed.

Another type of PET is multi-party computational privacy (“MPC”). This method permits users to share query outcomes without sharing the individual data inputs. Its first known usage was in 2008, when an auction was designed in Denmark to find the market-clearing price for sugar beets.¹⁰ Beet farmers wanted their individual pricing structures to remain confidential, but all agreed that the market-clearing price should be publicly available. MPC enables this analytical scenario by using encryption to essentially create an electronic bailiff for data inputs,¹¹ allowing third parties to access only query outcomes.

B. Existing Applications of PETs.

Apple and Google already use DP in their respective mobile operating systems. Apple built DP into iOS 10 for all data collection and uses it for improving pre-installed applications like Notes and the keyboard.¹² Google, credited with developing federated learning, uses it for word recommendations in the Android keyboard and possibly soon for ranking photos and modeling language.¹³ In addition, IBM and Uber have released open-sourced libraries for

⁷ See Fang Liu, *supra* n.2, at 482; see also Andrea Scrips Els, *supra* n.4, at 221.

⁸ Micah J. Sheller, *et al.*, Federal learning in medicine: facilitating multi-institutional collaborations without sharing patient data, *Scientific Reports* 10:12598, at p.2 (2020).

⁹ Andrea Scrips Els, *supra* n.4, at 223.

¹⁰ Multi-Party Computation: Private Inputs, Public Outputs, *Forbes.com* (Oct. 26, 2021), <https://www.forbes.com/sites/forbestechcouncil/2021/10/26/multi-party-computation-private-inputs-public-outputs/?sh=403c70521bb0>.

¹¹ See Multiparty computation as supplementary measure and potential data anonymization tool, IAP (Oct. 27, 2021), <https://iapp.org/news/a/multiparty-computation-as-supplementary-measure-and-potential-data-anonymization-tool/>.

¹² Andrea Scrips Els, *supra* n.4, at 221; Fang Lui, *supra* n.2, at 486.

¹³ Andrea Scrips Els, *supra* n.4, at 223; Fang Lui, *supra* n.2, at 486.

experimenting with various DP applications.¹⁴

The U.S. Census Bureau adopted DP in 2018 to protect publications of data and statistics developed from the 2020 Census.¹⁵ Called the Disclosure Avoidance System (“DAS”),¹⁶ this technology relies on algorithms borrowed from 1965 Voting Rights Act enforcement – more specifically, drawing legislative districts – to “efficiently distribute the noise injected by differential privacy.”¹⁷ The Bureau published a [handbook](#) describing DAS and outlining the increased risks of data disclosure that caused it to adopt DP technology for the first time.

III. How the U.S. Government Can Foster PET Adoption and Innovation

Although PETs are substantially automated through reliance on AI, they nonetheless entail great cost.¹⁸ DP requires several processes to establish a privacy layer that is exactly porous enough to enable the user to answer questions while preserving the pristine raw data. Federated learning requires creation of theoretically infinite meta-updates, one for each question that can ever be asked, constructed from the point of collection. Presently, due largely to the absence of a uniform and final set of privacy regulation, there is no external incentive for developing and deploying these technologies. There are neither guideposts nor rewards for privacy technology. CCIA urges OSTP to establish, or recommend the establishment of, PET incentives.

A. PETs Can Be Expressly Incorporated Into Existing Federal Regulations.

Many federal agencies already have regulatory vehicles for encouraging the development and use of PETs. The Federal Trade Commission can review PETs in the context of its Section 5 authority to protect consumer privacy. In addition, and as the RFI notes at Paragraph 4, adding PETs as a criterion for selection under the Federal Acquisition Regulations could be a meaningful incentive for deployment. Additional regulatory applications can include making PETs a required component of data-breach notifications: the disclosing entity could be required

¹⁴ Fang Lui, *supra* n.2, at 486.

¹⁵ John M. Abowd, The U.S. Census Bureau Adopts Differential Privacy, Ass’n for Computing Machinery (July 19, 2018), <https://dl.acm.org/doi/10.1145/3219819.3226070>.

¹⁶ U.S. Census Bur., Processing the Count: Disclosure Avoidance Modernization, <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html>.

¹⁷ John M. Abowd, *supra* n.15.

¹⁸ Jonathan Mayer and Arvind Narayanan describe several impediments to adopting what they term “privacy substitutes” in their paper Privacy Substitutes, 66 Stan. L. Rev. Online 89 (2013-14).

to state whether any PETs were in use for the data believed to have been compromised, and if so, how those technologies might mitigate the risk that the unlawfully obtained data could harm any consumers.¹⁹

The FTC's rulemaking authority seems a particularly appropriate tool for incentivizing PETs at this time. American industry needs guidance on how PETs can be incorporated into their various business models. The FTC has expertise in collecting and harmonizing technological solutions for preserving privacy. An FTC proceeding devoted to assessing the feasibility and efficacy of DP, federated learning, and other forms of PET would give businesses the confidence to choose and implement privacy technologies that will benefit consumers.

A good deal of scholarship has arisen on how PETs can be used to comply with the Family Educational Rights and Privacy Act of 1974 (FERPA)²⁰ and the privacy laws that govern healthcare information.²¹ Education and health care both rely on personal data, within the context of larger sets of data, to make decisions that deeply affect an individual's life. And FERPA in particular protects not only PII (personally identifiable data) but also statistics built from PII.²² DP is well suited to enabling access to large data sets through the protective layer of targeted machine learning. Thus, the Department of Education and the Department of Health and Human Services might consider releasing guidance on implementing PETs as a means for privacy compliance within their respective data regimes.

B. Federal Agencies Can Advise Congress on Including PET Safe Harbors and Exemptions in Privacy Legislation.

In recent weeks, Congress has displayed a renewed interest in federal privacy legislation. This past June, H.R. 8152, the American Data Privacy and Protection Act, was given a hearing and a markup in the House Consumer Protection Subcommittee, and Senator Cantwell has

¹⁹ Examples of data-breach notification rules include the Consumer Proprietary Network Information (CPNI) rule for customer notifications, 47 C.F.R. § 64.2011, the forthcoming Cyber Incident Reporting for Critical Infrastructure Act (CIRCA) rules, and the Health Insurance Portability and Accountability Act Breach (HIPAA) Notification Rule, 45 C.F.R. §§ 164.400-.414.

²⁰ *E.g.*, Kobbi Nissim, *et al.*, Bridging the Gap Between Computer Science and Legal Approaches to Privacy, 31 Harvard J.L. & Tech. 687 (2018).

²¹ *E.g.*, Micah J. Sheller, *et al.*, *supra* n.8; *see also* Mirko Forti, The Deployment of Artificial Intelligence Tools in the Health Sector: Privacy Concerns and Regulatory Answers within the GDPR, 13 Eur. J. Legal Stud. 29 (2021). The need for technologically assured patient privacy is a chief inhibitor of moving toward healthcare information portability. W. Nicholson Price II, Ph.D., Risk and Resilience in Health Data Infrastructure, 16 Colo. Tech. L.J. 65, 71-73 (2017).

²² Kobbi Nissim, *et al.*, *supra* n.15, at 722.

introduced a slightly different privacy bill in S.3195, the Consumer Online Privacy Rights Act. It is expected that privacy will receive even greater attention in the next Congress. This activity provides a significant opportunity for encouraging PET innovation and adoption.

Expert agencies like OSTP can advise Congress on the technologies available for collecting and using consumer data in a responsible manner. More than that, they can work with Congress to establish statutory incentives, like safe harbors and exemptions, centered on PETs. A federal privacy statute can, in a technologically neutral way, encourage the use of reliable automation to protect individual data. In this way, federal privacy legislation would become a means of, as one professor put it, “recoding privacy law” to become a solution and not merely a punitive tool.²³

IV. Conclusion

CCIA applauds OSTP’s constructive approach to securing consumer data through technological innovation. Encouraging the adoption of PETs through existing federal regulation would be a meaningful way to preserve privacy and data security as we await the enactment of a federal legislative solution spurring adoption through statutory incentives.

Respectfully submitted,

Stephanie A. Joyce
Chief of Staff and Senior Vice President
Computer & Communications Industry Association
25 Massachusetts Avenue NW, Suite 300C
Washington, DC 20001

July 8, 2022

²³ See generally Urs Glasser, Recoding Privacy Law: Reflections on the Future Relationship Among Law, Technology, and Privacy, 130 Harvard L. Rev. F. 61 (2016).

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Computing Research Association (CRA) Computing Community Consortium (CCC)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



CCC

Computing Community Consortium
Catalyst

Computing Community Consortium's Response to RFI "[Request for Information on Advancing Privacy-Enhancing Technologies](#)"

July 8, 2022

Written by: Brian LaMacchia

This response is from the Computing Research Association (CRA)'s Computing Community Consortium (CCC). CRA is an association of nearly 250 North American computing research departments - academic, industrial, and professional societies. The mission of the CCC is to enable the pursuit of innovative, high-impact computing research that aligns with pressing national and global challenges. The CCC brings together a diverse set of individuals representing the broad community to lead initiatives and activities, such as this response.

Developing a national strategy around privacy enhancing technologies (PETs), and how such technologies may be harnessed to develop privacy-preserving data sharing and analytics, is a critical need for our nation. Every day both public and private sector organizations create, use, and store exceptional amounts of data and use it in exponentially new ways, but these uses are restricted to specific intra-organization silos or governed by one-off data-sharing agreements to which users have given explicit consent. PETs hold the promise to permit sharing, aggregation, and analysis of relevant data across organizations while maintaining the privacy of its subjects.

In this RFI response, we focus on two of the questions presented in RFI: Question #1 (research opportunities to advance PETs), and Question #5 (specific laws that could be used, modified, or introduced to advance PETs).

With respect to Question #1, we see two broad areas where an increased research focus would help the development and deployment of PETs and specifically advance the goal of furthering privacy-preserving data sharing and analytics. At present, there are not any widely accepted and adopted standards for sharing PET-protected information, even within a category of PET technology (e.g., secure multi-party computation, fully homomorphic encryption, etc.). If the Federal Government wants to use PETs to ensure the confidentiality of telemetry it desires to collect, there needs to be well-defined standards for how specific PETs are applied to various classes of telemetry and how PET implementations interoperate. Even though data is theoretically accessible (in the cloud) and shareable, it is difficult to preserve privacy of that data because PET technologies, data formats, encryption and signature techniques, etc. are not able to work together. A series of interoperability standards for PETs as applied to common data

categories of interest would help tremendously both for data use and the creation of new technologies.

A related problem which standards may help address is users' comfort and acceptance of sharing of their data through privacy-preserving mechanisms with the Government. While the proliferation of private and public clouds makes it possible to share collected data with Government entities, individual users still have to decide to opt-in to such sharing. For example, in the case of desired collection of citizen data, citizens need to (a) see the value to society of the data collection, (b) have confidence that data associated with them personally that is shared through a properly working PET cannot be re-identified by the Government, and (c) have further confidence in the security and privacy of the PET's security protocols.

The second broad area that needs increased research and development is that of privacy-preserving credentials. Just as users have expectations about the privacy of data related to them, they also have privacy expectations around the secure credentials that they use to authenticate to various cloud-based services. Privacy-preserving credentials have been an area of active research for well over two decades, yet they still have not seen widespread adoption and use. Now, however, we may have an opportunity to push research and adoption of privacy-preserving credentials enhanced with minimal disclosure technologies; that is, capabilities that allow users to selectively reveal aspects of their identity that they choose and control. Some research in these areas has already begun, particularly building on digital vaccine credentials (e.g., SMART Health Cards) that saw widespread adoption and availability by States, pharmacies, and medical providers during COVID-19 pandemic, but we have only scratched the surface of what is possible in this realm.

We turn now to Question #5, which asks for input regarding specific laws that could be used, modified, or introduced to advance PETs. Here we would note that many PETs are implemented using encryption technologies that are currently subject to regulation under the Export Administration Regulations (EAR) enforced by the Department of Commerce's Bureau of Industry and Security (BIS). This is true both for PETs built on classical public-key and symmetric key technologies as well as those built on quantum-resistant public-key schemes. For example, fully homomorphic encryption (FHE) schemes are often based on shortest vector or closest vector problems associated with lattices, and schemes based on these problems are now explicitly regulated under the EAR (see Title 15 Part 774, 5A002.a, and Technical Note 2.c.1 in the N.B. for 5A002.a.4).

Using PETs to enable privacy-preserving data sharing and analytics is going to necessitate using the encryption schemes that underlie many PETs and applying those schemes to data gathered both within the United States and around the world. A national strategy for privacy-preserving data sharing and analytics should examine the application of the existing EAR related to cryptography to encryption used within PETs, as the existence of export regulations that encompass PETs will hinder their commercial development and deployment. We would suggest that as part of the development of a national strategy that BIS consider whether the EAR should be amended to make the use of encryption within a PET scheme an uncontrolled use.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Confidential Computing Consortium

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



CONFIDENTIAL COMPUTING CONSORTIUM

July 8, 2022

To Whom It May Concern:

Please consider the following submission to the Request for Information on Advancing Privacy-Enhancing Technologies from the Confidential Computing Consortium. The Confidential Computing Consortium (<https://confidentialcomputing.io>) is a Linux Foundation project “to accelerate the adoption of Trusted Execution Environment (TEE) technologies and standards” and has a diverse membership of hardware and software vendors and cloud service providers (<https://confidentialcomputing.io/members/>). This response was prepared by the group’s Technical Advisory Council with participation from across the membership, and ratified by its Governing Board. The Linux Foundation is a non-profit organization registered in the United States as a 501(c)(6).

The Confidential Computing Consortium has a mandate to engage with governments, standards agencies and regulatory agencies to encourage adoption of Confidential Computing, as well as work with the larger ecosystem and engage with existing and potential end-users of the technologies. It also works with open source projects to further development of implementations. The Confidential Computing Consortium is committed to encouraging open source implementations of Confidential Computing technologies to ensure wide-spread adoption, scalable community involvement, transparency of process, increased security and ease of auditing by relevant interested parties and authorities.

The Confidential Computing Consortium welcomes collaboration with governmental and non-governmental organizations and has mechanisms in place to provide appropriate membership, as well as open technical participation without any membership requirement.

Sincerely,

Stephen R. Walli
Confidential Computing Consortium, Governing Board Chair

Office of Science and Technology Policy
[Request for Information on Advancing Privacy-Enhancing Technologies](#)

Response by the Confidential Computing Consortium on Confidential Computing and hardware-based Trusted Execution Environments

Confidential Computing is the protection of data in use by performing computation in a hardware-based Trusted Execution Environment. A technical description of Confidential Computing and the use of Trusted Execution Environments is available in the Consortium's white paper "*A Technical Analysis of Confidential Computing*": <https://confidentialcomputing.io/wp-content/uploads/sites/85/2022/01/CCC-A-Technical-Analysis-of-Confidential-Computing-v1.2.pdf>

Information Requested: Respondents may provide information for one or as many topics below as they choose. Through this RFI, OSTP seeks information on potential specific actions that would advance the adoption of PETs in a responsible manner, including on the following topics:

1. *Specific research opportunities to advance PETs:* Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.
2. *Specific technical aspects or limitations of PETs:* Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

Confidential Computing is the protection of data in use by performing computation in a hardware-based Trusted Execution Environment (TEE). A comparison with other technologies, including other PETs, is available in Section 4 of "*A Technical Analysis of Confidential Computing*". Key characteristics of TEEs include:

- **Data confidentiality:** Unauthorized entities cannot view data while it is in use within the TEE.
- **Data integrity:** Unauthorized entities cannot add, remove, or alter data while it is in use within the TEE.
- **Code integrity:** Unauthorized entities cannot add, remove, or alter code executing in the TEE.

In addition, implementations may include the following capabilities (also described in "*A Technical Analysis of Confidential Computing*"):

- Code Confidentiality
- Authenticated Launch

- Programmability
- Attestability
- Recoverability

Confidential Computing, by combining these capabilities, can provide two of the three characteristics of the *CIA triad*, Confidentiality and Integrity. It can provide this both for data and code, with the ability to have strong cryptographic assurances derived from attestation measurements and validation. These properties can be used to build applications which are privacy-enhancing, and the deployment of complementary PET software within TEEs may provide extra benefits. There is an existing and growing installed base of computer systems equipped with processors with TEE capabilities, that can support Confidential Computing, including availability in public clouds, both in the United States and globally. The general availability of such systems allows parties with an interest in deploying PETs easy and fast access to their benefits, typically with the ability to support general computation with small performance penalties, rather than having to design and deploy specific algorithms that may operate at significantly slower speeds. The document *A Technical Analysis of Confidential Computing* provides a comparison of some of the characteristics of Confidential Computing and other PETs.

Confidential Computing approaches that support attestation allow cryptographic assurances for every workload or application for each deployment, extending the software supply chain security beyond development and deployment into the runtime environment.

Current research and development in Confidential Computing promises opportunities for TEEs not just in CPUs but in other processors (e.g., GPUs) and the creation of a trust fabric of mutually trusting computational units, with the trust based in cryptographic assurances at runtime.

3. *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:* Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of [Executive Order 14058](#): Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government.

Since most TEEs run on general-purpose computing platforms with their associated performance and programming flexibility, the use cases for Confidential Computing span a wide range of sectors. The Confidential Computing Consortium's white paper "*Confidential Computing: Hardware-Based Trusted Execution for Applications and Data*" (https://confidentialcomputing.io/wp-content/uploads/sites/85/2021/03/confidentialcomputing_outreach_whitepaper-8-5x11-1.pdf) contains a section "Use cases for Confidential Computing?" providing some examples including cryptographic key management, Edge and IoT, blockchain, mobile and personal computing devices, public cloud, Point of Sale (PoS) and payment and Multi-Party Computation. These use cases span government, defense and security applications as well as all areas of private enterprise and non-governmental organizations including financial, healthcare, automotive, manufacturing, energy, and beyond. Given the increased global concerns around privacy protection for consumers (and associated regulatory regimes), and the

increasing focus of the cybersecurity industry on protection of data and intellectual property, there are few industries or sectors that would not benefit from Confidential Computing, alongside other PETs.

4. *Specific regulations or authorities that could be used, modified, or introduced to advance PETs:* Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking authorities under the Office of Management and Budget, the Federal Trade Commission, and financial regulatory bodies, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (e.g., Other Transaction Authority) to procure PETs for development.

There are a number of regulations and industry frameworks that currently lag behind the developments in PETs in general, and Confidential Computing in particular. These include, among others:

- NIST Cybersecurity Framework (<https://www.nist.gov/cyberframework>), which is widely used within the cybersecurity field. The *Protect Function*, one of the core constituents of the framework, is perfectly suited to benefit from the capabilities provided by Confidential Computing. NIST is currently looking to improve the framework: (<https://www.federalregister.gov/documents/2022/02/22/2022-03642/evaluating-and-improving-nist-cybersecurity-resources-the-cybersecurity-framework-and-cybersecurity>)
- IEC 62443

We would like to note that threats and regulations evolve at a more rapid pace than rule-making. The most effective laws and regulations are written from a principles- and risk-based perspective, and they should be able to withstand the test of time despite increasingly sophisticated threats and the emergence of new technologies. As an example, the Gramm-Leach-Bliley Act (GLBA) is two decades old, yet the principles of the law remain relevant. However, if rules are written with prescriptive requirements, they are in danger of becoming outdated quickly. This can put security at odds with compliance when security managers need to choose between staying in compliance or using new techniques that are better for the business.

The following are existing regulations and guidance that the financial sector relies on for data handling. Our recommendation is that any recommendations to update guidance principles are objectives-based and avoid prescriptive guidance.

- [Cyber Incident Reporting for Critical Infrastructure Act](#) (March 2022)
- [Gramm-Leach-Bliley Act \(GLBA\) Title V, Subtitle A: Disclosure of Nonpublic Personal Information](#) (1999)
- [SEC Regulation S-P: Privacy of Consumer Financial Information](#) (November 2000)
- OCC Comptroller's Handbook: Privacy of Consumer Financial Innovation (October 2011)
- [FRB Regulation P: Privacy of Consumer Financial Information](#) (December 2001)
- [FFIEC Sound Practices to Strengthen Operational Resilience](#) (October 2020)
- [FFIEC IT Booklets: Information Security](#) (September 2016)
- [FFIEC Guidance: Authentication and Access to Financial Institution Services and Systems](#) (August 2021)
- [FFIEC Security in a Cloud Computing Environment](#) (April 2020)
- [NIST Special Publication 800-53 \(Rev. 5\) Security and Privacy Controls for Information Systems and Organizations](#) (September 2020)

5. *Specific laws that could be used, modified, or introduced to advance PETs:* Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.
6. *Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:* This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

The Confidential Computing Consortium (<https://confidentialcomputing.io>) is a Linux Foundation project that “brings together hardware vendors, cloud providers, and software developers to accelerate the adoption of Trusted Execution Environment (TEE) technologies and standards”. The effort includes commitments from numerous [member organizations](#) and contributions from several [open source projects](#). The Confidential Computing Consortium has a mandate to engage with governments, standards agencies and regulatory agencies to encourage adoption of Confidential Computing, as well as work with the larger ecosystem and engage with existing and potential end-users of the technologies. It also works with open source projects to further development of implementations.

The Confidential Computing Consortium, as part of the Linux Foundation, is committed to encouraging open source implementations of Confidential Computing technologies to ensure wide-spread adoption, scalable community involvement, transparency of process, increased security and ease of auditing by relevant interested parties and authorities. Opportunities for grant-giving to open source projects (either within the Consortium or outside it) should be considered, to encourage commercial implementation and adoption of Confidential Computing.

The Confidential Computing Consortium welcomes collaboration with governmental and non-governmental organizations and has mechanisms in place to provide appropriate membership, as well as open technical participation without any membership requirement.

We believe that appropriate use of Confidential Computing can lower the security risks associated with deployment of smart devices in consumer and industrial scenarios, where insurance may play a role financially. We would thus invite discussion of how insurance regulations might be used to incent use of Confidential Computing technologies in various sectors, including devices ranging from IoT to Edge to public cloud and on-premises computing.

7. *Risks related to PETs adoption:* Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

8. *Existing best practices that are helpful for PETs adoption:* Information about U.S. policies that are currently helping facilitate adoption as well as best practices that facilitate responsible adoption. This includes existing policies that support adoption, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This also includes information about where and when PETs can be situated within tiered access frameworks for accessing restricted data, ranging from publicly accessible to fully restricted data.

9. *Existing barriers, not covered above, to PETs adoption:* Information about technical, sociotechnical, usability, and socioeconomic barriers that have inhibited wider adoption of PETs, such as a lack of public trust. This includes recommendations on how such barriers could be overcome. Responses that focus on increasing equity for underserved or marginalized groups are especially welcome.

10. *Other information that is relevant to the adoption of PETs:* Information that is relevant to the adoption of PETs that does not fit into any of the topics enumerated above.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Cybernetica

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

CYBERNETICA

Recommendations on Advancing Privacy-Enhancing Technologies Response to RFI

Version: 1.0

July 7th, 2022

5 pages

Contents

1 Introduction.....	3
1.1 Purpose.....	3
1.2 Background.....	3
2 Responses to topics of the RFI.....	4
2.1 Specific research opportunities to advance PETs.....	4
2.2 Specific technical aspects or limitations of PETs.....	4
2.3 Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs.....	4
2.4 Existing barriers, not covered above, to PETs adoption.....	5

1 Introduction

1.1 Purpose

This document is a response to the Request for Information on Advancing Privacy-Enhancing Technologies as published in a notice from A Notice by the Science and Technology Policy Office on 06/09/2022.

1.2 Background

Cybernetica is a research organization with 15 years of practice in developing Privacy-Enhancing Technologies and their applications. Cybernetica has engaged with research programs in United States of America (e.g., DARPA PROCEED, DARPA Brandeis, DARPA SIEVE) and European Union (various European Commission and Estonian programs)¹. Our researchers have published a range of papers on privacy technologies².

Cybernetica has also been successful in procurements on applications of privacy technology applications and training by multinational organizations (e.g., European Space Agency, EUROSTAT, OECD).

- 1 A list of recent projects is available at <https://cyber.ee/research/projects/>
- 2 A list of all papers <https://cyber.ee/research/library>, papers on secure multi-party computation papers up to 2021 available at <https://sharemind.cyber.ee/research/>

2 Responses to topics of the RFI

2.1 Specific research opportunities to advance PETs

We propose the following opportunities.

1. Training programs on (re)designing services and business processes for data minimisation, including with PETs. These should be parts of existing academic programs, but also retraining programs
2. Training programs about the usage patterns and re-identification risks of different PETs (what works in identity protection, cloud services, what works in data publishing, what works in data collaborations).
3. Piloting programs that support small enterprises and startups in prototyping services with PETs. The service developer will have to justify that they are performing service innovation, reducing unneeded data processing and they have a technology partner that is competent in supporting the development of new innovative service models with PETs.
4. Funding academic research into PETs in general, but also in specific domains (healthcare, finance, homeland security, government applications are good initial candidates). This can build on successful programs from DARPA, IARPA, NSF and private foundations.

2.2 Specific technical aspects or limitations of PETs

We would like to recommend the following materials (justification to why the materials are relevant is provided).

1. United Nations Handbook on Privacy-Preserving Computation Techniques. Available from: <https://unstats.un.org/bigdata/task-teams/privacy/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf>
Provides a helpful explanation of the trade-offs of several privacy-preserving techniques.

2.3 Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

We would like to recommend the following materials (justification to why the materials are relevant is provided).

1. THE PROMISE OF EVIDENCE-BASED POLICYMAKING Report of the Commission on Evidence-Based Policymaking. September 2017. Available from: <https://web.archive.org/web/20210207094854/https://www.cep.gov/> (copy attached). Provides examples of using data and privacy technologies in policymaking.
2. Social Impact Applications of Secure Multi-Party Computation. Asemio Available from: <https://www.asemio.com/social-impact-applications-of-secure-multi-party->

[computation/](#)

Descriptions of using secure computing in social impact applications.

2.4 Existing barriers, not covered above, to PETs adoption

We would like to recommend the following materials (justification to why the materials are relevant is provided).

1. A study of barriers to the adoption of secure computing techniques conducted by social scientists, among multiple industries.
 - a) Usable and Efficient Secure Multi-party Computation. Requirements specification based on the interviews. Available from: <http://uaesmc.cyber.ee/workpackages-and-reports/wp1-requirements-gathering/d12-requirements-specification-based-on-the-interviews.html>
This document lists possible preconditions and barriers that need to be addressed for the implementation of SMC, and possible use-cases brought out by the interviewees.
 - b) Usable and Efficient Secure Multi-party Computation. Expert feedback on prototype application. Available from: <http://uaesmc.cyber.ee/workpackages-and-reports/wp1-requirements-gathering/d14-expert-feedback-on-prototype-application.html>
The document further drills down on the barrier perception to adoption of secure computing by interviewees.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Data Freedom Foundation

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

**“If our digital property
can be infinitely reproduced
and instantaneously distributed
all over the planet without cost
without our knowledge
without its leaving our possession
how can we protect it?**

**How are we going to get paid
for the work we do with our minds?**

**And, if we can't get paid,
what will assure the continued creation
and distribution of such work?”**

John Perry Barlow



Alan Rodriguez – Founder Data Freedom Foundation

The Data Freedom Foundation is an open-source non-profit advocacy group and standards body creating and advocating for the adoption and advancement of Consent and Privacy Technologies.

[Summary Video](#)
(2 mins)

[Google Selfish Ledger Commentary](#)
(12 mins)

[Creating Digital Rule of Law](#)
(39 mins - MIT CDOIQ Symposium)



Creating the Economy of Ideas

Who will Own the Third Generation of the Internet?

We are witnessing the emergence of an Economy of Ideas, or the third generation of the internet. Web3—a group of technologies that encompasses Privacy Enhanced Technologies (PETs), blockchain, cryptographic protocols, digital assets, decentralized finance, social platforms, NFTs, and DAOs, combined with smart glasses that merge physical and virtual reality into one inescapable Metaverse – is the third generation of the Internet. Together, these technologies will serve as the basis for new forms of economic and social interaction arising from self-governed communities of trust that allow people to collaborate, create, exchange, and take ownership of their digital identity and assets.

Volatile digital currencies and speculative, risky investments grab most media attention. That's simply the spectacle, the rockets' red glare and bombs bursting in air. It's a distraction from something much more important. What matters most are the flags and the communities and values those flags represent. What matters most are the self-governed and self-funded communities, currencies, marketplaces, ecosystems, standards, and protocols.

Groups of people worldwide are creating digital communities of trust imbued with specific values and goals. They're creating and self-funding their own decentralized financial system. Digital pioneers are creating a parallel digital economy that will co-exist and increasingly interoperate with the physical economy. A digital economy of ideas orchestrated by efficient, transparent, and accountable algorithms powering trustable intermediaries replacing our aging medieval institutions. A digital economy of ideas is coming to life and illuminating the way to an infinite digital frontier.

Few would debate that Web2, powered by massive global data monopolies, took a wrong turn along the way. Neither the public sector nor the private sector has figured out how to grapple with the Pandora's Box of privacy breaches, disinformation, monopolistic practices, and algorithmic biases that have come to define much of the Internet. Meanwhile, authoritarian governments have never had more data with which to surveil, censor, and manipulate their citizens and those of other nations.

We need to have a serious conversation about the role we want technology to play in open societies. In this sense, web3 isn't just a new wave of innovation. It's an opportunity for a reset that allows us to obtain profound new benefits while solving some of the thorniest problems arising out of the technologies of the past.

For the first time in human history, we look upon an infinite digital frontier combined with the technologies for communities of trust to organize, own, create, and self-govern a global Economy of Ideas. What it means to be human in a digital world will change based on what we chose to create. And Web3 technologies provide the foundation to organize, fund, and create anything we can imagine.

The Problem/Opportunity

The world faces a profound problem with data and trust. Data propagates online without essential metadata to document ownership, define terms of use, or record its history. This lack of trust (data quality) is the root cause of most data challenges facing organizations of all sizes.

Our cultural lack of trust in data is also fundamentally incompatible with an open, participatory, and free digital society. This lack of trust is the root cause of most data challenges facing organizations of all sizes.

Smart Data Protocol binds software to data as it moves to give our data self-awareness and intelligence to assert its ownership, to record its history, and to enforce the terms of any contained agreements. The Smart Data Protocol is the missing licensing solution to build practical applications with digital assets.

1. The Smart Data Protocol is the patent-holding missing licensing protocol for highly tailored and transient digital asset reuse equipped with secure access revocation by digital asset owners.
2. The Smart Data Protocol is the direct link between digital asset creators and their enforceable ownership enabling sophisticated licensing monetization.
3. The Smart Data is the Distributed Data Fabric and Digital Rights Management Protocol for interoperable Web3 applications and the future of the Internet.

Data Freedom Foundation has patented the first comprehensive solution to controlling how our data propagates online and is used by others, even after being shared.

Solution/Product

Data Freedom Foundation's patented Smart Data Protocol binds software to data as it moves to give our data self-awareness and intelligence, assert its ownership, record its history, and enforce the terms of any contained agreements. The Smart Data Protocol fundamentally changes the nature of data from the uncontrolled propagation of untrusted bytes to the precisely tailored propagation of high-quality and trusted information.

1. The Smart Data Protocol dramatically improves data quality, data security, data provenance, and data trust amplifying the value generation from all digital and data-driven initiatives.
2. The Smart Data Protocol automates most security, privacy, and regulatory compliance reducing the cost of data transactions by half.
3. The Smart Data Protocol improves data comprehension and democratization boosting digital transformation outcomes.

Smart Data Protocol - Target Industries: Smart Cities, Smart Homes, Smart Farms, Smart Transportation, Smart Manufacturing, Smart Energy, Smart Healthcare, Smart Wearables, Smart Retail, and Smart Supply Chains.

Partnerships:

1. **Global IDs: Our first paying customer/reseller.** Global IDs is the leading Data Governance PaaS used by Chief Data Officers and Data Governance teams in global Banking, Insurance, Pharma, Telco, and Retail industries.
2. **Sertainty:** Patented provider of Data Container Cryptographic Data Security that remotely controls data access.
3. **TodaQ:** Provider of Cryptographic Data Container Provenance that operates in a fully decentralized manner with zero transaction cost. TodaQ provides unforgeable credentials, immutable audit log, and verifiable proof of ID to any asset to assure ownership (tickets, records, agreements, tangible, and intangible assets).
4. **Reprivata:** Provides Universal Data Ownership via Patented Communities of Trust (CoTs) that enforces a legal basis for the ownership of our identities and data grounded in international patent and contract law.

Smart Data — Timeline of Intelligent Technology

The unstoppable progression to “Smart Everything” also known as “Software-Defined Everything”.

‘SMART’ means “Self-Monitoring Analysis and Reporting Technology”. Smart software-defined products are capable of environmental awareness, group intelligence, and can automatically respond to internal and external events. [1 Wikipedia]

A **smart object** enhances the interaction with not only people but also with other smart objects. Also known as **smart connected products** or **smart connected things (SCoT)**, they are products, assets, and other things embedded with processors, sensors, software, and connectivity that allow data to be exchanged between the product and its environment, manufacturer, operator/user, and other products and systems. [2 Wikipedia]

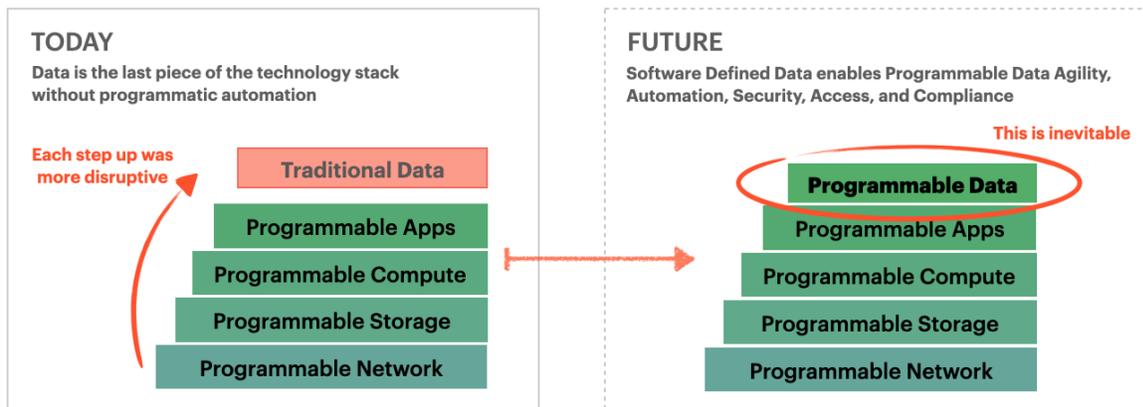
The smart object concept was introduced by Marcelo Kallman and [Daniel Thalmann\[4\]](#) as an object that can describe its own possible interactions. The focus here is to model interactions of smart virtual objects with virtual humans, agents, in virtual worlds.

Software-Defined “Smart” Everything

We’ve witnessed the slow and steady march of software control of complex hardware up the technology stack over the last twenty years from [Software-Defined Networking](#) and [Network Virtualization](#) to [Software-Defined Storage](#) and [Storage Virtualization to Server Virtualization](#) to [Software-Defined Data Centers](#) and [Cloud Computing](#) to [Application Containerization](#) and [Application Virtualization](#).

Surprisingly, data remains the final tier of the technology stack that has not completed this software “smart” transformation. Each step up the technology stack enabled automation which drove exponential cost efficiencies. Each step changed application architectures, allowing new business models, fundamentally altered technology roles, and the structure and operation of technology organizations.

Arguably, each step up the technology stack was more disruptive than the previous steps. We believe Smart Data (software-defined data, programmable data, data virtualization with containerization, or data as code) will be the most disruptive step up the technology stack thus far.



Each step followed [the Law of Diffusion of Innovation](#), with early adopters and innovators taking risks and benefiting immensely from first-mover advantages. We expect software-defined data to be more disruptive than previous steps.

As we look at each step up the technology stack over the last two decades, four repeating themes reveal what we can expect from The Smart Data Protocol:

1. Virtualization & Containerization as Software Abstractions
2. Application Programmable Interface (APIs) & Automation
3. Centralized Control & Distributed Enforcement from Policy Templates
4. Increasingly Self Aware, Adaptable & Environment Responsive

Virtualization & Containerization

Networking, Storage, and Server Virtualization all represent intelligent software and highly redundant hardware that was previously a single point of failure — a single hardware device. These are kinds of software-created network, or storage, or server abstraction.

Application Containers allow code to be packaged and distributed across different cloud platforms. These cloud platforms are software-defined data centers we can rent on demand that uses software-defined networking, storage, and compute as underlying technologies.

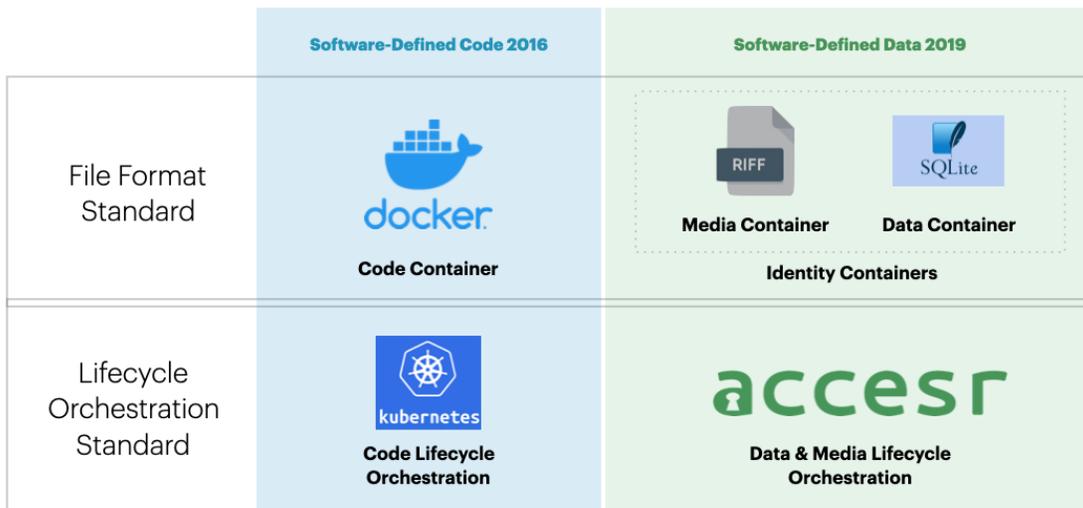
Data Virtualization is an existing enterprise technology that helps data management professionals understand and manage data flow within an enterprise. It enables internal applications to retrieve and manipulate data without requiring technical details about the data, such as how it's formatted at the source or physically located. It can provide a valuable master data source for a single customer view.

Data Virtualization does not currently extend beyond the organization boundary, limiting its ability to manage all organizational data. We propose adding Data Containerization that extends Data Virtualization to the network edge anywhere data moves inside and outside an organization.

SMART DATA = KUBERNETES FOR DATA

Our thinking builds upon Code Containers, called Docker files which ensure application code portability across all cloud platforms. Kubernetes provides Docker Code Container Lifecycle Orchestration across all compatible clouds coordinating creation, deployment pipeline automation, status monitoring, automated reconfigurations, versioning, and eventual deletion.

Data Containers enable secure and trustable distributed data automation across cloud platforms much the same way Application Containers enable safe and trustable distribution and orchestration of application code across cloud platforms.



On a more granular level, Accesr’s Data Container Lifecycle Orchestration provides unprecedented data control for individual data owners, controlling creation, status monitoring, versioning, deactivation for non-compliance, reactivation for compliance, and finally deleted across all compatible applications and platforms.

Application Programmable Interface (APIs)

APIs are interfaces that allow different software programs to communicate and share data according to a set of clearly defined methods of communication. Developers can reuse and scale software architectures by making calls to available APIs of other programs. APIs are electrical sockets that connect programs and are the critical enabler of software automation.

APIs and their automation enabled the programmatic reconfiguration of networks and allowed traffic rules to adjust to changing networking needs. Or automate reconfiguration of server file systems to adapt their storage capacity based on dynamic and changing storage needs. APIs and automation enable entire data centers to be created within software, automated, and dynamically adjusted based on changing user and application needs.

All of this occurs with a fraction of the people it took to manage these systems without automation. Automation creates organizational agility and cost efficiencies while dramatically improving user experiences through enhanced consistency and quality. Data Container lifecycle APIs enable distributed Data Container orchestration and management for all targeted containers at each step in the data lifecycle.

Centralized Control & Distributed Enforcement

We repeatedly see centralized control paired with distributed policy enforcement across many technology layers. Organizations can set and automatically enforce policy constraints. They can change these constraints across potentially thousands of distributed systems by changing a single policy statement.

Increasingly Self Aware & Environment Responsive

Networks, servers, and applications became increasingly self-aware, adaptive, and responsive to their environment while operating within predefined administrator policies with instant recovery from every kind of hardware failure.

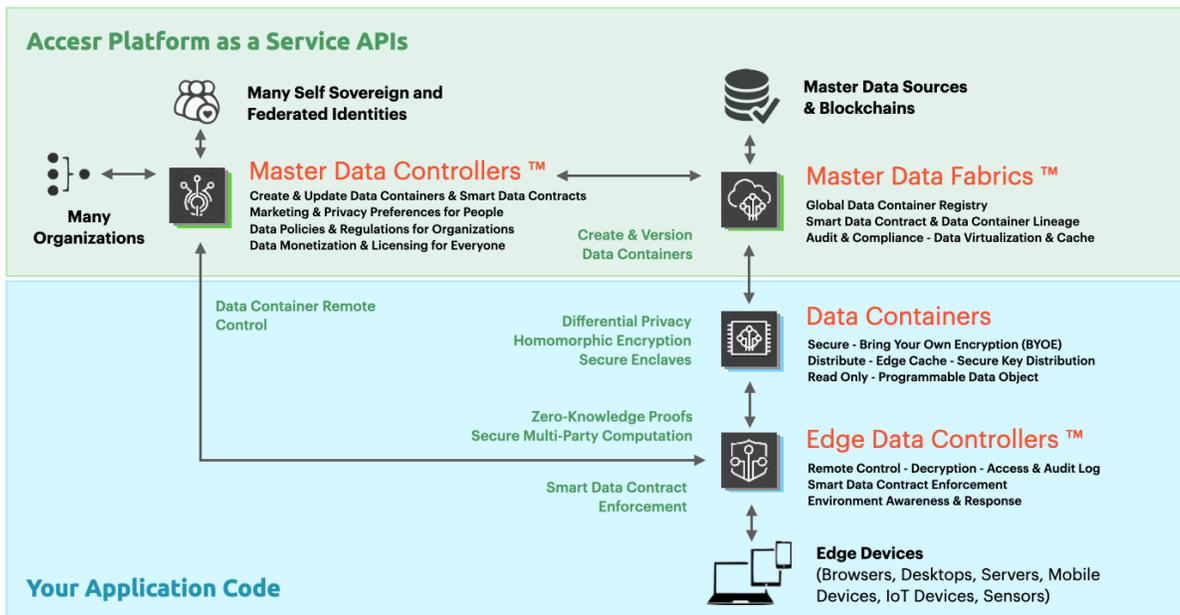
What if our data were self-aware and responsive to its environment? Imagine small open-source programmable data containers, holding only your data, encrypted with keys you control, moving around with Smart Data Contracts you create that constrain or guardrail the uses of your data — your personal data and your organizational data.

Applying these themes to data enables unprecedented data self-awareness; this allows a new era of data agility, regulatory and policy automation, distributed and dynamic data architectures, transparent and accountable data usage, with trustable data security, provenance, and quality

Open Source Architecture for Data & Digital Rights - Converting centralized monopoly platforms into decentralized protocols and standards

We decided we needed two core distributed technologies, and we propose them as standards:

1. **Centralized Control:** [Master Data Controller](#) aggregates all our data and controls its current and future uses from one convenient location.
2. **Distributed Enforcement:** [Programmatic Data Control & Automation](#) secures our data with keys and access rules we control. We can alter all our data and approved uses, as it flows through and is acted upon by external technologies.



Central Control of Master Data

We consistently and repeatedly see the theme of centralized control paired with distributed enforcement across network devices, storage devices, servers, data centers, and application code. Organizations can set and automatically enforce storage, compute, server, and application code policy constraints. They can change these constraints across all distributed systems by changing a single policy statement from a central location.

We imagine an unlimited number of Master Data Controllers; each focused on the data transactions and data contracting needs of their respective communities, groups, or ecosystems. Master Data Controllers are where people, organizations, and groups establish a “home” for their individual or group data. They also control the uses of that data by others.

The [Solid Protocol](#) proposes a decentralized standard for establishing centralized control of our data within a distributed architecture. The Solid Protocol is the mid-course adjustment for the Web by its inventor, [Sir Tim Berners-Lee](#). It realizes Tim’s original vision for the Web as a medium for the secure, decentralized exchange of public and private data.

A Solid Server hosts one or more Solid Pods. Pods are where you store your data:

1. Each Pod is fully controlled by the Pod owner (i.e., you).
2. Each Pod’s data and access rules are fully distinct from those of other Pods.
3. You can get a Pod from a [Pod Provider](#), or you may choose to [self-host](#) your Pod.

You can have multiple Pods. They can be hosted by the same Pod Provider or by different Providers or be self-hosted or any combination thereof. The number of Pods you have and which Solid Server or Servers you use are effectively transparent to the applications and services you use. In the Solid ecosystem, data is linked through your [Identity](#) and not through the specifics of your Pod.

The Master Data Controller for Organizations provides one location to control all organizational Master Data, and by extension, all downstream data flows. It does this by managing all organizational data policies, external regulations that apply to internal data, licenses from external data providers, and user-provided personalization preferences. It encodes these data agreements into Smart Data Contracts and automates their enforcement inside and outside the organization.

The Master Data Controller for People provides one location to control the uses of all your personal data, and by extension, all downstream data flow. It does this by managing your application, site, and device personalization settings,

accessibility preferences, as well as marketing and privacy preferences. It encodes these data agreements into Smart Data Contracts and automates their enforcement everywhere your data flows and grows.

Data Containers

Data Containers are open-source transferable data files containing a single person's data and media like images, video, audio, key/pair value store, and relational value store. The Smart Data Protocol combines [Sertainty.com](https://www.sertainty.com) Data Containers, with [TodaQ.net](https://www.todaq.net) cryptographic proof of provenance wrapped in an array of templated Privacy Enhanced Technologies.

The Sertainty data privacy protection technology encapsulates sensitive data inside self-protecting, self-authenticating, and self-governing computer files. This is accomplished by combining encryption keys, authentication credentials, and our intelligence engine with your data, creating impenetrable Sertainty files. Protection is adjacent to, not in conjunction with existing network security protocols. To add clarification to some phrases used above:

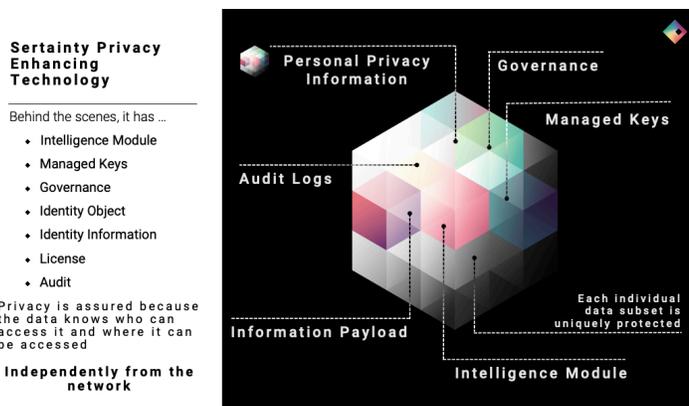
Intelligence Module – An embedded executable (think procedures inside a programming object) that is only accessible with an installed application having libraries and logic to access the files. Once the file is accessed, the executable controls data access, privacy and key management and auditing functions.

Self-Governing – The data owner determines the data's accessibility in terms of when, where, who, on what devices(s) and what network location(s). Those instructions are embedded in the file and enforced by the intelligence engine.

Self-Authenticating – Inside each file there is a Policy Implementer (PI) that contains information about each authorized user, enabling the intelligence engine to confidently authenticate users, processes, or devices, and thus ensure privacy.

Self-Protecting – The intelligence engine acts as an internal key management system (KMS). The keys are both protected and transported within the file itself, but NEVER as clear text, this is NOT protection by obscurity. Because we embed our own KMS, we produce multiple keys that allow us to independently protect multiple internal structures.

Self-Reporting – Each time the Sertainty file interacts with an Operating or is modified, the event is permanently recorded in the file, and those interactions can then be securely sent to a SIEM Device.



The file components are broken into multiple parts, with each part separately secured, providing independent governance rules and encryption keys for each component. Sertainty uses AES-256 encryption plus other obfuscating techniques. The goal is to ensure the secured file is: used by the right person; at the right time; in the right location (physical or logical); on the right device. Encryption is a means to a goal, not the ultimate outcome.

Bring Your Own Encryption (BYOE)

As we contemplated the uses of Data Containers in real-world scenarios, intrinsic and undeniable privacy became the most interesting property. We quickly realized data owners could, would, and should encrypt their data with keys they control before placing their data inside their containers for distribution. We now consider this behavior of using one's encryption unavoidable. It's very difficult to stop. It's potentially impossible to stop. Even if we, or others controlling the technology, wanted to prevent this behavior. It's an unavoidable privacy-preserving property of Data Containers. The intrinsic and unavoidable privacy preservation of Data Containers continues to fascinate our team.

Smart Data Contracts

[Smart contracts](#) have evolved into sophisticated distributed policy engines often associated with blockchains. [Smart Data Contracts](#) build on Smart Contract technology and apply the same ideas to automate data transactions and data agreements. They document data policies, regulations, licenses, and user preferences in language software can execute and automate.

Smart Data Contracts define and enforce data licenses, regulations, policies, and preferences, allowing data owners to remotely control each Data Container by defining, revising and revoking:

1. Who can access our data
2. When can they access our data
3. Where can they access our data
4. What questions do they ask of our data - [Zero-Knowledge Proofs](#)
5. How they need to protect our data – Differential Privacy, Synthetic Data, Security Standard Compliance, etc
6. How they can share data with others – [Secure Multiparty Computation](#).

Review, revise, or revoke your Smart Data Contracts at any time regardless of where your Data Containers are located.

Taken together, we start to see a standard for automating GDPR, CCPA, CPRA, and other regulations. Imagine fully automating the right to be forgotten or the right to data portability? Imagine automating internal data policies as data flows inside and outside an organization?

Key Point: While we can't automate all data usage terms that might exist in policies, regulations, licenses, and preferences, we now possess the technology to automate most of those terms. We believe that, with the proper foundation, future innovators will discover many new and interesting automation and encryption methods.

Edge Data Controllers

The [Edge Data Controller](#) gates and controls all access to Data Containers while automatically enforcing Smart Data Contract terms. Edge Data Controllers are like database executables interacting with Data Containers like database files. They are like data bodyguards or intelligent agents scanning and assessing their environment.

Data Containers are inaccessible without the cooperation of an Edge Data Controller providing secure key distribution and gating all data interactions per the terms of the attached Smart Data Contract defining the terms of the Zero-Knowledge Proofs and Secure Multi-party Computations.

Edge Data Controllers follow the familiar (CRUD) Create, Read, Update, Delete data lifecycle for all Data Containers. Edge Data Controller APIs in coordination with Master Data Controller APIs enable the Data Container lifecycle.

Intelligent Composable Applications

Programmers can think of Data Containers as Programmable Data Objects or modular software components that hold our data, our data usage rules, as well as presentation code. These rules allow and disallow data interactions. Presentation code can include dashboards for analytics data, a shopping cart within a social post, or a micro pay button to bypass a paywalled article.

Programmable Data Objects are a kind of [Intelligent Composable Application](#) with more agile, modular, and adaptable data architectures. They also enable automated preferences based on personalization at scale.

Edge Compute/Metaverse Integrations

Edge devices exist outside traditional IT boundaries without physical protection and workload integrity and without standards for data protection, detection, and remediation across potentially compromised networks and platforms.

Edge devices and data must interoperate across a growing number of organizations. The Internet of Things (IoT) includes a bewildering array of edge devices and use cases. The combined Global IDs/Access solution solves this missing standards problem and secures the remote data flowing across potentially insecure edge networks.

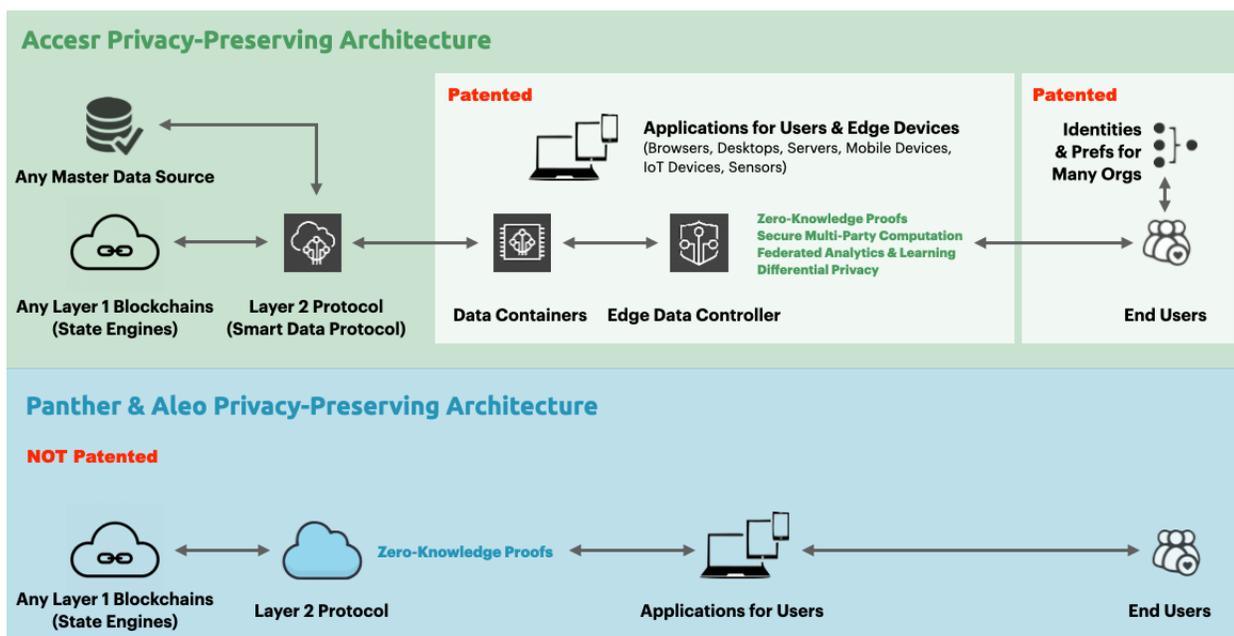
Non-Fungible Tokens/Blockchain Integrations

NFT Licensing doesn't work the way NFT advocates say. We cannot license, limit use, or revoke access to digital assets or data. NFT licensing is basically broken.

Our combined solution enables PETs, cryptographically secure data containers, and cryptographic data provenance to be used in combination with blockchains and NFTs solutions to finally solve digital asset and NFT licensing and monetization challenges.

Digital and data asset monetization require:

1. Transient or time-bound use with a commencement date, duration, renewals, and extension conditions.
2. Limitations on terms of use include responsibilities, geographic restrictions, rights of use, permissible channels of trade, sub-licensing, or assignability to another party.
3. Legal rights enforcement ensures the digital asset owner can take data assets back and revoke access per the terms of the license agreement.



Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Datavant

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



44 Montgomery St, 3rd Floor ♦ San Francisco, CA 94104

**Request for Information on Advancing Privacy-Enhancing Technologies
Office of Science and Technology Policy (OSTP)**

submitted via email PETS-REFI@nitrd.gov

Organization: Datavant, Inc.
Respondent Type: Industry
Contact: Doug Fridsma, Head of Government Partnerships, doug@datavant.com

Organizational Details: Datavant is the leader in privacy preserving data exchange, working with over 500 institutions to connect health data. Our mission is to connect the world's health data to improve patient outcomes and bring new treatments to patients faster. To accomplish this, we are connecting a network of companies, non-profits, and government entities that utilize our common infrastructure for the safe exchange of patient-level health information.

At Datavant, we believe that data fragmentation is the largest challenge facing the health data industry, and protecting patient privacy is paramount when using health data to improve health and health care. We are focused on building an open data ecosystem that allows stakeholders in the healthcare system to freely exchange data while protecting patient privacy.

Datavant provides three key solutions within the health sector privacy-enhancing technologies (PET) space:

- Privacy-preserving record linkage (PPRL), which enables disparate records to be linked in a de-identified manner,
- Data de-identification and redaction tools and services, which enable data to be redacted and modified to meet the definition of de-identification within HIPAA,
- HIPAA Expert Determination and data risk disclosure tools and services, applying statistical and cryptographic expertise to ensure datasets formulated meet the definition of the HIPAA Privacy Rule for the Expert Determination Standard §164.514(b)(1).

Datavant's privacy-preserving record linkage and de-identification technology is a foundational, neutral privacy enhancing technology. It has been used to power innovative solutions that enable scientific advancement while preserving individual privacy. The use cases that this technology powers includes, but is not limited to:

- The formation of registries and data collaboratives such as the NIH National COVID Cohort Collaborative, N3C.
- Linkages between trial data and real world data sources (e.g. claims, EHR data) to form more complete longitudinal views of clinical trial cohorts for long term safety and effectiveness studies.
- Discovery of shared patient cohorts across disparate datasets to form more complete longitudinal medical records for patient cohorts of interest.
- Real World Data repositories to power large scale evidence generation studies.

- Linkages between data sources that fall under differing privacy frameworks such as health data and social determinants of health data.

We draw on our experience across all these various use cases in response to this request for information regarding advancing privacy-enhancing technologies.

1. Specific research opportunities to advance PETs

Feasibility Studies: We encourage publication and knowledge-sharing of feasibility studies that build the evidence base for PETs. Peer-reviewed feasibility studies outline validated methods and results that build confidence in novel strategies that can produce equal if not better ultimate outcomes. One example of a gold standard validation study is the CDC National Center for Health Statistics comparison of performance for privacy-preserving record linkage to clear-text identifiable data linkage.¹ The study showed that the method using a PET can effectively link datasets without using identified data and produced results similar to identifiable data linkages. Importantly, this study outlined areas of future research and several important hypotheses.

Standardized, Comprehensive Evaluation Framework: One weakness in feasibility studies for PETs is the lack of standardized framework for conducting these studies and clear benchmarks for feasibility that will be widely accepted by the community of practice. There should be clear measures of feasibility that include privacy, security, neutrality, interoperability, utility and usability. There should be official standards for scoring feasibility across different PET technologies that can provide guard rails for feasibility studies such to ensure that the evidence base for these novel technologies is both rigorous and fair. In particular, we encourage the publication of evaluation frameworks for the assessment of privacy and security since the dimensions of methods used to enable PETs are not well understood. For example, evaluators and decision-makers frequently confound encryption with secure hashing techniques.

Sector-Specific Definitions of De-identification and Re-identification Risk: Assuring the PET meets regulatory requirements and protects privacy is critical to the use of PETs. Feasibility studies and standardized frameworks may be sector specific and tied to specific regulatory frameworks. For example, within the healthcare sector, there are specific standards in HIPAA which outline the definition of de-identification. PETs exist in a constantly shifting landscape of new methods and new data, requiring a continuous monitoring program to ensure de-identified or anonymous data assessments evolve lockstep with landscape shifts. Within healthcare, understanding the base rate frequencies of certain kinds of data (genetic mutations, rare diseases, etc) will be critical to using statistical methods to assure data has been properly de-identified. Additional studies to understand base rates of certain kinds of data (genetic, rare diseases, orphan drugs, etc) will be an important adjunct to the use of PET and to assure that data has a low risk of re-identification.

¹ Mirel, Lisa. "Privacy Preserving Techniques: Case Studies from the Data Linkage Program". (2021) Presentation for the Board of Scientific Counselors, National Center for Health Statistics. <https://www.cdc.gov/nchs/data/bsc/bscpres-Mirel-Data-Linkage-20210519-508.pdf>

Official Reviews of PET Best Practices & Applications: There should be official reviews of best practices in the use of PET. These official reviews should utilize impact frameworks evaluating the inputs, outputs, and outcomes of their application; they should clearly document the exemplars of the usage of PET for a given outcome. Again these official reviews should have a set of standard methods and formats that can provide structure and rigor to best practice reviews.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs;

Health Sector Focus: Datavant specializes in privacy-preserving record linkage (PPRL) in the health sector. The fragmented data across the health sector requires this privacy preserving methodology in order to connect data at the individual level across sectors while complying with HIPAA privacy regulations. The core technology enables Datavant's data partners to de-identify data on-premise using cryptographically certified hashing methods and then link data outside of their institution without revealing personally identifiable information. Datavant has paired core privacy-enhancing technology with technology-assisted statistical risk and data disclosure methods within the Expert Determination standard of the HIPAA Privacy Rule to deliver a comprehensive package of data connectivity products and services to our industry and government partners.

Data Inclusivity & Completeness: The health sector's lack of available data due to privacy concerns tends to have a chilling effect on data-sharing. This in turn exacerbates the problem of representative and inclusive data. By enabling PPRL across a broad, inclusive, interoperable ecosystem of 600+ organizations including academic medical centers, Federally Qualified Health Centers (FQHCs), and community organizations, beyond traditional data partners encompassing claims, registries, lab, -omics, and prescription data, PETs are particularly suited to close data completeness gaps. Data-sharing initiatives are ultimately about trust, and disadvantaged, and underserved populations tend to be wary about how their individual data can be used for non-sanctioned purposes. By increasing adoption and literacy on how PETs facilitate privacy-preservation of individuals, as an industry we can collectively work towards a more representative, inclusive, and complete data analysis landscape.

Enabling Data-Sharing Across Regulatory Frameworks: Privacy preserving record linkage has shown particularly salient utility in the face of the data sharing challenges between jurisdictions (e.g. state-to-federal, state-to-state) and between HIPAA covered entities and non-covered entities (e.g. social determinants or community-based linkages). We have outlined below several examples in each of these categories to illustrate the potential benefits of PETs to address data sharing issues and obstacles.

(A) PPRL Between Different Legal Jurisdictions

Data sharing between jurisdictions can be complex due to the differences in regulations and entities making decisions when identifiable data is involved even with PETs. Certain types of approaches, such as PPRL facilitate data sharing even across different regulatory

boundaries. For example, federal agencies requesting data from states, state agencies sharing data with their neighboring states, local government sharing data with states.

- **HIV Data Enclave** The City of Chicago Department of Public Health has to coordinate with healthcare providers, county public health, and state public health to ensure the correct jurisdiction is able to follow-up with the patient. While all these entities share relevant data in identifiable data form, the compliance requirements and increasing risks of holding sensitive HIV data centrally precludes many centralized data approaches. The Chicago Department of Public Health is using PPRL in a central data enclave, as a means to aggregate HIV data across multiple systems and databases that contain patients that are solely managed by Chicago, and patients that have been diagnosed with HIV, or have received HIV care in other jurisdictions.

(B) PPRL Between HIPAA Covered Entities and Non-Covered Entities

When an entity holding HIPAA protected health information (PHI) needs to share data with an entity that is not subject to HIPAA (e.g. the FTC governs consumer data-sharing such as consumer wearables), or community social work organizations, there tends to be a lengthy legal process to determine how to formulate the appropriate data-sharing agreements, and how both entities can meet the regulatory requirements. This lengthy process in an already complex data-sharing landscape particularly impacts the ability for healthcare organizations to collaborate on non-traditional medical factors, frequently termed the social determinants of health — examples of these are homelessness advocacy organization, community preventative care organizations.

- **Clinical trial design for the ADAPTABLE study:** The ADAPTABLE study was highlighted in one of the first guidance documents FDA issued on the Real-World Evidence Program. In the Chicago-based trial recruitment for ADAPTABLE, we employed a different approach to patient trial recruitment by meeting patients in their communities to ensure a sufficiently diverse population. Specifically, Datavant partnered with community-based organizations such as the Pastors 4 PCOR, and the Sinai Urban Health Institute. RWD originating from EHR data from the areas' health systems were tokenized using Datavant privacy-preserving tokens. When a patient expressed interest and met a small set of inclusion/exclusion criteria, their data would be linked using the privacy-preserving tokens to the pool of EHR RWD to verify eligibility based on the medical record. This is a key innovation in ensuring comprehensiveness of trial participants, and continuing to use further RWD from the medical record to disqualify enrolled participants, e.g. new unrelated complications arise during the period of the study, new indications that would disqualify their participation.²
- **Privacy preserving feedback loops between health and social services for vulnerable populations:** The University of Texas at Austin has leveraged their existing Breathe Austin platform to incorporate privacy-preserving linkages such that when a patient visits a clinic location they can consent to have their data shared between the clinical system and the findhelp.org social services referral and management platform.

² Zimmerman, L., Goel, S., Sathar, S., Gladfelter, C., Onate, A., & Kane, L. et al. (2018). A Novel Patient Recruitment Strategy: Patient Selection Directly from the Community through Linkage to Clinical Data. *Applied Clinical Informatics*, 09(01), 114-121. doi: 10.1055/s-0038-1625964

This connectivity helps close the feedback loop between social services and clinical care. Currently, when social service referrals are made, health care providers don't know whether the patient used those services, or if those services were effective. By developing a feedback loop through privacy-preserving data linkages between a HIPAA covered entity and community organizations which are non-covered entities, better care coordination can occur between providers, social service providers, and downstream community coordinators who work to engage patients in their own care.

- **Data connectivity for a community-based approach to care coordination and triage of homeless and housing-insecure populations:** Data from geographically proximal health systems, including VA hospitals were pooled and linked with data from two community-based homeless housing organizations in Chicago, Illinois. This linked data source enabled researchers to analyze health care utilization, access, and medical conditions for individuals, including Veterans, experiencing homelessness or housing-insecurity. Data was combined using the Datavant privacy-preserving record linkage solution which enables disparate datasets to be linked longitudinally without needing to share personal identifiers. The community linked datasets were used as a basis to inform coordination amongst area health systems, their community partners, and public health entities. The area health systems formulated support processes including a homeless housing subsidy to facilitate a path for individuals to receive housing. The Veterans Health Administration used this community-based linkage to gain a better understanding of their homelessness support programs and record systems by understanding which civilian sector health systems were most likely providing care for at-risk Veterans. This data connectivity has been cited in two widely circulated publications.

William E. Trick, Jennifer C. Hill, Peter Toepfer, Fred Rachman, Beth Horwitz, and Abel Kho, 2021: Joining Health Care and Homeless Data Systems Using Privacy-Preserving Record-Linkage Software American Journal of Public Health 111, 1400_1403, <https://doi.org/10.2105/AJPH.2021.306304>

Jason H Raad, PhD, Elizabeth Tarlov, PhD, RN, Abel N Kho, MD, MS, Dustin D French, PhD, Health Care Utilization Among Homeless Veterans in Chicago, Military Medicine, Volume 185, Issue 3-4, March-April 2020, Pages e335–e339, <https://doi.org/10.1093/milmed/usz264>

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:

Enact data portability and interoperability guidance: The Government should require responsible data sharing strategies to be submitted when digital health solutions funded by the Government are deployed. These data-sharing policies should be a parallel to the Information Blocking Rule in the 21st Century Cures Act from the HHS Office of the National Coordinator (ONC), by ensuring that digital health solutions deployed in the community offer some form of data sharing portability.

Ensure personal privacy: As part of this process, it is imperative for the Government to establish data sharing practices that protect the privacy of individuals. Without this guidance, digital health solutions intended to benefit and reach underserved and underrepresented populations that are already distrustful of Government initiatives may be underutilized, leaving a gap in enabling communities to have a comprehensive connected health strategy.

10. Other information that is relevant to the adoption of PET

As the government defines a National Strategy to advance the appropriate use of PETs, we recommend the following:

- **Invest in technology and infrastructure that allows for interoperability between different PET solutions.** We need neutral PET solutions that can work with multiple analytics platforms, different data aggregation solutions, and provide bridges between different data ecosystems. Continued investment in interoperability within the PET is critical to the success of these approaches.
- **Support studies that provide baseline frequency information for certain data types, to improve the accuracy of expert determination.** PET can only be effective if we have ways to assure the technology is protecting privacy. For expert determination (and other statistical methods of assessing re-identification risks), baseline frequencies of certain kinds of data would be needed. Additional studies to identify a library of vulnerabilities and attack scenarios on each type of PET to assess the risk of privacy risks and potential of disclosure or information leak.
- **Incentivize the private sector to adopt PET.** Providing safe harbors for liability to organizations that use PET that has been certified to meet regulatory requirements.
- **Formulate intergovernmental guidance to provide clarity for use of PETs across federal agencies when linking data from different sources.** As more data enclaves are used to protect sensitive data, it is imperative that mechanisms to link across enclaves are used. PET can provide a safe and certifiably de-identified mechanism to aggregate data across enclaves while assuring patient privacy.
- **Continue to invest in research that enhances the performance of PET.** Investments from ARPA, ARPA-H, NIH, NSF, and other agencies can be used to advance the science and evaluation of PET.
- **Provide a regulatory framework that enhances data sharing and linking using PET.** This can include agency level incentives (scored data sharing plans for the NIH, OCR rules for use of PET), consistent frameworks for evaluation of effective PET, requirements to assure PET is compliant with regulatory requirements, and other measures to ensure that PET is protecting privacy while making the data useful for policy and research.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Deloitte Consulting LLP

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



Advancing Privacy-Enhancing Technologies

July 8, 2022

In response to Office of Science and Technology Policy Request for Information

July 8, 2022

Jeri Hessman
on behalf of the
Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, D.C. 20504

RE: RFI Response: Privacy Enhancing Technologies

Dear Ms. Hessman:

Deloitte¹ Consulting LLP (Deloitte) is pleased to submit a response to The White House Office of Science and Technology Policy (OSTP) Request for Information (RFI) on Privacy Enhancing Technologies (PETs).

We applaud OSTP's efforts to inform a national strategy on accelerating the development and adoption of PETs. Such technologies have the potential to lead to more trustworthy artificial intelligence systems, while also preserving the privacy of sensitive data.

Despite realizing the importance of privacy protection, policymakers and technical experts alike have struggled to overcome the challenges that current approaches to data privacy have created. Appropriate legislative and regulatory environments may help to reduce some of these frictions, but to be effective they should be complemented by the development and application of new methods for data analysis, such as PETs.

We would be pleased to have an open dialog with OSTP as it continues to explore PETs and reviews responses from this and future RFIs. Please do not hesitate to contact me at (703) 216-4581, should you have any questions.

Sincerely,

Ed Van Buren

Principal

AI in Government Leader

Deloitte Consulting LLP

¹ As used in this document "Deloitte" means Deloitte Consulting LLP. Please see www.deloitte.com/us/about for a detailed description of the legal structure of Deloitte LLP and its subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.

Table of Contents

Company Profile 2
 Executive Summary 2
 Introduction 4
 Response to Topic 2 4
 Response to Topic 3 6
 Response to Topic 6 8
 Conclusion 10
 References 11

Company Profile

Company Name	Deloitte Consulting LLP
Headquarters Location	New York City
Contact Name	Ed Van Buren
Contact Title	Principal, Deloitte Consulting LLP
Contact Email Address	_____
Contact Phone Number	
Primary Type of Service(s) Provided	Software Development, Professional Services, Management Consulting, Technical Support, Maintenance, and Support Services

Executive Summary

Artificial Intelligence (AI) has been widely recognized as having the potential to transform how services are delivered to citizens and to help government leaders make better informed, data-driven decisions. The National AI Initiative Act represents a significant step in ensuring that the United States continues to lead the world in AI research and development, as well as preparing the nation for the adoption of AI systems across the federal government and, more broadly, society. More recent developments, like the White House Office of Science and Technology Policy (OSTP)’s efforts to engage the public on a “Bill of Rights for an Automated Society,” illustrate the importance of aligning data-driven technologies (e.g., AI) with the values that underpin American society.

In preparing a response to this RFI, Deloitte draws upon extensive experience in AI strategy, implementation, and operational projects across the public and private sectors. More than 1,200 Deloitte practitioners are currently engaged with federal, state, and local government clients to harness the power of AI, machine learning (ML), and data analytics. In the commercial sector, Deloitte’s comprehensive range of technology solutions and service offerings have helped accelerate AI adoption at the world’s leading companies. In addition, Deloitte’s AI Institute for Government [1] has served as a prominent hub for AI research, eminence, and applied innovation. Deloitte’s Trustworthy AI™ (TAI) framework [2] has also been used in three executive agencies as a blueprint to maximize the benefits of AI, while also effectively managing potential risks and ethical imperatives.

One critical dimension of our TAI framework is privacy protection, a concept that has also been identified by the federal government as a strategic pillar of trustworthy AI systems. Despite realizing the importance of privacy protection, policymakers and technical experts alike have struggled to overcome the challenges that current approaches to data privacy have created. For example, parties that want to collaborate on building AI systems using shared data must first negotiate complex data use agreements. These data use agreements often take many weeks or months to finalize, creating delays and limiting the ability of organizations to fully harness the power of AI.

Appropriate legislative and regulatory environments may help to reduce some of these frictions, but these efforts should be complemented by the development and application of new methods for data analysis, such as privacy enhancing technologies (PETs). PETs refer to a broad set of algorithms, software, and hardware tools that preserve the privacy of sensitive data during analysis. Many of these technologies, like homomorphic encryption and zero knowledge proofs, have been researched and applied extensively in the fields of mathematical cryptography and computer science since the early 1980s. Other PETs, such as federated learning and differential privacy, were formalized more recently to help securely analyze the massive amounts of sensitive digital data generated from smartphones and wearable electronics.

We believe that successful use of PETs for trustworthy AI will require the federal government to take a holistic, “systems-level” perspective. Proper application of PETs will involve an understanding of the interconnections between AI and a host of related topics, including digital identity, the Internet of Things (IoT), and quantum computing. Moreover, by treating PETs as modular components that can be integrated into existing information systems, the government can lower some of the technical barriers to adoption. Identifying opportunities to combine different PETs into the same application or technology stack could also help to increase data privacy protections.

One notable PET that can serve as a useful template for integrated, privacy preserving data analytics systems is federated learning (FL). FL refers to a setting in which machine learning models are trained collaboratively by a network of participants using a distributed network of datasets that are independently owned and maintained. In a FL system, no raw source data is explicitly shared between the collaborators.

FL can be applied to address important challenges in key areas, such as public health, national security, and civil infrastructure. Using data analytics and AI to improve disease surveillance, combat financial crimes, enhance military interoperability at the tactical edge, and build the next generation of “smart” cities will require a whole-of-society response. Different government agencies, departments, and even private sector parties can use FL to collaborate on critical data analytics efforts, while also mitigating risks to privacy.

To realize the full potential of PETs in creating trustworthy AI systems, the federal government can focus on the mechanisms that are proven to support scientific and technological innovation. Among these are workforce training and educational programs, incubation of public-private partnerships, and the development of common technical standards. By encouraging public-sector workforce training programs to integrate PETs into their curriculums, OSTP can help lead with fellow federal partners to ensure the next generation of technical practitioners and administrators are well equipped to apply PETs effectively. Similarly, by expanding the National Science Foundation (NSF)-led National AI Research Institutes to include PETs as a research area, the federal government can establish adequate funding for these important technologies.

An effective national strategy to accelerate the adoption of PETs should also address the key roles that external non-government partners can help play in implementing these technologies effectively. By creating an ecosystem that incubates relationships between academia, early-stage start-ups, and leading technology companies, the federal government can build on the collective strengths of a robust group of stakeholders. Low-risk collaboration opportunities, like agency-specific technology sprints, public competitions, and state-wide challenges can help to create a robust stakeholder ecosystem. In turn, this ecosystem can help shape common standards for how PETs are designed, developed, and implemented as part of more complex data analytics systems. As a leader in both AI and private-sector consulting, Deloitte has deep experience in developing commercial technology ecosystems, managing alliance partnerships, and fostering productive cross-sector collaboration on AI and data analytics.

Deloitte stands ready to support OSTP in creating a national strategy to advance PETs. We are confident in the vision put forth by Dr. Alondra Nelson for “a world where our technologies reflect our values and innovation opens the door to solutions that make us more secure.” [3] We believe that such a future, designed with our society’s core values in mind, is well within reach.

Introduction

OSTP's inquiry into PETs marks a significant milestone for the government in its efforts to advance technological innovation that adheres to our society's core values. In support of OSTP's efforts, we are pleased to share our knowledge from our historical and ongoing experience advising public sector and commercial organizations on the implementation of data analytics, AI, and privacy solutions.

In this document, we focus on three topics: specific technical aspects and limitations of PETs (RFI Topic 2), sectors that would benefit from the adoption of PETs (RFI Topic 3), and specific mechanisms that could be used, modified, or introduced to advance PETs (RFI Topic 6). The scope of our response in the first two sections is restricted to a detailed discussion of federated learning (FL), a PET with great potential for rapid adoption across the public and private sectors. While these three topics are the focus of this response, Deloitte's expertise spans across a range of related technical and policy areas. We would be happy to offer our perspective on additional topics in another forum.

Response to Topic 2

Specific Technical Aspects and Limitations of Federated Learning

In this section, we begin with an overview of the technical aspects of federated learning. Next, we illustrate the aspects of FL *systems*, as they pertain to driving adoption of multiple PETs. We then outline the technical limitations and operational risks of FL systems and conclude with a discussion of recent advancements which help to mitigate some of the risks of adopting FL systems.

Federated Learning enables privacy-preserving collaborative analytics

FL is a form of multi-party machine learning that has been used in the consumer technology sector since 2017. FL techniques have been applied to train text-prediction models, improve voice-recognition models, and build other machine learning models in settings where training data could not be centralized due to privacy concerns and bandwidth limitations [4] [5]. FL allows for a distributed network of data providers to collaboratively train a global machine learning model without revealing any of the underlying training data. Each participant performs a learning or analytics task on its own local (private) dataset, and only shares information about the result of that task, rather than the underlying dataset itself.

This approach to machine learning presents clear opportunities to preserve data privacy by bypassing the direct access or transfer of raw data. As a result, organizations can choose FL as a strategy to collaborate on machine learning tasks even when privacy, security, or trust concerns prevent data centralization. The ability of FL to effectively expand the scope of data available for mission-critical AI systems makes it especially relevant in the public sector. Furthermore, when organizations find the same type of data spread across many locations or providers, FL can be used to build models which are more robust, generalizable, and representative. By allowing data providers from different locations and organizations to collaborate in a privacy-preserving way, FL can serve a key role in promoting equitable benefits of AI for participants that may not have access to large volumes of high-quality data.

It is important to note that the key principles that underpin FL can be extended to create more general purpose "federated analytics" systems. Such systems can be used for arbitrary data analysis tasks, like calculating metrics across a distributed network of datasets.

Federated Learning can serve as a vehicle for broader PET adoption

Like all forms of machine learning and data analysis, FL is not applied in a vacuum. FL models are trained, tested, and deployed as part of a complex *system*. Organizations that wish to use FL for collaborative analytics must design and develop an integrated system that involves data storage mechanisms, secure messaging and communication protocols, and distributed compute capabilities. Each of these system functions will involve processing sensitive data, creating a

need for security and privacy protections. These privacy and security needs, in turn, can be addressed with other PETs. Hence, an FL system designed for the purpose of high-security, privacy-preserving collaborative data analytics can serve as a platform to integrate a broader set of PETs [6].

Limitations of Federated Learning

- **FL model development can involve significant technical “overhead”:** Performing a ML or analytics workflow in an FL setting requires FL-specific aggregation and optimization algorithms suitable for the desired task [7, p. §3]. In many cases, FL involves system-level parameter tuning, code-refactoring, and new algorithm development, leading to longer development cycles and a need for specialized technical expertise.
 - *Advancements:* Technical overhead can be reduced by relying on open research and open-source tools. Researchers are developing and providing code for algorithms which are tailored to specific settings or needs [8] [9], or to support new families of models [10]. At the same time, popular toolkits for FL systems have common algorithms supported out-of-the-box [11] and are increasingly focused on streamlining the developer experience for implementing arbitrary workflows [12].
- **FL model performance can be degraded by highly variable data:** The accuracy, generalizability, and overall performance of FL models are heavily influenced by the quality, heterogeneity, and structure of training data. Different organizations may sample different populations, employ different data collection methods, and have different quality control resources available. Ensuring that an aggregate model will converge with high performance on data distributions which are not necessarily independent and identically distributed (non-IID) is a major challenge in FL.
 - *Advancements:* Strategies to improve FL performance on non-IID data have emerged in various settings including implementing new FL optimization algorithms [13], relying on synthetic data generation [14], augmenting FL workflows [15], or even redefining the target FL workflow to prioritize personalized models over a global model [16].
- **Heterogeneous computing environments can make it difficult for participants to coordinate:** Since FL models are trained collaboratively by a distributed network of participants, it is essential that there are common operating standards across the network. Each participant must have machines with adequate compute power, sufficient data storage capacity, and reliable communication channels. Depending on the setting, there may be significant effort needed to configure computer systems consistently across different participants.
 - *Advancements:* The rapid growth in adoption of containerization tools like Docker is continually making it easier to standardize software environments on different machines [15]. In addition, newly released FL software toolkits are making use of well-established protocols such as gRPC to simplify communication between participating machines. Discrepancies in compute power and availability of participants is also being addressed by researchers exploring techniques of asynchronous FL, active sampling, and fault tolerance [18].

Privacy and Security Risks associated with Federated Learning

- **Model training can be undermined by malicious participants:** The inability to directly inspect the training data from each participant leaves FL models susceptible to data poisoning attacks. These attacks occur when the private datasets or updates returned during federated model training are crafted by a malicious participant to negatively impact the model’s quality. Such attacks may come from a single malicious participant or a coordinated group of participants forming a Sybil Attack [19]. If not properly addressed, these attacks can be used to degrade model performance or embed backdoors to make future adversarial attacks more effective.
- **Sensitive information can be reverse engineered:** Models created during federated learning are usually shared between participants during a learning task. In the academic literature, it has been shown to be possible for

participants to perform data reconstruction or membership inference attacks on these models [20]. Without sufficiently strong protections in place, such attacks make FL systems vulnerable to leaking sensitive data.

Risk Mitigation Strategies

Well-designed FL systems typically have strong data privacy requirements, providing an opportunity for the integration of complementary PETs. This integration not only provides the privacy guarantees of each individual PET but also serves to mitigate some of the risks of FL outlined above. Below, we present examples of PETs that have been used to further enhance the privacy protections of a FL system:

- ❑ **Differential Privacy** can be applied to participant information before it is shared so that the contributions have additional privacy guarantees, mitigating the effectiveness of data reconstruction or membership inference attacks [5].
- ❑ **Homomorphic Encryption** can be applied to participant information before it is shared so that the contributions have additional security guarantees, allowing for all participant contributions to remain encrypted throughout the workflows [4].
- ❑ **Secure Aggregation** can be used to enforce cryptographically secure aggregation of participant information so that no contribution can be inspected individually, preventing attacks from targeting specific participants [2].
- ❑ **Zero Knowledge Proofs** can be used to defend against poisoning attacks hidden in participant contributions without violating secure aggregation or revealing participant information [17].
- ❑ **Private Set Intersection** can be used in FL settings to perform entity resolution between participant datasets in a privacy-preserving manner [18].
- ❑ **Trusted Execution Environments** can provide confidentiality and integrity guarantees on local data processing for participants in a FL system [20].

Since FL system architects often turn to other PETs to address a spectrum of privacy requirements and security considerations, it is likely that encouraging greater adoption of FL systems will in turn increase understanding, adoption, and advancement of PETs more generally.

Response to Topic 3

Sectors that would benefit from adoption of Federated Learning

In this section, we seek to direct OSTP's attention to the key areas in which FL systems are especially useful. Data analytics technologies have long been utilized by public and private sector organizations, but increasing desires for consumer privacy and control, coupled with a growing body of privacy regulation, means that the need for managing risk is growing more complex every day. We believe that adoption of FL systems has the potential to drive transformative change across many sectors, including healthcare, national security, finance, and civil infrastructure.

Healthcare

Policymakers and medical practitioners alike have long envisioned a public health system focused on predicting and proactively preventing illnesses, rather than providing services retroactively. Realizing this vision will require a diverse group of public and private sector partners working towards a unified goal of sharing data-driven insights with one another in real time. The future of public health will depend on an ability to fuse real-time data from multiple sources to empower health leaders to identify health trends and respond to potential threats. In such a future, partners can both contribute to and benefit from predictive data analytics.

FL systems and other PETs would allow for patient information to stay protected on the de-centralized data silos in which they are usually stored, while also allowing for collaboration between hospitals, insurers, researchers, and health departments on critical public health issues.

In 2021, an applied research test used data from a global network of institutes to train a FL model that predicted future

oxygen requirements of symptomatic patients with COVID-19 using inputs of vital signs, laboratory data, and chest X-rays [24]. The model achieved an average area under the curve (AUC) > 0.92 for predicting outcomes at 24 and 72 h from the time of initial presentation to the emergency room. Additionally, it provided 16% improvement in average AUC measured across all participating sites and an average increase in generalizability of 38% when compared with models trained at a single site using that site's data. This study showed that FL resulted in collaboration without loss of identifiable data and produced a result that generalized across heterogeneous, decentralized datasets for prediction in patients with COVID-19.

Project MELLODDY, which concluded in May 2022 after three years of study, is another example of how FL has been successfully applied in the healthcare sector [25]. In this project, pharmaceutical companies collaborated to find a way to combine resources and research – without the loss of intellectual property (e.g. data) – to facilitate the development of new biomedical drugs. Using a FL system, these companies collaborated to improve performance of predictive models and reduce decision time needed to research next steps in drug discovery.

National Security

The benefits of a fully integrated FL system are already known to the Department of Defense (DoD). A 2021 call for industry response (Opportunity ID: TW-22-0003) was issued to address complications associated with integrating ML and AI in assets with Joint All-Domain Command and Control (JADCC). US Naval fleets are frequently inconsistent in both hardware and software, due to the changing landscape of asset manufacturers, operators, contracts, and department policy. Consequently, assets may have different network bandwidths, memory limits, and local data distributions. These are problems addressed by a FL system, as learning can occur asynchronously and independently between node machines. Base requirements for learning are configurable between Naval assets and allow for heterogeneous data sources, heterogeneous hardware bandwidth limitations, and secured communication between assets of JADCC. The result of building an integrated FL system could enable a greater degree of interoperability to increase mission capabilities.

FL systems would benefit national security beyond the tactical edge, as well. In March 2021, the Biden administration issued an executive order identifying supply chain reforms as critical to protecting US economic and national security interests [26]. Current supply chain management is set up on point transactions between top-level suppliers and buyers. The vulnerable upstream supply chain network is more opaque, largely due to hesitation by organizations to share information that could compromise competitive position, reveal compliance posture, or highlight security concerns. FL systems could allow for analysis directly on siloed data that can never be pooled directly due to concerns over privacy, intellectual property, and sovereignty. The result could be a federated global supply chain model built on top of a network of protected data, giving the federal government a more accurate and real-time assessment of supply chain strength and resilience.

Financial Crime Compliance

Private financial institutions sit at the front line of defense against financial crimes by detecting and flagging transactions based on suspicious activity. However, this process is costly and time-intensive due to the need for manual validation. A recent study sought to provide support by developing a FL system with additional PETs (i.e., homomorphic encryption and differential privacy) to increase model accuracy of detecting suspicious activity [27]. Leveraging this FL system, researchers showed accuracy improvements by as much as 20% in comparison to individual financial institutions. These results suggest that private financial entities can use FL systems to collaboratively investigate financial crimes without exposing private data.

FL systems can also be applied in similar settings to detect sanctions evasion, tax fraud, and other financial crimes without compromising the privacy of individual citizens or companies. The potential for PETs like FL to help investigators detect these types of financial crimes has been clearly recognized by the White House, which recently announced PETs-focused prize challenges involving the U.S. Financial Crimes Enforcement Network [3].

Civil Infrastructure

The bipartisan \$1.2 trillion Infrastructure Investment and Jobs Act (IIJA), signed by President Biden in November 2021, indicates a clear national priority that both Congress and the President support [28]. With this infusion of support and funding to update and modernize civil infrastructure, the need for PETs to protect citizens' privacy and ensure efficient access to national resources (i.e., power, water, transportation, internet) is paramount.

Over 60% of infrastructure officials – local and federal – who responded to Deloitte's Future of Infrastructure survey identified AI as having the largest impact on infrastructure plans over the next three years [29]. Today, even physical infrastructure increasingly entails a digital component, evidenced by recent cyberattacks against critical infrastructure operators, schools, hospitals, manufacturers, banks, and others [30]. Additionally, as autonomous vehicles become more common, they will require compatibility with vehicle-to-vehicle, vehicle-to-grid, and Internet of Things (IoT) requirements, all of which come with an increased need for controlled learning and communication so that privacy considerations can be identified and protected. PETs, like FL, can play an important role in the development of new "smart" infrastructure, as communication and learning will be vital as new technologies are adopted with privacy in mind.

Among other provisions, the IIJA directs the Environmental Protection Agency to establish grant funding programs for the purpose of reducing cybersecurity vulnerabilities. In addition, it contains provisions related to high-performance computing and smart manufacturing [28]. This funding represents a potential opportunity to pursue the advancement of FL in identifying cybersecurity threats and facilitating smart manufacturing capabilities. In alignment with White House goals for transparency and improving citizens' access to services and service delivery experience [31], FL systems could address rising concerns surrounding security, access, and efficiency of critical infrastructure.

Response to Topic 6

Specific mechanisms that could be used, modified, or introduced to advance PETs

FL systems and other PETs can only succeed in a well-designed and efficiently functioning ecosystem that serves to improve the lives of US citizens. Such an ecosystem should focus on stakeholder organization and communication to prioritize transparency, privacy, and security. In this section, we describe some core mechanisms that could be introduced, modified, or utilized to achieve the privacy preserving and collaborative analytical benefits of FL systems and other PETs.

Integrate PETs into education and workforce training programs

Public sector practitioners should have extensive technical training on PETs to ensure they can successfully apply them at work. Agency leaders also need to understand PETs and their strategic applications for cross collaboration that range in requirements and feasibility. This can be achieved through integrating PET curriculums into existing educational and training programs.

For example, the Public Health Informatics and Technology (PHIT) Workforce Development Program could include additional technical training on PETs for future system engineers. Existing programs, such as National Institute of Standards and Technology (NIST)'s Differential Privacy Program, could be modified to increase PET education. At the academic level, scholarship programs for graduate studies could be tailored to PET research and include employment opportunities. For example, the National Science Foundation (NSF) AI Institutes could also be expanded to support PET researchers. Academic-driven initiatives would also enable students – America's future workforce – to experience the possibility of careers in the science, technology, engineering, and mathematics (STEM) fields. This could ensure a future pipeline of workers from diverse backgrounds who are both technically prepared and privacy-aware.

Encourage innovation and organic partnerships through a stakeholder ecosystem

PETs are an expanding field of privacy and security methods that benefit from the latest techniques to remain up to date and effective. Decision makers in the public sectors should recognize that collaborating with the private sector on PET development could help avoid strategic gaps in knowledge or capabilities. Encouraging innovation, maintaining strong partnerships, and referencing open-source tools are critical for this purpose. By leveraging agency-specific “tech sprints,” public competitions, and managing procurements through dedicated Other Transaction Authorities (OTAs), the federal government can help foster innovation.

By replicating the Department of Veterans Affairs’ (VA) National AI Institute’s AI Tech Sprints across other agencies, the federal government could provide low-risk testbeds for PETs in the public sector. Successful outcomes could then be followed by rigorous testing and evaluation for production use. Similarly, current programs such as SBIR Catalyst and the Growth Accelerator Fund could provide a pipeline for constant advancement in implementation and security, and government focused venture firms such as In-Q-Tel or NSF’s America’s Seed Fund could further ideation development. Internally, establishing research partnerships amongst academic and research institutions provides necessary advisory reference and quality assurance support.

To ensure strong partnerships, we encourage the government to provide select smaller private entities access to its own large data resources, when appropriate. When private institutions access large pools of public data to train their own models, they can expedite results and make potentially large impacts. For example, many private health institutions might have the expertise to develop advanced models but lack robust data to train their models. Access to robust and plentiful healthcare data collected by the government could expedite healthcare research and discovery.

Develop a standardized set of best practices for collaboration

We see the value in having a foundational set of standardized best practices to foster innovative ideas, talent, and resources. For instance, the NIST developed its Cyber Security Framework (CSF), consisting of standards, guidelines, and best practices to manage cybersecurity risk. The CSF has proved effective at managing a cohesive set of standards when implementing cyber security programs. Establishing similar guidelines to the CSF, among others [32], could provide the framework and foundation necessary to develop, use, and evaluate FL and PET products, systems, and services. When Deloitte supported NIST to develop the CSF, we conducted interviews with the private sector on NIST’s behalf and one of the most important aspects of the CSF that was highlighted was the flexibility to achieve the desired outcomes in a number of pre-existing ways.

We also helped to establish NIST’s Privacy Engineering Program (PEP), which includes an operating framework for contributing relevant tools and use cases. Such a framework regarding privacy techniques and standard practices could provide a template for PET pilot programs. Additional considerations, such as developing a common set of approved software tools, could potentially mitigate privacy and security risks—as well as help to uncover any unanticipated consequences of normally-operating PETs.

Focus on transparent and clear communication about PETs

Much of the data used to improve the daily lives of US citizens contains sensitive and private individual information. As such, it is vital to provide assurances and instill accountability when utilizing individual-level data. We recommend internal awareness communiques regarding PET impact to be circulated among government practitioners and leadership. These communications could take the form of newsletter and blogs, like the NIST’s PEP Program. Externally, easily explainable public awareness campaigns can help highlight the use cases and benefits of PETs. Consistent communication can help reinforce transparency and ensure that safeguards remain vigorous and effective.

Conclusion

Deloitte is excited about a future that is enhanced through data analytics and AI, especially when the appropriate privacy protections are identified and integrated. PETs have the potential to enable our society to operate more safely and efficiently, while also protecting individual liberties and privacies. Across a broad swathe of sectors, PETs can be used to build more trustworthy and secure data analytics systems. It is with this promise in mind that we look forward to supporting OSTP and other federal government partners in establishing a national strategy to accelerate the development and adoption of PETs.

References

- [1] "Deloitte AI Institute for Government," Deloitte US, [Online]. Available: <https://www2.deloitte.com/us/en/pages/public-sector/articles/artificial-intelligence-government-sector.html>.
- [2] "Trustworthy Artificial Intelligence (AI)™," Deloitte US, [Online]. Available: <https://www2.deloitte.com/us/en/pages/deloitte-analytics/solutions/ethics-of-ai-framework.html>.
- [3] U.S. White House Briefing Room, "U.S. and U.K. Governments Collaborate on Prize Challenges to Accelerate Development and Adoption of Privacy-Enhancing Technologies," Press Release, 13 June 2022. [Online]. Available: <https://www.whitehouse.gov/ostp/news-updates/2022/06/13/u-s-and-uk-governments-collaborate-on-prize-challenges-to-accelerate-development-and-adoption-of-privacy-enhancing-technologies/>.
- [4] B. McMahan and D. Ramage, "Federated Learning: Collaborative Machine Learning without Centralized Training Data," Google AI Blog, April 2017. [Online]. Available: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>.
- [5] K. Hao, "How Apple personalizes Siri without hoovering up your data," MIT Technology Review, December 2019. [Online]. Available: <https://www.technologyreview.com/2019/12/11/131629/apple-ai-personalizes-siri-federated-learning/>.
- [6] D. Stewart, A. Bucaille and G. Crossan, "Homomorphic encryption and federated learning can underpin more private, secure AI," Deloitte Insights, December 2021. [Online]. Available: <https://www2.deloitte.com/xe/en/insights/industry/technology/technology-media-and-telecom-predictions/2022/homomorphic-encryption-federated-learning.html>.
- [7] P. Kairouz, B. McMahan and et al., "Advances and Open Problems in Federated Learning," 2019. [Online]. Available: <https://arxiv.org/abs/1912.04977>.
- [8] S. Reddi, Z. Charles, M. Zaheer and et al., "Adaptive Federated Optimization," 2020. [Online]. Available: <https://arxiv.org/abs/2003.00295>.
- [9] T. Li, M. Sanjabi, A. Beirami and V. Smith, "Fair Resource Allocation in Federated Learning," 2019. [Online]. Available: <https://arxiv.org/abs/1905.10497>.
- [10] Y. Wu, S. Cai, X. Xiao and et al., "Privacy Preserving Vertical Federated Learning for Tree-based Models," 2020. [Online]. Available: <https://arxiv.org/abs/2008.06170>.
- [11] "NVIDIA FLARE," [Online]. Available: <https://developer.nvidia.com/flare>.
- [12] "Flower: A Friendly Federated Learning Framework," [Online]. Available: <https://flower.dev/>.
- [13] L. Zhang, Y. Luo, Y. Bai and L.-Y. Duan, "Federated Learning for Non-IID Data via Unified Feature Learning and Optimization Objective Alignment," 2021. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/papers/Zhang_Federated_Learning_for_Non-IID_Data_via_Unified_Feature_Learning_and_ICCV_2021_paper.pdf.
- [14] H. Chen and H. Vikalo, "Federated Learning in Non-IID Settings Aided by Differentially Private Synthetic Data," 2022. [Online]. Available: <https://arxiv.org/abs/2206.00686>.
- [15] Y. Zhao, M. Li, L. Lai and et al., "Federated Learning with Non-IID Data," 2018. [Online]. Available: <https://arxiv.org/abs/1806.00582>.
- [16] Y. Huang, L. Chu, Z. Zhou and et al., "Personalized Cross-Silo Federated Learning on Non-IID Data," 2021. [Online]. Available: <https://www.aaai.org/AAAI21Papers/AAAI-5802.HuangY.pdf>.

- [17] "Docker," [Online]. Available: <https://www.docker.com/>.
- [18] T. Li, A. Kumar Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," 2019. [Online]. Available: <https://arxiv.org/abs/1908.07873>.
- [19] C. Fung, C. J. Yoon and I. Beschastnikh, "The Limitations of Federated Learning in Sybil Settings," 2020. [Online]. Available: <https://www.usenix.org/system/files/raid20-fung.pdf>.
- [20] V. Mothukuri, R. M. Parizi, S. Pouriyeh and et al., "A survey on security and privacy of federated learning," 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167739X20329848>.
- [21] T. Nguyen and M. T. Thai, "Preserving Privacy and Security in Federated Learning," 2022. [Online]. Available: <https://arxiv.org/abs/2202.03402>.
- [22] L. Lu and N. Ding, "Multi-party Private Set Intersection in Vertical Federated Learning," 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9343209>.
- [23] F. Mo, H. Haddadi, K. Katevas, E. Marin, D. Perino and N. Kourtellis, "PPFL: Privacy-preserving Federated Learning with Trusted Execution Environments," 2021. [Online]. Available: <https://arxiv.org/abs/2104.14380>.
- [24] I. Dayan, H. R. Roth, A. Zhong and et al., "Federated learning for predicting clinical outcomes in patients with COVID-19," 2021. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/34526699/>.
- [25] Project MELLODDY, [Online]. Available: <https://www.melloddy.eu/>.
- [26] Executive Office of the President, "Executive Order 14017: America's Supply Chains," 24 February 2021. [Online]. Available: <https://www.federalregister.gov/documents/2021/03/01/2021-04280/americas-supply-chains>.
- [27] T. Suzumura, Y. Zhou, N. Baracaldo, G. Ye and et al., "Towards Federated Graph Learning for Collaborative," IBM Research, 2019. [Online]. Available: <https://arxiv.org/pdf/1909.12946.pdf>.
- [28] 117th Congress, "H.R. 3684 - Infrastructure Investment and Jobs Act," 2021-2022. [Online]. Available: <https://www.congress.gov/bill/117th-congress/house-bill/3684/text>.
- [29] Deloitte Center for Government Insights, "The future of United States infrastructure: A survey of infrastructure trends," Deloitte Insights, 2022. [Online]. Available: <https://www2.deloitte.com/us/en/insights/industry/public-sector/future-of-infrastructure.html>.
- [30] U.S. White House Briefing Room, "Press Briefing by Press Secretary Jen Psaki and Deputy NSA for Cyber and Emerging Technologies Anne Neuberger, March 21, 2022," Press Briefings, 21 March 2022. [Online]. Available: <https://www.whitehouse.gov/briefing-room/press-briefings/2022/03/21/press-briefing-by-press-secretary-jen-psaki-and-deputy-nsa-for-cyber-and-emerging-technologies-anne-neuberger-march-21-2022/>.
- [31] Executive Office of the President, "Executive Order 14058: Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government," 13 December 2021. [Online]. Available: <https://www.federalregister.gov/documents/2021/12/16/2021-27380/transforming-federal-customer-experience-and-service-delivery-to-rebuild-trust-in-government>.
- [32] National Institute of Standards and Technology, "AI Risk Management Framework: Initial Draft," 17 March 2022. [Online]. Available: <https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf>.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Diveplane Corporation

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to Request for Information on Advancing Privacy-Enhancing Technologies



July 2022

Office of Science and Technology Policy (OSTP)

Document Number: 2022-12432

Re: Diveplane Corporation's Response to Request for Information on Advancing Privacy-Enhancing Technologies, OSTP 2022-12432

Introduction and Overview of Diveplane's Response

Diveplane applauds the OSTP's attention to this important topic and welcomes the chance to respond to the RFI on the use of privacy-enhancing technologies. Privacy and AI fairness are at the heart of Diveplane's (www.diveplane.com) mission and are the focus of its product development. Diveplane is led by multiple computer science PhDs and a team that has decades of relevant AI, technical, business, and legal experience.

The aim of this response is to present a summary of the potential risks associated with using a single method or metric for privacy-enhancing technologies ("PETs"), and to present an alternative approach using a k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory that can significantly increase the potential to preserve privacy and accuracy while enabling the safe distribution of data. You will see below in our response the theme that current PETs, like differential privacy and k-anonymity can be unraveled or "attacked" in a way that can quickly erode the very privacy they are trying to protect. This is to say nothing of the possibility of accidentally synthesizing data that recreates PII or PHI, a mildly statistically improbable, but very grave, risk. As such, these single method or metric PETs can be false security blankets. We at Diveplane believe that a more comprehensive approach to creation of synthesized data is needed, including using a k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory, as well as building tools to check the privacy of the synthetic data created vis-à-vis the original data. We believe that PETs as we describe promote continued innovation in emerging technologies in a manner that supports human rights and shared values of democratic nations.

Responses to the questions from the RFI:

The bulk of our response will be directed to question number 2, *Specific technical aspects or limitations of PETs*. After our response to question number 2, we briefly respond to a few of the other questions.

2. Specific technical aspects or limitations of PETs:

2(A) BACKGROUND: PRIVACY IS NOT ONLY ABOUT CHECKING THE BOX TO MEET CURRENT LAWS

Ensuring privacy of data is about more than masking data, whether that masking is done with differential privacy or another method, such as those discussed herein. The risk of reidentification of masked data is real. Consider the issues Netflix had releasing anonymized data in 2007. “Anonymizing data still enjoys a good reputation despite an abundance of evidence that it is too easy to defeat. In 2007, Netflix offered a million-dollar prize to the first algorithm that could outperform their collaborative filtering algorithm. The dataset they supplied was anonymized, but one group de-anonymized it by joining it with information from the IMDb database. An anonymized database can happily expose the PII (personally identifiable information) by combining it with a PII data source and matching other criteria (so-called latent values).”¹

Further, it is important for every business to abide by the laws where it does business. The expansion of privacy laws and expectation of privacy from some consumers has led to class-action lawsuits seeking \$5000 per individual.² Data can be incredibly valuable but keeping around the original data for future analysis or innovation can lead to additional liabilities. The average cost of a data breach is just under \$4M USD per breach,³ with some of the worst breaches getting in the realm of billions of USD.⁴

In addition to the legal liabilities, there are also reputational and customer selection effects that can impact the bottom line. Having a strong privacy stance can attract additional customers.⁵ Conversely, not having thoughtful privacy practices can lead to undesirable self-selection of customers. Privacy can be a buffer to prevent misalignment of incentives between an organization and its customers. For example, an insurance company that collects data about how its customers drive may learn a lot, especially about how to price the auto insurance for different customer segments and improve profitability. However, if this data can directly tie back to customer behavior, then customers become incentivized in ways that are harmful to the business. Some safe drivers may think they will be unfairly discriminated against and so choose a competitor, whereas other drivers may attempt to game the system and perform undesired behaviors, like unnecessary excessive driving to lower premiums,⁶ or even second-guessing their driving decisions when faced with split-second life-or-death situations with pedestrians.⁷ Privacy safeguards, such as the use of synthetic data, can help ensure these misalignments of incentives are reduced or avoided, for example by softening the impact of bad luck on customers and reducing the chances of the customer having a bad experience, while still retaining the

¹ The fragility of privacy – can differential privacy help with a probabilistic approach? Neil Raden, 2020,

<https://diginomica.com/fragility-privacy-can-differential-privacy-help-probabilistic-approach>

² <https://www.forbes.com/sites/daveywinder/2020/06/03/google-chrome-privacy-lawsuit-could-you-get-a-5000-payout-incognito-mode-class-action/>

³ <https://www.ibm.com/security/digital-assets/cost-data-breach-report/>

⁴ <https://www.eweek.com/security/epsilon-data-breach-to-cost-billions-in-worst-case-scenario/>

⁵ <https://www.geekwire.com/2019/privacy-becomes-selling-point-tech-companies-apple-microsoft-leading-way/>

⁶ <https://blog.joemanna.com/progressive-snapshot-review/>

⁷ https://www.reddit.com/r/Insurance/comments/7ck6kh/dont_sign_up_for_progressive_snapshot_if_you_have/

insights, analytics, and general business model from data that can drive the organization to further success.

2(B) THE RISK TO DATA PRIVACY FROM MULTIPLE ATTACK VECTORS WHEN USING COMMON PETS

Given the importance of privacy, it is important to select an effective solution to ensure privacy of the data while maximizing utility, insights, and accuracy. It can at first seem appealing to select a technique that makes intuitive sense regarding privacy, like **k-Anonymity**, where you ensure that the data released has at least some k records that are sufficiently similar and not uniquely identifying. Or it may seem appealing to apply a technique that is well-known, like **differential privacy** and be done. However, using just one of these anonymity or privacy techniques without examining the different ways the data and the privacy can be attacked can yield unintentional, unknown, and potentially vast privacy leaks. For example, the principle of differential privacy is mathematically solid, but when used alone, it is easy to make egregious mistakes without knowing it.

Further, while well-known techniques, like k-anonymity or differential privacy, can measure or ensure a certain aspect of privacy, there are no globally accepted standards that can tell you how good your privacy actually is or that can characterize the risks of reidentification. This leaves key governance stakeholders to assert their own definition of “good” privacy. The risk of poor privacy practices is a long-term existential risk, especially as new attacks on privacy and personal autonomy continue to emerge. Masked or privatized psychographic, activity, location, financial, or preference data points are often considered to be nonidentifying in some contexts and organizations today. These “anonymized” data sets may be considered safe to distribute, and in some cases they are. However, as data-based insights continue to escalate in ubiquity and efficacy, privacy models based on a one-to-one mapping between the original data and the outputs of analysis and data sharing have potential risks when combined with other data sources, known as auxiliary data, regardless of whether they were anonymized. The ability to combine these pieces of data, which may be unforeseen when the analysis is performed or data released, increases the potential for attack vectors to privacy. These privacy leaks can even yield information far beyond what was intended or included, such as mental health patterns suggested by a user’s historical patterns.⁸

Here are some discussions from the literature:

“The rapid decrease in the sequencing technology costs leads to a revolution in medical research and clinical care. Today, researchers have access to large genomic datasets to study associations between variants and complex traits. However, availability of such genomic datasets also results in new privacy concerns about personal information of the participants in genomic studies. Differential privacy (DP) is one of the rigorous privacy concepts, which received widespread interest for sharing summary statistics from genomic datasets while protecting the privacy of participants against inference attacks. However, DP has a known drawback as it does not consider the correlation between dataset tuples. Therefore, privacy guarantees of DP-based

⁸ E.g., https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2483426

mechanisms may degrade if the dataset includes dependent tuples, which is a common situation for genomic datasets due to the inherent correlations between genomes of family members.”⁹

“Four spatio-temporal points are enough to uniquely identify 95% of individuals.”¹⁰

“Our study shows that sharing anonymized location data will likely lead to privacy risks and that, at a minimum, the data needs to be coarse in either the time domain (meaning the data is collected over short periods of time, in which case inferring the top N locations reliably is difficult) or the space domain (meaning the data granularity is strictly higher than the cell level).”¹¹

THEORETICAL EXAMPLE 1: A restaurant chain collects customer data to be analyzed by a consumer analytics firm as part of a preference-based marketing campaign and applies differential privacy. Unrelatedly, the consumer analytics firm also collects data from a separate group that collects data about the retail area around the restaurant, and this group only uses data masking to remove apparently identifying features. Individually, these data sets may pose little risk to privacy leaks. However, because the differential privacy was applied per transaction and not per customer, the customers who frequently visited the restaurant had a vastly eroded privacy. Further, the retail analytics firm didn’t mask out certain fields for one of the boutique retailers, which sometimes contained names and addresses. When the restaurant data is coupled with the retail data by the consumer analytics firm, it is possible to not only match up and reidentify most of the frequent customers who also visited this boutique retailer, but to obtain the person’s name and address for some of those records. A customer of the third party had poor data security and the analytics are leaked to the internet. One of the reidentified restaurant and boutique store’s patrons was a high-profile individual who had a picture taken at the restaurant. A savvy internet follower combines this picture with the leaked data and posts private information about the high-profile individual, who then files suit against all three firms. Even though this specific combination of events sounds rare, when enough data is collected and combined, such rare events become common and can have lasting horrible effects for individuals and firms.

Each individual privacy model bears several weaknesses when pitted against various known attack vectors, so no single privacy model or single privacy measurement is sufficient to defend and identify vulnerabilities in the presumed privacy of data. Regulatory standards like CCPA, GDPR, and COPPA are all, generally speaking, liability regimes and do not outline safe harbor practices to either shield you from liability or to protect your data. As such, compliance with any particular privacy regime provides incomplete defense against experienced actors and increasing volumes and types of data collection. The following table details a number of popular privacy techniques and measures and the types of attacks that may be effective against them.

⁹ Almadhoun et al., *Bioinformatics*. 2020, <https://academic.oup.com/bioinformatics/article-abstract/36/6/1696/5614817>

¹⁰ de Montjoye et al., *Sci Reports*, 2013, <https://www.nature.com/articles/srep01376>

¹¹ Zang & Bolot, *ACM MOBICON*, 2011, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.651.44&rep=rep1&type=pdf>

Here is a summary of *known* attack vectors against PETs¹²:

Privacy Model	Attack Model			
	Record Linkage	Attribute Linkage	Table Linkage	Probabilistic Attack
k -Anonymity	✓			
MultiR k -Anonymity	✓			
ℓ -Diversity	✓	✓		
Confidence Bounding		✓		
(α, k) -Anonymity	✓	✓		
(X, Y) -Privacy	✓	✓		
(k, e) -Anonymity		✓		
(ϵ, m) -Anonymity		✓		
Personalized Privacy		✓		
t -Closeness		✓		✓
δ -Presence			✓	
(c, t) -Isolation	✓			✓
ϵ -Differential Privacy			✓	✓
(d, γ) -Privacy			✓	✓
Distributional Privacy			✓	✓

2(C) ENHANCING PRIVACY THROUGH A COORDINATED, ROBUST APPROACH, USING A K NEAREST NEIGHBOR PET THAT INCLUDES ROBUST, PROBABILITY-THEORY-DRIVEN KERNELS WITH INFORMATION THEORY

To address the potential problems from using single privacy models it is possible to use multiple privacy models to significantly increase the robustness of the of the synthetic data while maintaining high levels of accuracy. We at Diveplane believe that the best approach is to mathematically unify a variety of privacy enhancing techniques, thus enabling a k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory. This technique enables maximum entropy noise, achieves various forms of differential privacy, and simultaneously utilizes the sparse space of multidimensional data to only synthesize new data that is sufficiently “surprising” (by the information theoretic definition of surprisal) relative to any original data. By doing this, one can generate new data that maintains both the underlying distributions of original data and accuracy under analytic techniques, while also allowing privacy to be maintained such that none of the synthetic data is too close or similar to the original data, all while not relying on a one-to-one relationship between the original data and the synthesized data. Not only must these techniques be used at generation, but we also believe the generated data should also be tested, using a data quality tool evaluation suite, after synthesis to make sure it maintains the expected privacy. See Section 2(D) below.

At Diveplane, we believe that those hoping to truly advance privacy using PETs should synthesize data using a PET that can create a verifiable synthetic ‘twin’ dataset with the same statistical properties of the original data, but without including the real-world confidential or personal information. This results in a data set rich in value but with no risks of re-identification. We believe this can be accomplished with a k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory, and that:

¹² Fung et al., ACM Computing Surveys, 2010, <https://dl.acm.org/doi/abs/10.1145/1749603.1749605>

- These synthetic data sets are far more robust, accurate, and safe compared to other masking and privacy techniques that may still be susceptible to reidentification, based on our experience with thousands of data sets.
- These synthetic data sets help companies navigate through national international privacy laws including GDPR, CCPA, CPRA, and HIPAA.
- These PETs can be used to create datasets that look, act, and feel realistic for the purposes of data modeling and analysis, but do not contain information to identify an actual person from the original dataset.
- These PETs can create synthetic data from all structured data, including time series data and relational databases.
- These PETs should include a data quality tool that enables users to identify the levels of privacy and accuracy of the synthetic data vis-à-vis the original data.

Benefits over differential privacy of a k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory

Differential privacy is a pillar of modern privacy enhancing technologies. It is a mathematical technique that ensures that, for a given probability as modulated by a “privacy budget”, the statistical results of certain types of analysis of a database with an individual’s record in it should in expectation be indistinguishable from the statistical results as if the individual’s record were not in the data. This creates a strong plausible deniability for analysis and generally prevents an individual from proving that they were in the data. As strong as differential privacy is, it is equally as easy to misuse in practice. If data has multiple records for individuals, or there are some strong correlations in the data such as family records or geographic information data, the actual privacy achieved may be far less than planned. Further, the typical applications of differential privacy either require upfront planning of analytics or use ML systems that train via a differential privacy budget but may extract private information into the model via correlations.¹³

Further, because each feature or combination of features could potentially be statistically identifying when combined with future data, all fields must be considered as potentially sensitive and have some form of differential privacy applied. If differential privacy is used in this manner, it leads to noisy, unusable data sets. An additional concern, though rare, is that certain queries with differential privacy can make one person appear like another, so synthetic data solutions that only rely on differential privacy may have more than one chance to synthesize data that appears almost exactly like the original data.

A k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory applies mechanisms that are differential privacy to all relationships in the data, not just the sensitive fields. It scales the noise based on the predictability in the data, ensuring that privacy and anonymity are maintained in every area, from dense clusters to far outliers. This combination of differential privacy

¹³ E.g., Hitaj et al., ACM SIGSAC, 2017 (<https://arxiv.org/pdf/1702.07464.pdf>) and Papernot et al., ICLR 2017 (<https://arxiv.org/pdf/1610.05755.pdf>)

techniques acts as a sensitive data cross-shredder, while making sure it still generally follows the underlying distribution of the data.

2(D) PETS SHOULD NOT STOP AT SYNTHETIC DATA GENERATION, BUT SHOULD PROVIDE A WAY TO CHECK THE PRIVACY OF THE GENERATED SYNTHETIC DATA VIS-À-VIS THE ORIGINAL DATA

Benefits of using an anonymity preservation check as a data quality measure after creating synthetic data with PETS

We at Diveplane believe that PETS should include an anonymity preservation check, which is an approach that finds for each data point in the synthesized data set, the closest corresponding data point of the original dataset, and measures how easy it would be for an original data point to be recognized in the synthetic data. We believe that a k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory can generate data points that are just outside of the certainty manifold of predictability from the original data, ensuring that it doesn't generate anything too predictable within the data set.¹⁴ A k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory can find the nearest data so that a user can audit and determine what is the closest synthesized datapoint to any original datapoint, thereby enabling the user (i.e. data scientist) to check whether the synthesized is too close to the original data. This measurement is a ratio between 1) the distance between a synthetic data point and the closest original data point and 2) the average or minimum distance between a data point and its closest neighboring data point in the original data, among the locally relevant data points and also all global data points. Distance in the can be scaled according to the data's range for each feature. A ratio greater than 0.5 for minimum distance ratio is a good indicator that privacy being preserved even in the worst case, as it is possible that any given data point is sufficiently could potentially be mistaken for another. A ratio of 1.0 or higher means that privacy is being preserved well enough such that in the worst case, an individual record from the synthetic dataset looks at least as different as any two different cases in the original dataset, and any higher value means that privacy is even stronger. The indices of the closest matches are reported so that an auditor could inspect the data to ensure that privacy is maintained.

THEORETICAL EXAMPLE 2 – Using anonymity preservation to check privacy for synthetic data. Consider a theoretical company that uses differential privacy with a GAN (generative adversarial network) to produce synthetic data, and distributes it to several parties. Even though the data is synthetic and used differential privacy properly, somebody believes they are recognizable in the data and files a lawsuit against the company. Because a dozen other people recognize themselves in the millions of synthetic data points, this lawsuit turns into a class action lawsuit. Had the company used a model that included anonymity preservation checking, the system would have inherent guardrails that would deny any synthetic data from being created that did not meet the minimum distance criteria.

Benefits of using entropy comparison as a data quality measure after creating synthetic data with PETS

¹⁴ This can be easily parameterized for workflow needs. For example, if there are default data point that is repeated frequently and identically in the data but does not match an individual and thus won't leak privacy, these can be optionally included.

Another technique that is often used is entropy comparison. Entropy comparison compares the entropy and KL-divergence to provide a measure of disorder/surprisal within a data set. Entropy of a dataset is a measure of how compressible or predictable a given feature is, which is generally correlated with the skewedness of the class distribution. Similar values for original and generated entropy are an indication of similar class-wise distribution. KL-divergence values measure how similar the distributions of the generated and original dataset are.

A dataset which fails this measure is an indicator that the original data is very noisy and unpredictable or that the privacy model/algorithm tools being applied is underperforming or not working. If the privacy system adds significant noise without degrading accuracy, the original data set has a strong signal and with predictable values.

Benefits of using equivalence class measurements (k-Anonymity, l-Diversity, and t-Closeness) as data quality measures after creating synthetic data with PETs

The three equivalence class measures track equivalence classes, records in which many people share one or more characteristics (zip code, ethnicity, eating preference, etc.). For privacy, one person should not be an island. Any outlier, the only or one of very few people in an equivalence class, may be discoverable.

These measures are designed more specifically for anonymized or masked data rather than synthetic data. They are still a useful heuristic for exploring characteristics of the data set. These concepts are closely related to how a k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory adds noise; in addition to the other methods mentioned, a k nearest neighbor PET that includes robust, probability-theory-driven kernels with information theory adds noise from the global distribution into every area of the data to ensure that these metrics are met in expectation, even if the metrics were not built to apply to synthetic data. For example, if a synthetic data point has insufficient anonymity under k-Anonymity, but it does not remotely correspond to any original data point, then the only privacy loss is that of a fictitious data point.

k-Anonymity measures the number of values within an equivalence class. If an equivalence class contains fewer than k values, that is considered a violation. k-anonymity is included to identify cases where there are low number of observations in an equivalent class (or outliers). It is useful for identifying cases where PII may be leaked if the equivalent class is unique. This protects the user from sharing data where a real person is potentially identified from auxiliary information.

l-Diversity measures the entropy within an equivalence class. If an equivalence class contains fewer than l unique values, that is considered a violation. This metric is not designed for determining privacy for synthetic data but is still a useful heuristic for certain characteristics of the data set. This metric is included to provide a measure of the diversity within a subpopulation (equivalence class) and ensure each type / outcome is represented in the class. It is useful for detecting the prevalence of an outcome throughout the dataset (e.g., is noise injected everywhere?) and protects against the scenario where a group is identified through a unique set of features criteria.

t-Closeness measures the distance t between distributions of an equivalence class and the global distribution. As t shrinks, a data set can be considered more private. This metric is not designed for determining privacy for synthetic data but is still a useful heuristic for certain characteristics of the data

set. This metric is a measure of the distribution of a subpopulation vs. the overall population. It is useful for detecting cases where a group may be identified from a set of unique feature values (equivalence classes).

Responses to additional questions from the RFI

Question 1. *Specific research opportunities to advance PETs.*

We at Diveplane believe the government has the opportunity to advance PETs by funding development of k nearest neighbor PETs that include robust, probability-theory-driven kernels with information theory as well as post-synthesis data quality tools that help data handlers assess the privacy of the synthetic data that they have created.

Question 2. *Specific technical aspects or limitations of PETs:* See extensive response above.

Question 4. *Specific regulations or authorities that could be used, modified, or introduced to advance PETs.*

As we have previously expressed, <https://www.nextgov.com/ideas/2020/09/need-unified-data-protection-us/168643/>, we believe that it would increase the certainty, understandability, and compliance by data-handling companies if the US had federal privacy regulations instead of a scattershot of laws and enforcements at the state and local levels. Further, and while we understand that creating safe harbors is difficult, the inclusion of safe harbors in the federal privacy regulation that outline approaches that companies could take to comply with the law would allow the vast majority of companies to comply by simply using these approaches.

Question 7. *Risks related to PETs adoption:* Our response to #2 also includes extensive discussion of the risks related to common PETs, such as differential privacy.

Question 8. *Existing best practices that are helpful for PETs adoption:* As we note in our response to #2, we believe PETs should have accompanying data quality tools to help assess the privacy of the generated synthetic data *after* it has been generated.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Duality Technologies

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Duality Technologies Response to OSTP Notice of Notice of Request for Information on Advancing Privacy-Enhancing Technologies

Dr. Ahmad Al Badawi, Senior Scientist
Ronen Cohen, VP Strategy
Dr. David Bruce Cousins, Director Duality Labs
Dr. Nicholas Genise, Scientist
Dr. Kurt Rohloff, CTO and Cofounder

Organization Type: Industry

July 6, 2022

Executive Summary

Duality Technologies Inc., a US-based corporation, thanks the OSTP for the opportunity to contribute the following comments regarding the request for information on advancing Privacy-Enhancing Technologies (PETs).

Duality is a leading PETs provider, enabling organizations to collaborate on personal data or other sensitive data without anyone but the providing data controller having unencrypted access to the data. Our software platform leverages several PETs, including Homomorphic Encryption (HE), Secure Multiparty Computation (SMPC), Federated Learning (FL), and more, and utilizes PALISADE, an open-source, standards-compliant Homomorphic Encryption library built with US Government funding. Duality's founding team is comprised from world renowned cryptographers, including Turing Award winner Prof. Shafi Goldwasser, and data science experts. We enable sharing of data with one or more entities processing the data in a robustly encrypted, fully private manner that protects the data from access by the processors or any other party except for the data controller who encrypts the data.

1. Specific research opportunities to advance PETs

There are several research opportunities that can be addressed to advance the current state of PETs. Following is a non-comprehensive list of opportunities, however we think the below are both immediate and crucial to the advancement of PETs.

- **Utility, privacy, and performance tradeoffs**

There is almost always a tradeoff between data utility, data privacy, and performance that should be considered in defining the best solution for a use case in question. Maximum utility requires full access to the original data without any added noise, masks or concealments (such as encryption), but this can violate data privacy requirements. Thus, several PETs apply some transformation to the

original data before sharing it for computation. This transformation can affect the extent to which data can be utilized. For instance, differential privacy-based solutions transform data by adding noise that renders the data subjects hard to re-identify, yet the transformed data remains useful for performing some computations like aggregate queries. With this technology, the added noise can affect the precision of some queries, which creates complications for high-precision applications. On the other hand, there exist PETs that apply noise-free transformation on the data (such as homomorphic encryption or multi-party computation), while allowing arbitrary computation to be performed at almost zero loss of precision. While initial realizations of these PETs were impractical, they are currently considered practical, and in some particular cases, very efficient. Further research is still needed in this domain to make them more efficient for arbitrary computations. As an example, this is an active research area for DARPA, who have funded several PET based research programs in the last ten years, including the Cooperative Secure Learning (CSL) program for PET-based machine learning, and the Data Protection in virtual Environments (DPRIVE) program to accelerate the performance of fully homomorphic encryption (FHE) with novel hardware, both of which have Duality as a participant [1] [2].

- **Quantifying privacy for risk assessment**

One of the main issues in most PETs is their lack of rigorous security proofs. In general, such proofs stem from the computational hardness of well-defined and well-studied mathematical problems like integer factorization, discrete logarithm, and some lattice problems. Cryptographic systems based on these techniques increase the effort an attacker needs to compromise the security of the system to a point where unrealistically large computers would need to run for years or even decades to crack them. In non-cryptographic PET systems, this effort cannot be quantified in the same way. Despite this complexity, quantifiable analysis is necessary for many of today's data-centric applications, as well as to satisfy the requirements of many information and data security organizations.

The literature is rich with several studies and incidents that showed the weakness of approaches with non-quantifiable security, and how they should be avoided [3] [4] [5] [6]. Yet, these PETs are still being used nowadays in several applications and their usage is even recommended in some data protection acts [7]. In contrast, more robust and concrete PETs such as FHE and SMPC have rigorous mathematical proofs of quantifiable security. The security level of these technologies is well defined and is usually expressed in the maximum number of elementary operations an attacker must do to certainly compromise the security of the underlying system. Therefore, an area of paramount importance to the advancement of PETs is to define and quantify different degrees of privacy in terms of the effort needed to compromise them, standardize these definitions, and focus research on applying them to non-cryptographic PETs, an effort that will make quantitative risk assessments more feasible, reliable, and more relatable to non-technical users.

- **Privacy in light of emerging technologies**

Research should be furthered to advance existing PETs while addressing arising challenges brought on by emerging data-driven technologies such as 5G/6G, Artificial Intelligence, Big Data, IoT, and Blockchains. Take Artificial Intelligence, for example. Over the past decade, a number of scientific branches in AI, such as machine learning and deep learning, have made huge leaps in numerous data analytics tasks. Fundamentally, these systems are trained on huge amounts of data used to develop models that can make informative predictions or take actions when deployed. Although existing data protection and PETs can be used to enforce privacy in the training phase, it has been found that this can be more challenging when it comes to the deployment phase. While still secure enough to prevent unauthorized access to the underlying data or model, attacks such as membership inference allow passive adversaries to learn whether a deployed model has been trained on a designated data subject or not [8]. This is an example of the kind of attack that can be difficult to predict while developing new PETs that can compromise both privacy and security.

- **The right PETs for the right job**

We can comfortably say that there is no one-size-fits-all PET that can solve all privacy problems, and as such, this requires a thorough analysis of any targeted use case in order to define the best viable option. Moreover, a PET that might work in one application may not be suitable for other applications. In fact, some use cases might require the employment of multiple PETs within one solution. This area has recently been an area of attention in the PET space; for instance, SMPC and HE have been used to devise more efficient multi-party HE schemes [9] and differential privacy has been married with HE schemes to alleviate some limitations of HE-based inference applications [10].

2. Specific technical aspects or limitations of PETs

As stated in the previous section, different PETs are often best suited for different applications, and may in fact work together in a complementary manner. The two PETs with broadest applicability in terms of functionality are FHE and multiparty computation, since they can securely compute mathematical functions on encrypted data. Therefore, one can build systems to implement any computer algorithm securely using FHE, SMPC, or a combination thereof. FHE allows all portions of a computation to be done on data at rest through computationally intensive cryptographic transformations, whereas SMPC requires passing multiple messages between all parties throughout the computation. Historically, both of these approaches had inefficiencies (compute-limited or bandwidth-limited) that impacted the feasibility of using these PETs for certain applications, although recent advances have improved performance by orders of magnitude. Ultimately, FHE combined with hardware acceleration could be an effective and provably secure solution

for most applications despite the initial expense of specialized hardware. Additionally, the combination of FHE and SMPC together is a new research area where the best of both approaches can be combined for better performance. Both areas are worthy of further research, since their security is the strongest among PETs.

Differential privacy (DP) is a statistical encoding that requires the input to conform to a known distribution *a priori*. This is a strong restriction on the input, but makes it efficient since the encoding is usually simple. Unfortunately, if the full statistics of the data are not known in advance, DP could make it difficult to uncover previously unknown relationships in the data.

A trusted execution environment (TEE) is a physically isolated execution environment, usually within a CPU. TEEs are very efficient, and multiple generations of the technology have been developed by various CPU vendors. However, active research has been able to uncover numerous security flaws due to side-channel attacks, so caution needs to be used when implementing systems with this technology in order to avoid these issues. [11]

Other forms of PETs, such as federated learning (FL), synthetic data, and anonymization have significant security drawbacks and are unlikely to pass a rigorous standardization process (e.g., NIST) for broad functionalities. They can, however, carefully be used in limited settings, or alongside other PETs. For example, FL suffers from a reverse-engineering flaw since the model is known to all collaborating parties, and synthetic data or anonymization eliminates the possibility for correlation studies among different datasets. This is why the combination of FL and FHE has been a research area for DARPA and others (for example, as part of DARPA's Cooperative Secure Learning (CSL) program, which was referenced above).

As mentioned previously, a limitation across all PETs is the lack of a shared set of threat models and a shared, quantifiable measure of security. This can be overcome with more research and a standardization thrust. Lastly, the best long-term solution for most applicable problems may be achieved by mixing PETs under these unified models.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

There are numerous challenges and use cases in both the public and private sectors that would benefit from the adoption of PETs. PETs are designed to enable the use of data and derivation of insights in challenging circumstances, such as when competitive or proprietary data concerns exist; when regulated data must be used (e.g., HIPAA-protected data, or financial data); when there are privacy laws and frameworks that must be adhered to (e.g., CCPA); when cross-border data transfers are needed in a way that respects the law of each country (e.g., when data localization / data residency laws exist), and many more. Some examples are below:

- Public Sector

- Investigating crimes while protecting investigation integrity (e.g., for money laundering, counter terror financing, tax fraud, corruption, and more)
- Secure distributed control of Critical Infrastructure, Networks, IOT
- Processing and/or distributing sensitive information in a military or security context
- Collaborating with the private sector in a Public/Private Partnership setting
- Cross-border collaboration with international counterparts and allies
- Collaborating and processing data in Zero Trust environments
- Analyzing sensitive data for public health scenarios
- Understanding provenance and quality of manufactured items to enable better maintenance / fleet management, and protecting supply chain and logistics
- Private Sector - Healthcare, Pharmaceuticals, and Life Science
 - Accessing, linking, enriching, and analyzing multi-center data to better understand and develop treatments and drugs
 - Leveraging Real World Data for clinical trials and drug development
 - Clinical-Genomic Analysis - Joining genetic information with clinical data to find correlations
 - Health and lifestyle insights leveraging data from wearables
 - Conducting patient recruitment for trials
- Private Sector - Financial Services
 - Cross-institution collaboration to prevent, detect, and investigate financial crimes (including money laundering, terrorist financing, and predicate offenses like fraud, cyber crimes, human trafficking, etc.)
 - Cross-border collaboration (internal to a single institution) on financial crimes, customer service, marketing, product development, etc.
 - Third party collaboration for personalized marketing and service development (e.g., for granting credit)
 - Offering personalized rates and services based on consumer activities (e.g., insurer offering better rates based on driving habits)
 - Benchmarking across organizations (e.g., around cyber security posture)
- Private Sector - Telecom
 - Securing 5G and other critical infrastructure software to run on untrusted hardware (note, DARPA is funding research in this area) [12]
 - Acting as a secure access data broker (for publish/subscribe scenarios)
- Private Sector - Other
 - Collaborating across retailers to better manage inventory and supply chain
 - Collaborating across manufacturers and vendors to increase quality
 - Enabling secure computations in the cloud

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs

This response combines both the question on regulations (#4) as well as on specific laws (#5). There are several aspects to consider regarding these topics.

First, regulators and lawmakers should define priority areas for collaboration and give organizations a “safe harbor” for collaboration. Good candidates for priority areas may be “public good” use cases like stopping fraud and cyber crimes, money laundering, terrorism, and even enhancing pharmaceutical or medical research.

Second, regulators and lawmakers can be more forthcoming on how and when PETs can enable collaboration in compliance with privacy and industry-specific laws. This also includes a few different facets:

- Harmonizing approaches and interpretations across regulators. As an example, the US alone has over a dozen different financial services regulators at the federal level, and each state often has its own authorities as well. A single financial institution may be regulated by several of them, and all may interpret the law differently, or issue guidance which is unharmonized. This makes it challenging to remain compliant. As an example, in December 2020, FinCEN issued a clarifying remark on information sharing under the PATRIOT Act section 314(b), which they interpreted as including “predicate offenses” to money laundering and terrorist financing (like Fraud, for example) [13]. Unfortunately, other regulators have not formally issued this same guidance, and as such, fraud information sharing under 314(b) amongst US financial institutions remains under-utilized.
- Harmonizing approaches and interpretations across borders (e.g., with US Allies). Many industry stakeholders have multinational operations, and a lack of harmonization across the jurisdictions in which they operate adds to complexity and cost.
- Encouraging innovation and collaboration. Regulators should consider allowing responsible organizations to freely innovate. Organizations that meet criteria for this should be allowed to try new and emerging technologies, with the input and observation of regulators, without any added risk.
- Clarifying, for any given industry, what data can be shared, under what circumstances, and how.
- Clarifying how and when PETs can be used in the context of both cross-industry privacy laws (e.g., CCPA) as well as industry-specific ones (e.g., HIPAA, Gramm-Leach Bliley, etc.), and explaining how to leverage PETs in a legally-compliant manner. For example, data de-identification is often used in the context of HIPAA. This comes with major drawbacks, including an inability to link data sets as well as a reduction in accuracy and precision of results. If it were clearly understood which PETs could be used in here in a regulatory compliant manner, the speed and accuracy of analyses would be significantly improved, which in turn would positively impact patient outcomes.

Third, it is important to understand that collaboration comes at a cost - for an organization to engage in collaboration and use PETs, they may have to train / retrain resources, change business processes, purchase and implement technology, and integrate technologies with legacy systems. To this end, regulators and lawmakers must understand that without laws which simply encourage collaboration (rather than mandate

it), will yield less participation than desired. One can again look to the USA PATRIOT Act as an example. Whereas Section 314(a) is mandatory and has full industry participation, Section 314(b) is not. As such, it is often the case that participating institutions devote less resources to 314(b) information sharing requests, meaning that requesting institutions may not receive timely responses. One reason for this is because 314(b) is not mandated, and participation is not incentivized in any way (e.g., by aiding in supervisory exam scores, which today only look at technical compliance to the law). In summary, regulators and lawmakers should consider both regulatory mandates and safe harbors, as well as incentives, to encourage the use of PETs in any given industry.

5. Specific laws that could be used, modified, or introduced to advance PETs

See #4 above.

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

In addition to the legal and regulatory considerations discussed previously, other supporting mechanisms could include:

- Subsidizing the testing and advancement of PETs in priority areas in both the public and private sector, similar to what the Singaporean Government does when it identifies technologies that it would like the private sector to leverage
- Simplifying government procurement processes for priority technologies in priority areas, and/or creating testing “sandbox” environments for both the public and private sectors to utilize, similar to what the ICO does in the United Kingdom
- Challenge prizes (similar to the US/UK PETs Challenge Prize)
- Techsprints (similar to the recent Anti-Corruption Tech Sprint, where Department of State and Department of Treasury were involved)

7. Risks related to PETs adoption

Some of the main risks related to PET adoption are that today’s stakeholders may not know how to adequately evaluate PETs for security, interoperability, and performance. Given that these are relatively new technologies tackling the most sensitive of data and analytics, using them without the adequate evaluation could introduce significant risks, and is exactly why open-sourced, standards-compliant security and interoperability, as well as a well-understood evaluation methodology, are so important.

As related to FHE, there are open-sourced encryption libraries like [PALISADE and its second generation version, OpenFHE](#). PALISADE supports the major publicly known and accepted encryption schemes, and has been publicly released and available for inspection for several years. It has wide use and adoption in the public and private sector, as well as academia. This is indicative of trust in the encryption technology. A key enabler

of this adoption is that all the source code is available for third party security inspections. Other PETs and some other homomorphic encryption providers do not open source their technology, meaning it is difficult to understand and verify their technological and security claims and capabilities. This creates risk, as related to the adoption of PETs.

Connected to this topic is that of security and interoperability standards. The leading standards body in this space is HomomorphicEncryption.org, which was co-founded by IBM, Intel, Microsoft, and Duality, and focuses on security and interoperability of Homomorphic Encryption. Standards written by this group have been leveraged by the ISO to create their draft Homomorphic Encryption standards. Not all PETs have such standards, which increases the risks and costs associated with them.

Finally, there has not historically been a generally accepted approach to evaluate and benchmark PETs. To this end, the HEBench Organization was recently founded by Duality, Intel, Deloitte, IBM, Microsoft, and others to assist in this matter by addressing the lack of structured and consistent measurements of full stack performance. This again helps market stakeholders better understand the PETs they are evaluating, which reduces risk.

8. Existing best practices that are helpful for PETs adoption

Many of the best practices which are helpful for PETs adoption all revolve around standardization. For example, some best practices helpful for PETs are:

- Making code and algorithms publicly available for cryptanalysis.
- Identifying and formalizing various explicit security threat models across multiple PETs.
- Developing with standardized quantifiable (mathematically based) security models and levels of security across PETs.
- Mutual understanding and agreement of relevant regulations and/or legal basis to collaborate on data, and
- Explicit regulatory government approval and even participation (e.g., in Estonia, regulators have actively been involved in efforts to facilitate info sharing to fight money laundering, which has yielded positive results).

9. Existing barriers, not covered above, to PETs adoption:

The main barriers have been covered above.

10. Other information that is relevant to the adoption of PETs

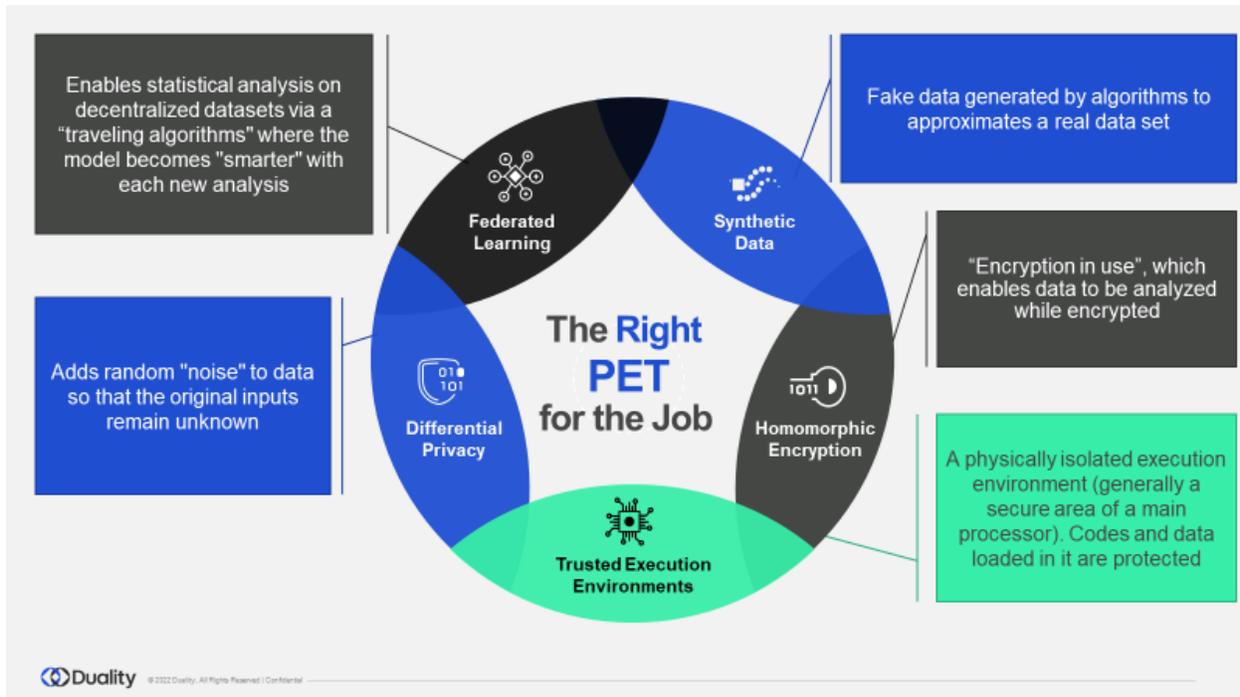
The key points relevant to the adoption of PETs has been covered above.

References and Extra Material

- [1] Baron, Dr. Joshua. Cooperative Secure Learning (CSL). DARPA, <https://www.darpa.mil/program/cooperative-secure-learning>.
-
- [2] DARPA Selects Researchers to Accelerate Use of Fully Homomorphic Encryption. DARPA, March 2021, <https://www.darpa.mil/news-events/2021-03-08>.
-
- [3] Rocher, L., Hendrickx, J. M., & De Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, 10(1), 1-9.
-
- [4] Sweeney, L. (2013). Matching known patients to health records in Washington State data. arXiv preprint arXiv:1307.1370.
-
- [5] Sweeney, L., Abu, A., & Winn, J. (2013). Identifying participants in the personal genome project by name (a re-identification experiment). arXiv preprint arXiv:1304.7605.
-
- [6] Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, 671(2000), 1-34.
-
- [7] 2018 reform of EU data protection rules. European Commission. May 2018. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>
-
- [8] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks against machine learning models. In 2017 IEEE symposium on security and privacy (SP) (pp. 3-18). IEEE.
-
- [9] Mouchet, C., Troncoso-Pastoriza, J., Bossuat, J. P., & Hubaux, J. P. (2020). Multiparty homomorphic encryption from ring-learning-with-errors. *Cryptology ePrint Archive*.
-
- [10] Li, B., Micciancio, D., Schultz, M., & Sorrell, J. (2022). Securing Approximate Homomorphic Encryption Using Differential Privacy. *Cryptology ePrint Archive*.
-
- [11] van Schaik, Stephan, Andrew Kwong, Daniel Genkin, and Yuval Yarom. "SGAxe: How SGX fails in practice." 2020. <https://sgaxe.com/files/SGAxe>. (2020).
-
- [12] Improving 5G Network Security. DARPA, February 2020, <https://www.darpa.mil/news-events/2020-02-05>.
-
- [13] Section 314(b) Fact Sheet. Financial Crimes Enforcement Network, Dec. 2020, <https://www.fincen.gov/sites/default/files/shared/314bfactsheet.pdf>.

Appendices

Appendix A – Privacy Enhancing Technologies: Definitions



Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Electronic Privacy Information Center (EPIC)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

COMMENTS OF THE ELECTRONIC PRIVACY INFORMATION CENTER

to the

White House Office of Science and Technology Policy

RFI on Advancing Privacy-Enhancing Technologies

87 Fed Reg. 35,250

July 8, 2022

The Electronic Privacy Information Center (EPIC) submits these comments in response to White House Office of Science and Technology Policy’s (OSTP) June 9, 2022 Request for Information on Advancing Privacy-Enhancing Technologies.¹ OSTP is interested in privacy preserving technologies for data sharing and analytics including “secure multiparty computation, homomorphic encryption, zero-knowledge proofs, federated learning, secure enclaves, differential privacy, and synthetic data generation tools.”² Through this RFI, OSTP is seeking guidance on adopting “a national strategy on privacy-preserving data sharing and analysis,” including “Federal laws, regulations, authorities, research priorities, and other mechanisms across the Federal Government that could be used, modified, or introduced to accelerate the development and adoption of PETs.”³

EPIC is a public interest research center in Washington, D.C. EPIC was established in 1994 to focus public attention on emerging privacy and related human rights issues and to protect privacy,

¹ 87 Fed. Reg. 35,250, *available at* <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>.

² *Id.* at 35,251.

³ *Id.*

the First Amendment, and constitutional values. EPIC has a longstanding interest in federal efforts to develop privacy-enhancing technologies and regularly comments on proposed federal planning efforts at the intersection of technology and privacy.⁴ EPIC has also repeatedly intervened to ensure proper privacy protections, including differential privacy, are used on the Census.⁵

EPIC urges OSTP to (1) prioritize the adoption of differential privacy in its national research plan and (2) direct federal agencies to shift and increase funding toward the development of privacy-enhancing technologies.

I. The federal government should invest more in differential privacy instead of relying on deidentification techniques that do not work.

Traditional techniques for deidentifying and anonymizing datasets are ineffective and do not account for the ease with which information in multiple datasets can be combined to reidentify individuals. Differential privacy is the intentional injection of controlled amounts of statistical noise into data products to provide a mathematical guarantee of privacy while preserving the ability to use

⁴ See, e.g., Comments of EPIC, Public and Private Sector Uses of Biometric Technologies, Office of Sci. & Tech. Policy (Jan. 15, 2022), <https://epic.org/documents/epic-comments-to-ostp-on-public-and-private-sector-uses-of-biometric-technologies/>; Comments of EPIC, Artificial Intelligence Risk Management Framework, Nat'l Inst. of Standards & Tech. (Aug. 18, 2021), <https://epic.org/documents/regarding-the-artificial-intelligence-risk-management-framework/>; Comments of EPIC, Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource, Office of Sci. & Tech. Policy & Nat'l Sci. Found. (Oct. 1, 2021), <https://epic.org/wp-content/uploads/2021/10/EPIC-Comment-NAIRR-Oct2021.pdf>; Comments of EPIC, Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning, Comptroller of the Currency et al., (July 1, 2021), <https://archive.epic.org/apa/comments/EPIC-Financial-Agencies-AI-July2021.pdf>; Comments of EPIC, Solicitation of Written Comments by the National Security Commission on Artificial Intelligence, 85 Fed. Reg. 32,055 (Sep. 30, 2020), <https://epic.org/wp-content/uploads/apa/comments/EPIC-comments-to-NSCAI-093020.pdf>; Comments of EPIC, Request for Comments on a Draft Memorandum to the Heads of Executive Departments and Agencies, "Guidance for Regulation of Artificial Intelligence Applications," (Mar. 13, 2020), <https://epic.org/apa/comments/EPIC-OMB-AI-MAR2020.pdf>.

⁵ See Br. EPIC as Amicus Curiae, *Alabama v. Dep't of Commerce*, 546 F. Supp. 3d 1057 (M.D. Ala. 2021), <https://epic.org/wp-content/uploads/amicus/census/2020/Alabama-v-Commerce-21-cv-211-EPIC-Amicus-Brief.pdf>; EPIC, *EPIC v. Commerce (Census Privacy)* (2019), <https://epic.org/documents/epic-v-commerce-census-privacy/>; Br. EPIC as Amicus Curie, *Dep't of Commerce v. New York*, 139 S. Ct. 2551, (2019), <https://epic.org/wp-content/uploads/amicus/census/2020/Commerce-v-NY-EPIC-Amicus.pdf>.

the resulting data. Differential privacy is a more robust means of protecting individual privacy that should be prioritized in OSTP's national strategy.

a. Traditional deidentification and anonymization techniques do not work.

It has been clear since the early 2000s that basic database deidentification techniques which rely on removing personally identifiable information like names and addresses are insufficient to stop reidentification attacks achieved by combining information from multiple databases. Pioneering work by Latanya Sweeney demonstrated that “87% of the U.S. Population are uniquely identified by {date of birth, gender, ZIP}.”⁶ Sweeney showed that re-identification was broadly possible with diverse and supposedly anonymized datasets including survey data,⁷ pharmacy data,⁸ data from clinical trials,⁹ public health registries,¹⁰ and partial Social Security Numbers.¹¹

Subsequent research has confirmed that supposedly deidentified datasets are at best weakly anonymized and are subject to increasingly easy reidentification.¹² For example, Netflix data stripped of names can be used to reidentify more than 80 percent of users.¹³ The work of Cynthia Dwork has demonstrated how broadly applicable mathematical reidentification techniques are.¹⁴ Newer mathematical methods for reidentification rely on large, aggregate databases, allowing the

⁶ Latanya Sweeney, *Simple Demographics Often Identify People Uniquely*, Carnegie Mellon Data Privacy Lab (2000), <https://dataprivacylab.org/projects/identifiability/paper1.pdf>.

⁷ Latanya Sweeney, *Re-identification of De-identified Survey Data*, Carnegie Mellon Data Privacy Lab (2000).

⁸ Latanya Sweeney, *Patient Identifiability in Pharmaceutical Marketing Data*, Carnegie Mellon Data Privacy Lab (2011), <https://dataprivacylab.org/projects/identifiability/pharma1.pdf>.

⁹ Latanya Sweeney, *Identifiability of De-identified Clinical Trial Data*, Carnegie Mellon Data Privacy Laboratory (2009).

¹⁰ Latanya Sweeney, *Iterative Profiler*, Carnegie Mellon Data Privacy Lab (1997).

¹¹ See Data Privacy Lab, SOS Social Security Number Watch, Harvard University, <https://dataprivacylab.org/dataprivacy/projects/ssnwatch/index.html>.

¹² Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701 (2010), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006.

¹³ *Id.*

¹⁴ Dwork, C., A. Smith, T. Steinke, & J. Ullman, *Exposed! A Survey of Attacks on Private Data*, 4 Annual Review of Statistics and Its Application 61–84 (2017), https://privacytools.seas.harvard.edu/files/privacytools/files/pdf_02.pdf.

identification of sensitive features of individuals in datasets and allowing a bad actor to determine whether a particular individual is included.¹⁵ The most up-to-date reidentification models can identify virtually all Americans (99.87 percent) in any dataset containing 15 demographic attributes.¹⁶ In short, traditional deidentification techniques do not work and do not reliably preserve individual privacy.

b. Differential privacy allows for responsible use of datasets containing personal information while reducing the risk of reidentification.

Although traditional deidentification and anonymization techniques are fundamentally ineffective, there is a viable alternative for preserving the privacy of individuals contained in datasets while still preserving the research value of the data: differential privacy. “‘Differential privacy’ describes a promise, made by a data holder, or curator, to a data subject: ‘You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available.’”¹⁷ Differential privacy is achieved through the controlled injection of statistical noise into a study or analysis of a dataset, providing a mathematical guarantee of privacy while preserving the research value of the information.¹⁸ Applying differentially private algorithms to datasets allows researchers to perform one or multiple analyses on the data while minimizing the risk that such analysis will reveal the identity of any person in the dataset. Differential privacy has proven valuable in a broad range of applications, including statistical research, machine learning, emoji suggestions on Apple devices,

¹⁵ *Id.*

¹⁶ Luc Rocher, Julien Hendrickx, & Yves-Alexandre de Montjoye, *Estimating the success of re-identifications in incomplete datasets using generative models*, 10 *Nature Commc’ns* 3,069 (2019), <https://www.nature.com/articles/s41467-019-10933-3>.

¹⁷ Cynthia Dwork & Aaron Roth, *The Algorithmic Foundations of Differential Privacy* 5 (2014).

¹⁸ Daniel L. Oberski & Frauke Kreuter, *Differential Privacy and Social Science: An Urgent Puzzle*, *Harv. Data Sci. Rev.* (Jan. 31, 2020).

LinkedIn’s Labor Market Insights reports, Microsoft’s suggested replies in Office tools, and Google’s reporting of COVID-19 search trends.¹⁹

Perhaps the most prominent example is the U.S. Census Bureau’s adoption of differential privacy for the 2020 Census disclosure avoidance system.²⁰ In the lead-up to the 2020 Census, the Bureau determined that existing census data products were alarmingly vulnerable to reconstruction and reidentification attacks. Specifically, the Bureau found that the sex, age, race, and ethnicity of 142 million individuals could be inferred from publicly available 2010 Census data and that 52 million census respondents could be reidentified with the added use of commercial datasets.²¹ And the potential harms of these reconstruction and reidentification attacks are significant:

Anyone could construct a linkage attack by purchasing commercial data[.] . . . Most people do not view the characteristics in the decennial census as particularly sensitive, but those who are most at risk to having their data abused (and are typically also the hardest to count) do. People who are living in housing units with more people than are permitted on the lease are nervous about listing everyone living there, unless they can be guaranteed confidentiality. Same-sex couples are nervous about marking their relationship status accurately if they feel as though they could face discrimination. Yet, the greatest risks people face often stem from how census data can be used to match more sensitive data (e.g., income, health records, etc.).²²

In order to fulfill its Title 13 confidentiality obligations, the Bureau turned to a new disclosure avoidance system for the 2020 Census based on differential privacy—one which ensures both useful statistics and a mathematical guarantee of privacy. The Bureau’s adoption of and experience with differential privacy should serve as a guide for other federal agencies engaged in the collection and analysis of large datasets containing personal information.

¹⁹ Damien Desfontaines, A list of real-world uses of differential privacy, desfontaines.es (last updated Jan. 27, 2022), <https://desfontain.es/privacy/real-world-differential-privacy.html>.

²⁰ JASON, *Formal Privacy Methods for the 2020 Census*, <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/privacy-methods-2020-census>

²¹ Michael Hawes, U.S. Census Bureau, *Differential Privacy and the 2020 Decennial Census* 13 (Mar. 5, 2020), <https://www2.census.gov/about/policies/2020-03-05-differential-privacy.pdf>.

²² danah boyd, *Balancing Data Utility and Confidentiality in the 2020 US Census* 15–16 (Apr. 27, 2020), https://datasociety.net/wp-content/uploads/2019/12/Differential-Privacy-04_27_20.pdf.

EPIC urges OSTP to look especially to the work of Cynthia Dwork,²³ one of the co-inventors of differential privacy, as a starting point for the further adoption and application of differential privacy:

- Dwork, C., and V. Feldman. “Privacy-preserving prediction.” In *Conference on Learning Theory*, 1693-1702, 2018, 1693-1702.²⁴
- Bun, M., C. Dwork, G. N. Rothblum, and T. Steinke. “Composable and versatile privacy via truncated cdp.” *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, 2018, 74-86.²⁵
- Dwork, C. “Differential privacy.” *International Colloquium on Automata, Languages, and Programming*. ICALP, 2006.²⁶

II. Agencies across the federal government should dedicate more research and development funding to privacy enhancing technologies.

Directing more funding towards privacy-enhancing technologies would create substantial benefits for individuals, industry, and the federal government. Better and more widely available PETs could help protect the privacy of individuals by reducing the harms from data breaches that are common across the federal government. And investing in PETs is worthwhile and cost-effective because PETs can enhance and preserve other research opportunities. Last year the National Security Commission on Artificial Intelligence listed funding privacy-preserving technologies as a priority in the broader rollout of AI and a necessity in democratic regimes.²⁷

Privacy-enhancing technologies can provide additional protection for individuals when data breaches and hacks occur. Federal agencies maintain vast amounts of personal information in databases across the government. And data breaches are both increasingly common and increasingly severe. As an example of this trend across the federal government, a 2015 data breach at the Office of Personnel Management (OPM) exposed social security numbers and other personal data from

²³ See *Cynthia Dwork*, Harvard University (2022), <https://dwork.seas.harvard.edu>.

²⁴ <https://dwork.seas.harvard.edu/publications/privacy-preserving-prediction>.

²⁵ <https://dwork.seas.harvard.edu/publications/composable-and-versatile-privacy-truncated-cdp>.

²⁶ <https://dwork.seas.harvard.edu/publications/differential-privacy>.

²⁷ Nat'l Sec. Comm'n on Artificial Intelligence, *Chapter 15: A Favorable International Technology Order* (2021), <https://reports.nscai.gov/final-report/chapter-15/>. **Error! Hyperlink reference not valid.**

21.5 million individuals.²⁸ Around the same time, OPM reported another major data breach exposing records on about 4 million federal employees.²⁹ Just a year before, a breach at the U.S. Postal Service led to the loss of personal information from more than 800,000 employees.³⁰ On August 24, 2020, a cyber-attack compromised a federal agency and documents were stolen.³¹

The greatest risks of data breaches come from the government holding large volumes of personal information that can have lasting financial and security impacts when wrongfully divulged. For example, The Federal Emergency Management Agency (FEMA) unnecessarily disclosed sensitive information from victims of the 2017 California wildfires, exposing up to 2.3 million people.³² FEMA shared details of victims' financial institutions and personal lives, including EFT and bank transit numbers and complete addresses.³³

While traditionally the focus on protecting federal agency databases has settled on improving cybersecurity practices, implementation of best practices has been uneven at best. In 2018 for example, the GAO found that over 700 of its cybersecurity recommendations since 2010 had not been implemented by federal agencies.³⁴ Privacy-enhancing technologies can complement

²⁸ U.S. Gov't Accountability Office, *DHS Needs to Enhance Capabilities, Improve Planning, and Support Greater Adoption of Its National Cybersecurity Protection System* (Jan. 2016) at 8, <https://www.gao.gov/assets/680/674829.pdf>.

²⁹ *Id.*

³⁰ *Id.*

³¹ Cybersecurity and Infrastructure Security Agency, *Federal Agency Compromised by Malicious Cyber Actor*, AR20-268A, Dep't. of Homeland Sec. (Sept. 24, 2020), <https://us-cert.cisa.gov/ncas/analysis-reports/ar20-268a>; Duncan Riley, *DHS discloses data breach of US agency but doesn't name which was hacked*, SiliconAngle (Sept. 24, 2020), <https://siliconangle.com/2020/09/24/dhs-discloses-data-breach-us-agency-doesnt-name-hacked/>.

³² Christopher Mele, *Personal Data of 2.3 Million Disaster Victims Was Released by FEMA, Report Says*, N.Y. Times (Mar. 22, 2019), <https://www.nytimes.com/2019/03/22/us/fema-data-breach.html>; John V. Kelly, *Management Alert – FEMA Did Not Safeguard Disaster Survivors' Sensitive Personally Identifiable Information*, OIG-19-32, Dep't of Homeland Sec. Off. of Inspector Gen. (Mar. 15, 2019), <https://www.oig.dhs.gov/sites/default/files/assets/2019-03/OIG-19-32-Mar19.pdf>.

³³ *Id.*

³⁴ U.S. Gov't Accountability Office, *GAO-19-105 Information Security: Agencies Need to Improve Implementation of Federal Approach to Securing Systems and Protecting Against Intrusions* (Dec. 18, 2018), <https://www.gao.gov/assets/700/696105.pdf>.

cybersecurity practices by making it harder to derive PII from federal databases, disincentivizing malicious hacks, and reducing the harms caused by data breaches when they do occur.

Investing in further research of PETs is a cost-effective strategy because these technologies are rapidly becoming a prerequisite to safe and ethical research. Developing better PETs and making them widely available can spur innovation across sectors by expanding the research possible for both government-supported scientists and industry. Increasing and shifting agency funding towards PETs is an investment in innovation.

III. Conclusion

EPIC applauds OSTP's development of a national strategy for advancing PETs. OSTP should focus funding on the development and adoption of differential privacy to address the shortcomings of traditional deidentification and anonymization techniques. EPIC also urges OSTP to direct federal agencies to increase funding toward PETs to protect individual privacy and promote innovation. If OSTP has any further questions, please reach out to EPIC Senior Counsel John Davisson at .

Respectfully Submitted,

John Davisson
John Davisson
EPIC Senior Counsel

Jake Wiener
Jake Wiener
EPIC Law Fellow

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Fortanix Inc.

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

From:
To: [PETS-RFI](#)
Subject: < RFI Response: Privacy-Enhancing Technologies >
Date: Friday, July 8, 2022 2:59:06 PM
Attachments:

Response of Fortanix, Inc. to the OSTP RFI "Notice of request for information on Advancing Privacy-Enhancing Technologies".

1. *Specific research opportunities to advance PETs:* Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.

Confidential Computing, defined as “the protection of data in use using Trusted execution Environments”^[1] (TEEs), is an emerging PET that is forecast to experience exponential market growth to 2026.^[2]

Confidential Computing has been successfully applied by Fortanix to a variety of use-cases where enforcement of data privacy and application security is necessary including: cryptographic key management (e.g. the Fortanix Data Security Manager™ product, with FIPS 140-2 Level 1 software certification^[3] and FIPS 140-2 Level 3 hardware certification^[4]), federated machine learning (FML), multi-party data analytics, private information retrieval, confidential artificial intelligence (AI) – model training and inference, and confidential blockchains. These use-cases are applicable to a range of industry requirements, encompassing healthcare, financial services, retail, supply chain management, decentralized applications (Web 3.0), and U.S. Federal Government agencies.

The announced bilateral innovation prize challenges focused on privacy-enhancing-technologies (PETs)^[5] could include specific challenges associated with the development or implementation of Confidential Computing.

Since publication of the National Privacy Research Strategy in 2016^[6], Confidential Computing has become available at scale through integration within general purpose computing platforms provided by public cloud service providers and OEM hardware vendors. Initiation of a technical comparison of different PETs and their appropriate application by the NSTC and/or NIST would encourage evaluation by potential end-users. Such an evaluation would also provide contemporary information to lawmakers and a basis for communication of public information on how Confidential Computing can protect private data while innovation through the secure integration of discrete datasets for in-depth analysis.

Confidential Computing represents one of the most promising PETs under development today, due to its flexibility of deployment, functional scope, and suitability for multi-party analytics and artificial intelligence applications. In the latter case, the ability for a regulatory agency to interrogate and verify the deployed software code is important for the “explainability” and transparency of models. Alternative, cryptography-based, PETs to Confidential Computing obfuscate software code, preventing analysis of system functionality.

2. *Specific technical aspects or limitations of PETs:* Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections or reduce the risks or costs of adopting PETs.

Confidential Computing incorporates attestation of Trusted Execution Environments (TEEs). Attestation provides auditable cryptographic validation of the integrity of a TEE and the software code deployed inside it. Mutual attestation of applications and enforced verification of attestation credentials creates segmentation of processes and supports robust software supply chain security.

Extension of Confidential Computing beyond CPUs to include GPUs will provide broad capability for secure machine learning and deployment of artificial intelligence, including deep learning with massive neural networks.

Confidential Computing represents one of the most promising PETs under development today, due to its flexibility of deployment, functional scope, and suitability for multi-party analytics and artificial intelligence applications. In the latter case, the ability for a regulatory agency to interrogate and verify the deployed software code is important for the “explainability” and transparency of models. Alternative, cryptography-based, PETs to Confidential Computing obfuscate software code, preventing analysis of system functionality.

3. *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:* Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of Executive Order 14058: Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government.

Confidential Computing can provide sufficient data and applications security to enable recommendation of the technology by Federal Government, in accordance with the provisions of Exec. Order No. 14,034 (2021)^[7]. Where data privacy is required to protect national security, the use of Trusted Execution Environments (TEEs), with process-based identity verification and software integrity validation through attestation, is consistent with the concept of Zero-Trust Architecture, mandated under Executive Order No. 14,028 (2021)^[8].

Fortanix has demonstrate the application of Confidential Computing to support clinical AI validation in healthcare^[9], directed towards rare disease diagnosis and therapeutic design, and genome analysis using foreign data sources that are subject to different privacy legislation (e.g., EU GDPR). These use-cases highlight the benefits of Confidential Computing for healthcare, including applications within the internet-of-medical-things (IoMT) and telemedicine.

As part of an investigation related to the secure acquisition of real-world data (RWD) by the FDA, Fortanix demonstrated how Confidential Computing enables data portability through end-to-end encryption of private healthcare data (PHI) sourced from different healthcare providers, within a distributed network.^[10] The reference architecture developed could be transferred to meet the needs of other U.S. Government agencies and could incorporate individual-level interaction via secure IoMT devices or controlled access to electronic health records (eHR) – as has been demonstrated in Germany, using Fortanix Confidential Computing technology within a national framework for patient privacy.^[11]

Confidential Computing has also been successfully demonstrated for financial crime detection, including anti-money laundering^[12] and synthetic identity fraud^[13]. As decentralized finance (DeFi) solutions proliferate, it will be increasingly important for national and international regulatory agencies to support analysis of private financial data, hosted by independent financial institutions. Confidential Computing provides an effective method of facilitating scalable, transnational, data analysis without any requirement for sensitive data to be shared between collaborating parties. This capability of Confidential Computing offers law enforcement agencies the ability to detect financial crime, without any need to access the underlying data or compromise data sovereignty requirements of partner states.

4. *Specific regulations or authorities that could be used, modified, or introduced to advance PETs:* Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking authorities under the Office of Management and Budget, the Federal Trade Commission, and financial regulatory bodies, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (e.g., Other Transaction Authority) to procure PETs for development.

Organizations that have identified privacy and security benefits in the adoption of Confidential Computing are often slow to implement this form of PET due to perceived risks to compliance with Federal regulations. Fortanix has successfully demonstrated the capability of Confidential Computing to secure protected data and sensitive software code for customers in financial services, healthcare, and US Government Agencies. Industry customers, however, typically remain constrained in their adoption of PETs and their plans to deploy data and applications using untrusted cloud infrastructure, by concerns that PETs do not address legislative requirements with respect to protected data and applications. Such protected data may include PII and PHI that is subject to security and privacy controls under the provisions of the Fair Credit Reporting Act (FCRA)^[14] [15 U.S.C §1681: §602(a)(4)]^[14], the Gramm-Leach-Bliley Act [15 U.S.C. §6801(V)(A): §501(a), and 15 U.S.C. §6821(V)(B):^[15]

§521] , and/or The Health Insurance Portability and Accountability Act of 1996 (HIPAA) [12 U.S.C. 1811(II)(F): §264(c) (1)]^[16] .

Explicit reference to Confidential Computing and other PETs within existing and future legislation that incorporates, or amends, data privacy requirements would clarify the terms of compliance with respect to privacy obligations for industry. An example of one opportunity to encourage adoption of Confidential Computing would be recognition of the additional security provided by Trusted Execution Environments (TEEs) within the Federal Information Processing Standards (FIPS) defined by NIST. This is particularly relevant to the secure management and enforced privacy of cryptographic keys that underpin identity management within today's internet and that will be the foundation of tomorrow's, predicted, Web 3.0 decentralization.^[17]

The revision of existing legislation to clarify approved methods of data privacy protection, to include a list of potential cyberattack vectors to be mitigated by technical solutions, would be assist organizations in evaluating their compliance with specific industry requirements. Such clarifications would also support audit and enforcement of data privacy and definitions of "best practice" by Federal regulators and law enforcement agencies.

5. *Specific laws that could be used, modified, or introduced to advance PETs:* Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.

Organizations that have identified privacy and security benefits in the adoption of Confidential Computing are often slow to implement this form of PET due to perceived risks to compliance with Federal regulations. Fortanix has successfully demonstrated the capability of Confidential Computing to secure protected data and sensitive software code for customers in financial services, healthcare, and US Government Agencies. Industry customers, however, typically remain constrained in their adoption of PETs and their plans to deploy data and applications using untrusted cloud infrastructure, by concerns that PETs do not address legislative requirements with respect to protected data and applications. Such protected data may include PII and PHI that is subject to security and privacy controls under the provisions of the Fair Credit Reporting Act (FCRA) [15 U.S.C §1681: §602(a)(4)]^[18] , the Gramm-Leach-Bliley Act [15 U.S.C. §6801(V)(A): §501(a), and 15 U.S.C. §6821(V)(B): §521]^[19] , and/or The Health Insurance Portability and Accountability Act of 1996 (HIPAA) [12 U.S.C. 1811(II)(F): §264(c) (1)]^[20] .

Explicit reference to Confidential Computing and other PETs within existing and future legislation that incorporates, or amends, data privacy requirements would clarify the terms of compliance with respect to privacy obligations for industry.

The revision of existing legislation to clarify approved methods of data privacy protection, to include a list of potential cyberattack vectors to be mitigated by technical solutions, would be assist organizations in evaluating their compliance with specific industry requirements. Such clarifications would also support audit and enforcement of data privacy and definitions of "best practice" by Federal regulators and law enforcement agencies.

6. *Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:* This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

Recent publication of NIST IR 8320^[21] provides helpful guidance on the requirement for, and implementation of, Confidential Computing technology within cloud and edge-based deployments. However, it has been incumbent upon industry to elucidate practical applications of PETs to address real-world privacy and security needs. It would be helpful if NIST and agencies responsible for data privacy enforcement collaborated with leading technology and service providers to establish reference architectures that are applicable to compliance with specific legislation. Such architectures could be vendor agnostic or, in common with NIST IR 8320, a survey of applicable technologies and services could be provided to direct PET adopters to appropriate partners for implementation.

The development of reference architectures incorporating PETs might also drive technical convergence and standardization within industry to reduce barriers to implementation created by price differentiation and poor interoperability. In the case of Confidential Computing, there is active work within industry to address problems created by heterogeneous approaches to Trusted Execution Environment (TEE) attestation – this work is being led by the Confidential Computing Consortium of the Linux Foundation ^[22] and the Remote Attestation procedureS (RATS) working group of the Internet Engineering Task Force (IETF) ^[23], however, there is no national, nor international, policy reference to align these industrial initiatives. Establishing reference standards for PET deployment in support of data privacy protection, including open-source protocols for attestation and identity assertion, would provide a framework within which industry could direct future technical research and work with authorities to improve legislation and enhance technical standards.

7. *Risks related to PETs adoption:* Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

Implementation of PETs typically requires a sophisticated set of skills, including advanced software engineering, cryptography, and software solution architecture. In the case of PETs such as Fully Homomorphic Encryption (FHE), Multi-Party Computing (MPC), and Confidential Computing, the knowledge required to undertake and evaluate system design and implementation can prove a barrier to adoption. Increasing provision of packages service models, such as the Fortanix Confidential AI managed service ^[24], allow non-technical users (e.g., data analysts and end-users) to take advantage of privacy-preserving computation by providing hosted infrastructure and auditability of deployment, to demonstrate compliance with privacy regulations.

8. *Existing best practices that are helpful for PETs adoption:* Information about U.S. policies that are currently helping facilitate adoption as well as best practices that facilitate responsible adoption. This includes existing policies that support adoption, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This also includes information about where and when PETs can be situated within tiered access frameworks for accessing restricted data, ranging from publicly accessible to fully restricted data.

Executive Order No. 14,028 (2021) ^[25] and the publication of NIST Special Publication 800-207 “Zero-Trust Architecture” ^[26] (ZTA) has generated increased interest in Confidential Computing as a PET that supports identity authentication, software isolation, and network resilience. Confidential Computing can be used as a generic PET solution to encapsulate end-to-end workflows, such as in a federated machine learning (FML) architecture, or it can be implemented to secure specific data processing operations. Fortanix has demonstrated the application of a private-information retrieval (PIR) solution for U.S. Government agencies that supports secure search of public data and integration of unclassified data with restricted data for private query by data analysts. This solution protects the identity of the analyst, the content of the classified data query, and demonstrates performance advantages (i.e., reduced system latency) compared to alternative cryptography-based PETs.

9. *Existing barriers, not covered above, to PETs adoption:* Information about technical, sociotechnical, usability, and socioeconomic barriers that have inhibited wider adoption of PETs, such as a lack of public trust. This includes recommendations on how such barriers could be overcome. Responses that focus on increasing equity for underserved or marginalized groups are especially welcome.

Implementation of PETs typically requires a sophisticated set of skills, including advanced software engineering, cryptography, and software solution architecture. In the case of PETs such as Fully Homomorphic Encryption (FHE), Multi-Party Computing (MPC), and Confidential Computing, the knowledge required to undertake and evaluate system design and implementation can prove a barrier to adoption. Increasing provision of packages service models, such as the Fortanix Confidential AI managed service ^[27], allow non-technical users (e.g., data analysts and end-users) to take advantage of privacy-preserving computation by providing hosted infrastructure and auditability of deployment, to demonstrate compliance with privacy regulations.

10. *Other information that is relevant to the adoption of PETs:* Information that is relevant to the adoption of PETs that does not fit into any of the topics enumerated above.

The “CIA triad” of confidentiality, integrity, and availability is fully supported by Confidential Computing as a form of PET: Confidentiality of private data is maintained through encryption of related memory in use. Software integrity can be asserted using “measurement” of deployed application code as part of the intrinsic attestation process that underpins deployment of robust Trusted Execution Environments (TEEs). Availability is supported by Confidential Computing due to the isolation guarantees that enable approved software to be safely executed, even where the underlying hardware or network infrastructure has been compromised by a cyber threat actor.

Recent research has elucidated the extended loiter time of cyberattackers within networks before detection and mitigation of the intrusion.^[28] Confidential Computing provides a means to ensure data privacy where processing occurs on untrusted infrastructure – which should be assumed under a Zero-Trust Architecture (ZTA) approach to system design. Where computation takes place using distributed resources over which the application owner has no control, Confidential Computing can provide systemic resilience by supporting continued, secure, data processing until network reconfiguration is complete.

In response to identified threats to national security disclosed by CISA^[29] and the FBI^[30], alongside international partners, Confidential Computing can be employed as a PET to secure the data privacy of U.S. citizens, to protect the confidentiality of intellectual property, and to prevent unchecked network exploitation through attestation-based identity verification and isolation of software from privileged system users. These capabilities are of particular importance within the context of cyber-physical systems, where the risk to critical infrastructure and the potential for physical harm to U.S. citizens is acute.^[31]

[1] Available from: https://confidentialcomputing.io/wp-content/uploads/sites/85/2021/03/confidentialcomputing_outreach_whitepaper-8-5x11-1.pdf [Accessed: July 8, 2022].

[2] Available from: https://confidentialcomputing.io/wp-content/uploads/sites/85/2021/10/Everest_Group_-_Confidential_Computing_-_The_Next_Frontier_in_Data_Security_-_2021-10-19.pdf [Accessed: July 8, 2022].

[3] Available from: <https://csrc.nist.gov/projects/cryptographic-module-validation-program/certificate/3326> [Accessed: July 8, 2022].

[4] Available from: <https://csrc.nist.gov/projects/cryptographic-module-validation-program/certificate/3545> [Accessed: July 8, 2022].

[5] Available from: <https://www.whitehouse.gov/ostp/news-updates/2021/12/08/us-and-uk-to-partner-on-a-prize-challenges-to-advance-privacy-enhancing-technologies/> [Accessed: July 8, 2022].

[6] Available from: <https://www.nitrd.gov/pubs/NationalPrivacyResearchStrategy.pdf> [Accessed: July 8, 2022].

[7] Available from: <https://www.federalregister.gov/documents/2021/06/11/2021-12506/protecting-americans-sensitive-data-from-foreign-adversaries> [Accessed: July 8, 2022].

[8] Available from: <https://www.federalregister.gov/documents/2021/05/17/2021-10460/improving-the-nations-cybersecurity> [Accessed: July 8, 2022].

[9] Available from: <https://www.beekeeperai.com/blog/58111-securing-healthcare-ai-with-confidential> [Accessed: July 8, 2022].

[10] Available from: <https://www.intel.com/content/www/us/en/newsroom/news/leidos-fortanix-accelerate-clinical-trials.html#gs.583ork> [Accessed: July 8, 2022].

[11] Available from: <https://www.gesundheitsindustrie-bw.de/en/article/news/ehr-and-phr-digital-records-in-the-german-healthcare-system> [Accessed: July 8, 2022].

[12] Available from: <https://www.fortanix.com/company/pr/2022/04/intel-fiverity-and-fortanix-bring-confidential-computing-to-the-fight-against-digital-fraud-in-financial-services> [Accessed: July 8, 2022].

[13] Available from: <https://www.fortanix.com/company/pr/2022/04/intel-fiverity-and-fortanix-bring-confidential-computing-to-the-fight-against-digital-fraud-in-financial-services> [Accessed: July 8, 2022].

[14] Available from: https://www.ftc.gov/system/files/documents/statutes/fair-credit-reporting-act/545a_fair-credit-reporting-act-0918.pdf [Accessed: July 8, 2022].

[15] Available from: <https://www.govinfo.gov/content/pkg/PLAW-106publ102/pdf/PLAW-106publ102.pdf> [Accessed: July 8, 2022].

[16] Available from: <https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf> [Accessed: July 8, 2022].

[17] Available from: <https://www.darkreading.com/endpoint/me-my-digital-self-and-i-why-identity-is-the-foundation-of-a-decentralized-future> [Accessed: July 8, 2022].

[18] Available from: https://www.ftc.gov/system/files/documents/statutes/fair-credit-reporting-act/545a_fair-credit-reporting-act-0918.pdf [Accessed: July 8, 2022].

[19] Available from: <https://www.govinfo.gov/content/pkg/PLAW-106publ102/pdf/PLAW-106publ102.pdf> [Accessed: July 8, 2022].

[20] Available from: <https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf> [Accessed: July 8, 2022].

[21] Available from: <https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8320.pdf> [Accessed: July 8, 2022].

- [22] Available from: <https://confidentialcomputing.io/webinar-the-future-of-attestation-in-a-confidential-world/> [Accessed: July 8, 2022].
- [23] Available from: [https://datatracker.ietf.org/wg/rats/about/#:~:text=Remote%20attestation%20procedures%20\(RATS\)%20determine.links%20to%20the%20supply%20chain.](https://datatracker.ietf.org/wg/rats/about/#:~:text=Remote%20attestation%20procedures%20(RATS)%20determine.links%20to%20the%20supply%20chain.) [Accessed: July 8, 2022].
- [24] Available from: <https://www.fortanix.com/products/confidential-ai> [Accessed: July 8, 2022].
- [25] Available from: <https://www.federalregister.gov/documents/2021/05/17/2021-10460/improving-the-nations-cybersecurity> [Accessed: July 8, 2022].
- [26] Available from: <https://csrc.nist.gov/publications/detail/sp/800-207/final> [Accessed: July 8, 2022].
- [27] Available from: <https://www.fortanix.com/products/confidential-ai> [Accessed: July 8, 2022].
- [28] Available from: <https://news.sophos.com/en-us/2022/06/07/active-adversary-playbook-2022/> [Accessed: July 8, 2022].
- [29] Available from: <https://www.cisa.gov/uscert/ncas/current-activity/2021/07/19/us-government-releases-indictment-and-several-advisories-detailing> [Accessed: July 8, 2022].
- [30] Available from: <https://www.mi5.gov.uk/news/speech-by-mi5-and-fbi> [Accessed: July 8, 2022].
- [31] Available from: <https://www.darkreading.com/physical-security/outlining-risks-to-the-world-s-vital-cyber-physical-systems> [Accessed: July 8, 2022].

Dr Richard Searle
Vice President of Confidential Computing
Fortanix®, 800 West El Camino Real, Suite 180, Mountain View, CA 94040, USA
Fortanix' mission is to solve cloud security and privacy.



Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Future of Privacy Forum

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

July 8, 2022

Office of Science and Technology Policy (OSTP)
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, D.C. 20504
VIA EMAIL: PETS-RFI@nitrd.gov
Attention: Ms. Stacey Murphy

RE: RFI Response: Privacy-Enhancing Technologies, Doc. No. 2022-12432

The Future of Privacy Forum (FPF) welcomes this opportunity to share our thoughts regarding specific actions that would advance the adoption of privacy-enhancing technologies (PETs). FPF is a 501(c)(3) non-profit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies. Our primary office is in Washington, DC, and we work closely with our colleagues in Brussels, Singapore, Tel Aviv, and around the world. FPF seeks to support balanced, informed public policy.¹

In response to the Office's invitation for comments, and with regard for the particular categories of information requested,² FPF recommends that the OSTP include the following three recommendations in the national strategy on privacy-preserving data sharing and analytics:

1. Support the growing discipline of privacy engineering aimed at bridging the gap between technologies and policies through direct funding of academic research, building expertise within government, encouraging business-academia dialogues, and directing agencies to require federal contractors to incorporate PETs as appropriate to promote common standards in the discipline.;
2. Recommend the establishment of a trusted inter-agency and multi-stakeholder body, including the FTC, NIST, HHS, NSF, and experts from the private sector, civil society, and academia, to provide guidance and standards-setting for de-identification and the role of PETs, with particular regard to their utility for compliance with state and federal legislation; and

¹ The views herein do not necessarily reflect the views of our supporters or Advisory Board.

² Office of Science and Technology Policy, *Request for Information on Advancing Privacy-Enhancing Technologies*, Federal Register, (June 9, 2022), Accessed June 25, 2022, <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>.

3. Encourage the establishment of Administrative Data Research Networks (ADRN)s that offer de-identification tools to facilitate researcher access to data in a secure manner.

Each of the next three sections correspond to one of these recommendations, exploring a barrier to PET's adoption and providing further information about the recommendations that address these issues. Below each section header, the topics implicated by a recommendation are identified by the corresponding number in the RFI. FPF hopes that this filing provides the Office with an enhanced understanding of this space and informs solutions.

1. Supporting Privacy Engineering as a Discipline Through Agency Funding of Research, Building Expertise Within Government, Greater Industry-Academia Dialogue, and PETs Requirements in Federal Contracting.

Relevance: Topics 6 and 9.

A national strategy for advancing privacy-enhancing technologies (PETs) should first aim to support the privacy engineering discipline to build the applied technical expertise needed to support the growth of privacy-enhancing technology in the United States. Organizations are increasingly engaging in the extensive collection and processing of data to enable data-driven products and services. The privacy issues raised by data processing have coincided with governments passing new laws regulating this activity, which has created a new global industry for “privacy tech.”³

In support of the emerging privacy tech sector, there has been a sharp increase in demand for privacy engineering, a subfield within computer science committed to the development of technologies that facilitate compliant data flows under privacy and data protection law.⁴

³ Privacy Tech Alliance, Future of Privacy Forum, Tim Sparapani, and Justin Sherman, *Privacy Tech's Third Generation - A Review of the Emerging Privacy Tech Sector*, Future of Privacy Forum, 37, (2021), Accessed June 27, 2022, https://fpf.org/wp-content/uploads/2021/06/FPF-PTA-Report_Digital.pdf (noting that all seventeen respondents to the survey question “Has your company ever purchased privacy-enhancing technologies, whether software or hardware, from a third-party provider?” responded yes).

⁴ *Id.* at 32 (describing how the PETs market has recently expanded to create “products and services that assist businesses in making the personal data they encounter both maximally available and maximally valuable for various components throughout the business . . .”).

However, despite the high demand for privacy engineers in the private sector, the academic field of privacy engineering is not currently aligned with market needs.⁵ While there are some long-standing and some newly emerging privacy engineering conferences, very few academic institutions currently offer certifications or graduate programs in privacy engineering.⁶ The lack of opportunities for practitioners to obtain necessary skills creates a shortage of talent who can develop these technologies.⁷ Even when such talent is being developed or already exists, researchers may be pursuing certifications or skills that are not aligned with companies' PETs needs.⁸ Furthermore, there are typically few opportunities for meaningful interaction between industry practitioners and academic technologists studying privacy engineering.⁹

- **Recommendation:** A national strategy for PETs should support the growing discipline of privacy engineering aimed at bridging the gap between technologies and policies through direct funding of academic research, building expertise within government, encouraging business-academia dialogues, and directing agencies to require federal contractors to incorporate PETs as appropriate to promote common standards in the discipline.

A national strategy should seek to bridge the gulf between academic talent and market demand through funding privacy engineering research, increasing opportunities for business-academic dialogue, building capacity within the US Government, and fostering growth of the field through requirements of federal contractors. Federal agencies can directly support academic opportunities through funding of privacy engineering research,¹⁰ which would increase the demand for privacy technologists and make academic career paths in privacy engineering more attractive. Federal agencies can build

⁵ Privacy Tech Alliance et al., *supra* note 3, at 22, (“Several buyers with whom we spoke identified the nascency of privacy engineering as one constraint on in-house privacy tech development. Simply put, there may not be enough privacy engineering talent to go around in general.”).

⁶ Carnegie Mellon University, *Privacy Engineering Program*, Accessed July 6, 2022, <https://privacy.cs.cmu.edu>, (Carnegie Mellon University is one of the few examples of an academic institution that provides training and certifications for privacy engineers).

⁷ See Privacy Tech Alliance et al., *supra* note 3, at 22.

⁸ See Jules Polonetsky and Jeremy Greenberg, *NSF Convergence Accelerator Paper: The Future of Privacy Technology*, Future of Privacy Forum, 12, (2020), Accessed June 27, 2022, https://fpf.org/wp-content/uploads/2020/03/NSF_FPF-REPORT_C-Accel1939288_Public.pdf, (“In a separate meeting with 15 companies that provide privacy tech tools for compliance, few of these companies were familiar with academic research or scientific advances in their particular focus area. As one of the academic observers at our New York City meeting commented, we need to address the academic to practitioner (and reverse) relationship ‘if we are serious about convergence.’”).

⁹ *Id.*

¹⁰ Legislation that promotes research into PETs could play a useful role. For example, the “Promoting Digital Privacy Technologies Act” (H.R.847), introduced in February 2022 by Representative Haley Stevens (D-MI) passed the US House of Representatives in May 2022.

dedicated internal capacity for privacy engineering; both CISA and the CIO Council have pursued valuable capacity building efforts regarding cybersecurity expertise that can serve as partial models.¹¹ In addition, agencies can create increased opportunities through workshops, forums, competitions, and conferences, for businesses and academic researchers to exchange ideas and collaborate on innovation in privacy enhancing technology. This could help achieve greater alignment between skills development and market needs.

Finally, a national strategy driven by the OSTP can support widespread PETs adoption by encouraging agencies to require PETs as a precondition for federal contractors placing bids on certain projects. Agencies could use this method to establish standards for PETs that businesses must comply with. Privacy engineers may then create or adopt PETs based on the federal benchmark, leading to a national standard for the market.

2. Promoting Definitional Clarity and Creating Consensus Around What De-Identification Methods Satisfy Legal Requirements Using a Trusted Inter-Agency and Multi-Stakeholder Body.

Relevance: Topics 2, 6, and 7.

A national strategy to advance privacy-enhancing technologies (PETs) should also seek to promote greater definitional clarity and consensus around what de-identification methods satisfy legal requirements. Many—but not all¹²—PETs¹³ are designed to enable de-identification, or the reduction of the identifiability of datasets. De-identification methods are a central aspect of all modern data processing, including healthcare

¹¹ See e.g., Cybersecurity and Infrastructure Security Agency, *Capacity Enhancement Guides for Federal Agencies*, CISA, Accessed July 7, 2022, <https://www.cisa.gov/capacity-enhancement-guides-federal-agencies>; U.S. CIO Council, *Chief Information Officers Council Handbook*, 124, Accessed July 7, 2022, <https://www.cio.gov/assets/files/Handbook-CIO.pdf>.

¹² Cem Dilmegani, *Top 10 Privacy Enhancing Technologies (PETs) & Uses in 2022*, AI Multiple, (June 14, 2022), Accessed on July 7, 2022, <https://research.aimultiple.com/privacy-enhancing-technologies/>, (Although some PETs, such as federated learning models and homomorphic encryption, resolve a key aspect of data privacy by allowing for learning *without* the collection or sharing of data, many other privacy enhancing technologies are designed for the de-identification of existing datasets).

¹³ Office of Science and Technology Policy, *Advancing a Vision for Privacy-Enhancing Technologies*, OSTP Blog, (June 28, 2022), Accessed July 6, 2022, <https://www.whitehouse.gov/ostp/news-updates/2022/06/28/advancing-a-vision-for-privacy-enhancing-technologies/>, (defining PETs as “technologies that will allow researchers, physicians, and others permitted access to gain insights from sensitive data without ever having access to the data itself.”).

research,¹⁴ assessments of corporate diversity and inclusion efforts,¹⁵ and product and service improvements.¹⁶

In general, de-identification methods, from blurring or suppressing data to more robust techniques such as differential privacy, share a similar goal of preserving utility of data while promoting privacy. Sometimes, this balance is poorly struck. Strong de-identification may protect privacy but prevent organizations from uncovering important insights from data. For example, differential privacy may underestimate bias when the size of a dataset changes.¹⁷ In other circumstances, de-identification methods may not provide a sufficient level of privacy while retaining the underlying utility of data. Some data sets are particularly difficult to de-identify in ways that appropriately reduce the risk of re-identifying individuals while retaining information needed for socially beneficial research. For example, it remains difficult to de-identify individualized location data¹⁸ while

¹⁴ Simon L. Garfinkel, *De-Identification of Personal Information*, NIST, 1, (2015), Accessed June 28, 2022, <https://csrc.nist.gov/publications/detail/nistir/8053/final>, (“[S]ignificant medical research resulting in societal benefit is made possible by the sharing of de-identified patient information under the framework established by the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule . . .”).

¹⁵ SafePorter, Accessed June 28, 2022, <https://www.safeportersecure.com>.

¹⁶ Apple, *Differential Privacy*, 1–2, Accessed June 27, 2022, https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf, (noting that Apple uses a differential privacy technique to learn about user’s behaviors for purposes of improving device features, such as auto-correction and predictive text-entry, without connecting those behaviors to people).

¹⁷ Heng Xu and Nan Zhang, *Implications of Data Anonymization on the Statistical Evidence of Disparity*, *Management Science* (accepted), ec6, (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3662612 (stating that “when disparity was operationalized through separation . . . noise insertion [differential privacy is a noise insertion technique] likely produces false negatives only, with false positives being highly unlikely.”).

¹⁸ Simon L. Garfinkel, *supra* note 14, at 37, (“Without some kind of generalization or perturbation, there is so much diversity in geographic data that it may be extremely difficult to de-identify locations.”); Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel, *Unique in the Crowd: The Privacy Bounds of Human Mobility*, *Nature*, 4, (2013), <https://www.nature.com/articles/srep01376>, (“[T]he uniqueness of human mobility traces is high, thereby emphasizing the importance of the idiosyncrasy of human movements for individual privacy. Indeed, this uniqueness means that little outside information is needed to re-identify the trace of a targeted individual even in a sparse, large-scale, and coarse mobility dataset.”).

preserving its utility for a range of purposes, including public health surveillance and contact tracing.¹⁹

Overall, one of the greatest barriers to widespread adoption of de-identification technologies has been the lack of regulatory certainty about legal definitions and standards for de-identification. Because privacy and data protection laws typically do not apply to personal data that has been de-identified,²⁰ the legal standards related to “de-identified data” are a subject of considerable international legal and policy debate. For example, the Health Insurance Portability and Accountability Act (HIPAA) provides a legal Safe Harbor for de-identification of medical records that is typically thought to differ from the standard of “anonymization” under the European Union’s General Data Protection Regulation (GDPR), a standard which is again likely different from the emerging legal standards in state privacy laws such as the California Consumer Privacy Act (CCPA).²¹ Given this regulatory uncertainty, many US businesses express reservations about investing in PETs due to confusion about whether they will effectively shield them from regulator scrutiny.

In addition to generating legal uncertainty, the lack of consensus makes it harder for market participants to have conversations about these tools, frustrating adoption.²² Furthermore, many organizations do not understand which use cases would benefit from

¹⁹ Stacey Gray, *A Closer Look at Location Data: Privacy and Pandemics*, Future of Privacy Forum, (March 25, 2020), Accessed June 27, 2022, <https://fpf.org/blog/a-closer-look-at-location-data-privacy-and-pandemics/>, (“Because location data is sensitive and challenging to truly “de-identify” (i.e. to significantly reduce or eliminate all privacy risks), there is a serious concern that once collected by a public health agency for pandemic tracking, it could be retained or used for other purposes. . . . [Location data] should be clearly siloed for that purpose and not re-used or retained for other civil or law enforcement uses.”).

²⁰ Jules Polonetsky and Jeremy Greenberg, *supra* note 8, at 2; Simon L. Garfinkel, *supra* note 12, at 4, noting that some “U.S. laws and regulations specifically recognize the importance and utility of de-identification.” Examples include the Family and Educational Records Privacy Act (FERPA) and the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule).

²¹ See Privacy Tech Alliance et al., *supra* note 3, at 27, (“[P]seudonymization and de-identification . . . have different meanings in the US and the EU, including to what extent they are distinct. Even when regulatory or statutory terms start out with a common definition, differing interpretations of those terms by regulators or judges can cause divergence over time of the meaning of a common term in different states or countries.”).

²² Privacy Tech Alliance et al., *supra* note 3, at 12 (noting that “a lack of common, consensus privacy tech definitions” is limiting the adoption of privacy-enhancing technologies).



these technologies.²³ These questions are particularly salient for small or resource-constrained businesses, who may find it challenging to make the business case for adoption of de-identification PETs. In essence, siloed conceptions of de-identification create uncertainty surrounding how de-identification standards fit within existing legal requirements and when they should be used.

- **Recommendation:** A national strategy should recommend the establishment of a trusted inter-agency and multi-stakeholder body, including the FTC, NIST, HHS, NSF, and experts from the private sector, civil society, and academia, to provide guidance and standards-setting for de-identification and the role of PETs, with particular regard to their utility for compliance with state and federal legislation.

As part of a national strategy, the OSTP should support efforts that promote definitional clarity and create consensus around what de-identification methods satisfy legal requirements. To achieve these goals, the Office could direct an agency or group of agencies to establish a trusted body.²⁴ This body could take the form of a multistakeholder commission, including the FTC, NIST, HHS, NSF, and experts from the private sector, civil society, and academia. It would be responsible for providing guidance, recommendations, and standards for de-identification tools and other PETs.²⁵ As part of this work, the body should clarify how uses of PETs map onto legal requirements and analyze how these technologies are best utilized.²⁶ For example, this exercise could shed light on what constitutes de-identification under certain laws and which applications an organization should utilize a particular tool for.

²³ Jules Polonetsky and Jeremy Greenberg, *supra* note 8, at 3, (“Our initial survey included conversations at three meetings with privacy leads from a total of 40 companies. In each meeting, companies indicated significant interest in de-identification, but in most cases had limited awareness of the academic state-of-the-art. To the extent companies were versed in the basics of differential privacy or homomorphic encryption, **they were unclear which use cases would benefit from these technologies . . .**”) (emphasis added).

²⁴ Privacy Tech Alliance et al., *supra* note 3, at 6.

²⁵ *Id.*

²⁶ *Id.*

3. Facilitating Researchers' Access to De-Identification Tools Through Administrative Data Research Networks.

Relevance: Topics 6, 8, and 9.

Finally, a national strategy should advance PETs by further encouraging government agencies to implement Administrative Data Research Networks (ADRN). De-identification tools are sought by a diverse array of organizations across many sectors of the economy. Despite this widespread interest, many organizations cannot utilize de-identification tools due to the significant amount of expertise and resources required to adopt these technologies. Implementation and resource barriers can hinder beneficial activity, such as improvements to municipal services through open data initiatives.²⁷ Researchers that cannot obtain and deftly operate de-identification tools may encounter hurdles to accessing administrative data, or information held by organizations.²⁸ Concerns about sharing data with researchers²⁹ can frustrate innovation and evidence-based policymaking.

- **Recommendation:** Encourage the establishment of Administrative Data Research Networks (ADRN) that offer de-identification tools to facilitate researcher access to data in a protective manner.

A national strategy to advance PETs should further encourage government agencies to implement Administrative Data Research Networks (ADRN). ADRN are networks of Administrative Data Research Facilities (ADRF), which are intermediary institutions that offer secure computing platforms for researchers to access data held by other organizations, such as government agencies.³⁰ ADRN would support the growth of PETs by: establishing the appropriate de-identification standards needed to facilitate the

²⁷ Kelsey Finch, *City of Seattle Open Data Risk Assessment*, Future of Privacy Forum, 7, (2018), , Accessed June 27, 2022, <https://fpf.org/wp-content/uploads/2018/01/FPF-Open-Data-Risk-Assessment-for-City-of-Seattle.pdf>, (“Tremendous benefits in healthcare, education, housing, transportation, criminal justice, and public safety are already being realized as richer and more timely datasets are made available to the public. Open data can unite the power of city and private sector abilities to improve community health and lifestyles, from bikeshare systems and commercial apps harnessing transit data to community advocates shining the light on ineffective or discriminatory practices through policing and criminal justice data.”).

²⁸ Daniel Goroff, Jules Polonetsky, and Omer Tene, *Privacy Protective Research: Facilitating Ethically Responsible Access to Administrative Data*, 675 *Annals Am. Acad. Pol. & Soc. Sci.*, 49 (defining administrative data as data that one organization collected that “researchers access . . . for a purpose different from the one for which the data were originally collected.”).

²⁹ *Id.* at 48, (“[T]he ability of researchers to access significant government datasets is often limited by a range of concerns, in large part consisting of privacy and security objections.”).

³⁰ Coleridge Initiative, *Administrative Data Research Facility*, Accessed July 6, 2022, <https://coleridgeinitiative.org/adrf/>.

sharing of administrative data with researchers, demonstrating the value of investment in PETs; and creating a market for nascent privacy engineering.

More specifically, ADRFs can create procedures to improve accuracy, efficacy and privacy at the data input, computation, and output stages.³¹ These facilities would also institute accountability measures for auditing and monitoring compliance with data-sharing rules.³² ADNRs would function as forums where individual ADRFs share best practices and identity standards that apply across industry sectors. For example, ADRFs could create working groups on topics such as data security, private and proprietary data protections, and more.³³ To deter misbehavior, ADNRs could rescind a researcher or ADRF's status, privileges, and data access.³⁴

Thank you for this opportunity to provide input on the specific actions that could advance the adoption of privacy-enhancing technologies (PETs) in a responsible manner. We welcome any further opportunities to provide resources or information to assist in this important effort. For more information or to clarify any information provided here, please contact Daniel Berrick at dberrick@fpf.org.

Sincerely,

Daniel Berrick
Policy Counsel, Future of Privacy Forum

Jules Polonetsky
CEO, Future of Privacy Forum

Limor Shmerling Magazanik
Managing Director, Israel Tech Policy Institute

Future of Privacy Forum
1350 Eye Street. NW, Suite 350
Washington, DC, 20005
info@fpf.org

³¹ Daniel Goroff et al., *supra* note 28, at 61.

³² *Id.* at 60.

³³ *Id.*

³⁴ *Id.*

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Georgetown University Massive Data Institute

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

GEORGETOWN
UNIVERSITY

McCourt School *of Public Policy*

**MASSIVE
DATA
INSTITUTE**

The Massive Data Institute (MDI) at Georgetown’s McCourt School of Public Policy focuses on the secure and responsible use of data to answer public policy questions. MDI works with researchers in government, academia, and industry to solve societal-scale problems using novel and traditional large-scale data sources. MDI’s strategic partnerships promote community and innovation across the health, social, computer, and data sciences.

MDI draws on expertise from across Georgetown’s campus and beyond, including the social, natural, and computer science departments, and through strategic partnerships with organizations like the Beeck Center for Social Impact and Innovation, Lawrence Livermore Laboratories, the Institute for Social Research at University of Michigan, and the Pew Charitable Trusts. The U.S. Census Bureau has designated MDI a Federal Statistical Research Data Center, one of only 32 in the nation.

The MDI regularly awards seed grants, houses post-doctoral fellows, and hosts panels and faculty seminars on public policy and massive data.

Contact:

Dr. Amy O’Hara
Research Professor
Massive Data Institute
Georgetown McCourt School of Public Policy
Director, Georgetown Federal Statistical Research Data Center

July 8, 2022
Stacy Murphy
Operations Manager, OSTP
725 17th Street NW
Washington, DC 20503

Dear Ms. Murphy,

Thank you for the opportunity to respond to the Office of Science and Technology Policy’s Request for Information, entitled Advancing Privacy-Enhancing Technologies. I am writing on behalf on Georgetown’s Massive Data Institute in reference to Topic 3: **Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs.**

Sectors and applications where data are exceptionally decentralized or sensitive:

The K-12 and postsecondary education sector in the U.S. provides an excellent opportunity for national PET adoption and standardization.

Education is a policy space with exceptionally decentralized data owners whose coordination directly influences students’ outcomes and well-being. Education data systems such as school information systems and state longitudinal data systems are designed and enacted state by state, district by district. Furthermore, the data generated by state and local education agencies is very sensitive. School districts across the country retain vast amounts of personal information, not just on students, but on parents, teachers, administrators, and often alumni, including grades, test scores, written evaluations, disciplinary records, health records, sensitive communications, and demographic data. In addition to local and state education agencies and postsecondary institutions, other players hold data that are key to student success, such as government agencies providing social services and education technology (“edtech”) companies whose tools record a wealth of data on students’ learning processes. While education and data professionals have an interest in sharing and linking data with other data owners, this sharing heightens privacy concerns, particularly where the Family Educational Rights and Privacy Act (FERPA) is involved. PETs—such as secure multiparty computation, trusted execution environments, and synthetic data—can address these concerns by allowing for safer and greater data access and data insights without compromising the safety of students’ sensitive information or the utility of the data.

Where PETs could unlock insights or services of significant value to the public:

PETs can help education data owners, particularly agencies, improve student service delivery, research on student outcomes, internal agency operations, and compliance reporting. If educational agencies incorporated PETs into their technical infrastructure, such data could be linked with other agency data that could be used to identify both short-term problems and long-term trends, and better deliver wraparound services to students in need. For example, PETs could allow for the linkage of a state education agency’s data with that state’s housing agency data, in order to direct academic tutoring, mental health counseling, and other supports for students

whose families are facing housing insecurity. Some education agencies are already using PETs in their jurisdictions, such as the State of Washington’s Education Research Data Center (ERDC), which uses [a secure virtual enclave](#) to allow external researchers to access de-identified school and workforce data, and Oklahoma’s Birth through Eight Strategy for Tulsa ([BEST](#)) program, which [uses](#) secure hashing to link data across service providers in order to break the cycle of intergenerational poverty.

Where PETs can reduce the risk of unintentional disclosures:

Sensitive education data is accessed by agency officials, researchers, edtech tool managers, and service providers on a regular basis. Not only is students’ information accessed by various approved users every day, but it is analyzed in home-built enclaves, linked to outside data owners, and sent between agency officials for compliance reporting and service delivery. School information systems are unfortunately becoming more [frequent](#) targets for hackers and [ransomware](#). Additionally, edtech companies are concerned about the open-ended, short answer responses in their data. Such freetext fields are hard to anonymize (and easier to suppress or redact). PETs can help reduce the risk of unintentional disclosures by masking identifiers, automating access rules, and altering outputs to protect individuals and groups from harm. For example, query servers and trusted execution environments can allow for highly protected spaces through which members of the public can view their state’s education data trends, secure hashing reduces the amount of PII such as SSNs being exchanged between parties for data linkages, and functional homomorphic encryption encrypts the education data while its being computed upon. This all could reduce the likelihood of unintentional internal and external breaches as well as deliberate attacks.

PETs and equity:

There are important equity implications of using PETs in education. Using PETs to better protect student information can enable more granular stratification of data without jeopardizing privacy. For example, without PETs, a district may be hesitant to analyze average test scores by race, disability status, or other groups that yield small sample sizes, for fear of those students being able to be re-identified by bad actors. Agencies could also share encrypted data with partner agencies that provide complementary services in the community, as mentioned in the above homeless services-education data linkage example. Currently, agencies at all levels of government are often very reluctant to share data with other related agencies, and when they do, the processes to build trust and develop data sharing agreements are arduous. PETs could enable quick, safe sharing that does not violate federal privacy laws.

We strongly suggest that OSTP and NITRD consider education as a critical sector that would benefit from the adoption of PETs, and partner with offices within the Department of Education such as the Student Privacy Policy Office (SPPO) and the Privacy Technical Assistance Center (PTAC) to develop regulations governing the adaptation of PETs in education. National guidance and standards would enable state and local education agencies to adopt PETs that best fit their needs, protect their students’ privacy, and enable linkage across agencies to better serve their constituents.

We thank you for your time in reviewing these comments and are looking forward to the Office of Science and Technology Policy's efforts in the privacy space.

Sincerely,

Amy O'Hara
Research Professor
Massive Data Institute
Georgetown McCourt School of Public Policy
Director, Georgetown Federal Statistical Research Data Center

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Google

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



Office of Science and Technology Policy
Request for Information on Advancing Privacy-Enhancing Technologies
Doc. No.: 2022-12432; 87 Fed. Reg. 35250
July 8, 2022

Google welcomes the opportunity to provide comments in response to the Office of Science and Technology Policy Request For Information on Advancing Privacy-Enhancing Technologies (PETs).¹ Our comments describe Google’s approach to PETs, including current applications. We conclude with policy recommendations, including public support, funding for research, regulatory incentives, technical standards and practices, and expert guidance for consideration in development of a holistic and flexible national strategy that facilitates responsible use of PETs.²

Emerging Privacy Tools And Techniques

New and emerging tools and techniques offer effective ways to safeguard and enhance privacy and security while enabling society to unlock the immense benefits that can be obtained from responsible use of data and technology across contexts, including research and commercial applications. As part of our continuing investments to drive innovation and make a safer ecosystem for Internet users, Google has invested significant effort in developing PETs, implementing them across our products, and making most of our research and tools open source so that anyone, anywhere can benefit from our advancements and contribute to progress in the field of PETs.³ We also make a number of packaged solutions available to enterprise customers through Google Cloud, enabling them to take advantage of advanced privacy technology without having to make significant independent investments in research and tooling.⁴ We discuss some specific PETs Google is investing in below.

¹ Google uses the terms “privacy-enhancing technologies” (PETs) and “privacy-preserving technologies” (PPTs) interchangeably and may use either term in the research cited herein.

² <https://research.google/research-areas/security-privacy-and-abuse-prevention/>.

³ In the commercial context, advancing PETs that both improve privacy and provide utility, will lead to more businesses and organizations embracing PETs. This in turn can improve privacy and data protection outcomes, and further enhance trust in the organizations that use PETs and in the data sharing economy at large. This is one of the driving principles behind Google’s [Privacy Sandbox Project](#), which seeks to collaborate with industry by applying PETs to evolve existing digital ads practices.

⁴ For example, [confidential computing](#) extends encryption to situations where the data is in use, unlocking new possibilities for collaboration, while preserving confidentiality of underlying

Differential Privacy

Differential privacy (DP) is an anonymization technology that adds specifically crafted noise to data or computations and provides a mathematical framework to quantify and understand the privacy guarantees of a system or an algorithm. DP is flexible and can be applied at the point of data collection (e.g. during survey collection), release (e.g. sharing a dataset), and model training (e.g. learning from data) to prevent identification of individuals. Because of its flexibility and ability to guard against privacy attacks such as reconstruction and memorization, it is suitable for many applications.

DP also can be used to share and analyze data in a privacy-respecting way across silos to train more robust machine learning (ML) models, while retaining privacy guarantees. For example, medical researchers may want to develop a ML model that predicts the prevalence of a disease such as cancer. Combining medical x-ray data across multiple institutions could produce a more accurate model but expose private patient data. If a model is trained on the aggregate data with DP, it could be more accurate and also provide a mathematically strong assurance about the privacy of the individuals' data.

At Google, we use DP in different applications including [Android keyboard text prediction with provable privacy guarantees](#), [traffic optimization on Google Maps](#), and the open source release of [mobility metrics due to COVID-19](#). To democratize access and enable collaboration and use that advances the state of the art, we make our DP research and technologies available through open-sourced projects such as the [Differential Privacy Library](#) (for data analysis) and [TensorFlow Privacy](#) (for ML).

It is important to note that DP is not suited for every application: use of DP can contribute to inefficiency and slower processing and impact utility and model accuracy. Consequently, DP is not appropriate for long-tail analysis. For instance, if a researcher wants to answer questions on a small population, the relative negative impact of the noise that DP introduces will likely be large. The impact of DP on fairness and equity is an area that merits further research. The

data. We make this advanced technology accessible to customers through [Confidential VMs](#) and [Confidential GKE Nodes](#). Another tool, [Data Loss Prevention \(DLP\)](#) helps customers protect against privacy breaches by scanning customer data against over 150 known information types to automatically identify, classify, mask, tokenize, and transform sensitive elements. DLP can also help measure how well quasi-identifiers are preserving data privacy through properties such as k-anonymity and l-diversity. And, because privacy requirements are increasingly baked into compliance frameworks, Google Cloud also offers tools that make it easier for customers to easily and confidently apply requisite controls to their data sets. [Assured Workloads](#) provides customers an easy mechanism for [applying FedRAMP High, FedRAMP Moderate, IL4, and CJIS compatible controls](#) to their workloads. For more, please see our overview of [products and services relevant to data controllers](#) and our [privacy resource center](#).

long-tail problem becomes more acute when training ML models. By erasing the long-tail, DP prevents models from learning the behavior of small groups. Language translation provides a helpful example. If a small group speaks a rare language, then a model might not be able to learn that language, and would not be able to provide recommendations in that language. Thus, recommendations for advancing DP must consider how its application can interact with other important objectives, while privacy research continues to try to close these gaps.

Federated Learning And Analytics

[Federated Learning](#) (FL) is a data minimization technology developed at Google that enables state-of-the-art ML without centralized data collection. Using FL, organizations can train ML models with information from real-world interactions with people and improve their functionality over time, without needing to collect and store the underlying user data. Instead, the system computes model adjustments on the raw data wherever it resides (e.g. on users' devices) and only makes the aggregate model available to downstream systems and engineers.

In its original form, FL enables ML models to learn and improve over time without personally identifiable information ever leaving a user's device. Google uses FL to power [smart text selection features](#) in the [Android operating system](#), ["Hey Google" detection by Google Assistant](#), and [Smart Reply in Android Messages](#), enabling more powerful and personalized services, and features that are accessible in no- or low-connectivity settings, while limiting the data that leaves a user's device. New applications of FL extend these capabilities across enterprises. For example, FL can be used to enable collaboration among companies and institutions, by allowing them to run analyses on their combined data without requiring the raw data to be shared. It can also be used to power privacy-preserving ML and analytics in data centers, allowing ML models to be trained with data from multiple silos without combining the data into large, centralized datasets. Google makes FL accessible to researchers and developers through the open-sourced [TensorFlow Federated](#) framework.

Because data is not collected in a data center when applying FL, we cannot use traditional methods to assess the quality of the model on real-world data. Federated Analytics (FA) addresses this by using the same federated infrastructure to compute metrics on device, and average those metrics to get population-level summaries of how models perform. For example, FA is used to show how accurate next word prediction models are in Gboard. FA also can be used to compute aggregate answers to data science queries over decentralized datasets, such as the data distributions or event frequencies.

Federated technologies have significant data minimization benefits and can also be combined with techniques like DP and Secure Multiparty Computation (see below) to further enhance privacy. But, they also come with tradeoffs. In on-device deployments, federated technologies

are limited by the available storage and computing power, meaning they may not be appropriate for certain applications. Furthermore, not collecting data in a centralized location limits organizations' ability to run certain types of analyses that require access to the raw data such as debugging or some types of fairness audits. As for the limitations impacting DP, these considerations should be taken into account in recommendations and further research to advance PETs.

Fully Homomorphic Encryption And Secure Multiparty Computation

Google uses a number of cryptographic protocols⁵ to help users stay safe across the Internet. For example, fully homomorphic encryption (FHE) enables computation on encrypted data without revealing the underlying data. Some recent advancements in the field of FHE make it easier to implement. However, its use still requires cryptographic expertise which poses a barrier to greater access and use. Lowering the barriers to FHE research and adoption was a driving reason behind our work last year to open source a solution that enables any developer to create a program that can process encrypted data without decrypting it.⁶

Secure Multiparty Computation (SMPC) is another advanced cryptographic technique that provides utility from private data while strongly protecting the privacy and security of individual-level data. Generally it can provide guarantees that no entity learns anything more than their specified output, with proof that learning more is equivalent to breaking a cryptographically difficult problem (e.g. factoring, encryption), which is computationally infeasible and thus extremely unlikely to be achieved.

Google has explored use in several applications: [Private-Join-And-Compute](#) to allow two organizations to compute statistics across their private datasets; [Secure Aggregation](#) to enable a server to learn a federated learning model update from a group of users, while keeping each individual user contribution confidential; [Private Information Retrieval](#) to equip user devices to retrieve database entries from a server, while protecting the retrieved entry from the server; and [Private Set Membership](#) to equip user devices to check membership of an identifier in a set held by a server, while protecting the queried identifier from the server and the database from the users. We have also deployed SMPC through [Exposure Notifications](#) [Private Analytics](#) to allow health authorities to collect metrics about COVID-19 exposure notifications without learning the specific contribution from any individual device.

Generally, applications involving sensitive data or processing (e.g. financial, health), and especially those that would involve combining data from multiple sources, would all greatly benefit from the use of SMPC. As noted above, these could vary in operation between organization-to-organization SMPC, aggregate metric collection by a single organization, or

⁵ <https://security.googleblog.com/2019/02/protect-your-accounts-from-data.html>.

⁶ <https://developers.googleblog.com/2021/06/our-latest-updates-on-fully-homomorphic-encryption.html>.

devices using MPC to privately leverage a server's expertise. SMPC is especially a good fit for simple functionalities like retrieval, counting, or aggregate statistics like averages, regressions or low-depth ML.

However, SMPC can add computational and communication costs; additional engineering and implementation burden; and, similar to FHE, requires substantial subject matter expertise to implement accurately. While this is a steadily advancing research and development area, adoption to date is limited to very specific use cases and largely in the confines of academia and some startup businesses and large tech companies. Because of these limitations, complex functionalities that are difficult to run even without SMPC (e.g. building an internet-scale knowledge graph) are likely to be an infeasible fit for SMPC in the near and medium-term because of the overheads introduced.

Opportunities And Recommendations

We welcome the development of a comprehensive and multifaceted national strategy to significantly expand the use of PETs to benefit individuals and society. We agree with the RFI's assessment⁷ that despite PETs' significant benefits, they have yet to achieve widespread adoption. Below we describe opportunities for government action and provide recommendations to achieve the important goal of greater privacy of, and utility from, data. Specifically, we would encourage a national strategy to:

- Champion the use of PETs through public commentary, policies, and government applications;
- Support federal investments in open fundamental and application-specific research and development, through additional funding, technical research, additional prize challenges, and other mechanisms for driving the development of PETs;
- Adopt a flexible, risk-based approach, that accounts for factors such as the sensitivity of data, available protections, and costs;
- Incentivize the development and adoption of PETs through smart regulation;
- Ensure that recommendations align with recognized privacy standards, and support the development of standards and recommended practices for implementation of PETs; and
- Provide guidance to facilitate the responsible use of PETs.

Champion Use Of PETs

PETs should be a staple of the Administration's public commentary about privacy and security. Administration officials should use their public platforms to raise awareness of PETs, share

⁷ <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>.

helpful explanations, and encourage their use, particularly outside of the tech sector.⁸ A national strategy should also encourage the use of PETs in government contracts and acquisitions, where feasible, and incorporate the use of PETs in existing and future government data collection, processing, and sharing practices.

Support Open Fundamental And Application-Specific Research

A national strategy should support fundamental research to develop and refine PETs and technologies through research organizations like the National Science Foundation and Defense Advanced Research Projects Agency, educational institutions, and through prize challenges like the one recently created by the US and UK governments.⁹ Government funding should address the dearth of expertise required for continued development and use of PETs and expand pathways to enable more people from unserved and underserved communities to build the necessary skills.

A national strategy should also support a strong open-source ecosystem, including by helping to contribute libraries and frameworks, that ease the development and deployment costs of PETs.

Some PETs are promising yet relatively nascent, such as synthetic data generation and zero shot learning, which could encourage the use of less sensitive data, where feasible. Support for research into these technologies can contribute to their advancement and help make their use more widespread.

Adopt A Flexible, Risk-Based Approach, That Accounts For Factors Such As The Sensitivity Of Data, Available Protections, And Costs

While the diversity of PETs deployed across Google to enable secure and private data use helps illustrate the value of a national strategy, it also underscores the imperative of flexibility. No single PET is a panacea. A number of factors influence which PETs are appropriate for a particular product or use case. These can include the sensitivity of data, whether data is individual or aggregated, how data is being used, the intended outcome, the impact on functionality, and the size of and resources available to the implementer, among others. PETs

⁸ For example, if a national strategy were to encourage companies to use SMPC to "exchange" nothing more than the output (since the intermediates are provably protected), more companies would see the benefit of the technology, and invest in its development, (including by making open-source contributions).

⁹

<https://www.whitehouse.gov/ostp/news-updates/2022/06/13/u-s-and-uk-governments-collaborate-on-prize-challenges-to-accelerate-development-and-adoption-of-privacy-enhancing-technologies/>

may also impose costs, whether in the form of increased expense to the implementing party¹⁰ or reduced functionality to end users. Because research and development into PETs is ongoing, the cost-benefit analysis is evolving.

A national strategy to advance the development and adoption of PETs must be similarly flexible to account for the myriad of use cases and organizations seeking to implement them. Rather than prescribe particular PETs or applications of PETs, a strategy should be risk-based, and provide guidance on appropriate use cases and safeguards, taking contextual factors and national and international standards into account.

A strategy that focuses only upon the most advanced technologies or upon particular organizations or use cases would exclude most organizations and fail to provide a comprehensive vision for the embrace of PETs. On the other hand, overly prescriptive, technology-specific approaches can discourage organizations from taking advantage of the most advanced PETs if regulations fail to keep up with research and development.

Incentivize The Development And Adoption Of PETs Through Smart Regulation

Drafters of privacy and data protection law in the US and around the world have recognized the need to balance the protection of data and privacy with important societal goals that can be advanced through the use of data. This objective is reflected in the appropriate exceptions for de-identified or anonymized data across many different pieces of privacy and data protection legislation. We recommend that the national strategy reflect this approach and encourage the use of PETs by granting similar exceptions to those found in privacy and data protection regulations. This is important given the costs PETs can pose to organizations, particularly those less-resourced. If organizations do not have a reasonable degree of certainty that they will benefit from implementing PETs, such as through reductions in compliance costs for exempt data, then they may be discouraged from taking on the increased cost of implementing these technologies.

It is also important to ensure that laws and policies do not inadvertently prevent or disincentivize the use of PETs in products. For example, requirements for the auditing or disclosure of datasets effectively require that those datasets be collected and stored centrally, precluding the use of techniques like FL and data minimization. Similarly, requirements around data accuracy should include clear exceptions to allow the introduction of noise into data to utilize DP.

¹⁰ Such costs can include technical hardware, additional compute power and time, hiring of expert personnel, and the time and expense of changing organizational collection and use of data.

Ensure That Recommendations Align With Recognized Privacy Standards, And Support The Development Of Standards And Recommended Practices For Implementation Of PETs

National and international consensus standards are essential to the health of the global technology ecosystem, promoting cross-border and cross-application interoperability and minimizing barriers to trade and innovation. They can also help to ensure that PETs are implemented responsibly and build a common understanding of the privacy-enhancing benefits and tradeoffs of these technologies. A national strategy should align recommendations with widely accepted privacy standards, including the [National Institute of Standards and Technology Privacy Framework](#), and International Organization for Standardization and Institute of Electrical and Electronics Engineers standards,¹¹ and where needed, support further development of national and international standards and recommended technical practices.

Provide Guidance To Facilitate Responsible Use Of PETs

Additional guidance and recommended practices for using PETs and managing their potential tradeoffs (e.g. between privacy and utility) would help organizations use them effectively, and encourage broader adoption. These recommendations may be general or technique-specific. For example, differential privacy is governed by the privacy loss parameter, epsilon, which affects the accuracy of the data, but also the privacy properties of the result. A lower epsilon means more privacy, but increases data loss. A higher epsilon is good for data utility, but a significant degradation of privacy. Practitioners have struggled to find the right balance between utility and privacy in these situations, and some guidance on how to balance competing equities when using PETs, particularly across various applications and contexts, would be welcome. Additionally, more information and guidance on how best to use some underutilized PETs such as SMPC and other advanced cryptographic techniques would also very likely expand adoption.

In addition, guidance around joint use of PETs to protect against a broader variety of privacy risks (e.g. FL and DP used together provide both data minimization and anonymization benefits) would be particularly welcome. A national strategy should make every effort to avoid an either-or approach to PETs and encourage joint usage where appropriate.

Conclusion

PETs play a critical role in the digital ecosystem by offering effective ways to safeguard and enhance individual privacy and data protection while allowing society to unlock the immense

¹¹ ISO/IEC PWI 6089 Guidance on addressing privacy protections for artificial intelligence systems may be a particularly strong opportunity.

benefits that can be obtained from the increased use and study of data. The protection of individual privacy is critical, and must also be balanced with other important societal goals that can be advanced through the use of data; PETs like DP, FL, FHE, and SMPC can help to strike that balance. PETs remain a top priority for Google as we seek to offer more secure, private experiences to our users on Google products and services and across the Internet, while also seeking to unlock the value data can bring to improving our products and helping all Americans. We welcome the opportunity to support OSTP's continued work to prepare a national strategy, and facilitate the responsible development and deployment of PETs.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

HUB Security Limited

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



HUB Security's response to the Office of Science and Technology Policy's Request for Information on Advancing Privacy-Enhancing Technologies

July 8, 2022

Organization	Hub Security Ltd
Organization type	Industry - Cyber Security
Respondent Person and Contact	Name: Gaurav Sharma Email: Role: VP, North America



Table of Contents

About HUB Security	3
2. Specific technical aspects or limitations of PETs	4
3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs	6
7. Risks related to PETs adoption	8
10. Other information that is relevant to the adoption of PETs	9



About HUB Security

HUB Security Limited ("HUB") was established in 2017 by veterans of the 8200 and 81 elite intelligence units of the Israeli Defense Forces. The company specializes in unique Cyber Security solutions protecting sensitive commercial and government information. The company debuted advanced confidential computing solutions aimed at preventing hostile intrusions at the hardware level while introducing a novel set of data theft prevention solutions.

HUB Security, a frontrunner in confidential computing hardware and software offerings, brings a fundamentally unique approach to prevent and mitigate the impact of present sophisticated cyber attacks by protecting the entire computing stack. With its military-grade ultra-secure hardware and software, HUB Security is capable of preventing cyber threats in any kind of environment while protecting critical assets such as data for all market segments. With HUB's technology, the compute platform is fully secured in one appliance or can turn any other computing environment into a secure compute platform.

Central to HUB's confidential computing solutions is its secured execution environment. The solutions are based on our hardware and software design used to protect any application and any data. We leverage a range of security mechanisms and a new security paradigm, HUB's confidential computing, aiming to isolate any application and associated elements within its own secure execution environment. This includes the application, data, AI models, policies for access or approval, audit trails and logs, and all cryptographic keys, services and key management.

HUB operates in over 30 countries and provides innovative cybersecurity computing appliances as well as a wide range of cybersecurity services worldwide.



2. *Specific technical aspects or limitations of PETs*

The multiple existing approaches to PET constitute a patchwork of narrow solutions to the specific problems in the privacy domain of data and applications. Each technique creates a basic tradeoff between the capability to achieve the application goals and the level of privacy protection. Several leading examples are:

1. Homomorphic encryption techniques enable complete privacy in theory by enabling encrypted applications to work on the encrypted data sets. The main penalty is extreme computational load, which makes large scale use in the cloud and in real time impractical and prohibitively expensive. Moreover, the data and the application code must be prepared in advance to be compatible with the technique, thus creating cumbersome and specific processes as a barrier to automation and simplicity of use.
2. Another option is to keep the sensitive data at the location of its origin and bring the applications to the data location, as in federated learning technique for AI models. The AI model trains on the subsets of the entire dataset and is enhanced by exchanging the training anonymized metadata coefficients. In this case the data privacy is preserved relative to the cloud provider, but the AI model accuracy and quality are reduced, since it is not trained on the entire combined data set from all sources.
3. The third option is putting the data into proprietary hardware secure enclaves such as Intel SGX, AMD SEV, which are basically just RAM partition inline encryption by the CPU. All the data and application code in the encrypted space of the CPU memory can be accessed only from the inside of this space. CPU proprietary hardware enforces access control to the allocated memory address space. This approach is growing today in popularity at the public clouds (Microsoft Azure, Google) as the solution to the Confidential Computing challenge. This is a very attractive option as the CPU is already present in the servers. It protects against local malicious access attempts to the specific memory regions - and that's it. It does not handle the Admin insider threat or remote access to the encrypted partition via correct credentials. Furthermore, the proprietary non-public implementation by the manufacturer prevents community review for the protection and exhibits its own vulnerabilities (as has been found multiple times in the mentioned implementations). The end user does not control the encryption keys and does not know the quality of the random numbers used to generate these keys inside the CPU silicon. In any privacy technique controlling the obfuscation key is the basic means of controlling the privacy of the data.

Looking forward, the truly useful PET should not impact the original goals of working with private and sensitive data, by enabling full computational throughput on multiple types of CPUs and providing multi layered protection centered on the data and working against multiple threat types



simultaneously. The main principle of it is isolation between data and data consumers, data sources, data users, infrastructure, etc. The isolation on the finest available level enables an implementation of zero trust architecture and continuous enforcement of the proper access to the data.

Additionally, any PET tool must be easily upgradable, as technology is always advancing and the attackers get access to better tools. For example, the rapid development of quantum computing created a huge gap and clear danger of exposing past, current and future state/enterprise level secrets and all the world communications in the 10 to 30 years from now. The upgrade from classical cryptographic algorithms to post quantum cryptography is looking like a long complicated and expensive process, because the infrastructure was not built to be upgradable. The basic assumption should be that every PET will be broken eventually and so must support easy and quick updates to its core functionality to prevent breach of privacy.



3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

The following are the areas that will benefit and have high potential for adoption of PETs:

- **Multi-Party Analytics:** Analytics and in particular AI usage will be used everywhere to automate and handle the exponentially increasing amount of data. It is a fundamental truism of artificial intelligence that the more data it is fed, the better it performs. This technology can, for example, help hospitals to more efficiently allocate beds, staff, medicine, and accessories. It can provide practitioners with life-saving speed by delivering data & insights at the moment they need it. Similarly, such technologies can facilitate fraud analysis across financial institutions to fight cybercrime.

However, most data is siloed among different health systems, financial institutions, departments, etc to conform to regulations and meet internal security policies. The ability to safely collaborate on AI projects without compromising privacy & security and meeting regulatory requirements is a major challenge. For example, sharing medical images and AI models across multiple medical institutions or fraud analysis across financial institutions, requires a high level of security, privacy and processing power.

Confidential computing can provide the ability to create isolated environments to protect the integrity and privacy of models and data and can be used to safely unlock the data for multi-party analytics.

- **Edge Computing:** Edge computing holds massive potential to transform lives with new innovations and lead to massive economic growth. To give you an idea, IDC is projecting that worldwide spending on edge computing will reach \$250 billion by 2024. Massive amount of data will be generated and consumed at the edge - a stat from Gartner says that data generated at the edge will increase from 10% today to 75% by 2025. This data and the resulting intelligence will become the competitive advantage for enterprises.

With the dramatic increase in broadband due to the Infrastructure Investment and Jobs Act and the desire to reduce long haul bandwidth needs, real time and low latency services will collect and process a dramatic amount of data locally. These services will become a major value add revenue generator. But these are often places where privacy has been challenging yet increasingly critical such as regional healthcare, mobile communications, public safety infrastructure, transportation, financial systems, etc.



- **Industry 4.0:** The Fourth Industrial Revolution (4IR) is an amalgamation of advanced capabilities mixing artificial intelligence (AI), robotics, the Internet of Things (IoT), Digital Twins, quantum computing, genetic engineering, and other technologies. 4IR is driving major transformation in the Industrial sector with smart autonomous systems fueled by data and machine learning. Connected things with tremendous volumes of data have created smart factories and systems that can perform predictive maintenance, remote monitoring and optimize production without human intervention. IoT & Digital Twins have made simulations and maintenance extremely efficient. This area is a natural securing, this is a natural extension of edge computing and AI as discussed above. The massive amount of data and insights are natural starting points for PETs.

All 16 critical infrastructure sectors (as defined by Cybersecurity and Infrastructure Security Agency (CISA)) will be impacted by the above technological advancements and will benefit the most from adoption of PETs.



7. Risks related to PETs adoption

The following are the risks and challenges related to PETs adoption:

Data Utility: Since data serves as the key ingredient that will drive economies and improve lives using innovations mentioned previously, PETs should be implemented to make sure that the data does not get stripped of its expected utility. While implementing privacy and security controls to an individual's data is extremely important, it also needs to be able to be used to its maximum extent to serve the individual and the larger population directly. For example, age or gender might be an important attribute of a person that needs to be treated confidentially, but completely removing it while diagnosing the person for a medical condition can have a catastrophic impact.

Redundant Encodings: On the flip side, PETs can give a false sense of security that good controls have been implemented to preserve privacy but other attributes called redundant encodings can be used to deduce the attribute that has been protected. For example, given the location, condition and race of a person seeking medical treatment, her age can be deduced. A more comprehensive approach is needed to identify all relevant attributes if PETs that protect data privacy by obscurity are adopted and in many cases these approaches might not be feasible to be implemented correctly.

Explainability: When algorithms are used to make decisions, explaining the path used to make the decision can be extremely difficult if the attribute has been protected. For example, why a loan application was denied might be difficult to explain if the data was protected. Some PETs that use a black box approach for analytics or decision making, will pose a challenge both from technical and regulation perspectives. Sectors such as healthcare and financial services, will be heavily impacted by this limitation.

Expertise: Given the complex nature of many PETs, specific skills are needed and at the current moment the talent pool is extremely limited. There are many innovators but large scale adoption needs a concerted effort to promote these approaches and attract the right talent.

Quantum Threat: Many approaches to PETs use complex encryption algorithms but are susceptible to quantum computing based attacks. PETs should consider post quantum readiness to address imminent threats posed by other nation states and adversaries.

In addition to the above, there are many unknown unknowns that can introduce complexity and gaps that require a larger public-private collaboration.



10. Other information that is relevant to the adoption of PETs

Data is the new resource that will introduce many new economic opportunities and transform our lives. Technologies such as Artificial Intelligence, edge computing and others will leverage rich amounts of data to unleash new innovations. These innovations will improve our healthcare, offer equitable access to services, improve our safety and help us address some of the existential threats such as climate change.

However, the same data and technologies can be used to knowingly monitor, deny access, subjugate, oppress or incarcerate individuals or entire sections of a population by an authoritarian regime or unknowingly discriminate, misinform or lead to identity theft by a large corporation. Regulations like CCPA and GDPR have laid the groundwork but there are still too many loopholes that have to be addressed.

PETs have come a long way and there are many approaches that can be used individually or in combination to protect data privacy while facilitating data sharing for the innovations mentioned above. However, PETs need to be combined with the other best practices for data security for increasing their efficacy. One proposed approach is to combine data privacy with security and resiliency efforts led by CISA.

Another key threat is related to quantum computing. President Biden signed a National Security Memorandum in May 2022, outlining the Administration's plan to address the risks posed by quantum computers to America's cybersecurity. However, most PETs currently do not factor the impact of quantum computing to data privacy and there needs to be a concerted effort to factor that threat in.

Effective PETs need to be adopted quickly. Data will unleash many new capabilities that will improve our lives but without effective PETs, the same data can hold us hostage knowingly or unknowingly.

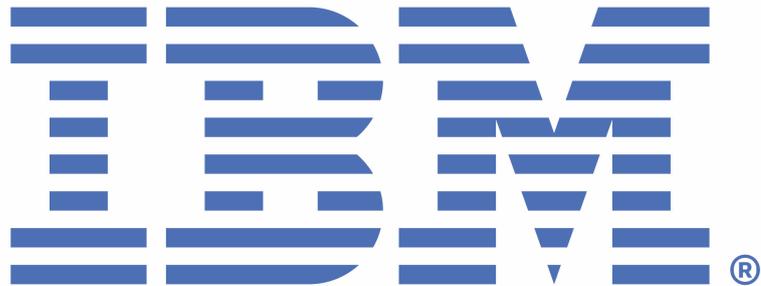
Request for Information (RFI) on Advancing Privacy Enhancing Technologies

IBM Research

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Request for Information on Advancing Privacy-Enhancing Technologies

IBM Research



International Business Machines Corporation

July 8, 2022



July 8, 2022

Subject: Request for Information on Advancing Privacy-Enhancing Technologies.

Dear Team:

IBM Research is pleased to offer the following preliminary information in response to the Office of Science and Technology Policy's Request for Information on Advancing Privacy-Enhancing Technologies.

Our response has been prepared by:

Omri Soceanu
Director AI Security
IBM Research - Security

Nir Drucker
Principal Researcher
IBM Research - Security

John Buselli
Offering Manager
IBM Research – Security

IBM Research has a long history of working closely with United States Federal Agencies to support the development of critical security initiatives. We recognize the importance of this initiative and welcome the opportunity to continue to provide input and guidance toward this effort.

We look forward to future collaboration.

Regards,

John Buselli
Offering Manager
IBM Research
jbuselli@us.ibm.com

Request for Information on Advancing Privacy-Enhancing Technologies

IBM Research

Overview

There are numerous privacy regulations that mandate organizations abide by certain security principles when processing personal information. These principles go beyond raw data. Recent studies have shown that a malicious third party, with access to a trained Machine Learning (ML) model, even without access to the training data itself, can still reveal sensitive, personal information about the people whose data was used to train the model. It is therefore crucial to be able to recognize and protect Artificial Intelligence (AI) models that may contain personal information.

In addition to regulation mandates surrounding personal identifiable information, many organizations are enacting stronger privacy related policies and frameworks to keep their confidential and sensitive data safe. Some of the major policies and frameworks already enacted by large organizations focus on Zero Trust, Key Management, Secure Multi-Party Computation (SMPC) and Access Management approaches.

IBM Research is working on several novel techniques and tools to both assess the privacy risk of AI-based solutions, and to help them adhere to any relevant privacy requirements. We have developed tools to address the different tradeoffs between privacy, accuracy and performance of the resulting models, and for addressing the different stages in the ML lifecycle. These developments include:

AI on encrypted data - Fully Homomorphic Encryption (FHE) allows data to remain encrypted even during computation. Using FHE we are able implement a wide variety of analytics and AI solutions over encrypted data.

Differential Privacy - This method allows queries to be executed on sensitive data while preserving the privacy of individuals in the data with its robust mathematical guarantees. Differential privacy relies on random noise to protect individuals' privacy while preserving accuracy on aggregate statistics and has applications in ML and data analytics more generally.

ML anonymization - This method creates a model-based, tailored anonymization scheme to anonymize training data before using it to train an ML model, enabling to create ML models that no longer contain any personally identifiable information.

Data minimization - This technique helps to reduce the amount and granularity of features used by machine learning algorithms to perform classification or prediction, by either removal (suppression) or generalization techniques.

Privacy risk assessment - We have developed ways to assess and quantify the privacy risk of ML models and to enable comparing and choosing between different ML models based not only on accuracy but also on privacy risk criteria.

Privacy in Federated Learning - Federated Learning (FL) is an approach used in machine learning in which a group of parties (data owners), work together to train a model collaboratively without sharing training data. FL enables the exchanging and merging the parameters of locally trained models. For increased privacy this approach can be combined

Request for Information on Advancing Privacy-Enhancing Technologies

IBM Research

with other techniques such as differential privacy, homomorphic encryption and secure multi-party computation.

We offer the following for additional consideration:

Fully Homomorphic Encryption

Fully Homomorphic Encryption (FHE) is a promising solution that has been getting significant attention because of its ability to perform an evaluation of certain functions on encrypted inputs. Specifically, FHE is using different analytics such as logistic regression, neural networks (NNs), or decision trees to provide inference results on private data (i.e., customers, patients, employees, transactions). The potential of FHE to handle private data was captured e.g., by Gartner, which predicted that fifty percent (50%) of large organizations will adopt such computational models for processing in untrusted environments for multiparty data analytics by 2025 [1]. Sectors that will most likely benefit from FHE are the Health, Finance and Government sectors who can collaborate and perform complicated functions over their private data, while keeping it confidential.

Several major barriers limit the wide adoption of FHE solutions to date namely: latency, memory usage, storage, key management, regulations, and standardization. IBM Research understands these barriers and invests in fundamental and applied research for ways to solve the first three dimensions, namely, latency, memory usage and storage, while collaborating with other companies on standardizing the available FHE solutions. We elaborate next on the standardization, key management and privacy over FHE barriers.

Standardization: NIST FIPS 800-57 [2] [Section 4] ('Recommendation for Key Management Part 1 – General') refers to three (3) types of approved cryptographic algorithms: hash functions, symmetric-key algorithms, and asymmetric-key algorithms. Nevertheless, it does not explicitly mention FHE or refers to it. One possible reason is that until a decade ago, FHE was considered impractical, and only a small number of organizations have experimented with it. In fact, standard organizations such as NIST were asking the cryptographic community to focus on other types of cryptographic algorithms, such as light-weight cryptography [3], or post-quantum cryptography [4]. Recently, this situation has changed, where organizations such as NIST [5], The Open Industry / Government / Academic Consortium to Advance Secure Computation [6], and ISO/IEC [7] already started to consider FHE standards. In IBM we believe that having such standards and including FHE primitives as approved primitives for FIPS 140-2 [8] based solutions, may speed up the adoption of FHE solutions by many organizations.

Key Management: Cryptographic keys play an important role in cryptographic algorithms. Having unique and well-formatted keys is a prerequisite for the security guarantees that these cryptosystems provide. However, once an adversary puts its hands on these keys, the associated cryptographic scheme can no longer guarantee the confidentiality or the integrity of the key owner's data.

Generalizing the Key Management Systems (KMS) recommendations [1] to support different FHE schemes such as CKKS, TFHE, and BGV with different key characteristics

Request for Information on Advancing Privacy-Enhancing Technologies

IBM Research

involves five (5) types of keys: secret, public, evaluation, rotation, and bootstrapping keys. Each type has a unique size and usage characteristics, which a KMS solution should support. In addition, access control mechanisms to these keys should be defined based on a valid trust model and corresponding security assumptions. For example, when a hospital would like to provide a cancer detection service to its clients using some untrusted cloud environment. The hospital (data owner) should encrypt the model and then either give access to the private key to the users to use the model, or should provide some other mechanism to do so, such as multi-key FHE or proxy-re-encryption. A KMS solution should provide definitions for the hospital on how and where they are allowed to store the different keys, who should get access, and how often the keys should be rotated. While IBM is working on developing such KMS solutions, we expect wide adoption only after standards are adopted for such exchanges.

Hardware Secure Module (HSM): Current HSMs are designed to meet the requirements of symmetric and asymmetric cryptosystems that uses much smaller keys compared to FHE keys. At IBM Research, we are investigating the means to leverage typical, existing hardware for building a FHE KMS service that uses HSMs (or to suggest new HSM designs that meet FHE requirements). Standardization will likely increase the adoption of such solutions.

Hybrid encryption (transciphering): To address the issue of the ciphertext size and computational overload on edge devices, a transciphering framework, also called hybrid encryption was suggested [9]. The idea is that the IoT/edge device use some symmetric encryption scheme to encrypt its data and send the ciphertexts, together with the encryption of the key under FHE, to the server which can decrypt the data "under" FHE. This allows faster communications (at the cost of extra latency) for the decryption on the server-side. One known barrier is that current symmetric encryption schemes are not considered efficient under FHE. One possible solution is to standardize new symmetric ciphers that behave more efficiently under FHE.

AI Privacy

There is a known tension between the need to analyze personal data to drive business outcomes and the need to preserve the privacy of data subjects. Many data protection regulations, including the EU General Data Protection Regulation (GDPR) and the California Consumer Protection Act (CCPA), set out strict restrictions and obligations on the collection and processing of personal data.

Many data processing tasks nowadays involve machine learning. In recent years, several attacks have been developed that are able to infer sensitive information from trained models, including membership inference attacks, model inversion attacks and attribute inference attacks. This has led to the conclusion that machine learning models themselves should, in some cases, be considered personal information.

In 2019 the British Information Commissioner's Office published an AI Auditing Framework which specifically mentions purpose limitation and data minimization, fairness, transparency, accountability and many more considerations. In 2020, the European

Request for Information on Advancing Privacy-Enhancing Technologies

IBM Research

Parliament published a study on the impact of GDPR on artificial intelligence, also mentioning purpose limitation and data minimization. In 2021, UNESCO published a draft Recommendation on the Ethics of Artificial Intelligence, mentioning privacy and data protection, privacy by design and privacy impact assessments as some of the recommended practices, and is currently in the process of designing an ethical impact assessment (EIA) tool for AI. Also in 2021, the European Commission proposed a draft regulation on trust in AI, becoming the first governmental body in the world to issue a draft regulation aimed specifically at the development and use of AI. The National Institute of Standards and Technology NIST (under DoC), on a directive from the American Congress, just published an initial version of an AI risk management framework (RMF).

Recent surveys indicate that organizations are currently struggling with building AI solutions that involve personal data. In addition, security and privacy of data for ML, as well as building trustworthy and ethical AI remain great challenges. This problem is exacerbated by reports predicting that privacy-preserving techniques for AI model training will unlock up to 50% more personal data for model training and 70% more AI collaborations in industry.

We have identified three (3) main areas of research and innovation we believe will be crucial in the next few years:

1. Privacy risk assessment of models - this is the first step to understanding which models pose a privacy risk, enable comparing between model alternatives based on privacy criteria (and not only accuracy), and to prioritize models for further action and possible mitigation strategies. Moreover, the increased use of third party or publicly available models, as well as the emergence of AI insurance companies, increase the need for AI privacy auditing. Risk assessment can be either theoretical or empirical, however we believe that a quantitative approach is critical to enable scaling and automation of this complex and time-consuming task.

2. Easy to consume privacy-preserving AI technology - many privacy practices and methods require deep expertise and/or consist of invasive techniques that require significant changes to existing workflows. We believe that privacy practices must be easily incorporated into existing ML-Ops pipelines to become widely adopted. This means supplying non-invasive techniques, with standard APIs that can be added as an additional step into existing pipelines, rather than replacing or disrupting existing practices. They should also support the principle of separation of concerns, for example by applying privacy it should not require deep data science expertise, and vice versa.

3. Compliance with data protection regulations for AI models - as mentioned earlier, ML models are not exempt from data protection principles such as purpose limitation, data minimization and the right to be forgotten. Therefore, specially tailored solutions for applying these principles in the domain of ML must be designed and implemented. Clear guidelines on how these principles should be regarded for ML models are still lacking. If a person requests deletion of their personal information, how does this apply to ML models that were trained on this data? How can deletion be measured or verified and what

Request for Information on Advancing Privacy-Enhancing Technologies

IBM Research

constitutes a "good enough" solution? Is exact deletion required or can approximate unlearning be applied?

For more information on this and related security topics please visit:

<https://w3.ibm.com/w3publisher/ibm-research-security>

- [1] Gartner. 2021. Gartner Identifies Top Security and Risk Management Trends for 2021. Technical Report. <https://www.gartner.com/en/newsroom/press-releases/2021-03-23-gartner-identifies-top-security-and-risk-management-t>
- [2] Elaine, B.: NIST Special Publication 800-57: Recommendation for Key Management Part 1 – General (2021). <https://doi.org/10.6028/NIST.SP.800-57pt1r57>
- [3] NIST: Lightweight Cryptography <https://csrc.nist.gov/projects/lightweight-cryptography> (2021), last accessed 30 Sep 2021
- [4] NIST: Post-Quantum Cryptography <https://csrc.nist.gov/projects/post-quantum-cryptography> (2021), last accessed 30 Sep 2021
- [5] NIST: Toward a PEC Use-Case Suite (2021) <https://csrc.nist.gov/CSRC/media/Projects/pec/documents/suite-draft1.pdf5>
- [6] Albrecht, M., Chase, M., Chen, H., Ding, J., Goldwasser, S., Gorbunov, S., Halevi, S., Hoffstein, J., Laine, K., Lauter, K., Lokam, S., Micciancio, D., Moody, D., Morrison, T., Sahai, A., Vaikuntanathan, V.: Homomorphic encryption security standard. Tech. rep., HomomorphicEncryption.org, Toronto, Canada (November 2018), <https://homomorphicencryption.org/standard/2>
- [7] ISO/IEC: ISO/IEC 18033-6:2019 IT Security techniques – Encryption algorithms – Part 6: Homomorphic encryption (2021), <https://www.iso.org/standard/67740.html10>
- [8] NIST: FIPS PUB 140-2: Security Requirements for Cryptographic Modules (2002). <https://doi.org/10.6028/NIST.FIPS.140-212>
- [9] Naehrig, M., Lauter, K., Vaikuntanathan, V.: Can homomorphic encryption be practical? In: Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop, pp. 113–124. ACM (2011)

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

IEEE Standards Association (IEEE-SA)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

7 July 2022

To:

Office of Science and Technology Policy (OSTP) on behalf of the Fast Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics of the Subcommittee on Networking and Information Technology Information Research and Development (NITRD).

National Science and Technology Council.

National Artificial Intelligence Initiative Office.

NITRD National Coordination Office.

PETS-RFI@nitrd.gov

RFI Response: Privacy-Enhancing Technologies

The IEEE Standards Association (IEEE-SA) is pleased to submit comments to the Office of Science and Technology Policy's Request for Information on Advancing Privacy-Enhancing Technologies.

IEEE SA is a globally recognized standards-setting body within IEEE, the largest organization of technology professionals in the world. We develop consensus standards through an open process that engages industry and brings together a broad stakeholder community.

As stated in the RFI, Privacy-Enhancing Technologies (PETs) present a key opportunity to harness the power of data and data analysis techniques in a secure, privacy-protecting manner.

However, the RFI noted to date, PETs have not achieved widespread adoption due to a variety of factors, among them, limited technical expertise, perceived risks, and financial cost. It can also be noted that there is a need for more collaborative research and application development.

In response to the OSTP request for information on potential specific actions that would advance the adoption of PETs in a responsible manner, IEEE SA provides the following comments:

One approach to advance PETs research opportunities is to use secure multiparty computation (sMPC) developed by Boston University (<https://multiparty.org/>). It is a proven research technique for privacy preserving data sharing and analytics to inform policy. The City of Boston used sMPC to conduct a study of 60 companies male-female pay gap [Multi-party computation helps address Boston's male-female pay gap](#), with the results of the study posted at <https://www.bu.edu/hic/2021/12/09/addressing-the-gender-racial-wage-gaps-with-data-science/>

Additionally, NSF funded the Northeast Big Data Innovation Hub headquartered at Columbia University, this program has a [National Student Data Corps \(NSDC\)](#) that provides free, open, online data science education, mentoring and career resources. The PETs can be taught through an open curriculum that can be created for the NSDC website, and taught in monthly NSDC webinars, serving the NSDC community of 3,000 individual members (mostly students) from 48 U.S. States and 20 countries.

The NSDC mentoring program can include sMPC and other PET content and mentors. There is an NSDC slack channel with 450 members including mentors and mentees.

Healthcare is a key area for potential use for PETs, as we work to leverage more contextual data to do precision medicine. Sharing of data across entities that do not have a central authority can be enabled securely and privately through PETs, enabling integration of intelligence from healthcare to national security, to provide insight and “wisdom computing”.

As Oliver Wendell Holmes said “... it is the privilege of wisdom to listen”. By listening to more data and metadata, we create more insight and wisdom to tackle difficult and complex challenges, from healthcare to national security.

In addition, IEEE SA has several standards under development that will help support PETs adoption.

- Developing standard IEEE P7012 **Standard for Machine Readable Personal Privacy Terms** will identify and address the manner in which personal privacy terms are proffered and how they can be read and agreed to by machines. [IEEE P7012](#)
- Developing standard IEEE P2895 **Trading Human-generated data** will define a taxonomy which can be used to describe the rules and categories of data rights in data contracts that govern the capture, use, sharing and trade of human-generated data. [IEEE P2895](#)
- Developing standard IEEE P2890 **Recommended Practice for Provenance of Indigenous People’s Data** details the rules by which the provenance of Indigenous People’s data should be described and recorded. [IEEE P2890](#)

We thank NIRD for considering these comments in response to the NITRD Request for Information (RFI) on Privacy-Enhancing Technologies.

We would look forward to further discussions with your agency on the continued use and application of Privacy-Enhancing Technologies. If you have questions, please do not hesitate to contact Karen Mulberry at .

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Immuta

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



Immuta's Response to OSTP's Request for Information on Advancing Privacy-Enhancing Technologies.

Immuta is the leading universal cloud data security platform, providing one platform to automate access control for any data, on any cloud service, across all compute infrastructure. Immuta also offers advanced privacy-enhancing technologies (PETs) to support data re-use and data sharing use cases.

At Immuta, we welcome the effort to advance privacy-enhancing technologies, which should be informed by a rigorous analysis of what PETs actually achieve in terms of privacy and security. We suggest that a risk-based approach to PET adoption is key to put data controllers or processors in a position to choose the right PET or set of PETs for each data re-use or data sharing use case with which they are confronted. In fact, more often than not, PETs should be considered as complementary rather than mutually exclusive. Through this response, we would like to share a summary of our PET evaluation framework¹ as one such risk-based approach. We have found such a framework helps stakeholders identify risks related to the implementation of a variety of PETs in the context of a specific use case.

Our PET evaluation framework relies upon threefold considerations supporting the claim that not all PETs are equal in all circumstances.

1. PETs don't offer the same formal (mathematical) guarantees. Generally speaking, soft PETs don't offer any formal guarantee, while hard PETs do. Still, these guarantees are not necessarily related to the same type of privacy or security risk, even if many of these PETs aim to achieve the same privacy or security goal, i.e. confidentiality.² By way of

¹ Presented at the Future of Privacy Forum masterclass on de-identification on 25 May 2022 at the Computer, Privacy and Data Protection Conference 2022, available at <https://www.youtube.com/watch?v=r1J93rTeeaY>.

² Privacy and Security requirements are usually broken down into subgoals in order to help with risk assessment and/or compliance. Security goals traditionally include confidentiality, integrity and availability, while privacy goals are usually considered to be more numerous and thus cover, on top of the



example consider two hard PETs, one PET could guarantee that every participant in a data set is in a position to deny the content of its record (this is the formal guarantee achieved by local differential privacy), while another could guarantee that within a data source there is at least a k number of identical records, so that no individual record is actually unique (this is the guarantee achieved by k -anonymization). It's important to remember that formal guarantees are intimately related to threat models. Each formal guarantee holds for a predetermined threat model.³

2. PETs may or may not assume a trusted data processor, i.e., an entity in direct contact with the data, for example because it is transforming the data or because it is running computation over the data. The hard/soft distinction is therefore a useful starting point to determine what a specific PET achieves in the context of a particular use case. Soft PETs often assume a trusted data processor. Assurances are thus mainly derived from organizational controls (including legal obligations, monitoring and auditing processes). On the other hand, hard PETs are considered robust to misbehaving, malicious, and negligent data processors. Assurances are therefore mainly derived from mathematics. Often, no second or third party protocol participant is deemed to be in a position to violate relevant privacy or security interests, such as confidentiality. As a rule of thumb, soft PETs include the following PETs: Synthetic Data, Access Controls, Encryption (internal, with key management challenges)⁴, Monitoring, and Auditing. Hard PETs include: Secure Multi-party Computation, Zero Knowledge Proofs, Partially Homomorphic Encryption, Fully Homomorphic Encryption, Differential Privacy, Shuffle Privacy, Local Differential Privacy, and Encryption (external, with no key management challenge).

security triad, data minimisation, purpose limitation or unlinkability, transparency, intervenability, fairness and accountability.

³ See e.g. Marsh, Catherine, et al. 'The Case for Samples of Anonymized Records from the 1991 Census'. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, vol. 154, no. 2, 1991, p. 305. DOI.org (Crossref), doi:10.2307/2983043.; Dankar, Fida Kamal, and Khaled El Emam. 'A Method for Evaluating Marketer Re-Identification Risk'. *Proceedings of the 1st International Workshop on Data Semantics - DataSem '10*, ACM Press, 2010, p. 1, doi:10.1145/1754239.1754271.

⁴ Operational risks of misusing the key are secondary to mathematical guarantees if an organization can decrypt. Therefore, internal use of encryption requires operational assurances.



3. PETs may be utilized in combination to enhance their properties. Let's take the example of Federated Learning, which attempts to train a shared global model from local model updates. It is a soft PET but there are ways to harden it. While updating a global model running upon an organization's server from local models running on edge devices could seem a good idea because only aggregate data is exchanged between the organization's server and the edge devices, there is no guarantee that the push to the organization's server does not reveal anything too private or personal. This is because aggregates can be rich, and can memorize or signal in completely unexpected ways the participation of an outlying individual for example. Federated learning can thus be hardened by applying a hard PET such as local differential privacy.⁵

Our PET evaluation framework produces a risk profile for each use case in order to inform the choice of the right PET or set of PETs for the use case at hand. Two to three types of stakeholders are considered to perform the risk assessment: a data processor, who performs tasks upon the data usually on behalf of the entity initiating the processing or its collaborator; a collaborator, who collaborates with the entity initiating the processing because it is interested in the outputs of the processing, i.e., the query results; and a model user, who is involved when the processing output is a model and the collaborator discloses the model to a third party that is different from the entity initiating the processing or the data processor. Our analysis makes use of the attack model used in differential privacy⁶ as a general framework for analyzing data

⁵ Differential privacy, which is a family of mathematical techniques that formally limit the amount of private information that can be inferred about each data subject. There are two main flavors of differential privacy, offering slightly different privacy guarantees: “global,” which offers data subjects deniability of participation, and “local,” which offers deniability of record content. Despite being slightly different, both operate by introducing randomization into computations on data to prevent an attacker from reasoning about its subjects with certainty. Ultimately, these techniques afford data subjects deniability while still allowing analysts to learn from the data.

⁶ See e.g., Dwork, Cynthia, and Aaron Roth. “The Algorithmic Foundations of Differential Privacy.” *Foundations and Trends® in Theoretical Computer Science* 9, no. 3–4 (2013): 211–407. <https://doi.org/10.1561/04000000042>; Stalla-Bourdillon, S., & Rossi, A. (2021). Aggregation, Synthesis and Anonymisation: A Call for a Risk-based Assessment of Anonymisation Approaches. In D. Hallinan, R. Leenes & P. De Hert (Eds.). *Data Protection and Privacy: Data Protection and Artificial Intelligence* (Computers, Privacy and Data Protection, pp. 111–144). Oxford: Hart Publishing. Retrieved July 5, 2022, from <http://dx.doi.org/10.5040/9781509941780.ch-005>.



privacy controls and expands it to allow for additional assumptions on the capabilities and resources of the model user. This allows our framework to be inclusive of other PET threat models.

Each PET is then assessed through nine dimensions, as defined below:

- **Formal guarantee:** guarantee that can be mathematically established, e.g. deniability of the attribute values included in a record or deniability of participation to a data set with a set number of records.
- **Threat model:** model through which relevant threats, or the absence of appropriate controls, can be identified, and controls prioritized. It usually relies upon the conceptualization of a situationally-relevant attacker (e.g., a journalist, a marketer, a prosecutor, or an untrusted analyst) with situationally-relevant characteristics, such as access to publicly available information, access to the protected data, expert skills, and/or prior knowledge.
- **Loss of fidelity:** loss of accuracy (in comparison with the original version of the data) due to the transformation of the data with a view to mitigate risks, e.g. through the injection of noise.
- **Processor inference:** the likelihood a processor could uncover the identity of an individual, the values of (sensitive) attributes attached to an individual record, the participation of an individual to a data set, or the relationship or link between one or more individual records.
- **Processor sabotage:** deliberate subversion of the processing by the processor, e.g. through collusion with another participant.
- **Collaborator inference:** the likelihood the collaborator could uncover the identity of an individual, the values of (sensitive) attributes attached to an individual record, the participation of an individual to a data set, or the relationship or link between one or more individual records.
- **Collaborator sabotage:** deliberate subversion of the processing by the collaborator, e.g. through collusion with another participant.
- **Model user inference:** the likelihood a model user could uncover the identity of an individual, the values of (sensitive) attributes attached to an individual record, the



participation of an individual to a data set, or the relationship or link between one or more individual records.

- **Practical barriers:** practical barriers to the implementation of the PET, e.g. barriers related to computation speed or set-up complexity.

Formal Guarantee	Threat Model	Loss of Fidelity	Processor Inference	Processor Sabotage	Collaborator Inference	Collaborator Sabotage	Model User Inference	Practical Barriers
------------------	--------------	------------------	---------------------	--------------------	------------------------	-----------------------	----------------------	--------------------

Table 1. The 9 dimensions of PETs

This framework thus shows that each PET usually has a narrow privacy focus, most of the time confidentiality, and often PETs must be deployed in thoughtful combinations in order to mitigate a variety of data processing risks, including processor inference and sabotage, collaborator inference and sabotage and model user inference.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Inpher

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



Dr. Jordan Brandt
Board Member, MPC Alliance
President and CEO, Inpher Inc.
36 West 25th St., Suite 300
New York, NY St. 300
CAGE: 8HGW2

Office of Science and Technology Policy (OSTP)

PETS-RFI@nitrd.gov

RFI Response: Privacy-Enhancing Technologies

To Whom It May Concern,

We applaud the NITRD initiative and thank you for the opportunity to provide a response to the RFI on Privacy-Enhancing Technologies (PETs). Inpher is a software vendor providing PETs capabilities to industry and government agencies, including our work with In-Q-Tel and the Intelligence Community (see [Inpher Announces Strategic Partnership with In-Q-Tel.](#))

Inpher is committed to the advancement of PETs adoption to increase American competitiveness in data analytics and AI while preserving privacy, individual liberties and democratic values. We have submitted public statements and recommendations advocating for responsible innovation through PETs to eight US Government offices, committees and agencies since 2019, including expert testimony to the U.S. House Financial Services Committee AI Taskforce. The full records can be found on the [Privacy Advocacy and Public Policy](#) page on our website.

Below we have categorized these statements and recommendations to the US and related international governing bodies with additional reference materials according to the RFI Topics list provided.

1. Specific research opportunities to advance PETs:
 - a. U.S. Office of Management and Budget (OMB)
 - i. Inpher public comment: [Identifying Priority Access or Quality Improvements for Federal data and models for Artificial Intelligence Research and Development \(R&D\), and Testing](#)



- i. Inpher expert testimony: [AI and the Evolution of Cloud Computing: Evaluating How Financial Data is Stored, Protected, and Maintained by Cloud Providers](#)
 - b. European Data Protection Board (EDPB)
 - i. Inpher public comment: [Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data](#)
 - c. [EU and U.S. Policymakers Emphasize Privacy-Enhancing Technologies as a Shared Priority in 2021](#)
 - d. Office of the Privacy Commissioner of Canada (OPC)
 - i. Inpher public comment: [Consultation on Transfers for Processing](#)
6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:
 - a. UK Information Commissioner's Office (ICO)
 - i. Inpher public comment: [ICO Consultation on the Draft Data Sharing Code of Practice](#)
7. Risks related to PETs adoption:
 - a. [The Privacy Risk Right Under Our Nose in Federated Learning](#)
 - b. [Risks of Privacy-Enhancing Technologies: Complexity and Implications of Differential Privacy in the Context of Cybercrime](#)
8. Existing best practices that are helpful for PETs adoption:
 - a. [How DataCo is Reinventing Data Partnerships by Putting Privacy-Enhancing Technologies to Work](#)

We sincerely hope that this information will aid in the development of a national PETs strategy and we remain at your service to answer any questions or participate in upcoming workshops.

Sincerely,

Dr. Jordan Brandt
President and CEO, Inpher
Board Member, MPC Alliance

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Intel Corporation

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Intel's response to RFI 35250 Federal Register / Vol. 87, No. 111 / Thursday, June 9, 2022 / <https://www.govinfo.gov/content/pkg/FR-2022-06-09/pdf/2022-12432.pdf>

July 8, 2022

Intel Corporation (Intel) appreciates the opportunity to respond to the Request for Information for Advancing Privacy Enhancing Technologies (PETs) published by the Federal Register at <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>. Intel agrees with the ideas on Privacy Enhancing Technologies expressed by OSTP in [the recent blog](#).

Intel's mission is to engineer solutions to address society's greatest challenges, providing our customers with reliable, cloud-to-edge computing, inspired by Moore's Law. Intel designs and manufactures advanced integrated digital technology platforms that power the connected world.

In this response we provide an introduction on PETs and address the following points in the RFI:

- Specific technical aspects or limitations of PETs.
- Specific research opportunities to advance PETs.
- Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs.
- Risks related to PET adoption.
- Existing best practices that are helpful for PET adoption.
- Other information that is relevant to the adoption of PETs.

Intel has researched and created PET's technologies for decades, including, but not limited to, research and standardization in fully homomorphic encryption, work in Multi-Party Security Computation and Federated Learning, and development of PETs friendly standards protocols, e.g., Direct Anonymous Attestation (DAA).

Introduction

According to a common definition, “[p]rivate-enhancing technologies (PET) are technologies that embody fundamental data protection principles by minimizing personal data use, maximizing data security, and empowering individuals and institutions to make decisions based on their data¹.” PETs identify systems, processes, and techniques that enable processing to derive value from data while minimizing privacy and security risks to individuals and institutions.

PETs deal with protecting the sensitive part of the data in an information system, e.g., private information about persons, institutions, or governments, through techniques such as data obfuscation (e.g., information theoretical metrics privacy, use of synthetic data), data confidentiality (with cryptography and security mechanisms), system architectures for federated

¹ https://en.wikipedia.org/wiki/Privacy-enhancing_technologies

analysis (e.g., secure multi-party computation, federated learning), and a combination thereof, without sacrificing the usability of the data.

PETs-based solutions help individuals and institutions to develop tools that assist with compliance with data protection regulations, enhance data integrity, mitigate cyber security risks, and preserve confidentiality while enabling data analysis and other functionality. PETs can be used in tandem with governance and operation systems capabilities to realize their benefits.

It is also important to note that PETs comprise a large portfolio of tools and techniques that are at different degrees of maturity, with some approaches already widely deployed, while others are in early stages of research. As we describe these technologies, the level of maturity will be specified if relevant. Greater investment in research in PETs can benefit both mature and emerging technologies in this area.

Specific technical aspects or limitations of PETs

PETs address privacy and data protection in a variety of ways. Important state-of-the-art techniques in this area include foundational technologies such as differential privacy, use of synthetic data, fully homomorphic encryption and other cryptographic mechanisms, secure multi-party computation, and federated machine learning, as well as hardware foundations for PETs, such as Confidential Computing and Zero-Trust technologies. We briefly explain some of these techniques below:

- **Differential Privacy (DP):** DP adds random data, or “noise,” to the real data, such that real information is hidden in the noise. DP allows for accurate analysis to be done on data in aggregate. The amount of “noise” is associated with the ability (or difficulty for an attacker) to re-identify the data. Hence, institutions using DP can set a privacy “budget” to tradeoff the difficulty of re-identifying data with the utility of the data.

Experiments in several use cases show that adding reasonable privacy guarantees does not meet utility requirements. Additionally, differential privacy guarantees depend heavily on the privacy objectives of the solution. Further work is required on DP mechanisms to lower the impact on performance, especially when being utilized to protect large scale AI models. Research in this area focuses on alternative empirical metrics, a shift to localized data analysis, compression methods to minimize data movement, as well as information theoretic methods to protect certain attributes of data.

- **Synthetic Data (SD):** SD replaces real data with altered or artificially generated data. It inserts additional, synthetic (not real or measured) information into real datasets. Synthetic data are commonly created through machine learning and mimics the characteristics of real-world data.

The data are generated by feeding real data into machine learning algorithms, which then identify data characteristics and trends, and replicate them in the form of synthetic data. Using synthetic data limits the need to collect and store real information about individuals. A main drawback is that the process of creating synthetic data using AI is not yet mature.

Additionally, the study of the limitations of synthetic data in comparison to real data is emerging, and a more thorough examination is needed to ensure adequate accuracy in analysis based on synthetic data.

- **Federated Learning (FL):** FL allows multiple owners of private data sets to jointly train a model based on the union of their data without sharing their data with other parties. First, a central server shares an initial version of the model with the data owners. Next, each data owner uses its private data to adjust and improve the model. Such adjustments are aggregated at the central server, which then sends the improved model to the data owners for further rounds of adjustments.

The technology has been applied successfully to a variety of problems (e.g., in healthcare²) but continues to be a subject of active research to expand its scope to allow data sharing and learning at a massive scale, with new mechanisms being devised to make it operate with inherent resource limitations in emerging applications such as networked IoT/immersive media applications.

- **Secure Multi-Party Computation (SMPC):** SMPC enables two or more parties to compute an output that depends on inputs that each party would like to keep secret, in a way that the parties learn the output but nothing about the secret inputs.

SMPC has been successfully deployed for key management in IT infrastructures, in point solutions in the finance, cryptocurrency, and code signing domains. There are many MPC cryptographic protocols, but they all tend to be communication intensive, that is, rate-limited by communication bandwidth between the parties, and memory footprint increase with the complexity of the computation.

- **Cryptographic primitives and Fully Homomorphic Encryption (FHE):** Encryption techniques are essential tools to protect data confidentiality. Classic encryption techniques (such as symmetric and asymmetric cryptography) protect confidentiality when data is temporarily not in use (e.g., at rest, and in transit). The shift to Post-Quantum Cryptography (PQC) driven by the advent of Quantum Computers and the NIST selection of candidate algorithms to standardize, urges to increase the key size used in symmetric cryptography (to resist to the Grover's algorithm), and to deploy quantum-resistant public key cryptography, e.g., lattice-based cryptography, for key encapsulation and key agreement (to resist the Shor's algorithm). Introducing quantum-resistance in future cryptographic suites implies an increase in cryptographic overhead, which in turn may favor hardware assisted acceleration to pure software solutions.

Modern cryptographic techniques, such as Fully Homomorphic Encryption (FHE) make it possible for data to be operated on and modified while in encrypted form and without having to disclose the associated decryption keys. However, among the implementations of FHE schemes to date, none are considered computationally efficient on existing hardware architectures to be widely utilized in commercial products.

² <https://newsroom.intel.com/news/intel-works-university-pennsylvania-using-privacy-preserving-ai-identify-brain-tumors/>

Unlike classic cryptography, FHE is still in its early stages of maturity from a deployment perspective, but it has the potential to be used widely in applications ranging from smart contracts to payment processing as the technology advances to become progressively less expensive.

FHE has several important properties unlike other methods used to process encrypted data. First, the deployment of FHE within a data processing pipeline does not disrupt the data lifecycle. As a result, when FHE is deployed in AI/ML data lifecycles, it can either preserve or amplify the value of data while reducing the risk of third-party data leakage during processing. Second, FHE can be deployed in various forms, including shared key, public keys, or secret sharing; this allows flexible deployments for different usages to elevate the level of confidentiality in existing security solutions while reducing the risk of third-party data leakage. Third, when FHE is deployed, the physical location of the servers where data is processed can be anywhere on earth or in space. Fourth, FHE can be deployed at both classical and post-quantum security NIST levels. Fifth, data can be encrypted before, during and after processing, providing protection throughout its lifecycle. In summary, using FHE can elevate the level of confidentiality of the security solution by preserving both data privacy and the value of data. At the same time, FHE can enable considerable business and public service opportunities through data sharing and collaboration.

The main challenges with homomorphic encryption include: the resources required for deployment, the lack of widely accepted standards, and practices to integrate FHE with other PETs and IT infrastructure components. Despite this, there are practical yet specialized uses of homomorphic encryption today. For example, FHE is being used to monitor whether passwords saved in Web browsers (e.g., Microsoft Edge) were ever exposed in a data breach.³

- **Confidential Computing as hardware foundation for PETs:** Confidential Computing is the protection of data in use by performing computation in a hardware-based Trusted Execution Environment (hereby TEE). A TEE is a key technology for data protection is a Trusted Execution Environment. A TEE is an environment that protects code and data confidentiality and integrity during execution. A hardware-based TEE typically provides these properties by mapping the TEE to a secure portion of memory and enforcing access-control from non-TEE executions. In addition, the hardware provides abstractions for attesting the state of a TEE to a third party that allows an external verifier to ensure that execution is happening in a TEE to gain trust. Developers can host portions of their applications into hardware-isolated programs or hardware-protected areas of execution to increase security, even on compromised platforms. Using TEEs, developers can protect the integrity of sensitive software and crucial configurations. Many cloud providers already offer hardware assisted trusted execution as part of their services, and companies can further increase protection by looking into third-party encryption solutions and providers. The technology is mature, so many of these hardware features are available today for both

³ <https://www.microsoft.com/en-us/research/blog/password-monitor-safeguarding-passwords-in-microsoft-edge/>

businesses and individuals, and accessible mainly thanks to the open-source efforts to deploy unmodified code in a TEE.

It is important to highlight that TEEs can enforce not only confidentiality, but also integrity of workload memory (typically using a combination of both access-control and cryptography mechanisms) even in the presence of privileged malware at the system software layers. TEEs in different forms have been the subject of standardization for two decades and are mature and adopted solutions.

- **Zero Trust and Trust as a Service (TaaS) as foundational service for PETs:** TEEs' functionality can be enhanced with the use of Zero Trust architecture. This type of architecture is based on the principle that all the resources must establish their trustworthiness before they are utilized, from the provenance of the platform to the execution of sensitive workloads in the infrastructure where the workload owner may have little control. Zero Trust architecture can support and supplement capabilities of PETs. Attestation of trustworthiness supports principles of Zero Trust and builds a stronger security foundation for PET-based systems.

Intel's vision of Trust as a Service establishes independent trust authority to attest to every computing asset in public cloud, edge, and private cloud. Trust as a Service (TaaS) can provide proof of trustworthiness within the key areas described below, strengthening the foundations of PETs or incorporating some features of PETs:

1. The platform has established correct provenance from the earliest point in the supply chain or software/data life cycle.
2. The platform has foundational integrity (for example, in its firmware).
3. The software on the platform has not been tampered with.
4. The integrity of the TEE (Trusted Execution Service) has been verified.
5. The relevant policies are enforced.

Specific research opportunities to advance PETs

With all the pros and cons about PETs, it is clear that the following types of initiatives have had the largest impact on advancing PETs:

- The establishment of public-private partnerships, e.g., funded by the US government through DARPA and IARPA (e.g., the DARPA PROCEED (archived), IARPA HECTOR (archived) and the DARPA DPRIVE⁴ (active) programs to advance FHE) and the National Science Foundation (NSF) (e.g., the partnership on "Resilient & Intelligent Networking Systems" (RINGS)⁵, and "Machine Learning for Wireless Networking Systems" (MLWiNS)⁶ to advance AI practices).
- The establishment of industry driven academic partnerships, e.g., led to the creation of the Gramine platform⁷ for wider adoption of hardware-based trusted execution by deploying unmodified applications.

⁴ <https://www.darpa.mil/news-events/2021-03-08>

⁵ <https://www.nsf.gov/pubs/2021/nsf21581/nsf21581.htm>

⁶ <https://www.nsf.gov/pubs/2019/nsf19591/nsf19591.htm>

⁷ <https://gramineproject.io/>

- The establishment of industry consortia and corresponding open-source initiatives to advance practices and uses of PETs and their hardware foundations, including but not limited to the formation of the Confidential Computing Consortium⁸ for hardware-based TEEs; the formation of the MPC alliance⁹ for SMPC; and the Open Federated Learning project for wider adoption of federated machine learning¹⁰.
- The development of standards, e.g., around FHE¹¹ and SMPC¹²; the NIST selection of post-quantum cryptography (PQC) candidates to standardize advances in the field by allowing institutions to accelerate the adoption of PQC algorithms in products.

Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

PETs can be and are applied in a variety of fields and use cases, e.g., AI modelling, cross-border data transfers, and data analytics to help security and risk stakeholders manage constraints while respecting individual privacy, and ultimately reduce direct (by direct access) and indirect (as a consequence of processing) third-party data leakage. Examples of applications include, but are not limited to:

- **Healthcare:** information including patient data can be pooled together for research, drug-development, and public health. The COVID-19 pandemic has highlighted the value of information in public health, from contact tracing to vaccine development. Given the importance and particularly sensitive nature of data in healthcare, this sector has been a leader in exploring the potential of PETs in emerging technology areas, such as FL¹³ and FHE¹⁴.
- **Financial services:** requirements such as Know-Your Customer (KYC), credit reporting, money laundering detection, and fraud mitigation drive the collection of data that is matched to specific individuals. Pooling this data together can help identify financial crimes and protect individuals from financial loss.
- **Government applications:** processing of super-resolution images; localization.

Since PETs cover data privacy and confidentiality, data aggregation use cases for PETs are not limited to situations where privacy of data subjects needs to be protected. They can also be used in data confidentiality cases, including information sharing in cybersecurity and other areas, such as intellectual property protection (e.g., AI models).

⁸ <https://confidentialcomputing.io/>

⁹ <https://www.mpcalliance.org/>

¹⁰ <https://github.com/intel/openfl>

¹¹ <https://www.iso.org/standard/83139.html>

¹² <https://www.iso.org/standard/80508.html>

¹³ <https://newsroom.intel.com/news/intel-works-university-pennsylvania-using-privacy-preserving-ai-identify-brain-tumors/>

¹⁴ <https://carnegieendowment.org/2021/08/17/introduction-how-korea-can-unleash-power-of-data-pub-85162>

Risks related to PET adoption

As in all technology areas, there are several challenges associated with the adoption of PETs that need to be overcome for successful adoption:

- **Internal capacity and expertise within organizations to deploy and manage PETs.** This is a common issue across technical deployments that require specialized expertise, but **since many current generation PETs are not widely used, their broad deployment can pose unique challenges.** Some techniques or systems may be able to function with limited support, while others need more oversight. Some technologies may also need to be used in conjunction with business enterprise systems, and therefore, require work to integrate and maintain those connections, which can increase costs.
- **PETs are in various stages of maturity.** While promising, some techniques and systems are still in early phases of development and require additional research before they are commercialized. The variability in maturity and research adds to the complexity of PET adoption and may make it harder for institutions to determine which PETs are appropriate and what resources they need to deploy them.
- In the systems that we have examined **it can be difficult to validate correct implementation of PETs.** As in the case of cryptography, some current PETs depend upon code review to ensure correctness, as correctness of privacy preservation cannot be validated by examining the outputs. This motivates the need for validation techniques and tools, such as heuristic measurements, to compliment the PETs themselves. It also motivates the value of standards and standard implementations such as cryptography standards.

Existing best practices helpful for PET adoption

The new techniques described above are broadly applicable to diverse datasets with multiple owners. While there are some existing standards which can support privacy while preserving data processing in these contexts, new standards, best practices, and guidelines will need to be created. This is an important area of pre-standardization research and development. Hence, best practices, including but not limited to the following, remain helpful for PET adoption:

- Transitioning to practice the results of research and development outcomes of public-private partnerships.
- Adopting the outcomes of best practices and usages developed by reputable industry consortia.
- Transferring the technical results of industry driven academic partnerships into industry, standards, and best practices.
- Supporting open international standards in the area.
- Building PETs on a secure foundation using up to date security models, e.g., Zero-Trust, secure hardware.
- Ensuring high-quality data and adequate data governance.

Mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

All of the PETs described above aim to address data confidentiality in various forms, and at different stages in the data lifecycle. However, integrity of the data, the computation, and the context is an important property that the integration of PET technologies in actual systems must support.

Computational integrity and attestation are important mechanisms for enforcing privacy policies on data. For many of the PETs listed above it is important to be able to validate that the technology is being correctly applied, since it is not currently possible to determine from the outputs whether the PET was instantiated. Systems that deploy PETs with integrity and attestation capabilities enable remote parties to have confidence that those PETs were correctly deployed and invoked.

Contextual integrity is a privacy concept in which the individual has the right not to be observed or disturbed when they don't want to be. Enforcement of contextual integrity on data use is another dimension of privacy that technology should enable – many privacy violations are contextual integrity violations of data that was accessed without violating confidentiality technologies. Through Confidential Computing, attestation, smart contracts, and, in general, verifiable computation, the beginnings of technical mechanisms to enforce contextual integrity policies on data can be constructed.

To conclude, Intel appreciates the opportunity to respond to the Request for Information for Advancing Privacy Enhancing Technologies. Please do not hesitate to contact Ro Cammarota, Intel Principal Engineer and Claire Vishik, Intel Fellow for further information.

Notices & Disclaimers

Intel technologies may require enabled hardware, software, or service activation.

No product or component can be absolutely secure.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

IQVIA Inc

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response for Request for Information (RFI) on Advancing Privacy-Enhancing Technologies

Submitted By: Kim Gray, Chief Privacy Officer, IQVIA Inc. (industry)

This response to the Request for Information on Advancing Privacy-Enhancing Technologies (“PETs”) is submitted on behalf of IQVIA Inc. IQVIA™ (NYSE:IQV) is a leading global provider of advanced analytics, technology solutions and clinical research services to the life sciences industry. IQVIA is a global leader in protecting individual patient privacy. The company uses a wide variety of privacy-enhancing technologies and safeguards to protect individual privacy while generating and analyzing information on a scale that helps healthcare stakeholders identify disease patterns and correlate with the precise treatment path and therapy needed for better outcomes. IQVIA’s insights and execution capabilities help biotech, medical device and pharmaceutical companies, medical researchers, government agencies, payers and other healthcare stakeholders tap into a deeper understanding of diseases, human behaviors and scientific advances, in an effort to advance their path toward cures. I am the Global Chief Privacy Officer for IQVIA.

The Office of Science and Technology Policy (“OSTP”) is seeking information concerning privacy-enhancing technologies. The focus of my comments is on one kind of privacy-enhancing technology – the use of de-identification or anonymization techniques as a means of disconnecting information about individuals from the identity of those individuals, so that this information, in the health care sector and others, can be used for a broad variety of publicly beneficial purposes without sacrificing privacy protections. This technique can be used across industries and can serve to benefit the public overall through providing a means of broader access to data for research and analytical purposes while still protecting individual privacy. We encourage OSTP to endorse this approach and to provide additional means for companies, in the health care industry and more broadly, to utilize experts to facilitate the development of appropriately de-identified information.

In particular, we want to focus OSTP’s attention on the de-identification standard of the Privacy Rule under the Health Insurance Portability and Accountability Act (“HIPAA”). This standard – one of the earliest legal standards in the United States related to this kind of technology – remains the gold standard for de-identification in the US legal system and beyond. We encourage the continued use of this standard in the health care industry generally and believe that its approach can serve as a model for other areas of our economic system. Under the HIPAA rules, the de-identification formula permits, for example, the linkage across data sets of patient level data in a way that permits longitudinal evaluation of the data (a patient’s journey through the health care process over time) without any identification of the patient. It is important when conducting medical research or public health activities to know that the health data applies to the same patient, but is not important for much of this analytical activity to know who the patient is. Individual privacy is protected while health data research benefits everyone.

Background for the HIPAA De-Identification Standard

Following the passage of the Health Insurance Portability and Accountability Act of 1996, Public Law 104 – 19, the Department of Health and Human Services developed the Standards for Privacy of Individually Identifiable Health Information (“the HIPAA Privacy Rule”).

Under the HIPAA Privacy Rule, HIPAA covered entities must provide privacy and security protection for “protected health information” (“PHI”) – individually identifiable health information about patients and insureds. The substantive provisions of the HIPAA Privacy Rule were designed to provide appropriate privacy protection while still permitting the health care system to work effectively and efficiently, for the benefit of both patients and the health care industry (and society at large). For example, provisions of the HIPAA Privacy Rule were designed to make the “common” uses and disclosures of this protected health information seamless, focusing on uses and disclosures for treatment, payment and health care operations. Uses and disclosures beyond these common purposes require explicit patient permission. The goal of this approach was to facilitate the efficient operation of the health care system while also protecting individual privacy.

A similar approach was taken in connection with the de-identification of this PHI, as set forth in 45 C.F.R. § 164.514(a-b). In drafting the regulations, the United States Department of Health and Human Services (“HHS”) recognized that the de-identification standard it developed could be even more patient protective, meaning almost no or no risk of re-identification of individuals rather than a low risk. However, when drafting these provisions and responding to public comments, HHS was explicit in acknowledging that this “no risk” additional protection offered only marginally more privacy protection and would come at the expense of a broad range of socially desirable uses and disclosures of this information. That led to the development of the HIPAA Privacy Rule de-identification standards, as a means of providing BOTH appropriate privacy protection and the socially desirable ability to use and disclose this “de-identified” information. Where this standard was met, “[r]egardless of the method by which de-identification is achieved, the [HIPAA] Privacy Rule does not restrict the use or disclosure of de-identified health information, as it is no longer considered protected health information.” Consistent with the goals of this RFI, we encourage a broader incorporation of this concept by OSTP, in the health care industry generally (both in and out of the scope of HIPAA) as well as in a broader range of other industries.

The de-identification standard is a core element of the HIPAA Privacy Rule. In its primary guidance concerning this standard, the HHS Office for Civil Rights stated that “[t]he increasing adoption of health information technologies in the United States accelerates their potential to facilitate beneficial studies that combine large, complex data sets from multiple sources. The process of de-identification, by which identifiers are removed from the health information, mitigates privacy risks to individuals and thereby supports the secondary use of data for comparative effectiveness studies, policy assessment, life sciences research, and other endeavors.”¹

The HIPAA Privacy Rule created two standards for de-identification – the “expert determination” method and the “safe harbor” method.

As set forth in this guidance, the HIPAA Privacy Rule “was designed to protect individually identifiable health information through permitting only certain uses and disclosures of PHI

¹ Department of Health and Human Service, Office for Civil Rights, “Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule,” https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf at 5 (November 2012).

provided by the [HIPAA Privacy] Rule, or as authorized by the individual subject of the information.” However, “in recognition of the potential utility of health information even when it is not individually identifiable, § 164.502(d) of the [HIPAA] Privacy Rule permits a covered entity or its business associate to create information that is not individually identifiable by following the de-identification standard and implementation specifications in § 164.514(a)-(b).” By utilizing these provisions, a covered entity is allowed “to use and disclose information that neither identifies nor provides a reasonable basis to identify an individual.”

Under this HIPAA de-identification standard, “health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual.” As HHS has noted, “[b]oth methods, even when properly applied, yield de-identified data that retains some risk of identification. Although the risk is very small, it is not zero, and there is a possibility that de-identified data could be linked back to the identity of the patient to which it corresponds.”

The first is the “expert determination” method. Under this approach, a covered entity may determine that health information is not individually identifiable health information only if:

A person, with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:

- (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
- (ii) Documents the methods and results of the analysis that justify such determination.

The second is the “safe harbor” method. This method involves removal of a specific set of defined identifiers. “[I]dentifiers of the individual or of relatives, employers, or household members of the individual, are removed . . . as well as absence of actual knowledge by the covered entity that the remaining information could be used alone or in combination with other information to identify the individual.”

HHS has consistently acknowledged the key purposes of this rule. In early commentary for the HIPAA Privacy Rule, in discussing public comments (See 67 Federal Register 53232 (August 14, 2002), HHS noted:

The Department is cognizant of the increasing capabilities and sophistication of electronic data matching used to link data elements from various sources and from which, therefore, individuals may be identified. Given this increasing risk to individuals’ privacy, the Department included in the [HIPAA] Privacy Rule the above stringent standards for determining when information may flow unprotected. The Department also wanted the standards to be flexible enough so the [HIPAA] Privacy Rule would not be a disincentive for covered entities to use or disclose de-identified information wherever possible. The [HIPAA] Privacy Rule, therefore, strives to balance the need to protect individuals’ identities with the need to allow deidentified databases to be useful.

Similarly, in the initial publication of the HIPAA Privacy Rule (see generally 65 Federal Register 82708 (December 28, 2000), HHS noted that:

We also disagree with the comments that advocated using a standard which required removing only the direct identifiers. Although such an approach may be more convenient for covered entities, we judged that the resulting information would often remain identifiable, and its dissemination could result in significant violations of privacy. While we encourage covered entities to remove direct identifiers whenever possible as a method of enhancing privacy, we do not believe that the resulting information is sufficiently blinded as to permit its general dissemination without the protections provided by this rule.

We agree with the comments that said that records of information about individuals cannot be truly deidentified, if that means that the probability of attribution to an individual must be absolutely zero. However, the statutory standard does not allow us to take such a position, but envisions a reasonable balance between risk of identification and usefulness of the information.

We disagree with those comments that advocated releasing only truly anonymous information (which has been changed sufficiently so that it no longer represents actual information about real individuals) and those that supported using only sophisticated statistical analysis before allowing uncontrolled disclosures. Although these approaches would provide a marginally higher level of privacy protection, they would preclude many of the laudable and valuable uses discussed in the NPRM (in § 164.506(d)) and would impose too great a burden on less sophisticated covered entities to be justified by the small decrease in an already small risk of identification.

In the almost two decades since these provisions became final, the HIPAA de-identification method has become both a core element of the HIPAA rules and has provided both appropriate privacy protections and the ability for those in the health care industry and otherwise to benefit from a wide range of uses for de-identified data, including public health purposes, medical research and a wide range of additional commercial purposes. These benefits have been able to be achieved without any identifiable risks to the privacy of this personal information. The information is protected, privacy is maintained, and the public (and individuals) can benefit from these uses. We encourage OSTP to both reconfirm the validity of this approach as an important privacy-enhancing technology and to evaluate how best to apply this standard in the context of activities outside of the health care system, where the same benefits of privacy protection and beneficial data uses can be achieved.

Specific Questions from the RFI

Beyond this general discussion of the benefits of the HIPAA de-identification standard and how this standard can be used as a privacy enhancing technology in the health care system and otherwise, I wanted to discuss some additional points related to specific questions in the RFI.

- Specific research opportunities to advance PETs; Existing barriers to PET adoption (Questions 1 and 9)

The government can assist with the development of privacy-enhancing technologies by providing additional education and training related to potential experts for this de-identification approach. The HIPAA de-identification standard is useful and important; at the same time, because of its complexity to ensure that data is both useful and privacy protective, the expertise necessary to provide an “expert determination” is limited. The government can assist by providing additional means of developing appropriate experts. This can include government education and training in this area, as well as identifying more specific means for organizations to identify and utilize experts for these determinations. The limited number of experts in this field and the associated costs for an expert determination may be barriers to additional adoption of the HIPAA de-identification approach, even within the health care industry.

- Specific technical aspects or limitations of PETs (Question 2)

The government may be able to assist in this area by developing standard or automating means of implementing these requirements under the HIPAA rules. These standards would serve to supplement any additional education and training for potential experts, as well as help assist with the current economic burdens involved in paying for these expert determinations.

- Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs (Question 3)

The formal elements of the HIPAA de-identification standard are used, at this point, primarily within the health care industry, focused primarily on covered entities and business associates subject to the HIPAA rules. These HIPAA rules also can reasonably be applied today in connection with various forms of health care research, whether within the scope of the HIPAA rules or not. The health care sector will benefit from additional experts in this area as well as additional education and training on means of ensuring appropriate expert determinations. We encourage consideration by OSTP of means by which these HIPAA principles can be applied more broadly in other sectors. The government should evaluate how these principles can be applied in a broader setting outside of the health care industry. The government should both endorse the principles of the HIPAA de-identification approach and seek to apply it in other settings.

- Specific regulations or authorities that could be used, modified, or introduced to advance PETs; Specific laws that could be use, modified, or introduced to advance PETs; Existing best practices that are helpful for PETs adoption (Questions 4, 5 and 8)

The primary legal provisions related to the HIPAA de-identification standard are set forth in 45 C.F.R. § 164.514(a-b). HHS’s most recent guidance for this standard is set forth in Department of Health and Human Service, Office for Civil Rights, “Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.”² This standard is designed explicitly to facilitate

² *Id.*

and encourage the use of de-identification in the HIPAA environment, for the benefit of patients, the health care industry and the public at large.

Conclusion

At IQVIA, we have worked regularly with the HIPAA de-identification standard since it was developed by HHS. This standard has allowed IQVIA, its customers and its partners in the health care industry to utilize de-identified health care data for a broad range of medical research and public health purposes, as well as the overall facilitation of improved medical care in the United States and around the world. We encourage OSTP to endorse the framework of the HIPAA de-identification standard, and to explore means of expanding the utilization of this standard to a broader range of industries around the country.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Ixup Solutions

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Office of Science and Technology Policy

Friday 8 July 2022

Reference: Request for Information on Advancing Privacy-Enhancing Technologies

Thank you for the opportunity to make comment on the commercial application of Privacy Enhancing Technology.

We are a small Australian Privacy Enhancing Technology development company specializing in encrypted data solutions, specifically to address collaborative data requirements across a variety of sectors, including Government, Banking, Insurance, Gaming, and Loyalty programs.

To that end, our commentary is specifically aimed at Category 3 of the topics requested by this RFI.

This is an important initiative being undertaken, and we wish you a very productive outcome.

If we can provide any further information or assistance, we would welcome an approach from the OSTP.

Yours sincerely,

Peter Hong
General Manager, Government
Ixup Solutions
Sub Base Platypus
Suite 1004, 118 High St.
North Sydney, NSW 2060
AUSTRALIA

Mobile: Email:

The volume of data that is being created and used continues to grow exponentially. In 2020, the world produced 64.2 zettabytes of information with that expected to triple by 2025. At the same time, there has been an increased focus on data protection and privacy that has resulted in encryption becoming the default for the storage and transmission of data.

The convergence of these two trends has created new challenges for organisations. Everything from digital transformation efforts and the growth of open data and online gaming to the increased oversight by regulators when it comes to the protection and use of personal data relies on the ability to understand data from multiple sources. But when that data is encrypted, the ability for it to be read and used can be hampered.

Encryption's purpose is to prevent the unauthorised use of information. This makes sense when data is being stored and moved. But what about when we want to use it? The typical approach has been to decrypt data when it needs to be transacted or analysed.

Encryption has significantly reduced the risk of doing business online at scale, improving productivity while also protecting sensitive commercial and personal information. But when we want data to be put to work it needs to be decrypted so that computers can transact, compute, analyse or report on that data. This is the weakest link in your data security chain but it's also where we turn data into real business value.

With about 70 per cent of the world's GDP digitised, securely managing data through the entire value chain is critical. That means industries as diverse as governments through to esports need to find ways to gain value from encrypted data. This is what homomorphic encryption makes possible. It is an enabler of consumer data rights, business integrity, compliance, and social impact, allowing encrypted data to be securely computed, analysed, processed, and reported on.

Encryption uses a public key to encrypt data so only someone, or a machine, with a matching private key can decrypt the data. Homomorphic encryption – “homomorphic” means “same shape” – uses algorithms to maintain the relationships between elements in data, which in turn, allows for encrypted computation. This supports businesses in their governance, risk, and compliance (GRC) obligations as they can share and process data without weakening security.

The resulting computations are also encrypted and can only be revealed using the secure, private key. This level of security means value can be added to data sets without compromising the integrity or privacy of the source data. The enriched data can be tracked, traced, and audited, ensuring a continuous chain of custody, building trust and provenance at the same time.

With environment, social and governance (ESG) now a major issue for boards, and increasing oversight from regulators all over the world, ensuring the confidentiality and integrity of data is vital. Homomorphic encryption is a game-changer that will enable organisations to protect data through its entire value chain at scale and in a form that supports large batch processing or real-time transaction processing, but organisations need to invest in the right tools and work with experienced partners.

As more organisations use public and private cloud solutions, privacy enhancing technologies like homomorphic encryption will become increasingly important.

For example, an organisation may choose to store an encrypted, sensitive data set on a cloud service and perform some calculations with that data. At no point does the operator of the cloud service decrypt the original data or the result of the calculations. Both the actual calculations and the result remain encrypted and can only be accessed by someone with the decryption key.

As data privacy legislation becomes more prevalent – Gartner says that by 2023, 73 per cent of the world will be subject to privacy rules such as the European Union’s General Data Protection Rules, the Australian Privacy Principles, and others – ensuring the confidentiality and integrity of personal data through the entire value chain is a significant compliance issue.

Securing private information has, up until now, been the responsibility of each organisation, but these regulatory changes are extending that responsibility across enterprises and industries – and businesses are needing to turn to privacy-enhancing technologies to fill the gap.

For example, **IXUP is collaborating with Microsoft to enhance the privacy of gaming operators to protect vulnerable gamblers and for sporting bodies to prevent impermissible betting (sports integrity)**. By ensuring information such as self-exclusion data always remains safe, gamblers will be able to protect themselves from the effects of gambling while enabling service operators to ensure vulnerable people are removed from marketing campaigns and prevented from betting.

This, we believe, was the first implementation globally of homomorphic encryption turning a privacy challenge into a social impact benefit.

More information is available here:

[Information for Regulators PlayPause v1.2.pdf](#)

<https://ixup.com/press/taking-responsible-gaming-and-sports-integrity-to-a-new-digital-level/>

A very similar model could be easily deployed to protect the privacy of individuals in any other register, including, for example, a gun owners or gun licensing register where checks may be required across disparate data sources (law enforcement, health, education, social services for example) before a license is granted.

Creating new approaches and extending the reach of financial audit processes also promises a long overdue positive change. Repurposing audits to include new service offerings such as the audit of machine learning algorithms or arrive at a realistic value for an IPO valuation using the measure of interest gleaned from a company’s own website. These are but two possibilities that would enhance existing offerings and help to create greater value for customers by the addition of complimentary offerings to the financial audit process.

Privacy enhancing technologies could provide an ideal technology platform to assist accountancy and financial services providers as well as fintech technology vendors an opportunity to rethink traditional audit and accountancy products and reinstitute trust in audit and assurance.

In the world of Open Banking/Open Finance, consumers demand, and deserve, ultimate privacy in all transactions that share their personal identifiable information. PET, and in particular homomorphic encryption, can underpin the integrity and security of those transactions and computations without compromise – the encrypted data cannot be reverse engineered to reveal the original information unlike hashing.

For more detail of this Use Case, please refer to [Privacy Preserving Homomorphic Encryption for Audit Practice v2.pdf](#)

As you know, organisations are under increasing pressure and scrutiny to ensure they protect sensitive data. Homomorphic encryption enables them to maintain security even when using data. The days of needing to decrypt data to extract value from it will be behind us as this proven and cost-effective technology helps keep personal data private.

End of Submission

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Juniper Networks

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



July 8, 2022

Ms. Jeri Hessman
Technical Coordinator
NITRD
White House Office of Science and Technology Policy
490 L'Enfant Plaza SW
Washington, DC 20024

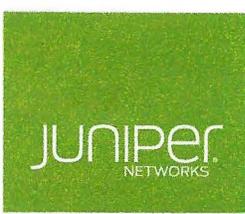
Re: Request for Information on Privacy-Enhancing Technologies (Docket 2022-12432)

Dear Ms. Hessman:

Juniper Networks submits these comments in response to the above-captioned Request for Information (RFI) regarding Privacy-Enhancing Technologies (PET). As a company dedicated to the security of networks, we appreciate NITRD's effort to develop a national strategy on privacy-preserving data sharing and analytics as well as associated policy initiatives. This submission provides background and recommendations in terms of networking technologies that can help ensure and enhance the privacy of data.

By way of background, Juniper Networks is a publicly-traded corporation headquartered in Sunnyvale, California. We develop high-performance networking solutions spanning routing, switching, security, wireless, and network management. Juniper Networks is a proud provider to the US government, major enterprises, cloud providers, and service providers around the world.

While Juniper does not collect personally-identifiable information as part of its business model, our customers rely on our solutions to transmit and protect their own data – personal, business confidential, and otherwise – that is critical to their operations, their employees,



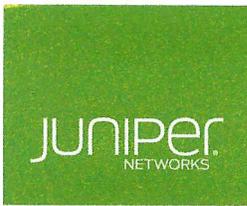
and their own customers. The nature of our business and solutions leads our submission to address a particular issue that the RFI raises:

Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

This paragraph of the RFI is critical because cybersecurity, cryptography, and computing architecture are key to maintaining privacy when sharing data over a network. In fact, these are primary factors in our product development cycle.

There are several security-related advances in networking technology that serve to enhance privacy:

- **Encrypted Traffic Analysis:** Any data being transmitted through a network needs to be analyzed for malware and other threats. Many such threat detection and analytical tools on the market today require the decryption of data, which ultimately can lead to a loss of privacy or may not even be feasible. Providers of networking solutions, including Juniper, have solutions available today that can effectively analyze network traffic without decrypting it. This means network operators and owners can often detect threats hidden in encrypted traffic without intercepting and decrypting it, thereby enhancing privacy.
- **Cloud-Based Threat Protection:** Cloud-based solutions currently on the market allow operators to monitor their traffic for patterns against known command and control/beaconing profiles. These protection offerings delivery three functionalities: (1) Identify and display information about scanned files and communication patterns whether clean or infected and identify communication identifiers that improve the threat intelligence networks many systems are tied to today (IP details, ports, C2, etc.); (2) create and view allowlists and blocklists that list safe or harmful network nodes, and profiles that define what file types to submit for investigation; and (3) view various reports, such as most infected file types, top malwares identified, and



infected hosts. Because this is a web-based service, customers can provision it on an as-needed basis.

- Security Director and Security Incident and Event Management (SIEM): These provide cloud-based policy orchestration as well as threat detection and mitigation across organizations including automation actions that enable scale and playbook execution permitting non-expert practitioners to act on advanced threat mitigation. This enables organizations to protect data privacy across devices, users, and networks.
- Secure Access Service Edge (SASE) and Zero Trust Data Center: Organizations can extend and adjust their network security beyond traditional firewalls and enable remote employees to work in a secure manner and protect their personal information from cyber threats. Provisioning a consistent security policy model that behaves and enforces efficacy to the highest levels achievable was previously difficult or impossible in the security industry, but recent innovations have changed that paradigm and these advances enable organizations to realize these benefits throughout their entire environment.

Juniper believes that OSTP/NITRD policy should recommend that technologies such as these that detect and address cyber threats while protecting and enhancing privacy. End users are able to deploy these as needed in a variety of environments – including if the data in question is financial, medical-related, or mission critical.

Thank you for your consideration of these recommendations.

Sincerely,

Mike Spanbauer
Field CTO Security Architecture

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

KarlsGate Inc

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Comments on Advancing Privacy-Enhancing Technologies

Respondent Karlsgate Inc.
3 Kobert Ave
Budd Lake, NJ 07828
<https://karlsgate.com>

Respondent Type Industry

Respondent Contact Brian Mullin
Founder & CEO

Responding to topic 2. (Specific technical aspects or limitations of PETs)

Privacy Enhancing Technologies (PETs) can become a great aid to responsible data management in both the public and private realms. Privacy and the control over who becomes aware of personal identity has protective, ethical, and legal implications.

NEW EMERGING TECHNOLOGIES

An emerging technique called Partitioned Knowledge Orchestration, which is in a family of secure multi-party computation technologies. This is particularly useful for data sharing operations where the **disassociation and confidentiality of identity** is critical. The defining feature of this approach is the deliberate fragmentation of any complete information that could lead to either identification or re-identification. This partitioning is created during a careful orchestration of transactions performed by at least three independent actors. Of the three actors, two will be data controllers that compute identity hashes, and one will be a blind facilitator, which receives identity hashes from each data controller for comparison but never receives the one-time-use formula used to build the hashes. This methodology is particularly useful when considering how to protect the data from re-identification risks.

THE HEART OF THE PROBLEM: DATA LINKAGE

Partitioned Knowledge Orchestration was purpose-built as a solution for the critical **linkage problem**, which can be defined as:

Given, two independent entities (public or private) are each managing a dataset about individuals. The understanding of each individual's identity is achieved using various identifiers such as name, postal address, email, and/or social security number. However, these components of personal data are sensitive and are tied to personal privacy rights, regulatory restrictions, and/or ethical handling concerns.

How to enable the 2 independent entities to share the understanding of the individuals in common between the 2 datasets without sharing any personal data and without inadvertently allowing reidentification of those individuals not in common (i.e., outside of the desired intersection)?

Some real-world examples of the linkage problem (where data privacy and data sharing are equally critical) include:

- Retrieving Protected Health Information (PHI) associated with a patient in the form of Electronic Health Records (EHR) from another health system.
- Detecting fraud or anti-money laundering activity between banking institutions
- Researching rare disease treatments by gathering longitudinal views of patient data
- Detecting duplication in voter registration databases
- Anonymizing contact tracing interactions for viral exposure monitoring

SHARING DATA SAFELY & SECURELY

The linkage problem causes many challenges when it comes to data collaboration. Whenever records with individuals as the data subject are joined together, solutions commonly used today require that one party needs to fully trust the identity of subjects with the other party. Sometimes that is due to direct, clear text data sharing. Other times, pseudonymization techniques are used. It is important to note that pseudonymization enables reidentification when combined with additional information and is defined clearly as "personal data" under GPDR regulations in the European Union. Several categories of PETs protect the privacy of data but still serve as a form of pseudonymized data when treating identifying data.

Another consideration for advancing PETs is the need for prolonged public and government sector evaluation of the safety, fidelity, and reliability of new cryptographic technologies. Since Partitioned Knowledge Orchestration reuses proven cryptography in an orchestrated manner, it has a key advantage in institutional settings. There is no need to review, analyze, and vet new and emerging cryptographic algorithms, each with its own weakness profile. Instead, the sequenced orchestration can rely completely on proven, FIPS-compliant cryptography that can be substituted with other approved algorithms over time in a very natural evolution. Again, **this is not a 'black box' technique** but an orchestrated framework of interactions with defined actions including secret key derivation, one-way hashing, and symmetric encryption transformations.

EMPLOYING FIPS COMPLIANT CRYPTOGRAPHY TO SOLVE THE LINKAGE PROBLEM

Some common choices for the cryptographic algorithms are: Elliptic Curve Diffie–Hellman (secp384r1) for the secret key derivation covered by FIPS PUB 186-4, HMAC-SHA-384 for the one-way hashing covered by FIPS PUB 180-4, and AES-256-CBC for the symmetric encryption covered by FIPS PUB 140-2.

The use of well-studied and FIPS-compliant cryptography can accelerate adoption and application of PETs in government settings.

Several observable attributes of Partitioned Knowledge Orchestration can satisfy the design goals of data sharing operations that are vulnerable to the linkage problem.

- Zero-trust required of exchange partners including actors in facilitating roles, since no other party ever receives information that can lead to re-identification. This forms a stronger information security stance that cannot be compromised due to never enabling exposure. There are no data artefacts or exhaust from an exchange that can be retained to leverage for re-identification attacks.
- Zero change in data custody preceding, during, or subsequent to the sharing transaction. This is vital to establishing provenance and permissioning over data assets without back-door allowances for data sharing operations.
- Zero obfuscation of the collaboration methodology including cryptography, matching rules, or expected outcomes. This leads to faster adoption of the PETs and higher levels of trust among government entities, private industry, and the public at large. New cryptography poses higher risk of undiscovered vulnerabilities. A transparent approach is superior to an opaque approach when cryptanalysis and wide-spread acceptance is required.
- Zero limitations on topology of the exchange network promote an open environment that can cross technological, industrial, and geopolitical boundaries. Since the data intersection architecture is completely distributed and decentralized, there is no limiting factor on capacity. Every transaction enrolls 3 actors: 2 encoding nodes and 1 comparison node between them. Each of these actors can be commissioned and decommissioned at will including once per transaction. This approach is a natural fit for global scale and open architecture.

COMPARING THE DATA PROTECTION METHODOLOGIES

The following table details the comparison of common PETs for personal identity linkage applications:

Encryption Using a secret key, information is scrambled until the key is re-applied	<ul style="list-style-type: none">• Only protects data in transit• The recipient will decrypt back to fully identified data upon processing• Full custody change
Hashing / Tokenization One-way scrambling of data that is very difficult to reverse	<ul style="list-style-type: none">• Full change of custody leads to uncontrolled, future re-identification attempts against an identity graph• Data owner has no agency over copies of pseudonymous data and that is why the practice is not GDPR-compliant

<p>Data Clean Room</p> <p>Full dataset is sent to a controlled and isolated environment that is typically managed by a third party</p>	<ul style="list-style-type: none"> • Each usage represents a full custody change event • Both parties must trust the same operator with full vetting, legally binding responsibilities, consent and security obligations • Difficult to have a single clean room service for all partners and there is no clean room-to-clean room exchange mechanism
<p>Fully Homomorphic Encryption</p> <p>An advanced form of cryptography that allows analysis of data without decrypting the payload</p>	<ul style="list-style-type: none"> • Does not solve the linkage problem; while the exchanged attributes are protected, the payload can still be attacked for re-identification through joining to a known identity graph • Performance problems make this technique max-out at ~5 million records practically speaking
<p>Federated Learning</p> <p>Building an aggregated model constructed from multiple, localized machine learning processes</p>	<ul style="list-style-type: none"> • Does not solve the linkage problem; while powerful in building audiences and propensities in a privacy sensitive manner, it does not produce a deterministic intersection between two identified data sets, therefore limiting its use cases
<p>Differential Privacy</p> <p>A process of adding randomly generated adjustments to data without introducing significant skew</p>	<ul style="list-style-type: none"> • Does not solve the linkage problem; while it can be helpful for obfuscating descriptive values, it has no purpose for directly linking one identity to another • May degrade accuracy for certain use cases
<p>Partitioned Knowledge Orchestration</p> <p>A coordination to share common identities while simultaneously blocking unwanted reidentification as a consequence of interacting</p>	<ul style="list-style-type: none"> • No identifying information, including ciphertext or hashes, ever flow directly from partner to partner • The facilitator is deprived of any cryptographic parameters, blinding it to any ability beyond simple comparisons • No trust is required to safely connect data sets and extract insights, since no identity information is put at risk

A NOVEL APPROACH DESIGNED BY KARLSGATE

KarlsGate has developed a practical implementation of the Partitioned Knowledge Orchestration technique called Triple-blind Identity Mapping Protocol (T-BLIMP), specialized for linking identity information at scale.

The orchestration procedure of T-BLIMP follows a simple sequence.

1. The orchestration of events starts with managing single-use cryptographic keys controlled solely by the trading partners.
 - a. Each party randomly picks a seed value (a very large number).
 - b. A shared secret is derived between the 2 partners via the Diffie-Hellman algorithm and is never shared with anyone else.
 - c. Identifier values are extracted from the raw data, e.g., email address.
 - d. Single use Hash Tokens are assembled from the identifier values + the seed values + shared secret + the selected cryptographic hashing algorithm. (Hashed Tokens are different for every trade.)
2. The next step enlists a third party, the facilitator, to stand between the trading partners to perform neutral and naïve adjudication of hash collisions.
 - a. Only Hashed Tokens are transmitted to the Facilitator by each trading partner.
 - b. The Facilitator compares the lists of Hashed Tokens and sends back a signal on each matching value directly to the trading partners. (The Facilitator never knows the math used to compute Hashed Tokens.)

This technology is fully developed, commercially available, validated by industry, and in international use at scale.

KarlsGate has been commercially offering privacy-enhancing dataset intersection services, based on T-BLIMP, since early 2020. The enterprise-scale solution is already processing data sharing transactions among very large data sets with over 1 billion identity comparisons per transaction. Usage of the service spans the globe including users in United States, Mexico, and Australia. The most common industries using the solution to date include healthcare, financial services, and digital marketing.

PRIVACY ENHANCING TECHNOLOGY FOR THE PUBLIC SECTOR

The use of PETs can have a major positive impact on the safe and ethical use of governmental and public institutional data. However, as described above, many of the methodologies that are commonly used today have made limited, if any, progress with respect to solving the linkage problem inherent in majority of the use cases where PETs could be of value. Partitioned Knowledge Orchestration is the ideal technology for meeting the public sector need. It enables the linking of identifying data without ever compromising privacy or security.

Responding to topic 3. (Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs)

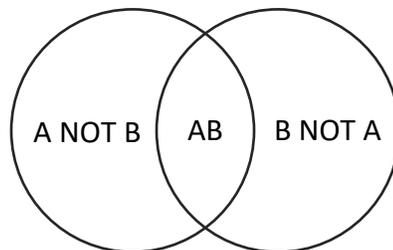
Data collaboration use cases continue to grow across all sectors and verticals as the value of data is unlocked with ever evolving technologies that enable leveraging data for data sharing, behavioral tracing/tracking, artificial intelligence, analytics, planning, reporting, targeting, measurement, personalization, etc. With data collaboration comes the need to be able to **link data without enabling any form of re-identification.**

With a challenge that is so pervasive, it would be unwieldy to create a comprehensive list of all potential applications that would benefit from adoption of PETs. However, the following applications illustrate a few of the types of applications that would particularly benefit from the adoption of PETs which specifically solve for the **linkage problem**:

NB: Special care is needed to avoid technologies that include or depend on pseudonymous keys. Many PETs only obfuscate identity through the use of pseudonyms (i.e., directly passing data values between partners that serve to uniquely reference individuals). The possession of pseudonymous keys inherently enables re-identification immediately or anytime in the future.

- Healthcare HIPAA compliance when sharing patient data while eliminating the transfer of specific Protected Health Information
- Cross-border data exchange where jurisdictions apply differing regulations on the Right to Privacy and use of Personal Data
- Financial fraud blacklisting and other domestic and international sanction lists
- Screening traveler identities against the FBI Terrorist Screening Center No Fly List
- Rare disease treatment research using Real-World Data (RWD), which can support and/or complement data collected from traditional clinical research
- Contact tracing operations for public health purposes
- Checking against FDA Debarment lists or other similar regulatory watchlists

A common requirement in many of the above use cases is that 2 organizations (with data sets **A** and **B** respectively) need to find the identities belonging to set **AB**, while never revealing the identities belonging exclusively to either set **A NOT B** or set **B NOT A**.



References:

- **GDPR** - EU General Data Protection Regulation <https://gdpr-info.eu/>
- **FIPS** - NIST Computer Security Resource Center <https://csrc.nist.gov/publications/fips>
- **T-BLIMP** - KarlsGate Identity Exchange <https://karlsGate.com>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Knexus Research Corporation

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Inequity Resulting from Partitioned Deidentification

Subject: RFI Response: Privacy-Enhancing Technologies

Agency: Office of Science and Technology Policy

RFI Description: The Office of Science and Technology Policy (OSTP) requests public comments to help inform development of a national strategy on privacy-preserving data sharing and analytics, along with associated policy initiatives.

Submission Date: July 8, 2022

Submitted To: PETS-RFI@nitrd.gov

Respondent Organization: Knexus Research Corp.
174 Waterfront Street, Suite 310
National Harbor, MD 20745

Respondent Type: Industry: Gov 8(a) Small business contractor

Respondent Description: KRC is an 8(a)-certified small business and a leading provider of Artificial Intelligence (AI), Privacy-Enhancing Technology (PET) and professional services to the US Government and the commercial sector. KRC has been a performer on Federal Contracts for 16 years since its inception in 2006, and its profile can be searched on the System of Award Management (SAM).

Respondent Technical POC: Dr. Christine Task

Respondent Administrative POC: Dr. Kalyan Gupta, President

Response Co-Authors: Dr. Christine Task, Lead Privacy Scientist
Aniruddha Sen, Intern

Topics Addressed:

1. Specific research opportunities to advance PETs
2. Specific technical aspects or limitations of PETs
8. Existing best practices that are helpful for PETs adoption

Inequity Resulting from Partitioned Deidentification

Introduction

Knexus Research Corporation (KRC) has worked with the National Institute of Standards and Technology (NIST) on benchmarking and evaluation of synthetic data generators, the UNECE Synthetic Data Working Group on defining best practices for synthetic data, and the US Census Bureau (USCB) on the research, development, and engineering of synthetic data generators. For this RFI, we wanted to share resources relevant to data deidentification and contribute a simple but important observation about the deidentification of diverse populations.

We begin by defining data deidentification and synthetic data as a category of PET and providing pointers to resources for both understanding the current state of practice and supporting future R&D in this field. [Topic 1]

We then identify a particular inequity issue that can arise due to distributional differences when deidentifying diverse populations. We provide a toy illustration showing that even a relatively large subpopulation (25%) can potentially be unintentionally erased from the data during deidentification, if its distribution differs too significantly from that of the majority population. This applies to all deidentification techniques that use partitioning as a preliminary step, including both traditional deidentification techniques (subsampling) and more recent formal privacy methods (differentially private histograms). [Topic 2]

We recommend, as best practice, that techniques should be evaluated separately on subpopulations as well as the overall population; these evaluations should be considered by stakeholders when configuring a data deidentification approach. [Topic 8]

Finally, we note complexities in this recommendation—distributionally distinct subpopulations may not be easily defined, similar inequity issues may arise more subtly in more complex (non-partitioning) deidentification approaches, and the problem of optimally addressing these issues for diverse populations is not fully solved. We recommend further research.

Deidentified and Synthetic Data

Topic 1: Specific research opportunities to advance PETs

Deidentified Data refers to anonymized data that has been processed with the intention of preventing the reidentification of the individuals in the dataset. Effective deidentification approaches provide both good *privacy* (successful defense against individual reidentification) and good *utility* (query results similar to the original data for population-level queries). Deidentification supports the following capabilities:

- Safe public release of sensitive data, enabling broad transparency and access.
- Safe internal use of sensitive data, reducing the risk of misuse and the burden on cybersecurity to prevent leaks.

Synthetic Data is a category of data deidentification that goes further by using modeling techniques to replace the original population with a new set of artificially generated synthetic individuals who have a similar data distribution at the population level. Because this new synthetic data does not contain real persons, we can refer to it as depersonalized data. It provides additional capabilities:

- Strong individual privacy protection.
- Retention of a data product for population-level analysis after the deletion deadline of the original personal data.
- Reuse of a data product for population-level analysis beyond the application scope of the original personal data.

We believe deidentified data plays an important role in the data privacy ecosystem. However, this is an evolving technology, and research and engineering challenges remain with regards to both privacy and utility. We provide references to the following resources on the current state of deidentification and recent research:

- **UNECE HLG-MOS Synthetic Data Test-Drive Website**
 - Over twenty national statistical agencies and interested organizations around the globe participated in a test drive of currently available synthetic data techniques.
 - Resource type: Directory of currently used synthesis techniques, as well as utility and privacy metrics chosen by data stakeholders. Challenge submissions include discussion of successes and failures for current techniques.
 - Location: https://pages.nist.gov/HLG-MOS_Synthetic_Data_Test_Drive/
- **NIST Differential Privacy Synthetic Data Challenges**
 - A sequence of eight challenge sprints over four years, scoring the performance of formally private synthetic data generators in difficult, real-world data contexts.
 - Resource type: Archive of challenge problems, techniques, performance results, and open-sourced solutions.
 - Location: <https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2020-differential-privacy-temporal>

Partitioning Plus Privacy Impacts Equity

Topic 2: Specific technical aspects or limitations of PETs

Data deidentification is intended to protect individual privacy and support population-level data analysis. As a result, it doesn't work on small groups--if a subgroup contains only a few individuals, such that they are recognizable as clearly distinct in the data, then as part of ensuring that no individual can be reidentified, the deidentification process will generally ensure poor utility for queries that would single out these individuals.

This raises an important question: What makes a group small?

We briefly introduce one behavior of deidentification algorithms on diverse populations. A diverse population is defined here to be a dataset containing more than one subpopulation, in which each subpopulation has a significantly differing distribution in the feature space. We draw a distinction (not typically done) between deidentification techniques that partition the individuals in the dataset and those that do not:

- **Partitioning Deidentification (table-based):** These approaches begin by considering the data in table form, where each column represents a feature (e.g., sex, race, or county) and each row has the count of individuals with a given combination of feature values (e.g., [Female, White, Alexandria]: 25,312). Deidentification is then performed by altering these counts. Because each individual contributes to only one count, we say that these techniques *partition* the data. Partitioning deidentification includes traditional approaches, such as

subsampling and cell suppression, as well as some recent techniques, such as differentially private histograms.

- **Non-partitioning Deidentification (query-based):** Non-partitioning approaches capture the data distribution using a sequence of queries in which one individual may contribute to multiple query results; different queries typically focus on different aspects of the data distribution. These techniques include model-based synthetic data.

An implicit assumption of partitioning-based deidentification is that small table counts represent outliers in the data, who (1) are more vulnerable to reidentification due to their unusual combination of feature values, and (2) as outliers can be altered or removed without substantially impacting analyses of interest. Deidentification then protects these individuals by using approaches such as redacting them (e.g., k-anonymity, cell suppression) or randomizing their counts (e.g., subsampling, differentially private histograms).

However, this assumption may be overly simplistic. In diverse populations, the choice of which features to include in the schema can potentially cause a relatively large subpopulation to be dispersed across many small table counts. This could result in the subpopulation being heavily altered or erased during deidentification. We illustrate this with a toy example below.

Figure 1: Subsampling

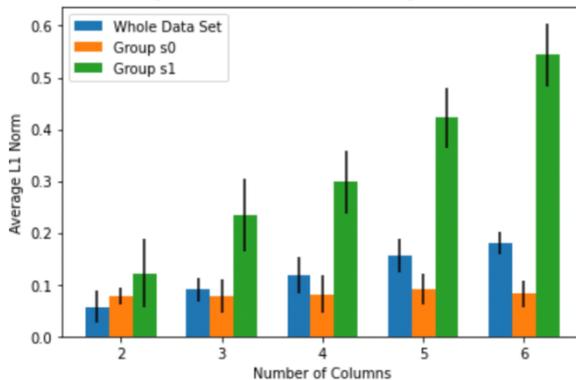
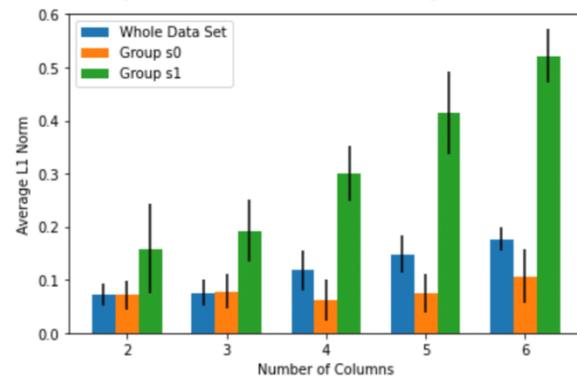


Figure 2: Simple DP Histogram



Our toy example contains a majority subpopulation s0 comprising 75% of the data, and a minority subpopulation s1 comprising the remaining 25%. We consider two partition-based deidentification algorithms:

- **Subsampling and weighting:** A 50% subsample of the original data is released, with each record given a weight of 2. Subsampling is one component of the privacy solution currently in use for the American Community Survey microdata
- **DP Histogram:** Laplacian noise with scale 1 (epsilon = 1) is added to every table count. This is arguably the simplest approach for differentially private data sharing and provides formal privacy guarantees against the release of individually identifying information.

Our experiment begins with a schema containing only two features (e.g., age and sex) and then iteratively adds additional features by appending new columns to the table. Each new column adds new information about the individuals in the data set. In general, releasing more information means individuals are more identifiable, and this increases the impact of deidentification on utility.

However, because the two subpopulations have different distributions, their utility is impacted differently. We use an extreme case for this example: In subpopulation s0 the new columns are strongly correlated with previous columns, whereas for subpopulation s1, the new columns are independent of the previous columns. This means that the distribution of the majority subpopulation will tend to remain concentrated and preserved with the addition of the new columns, whereas the

distribution of the minority will tend to be diluted and dispersed. When all six columns are added, the minority subpopulation becomes spread across many small table counts; the new features are effectively more identifying for s_1 than s_0 .

Figures 1 and 2 show the average error (expressed as L1 norm between the original and deidentified data density) while the number of columns increases. Error is shown separately for each subpopulation and for the group as a whole. In this toy example, we see that the error of the minority s_1 increases dramatically, whereas the error of the majority s_0 remains essentially constant. The overall population error only gradually increases as new information is added.

This type of utility inequity has the following properties:

- It arises when subpopulations have significantly differing distributions. It is especially a concern for partitioning deidentification algorithms, including traditional deidentification techniques.
- There are many ways this can occur beyond the example above; in general, a given partitioning scheme can have a disparate impact when subdividing different subpopulations.
- It can potentially impact even relatively large subpopulations (25% of the population in the above example).
- It may not show up clearly in utility evaluations that cover the full population, and so it may be overlooked when the deidentification approach is being designed/configured. This is true both when it's a person deciding what information to release, or an algorithm fitting a model (more below).
- It should be addressed by separately considering utility for each subpopulation.

Moving beyond our toy example, we'd like to add three more points:

- We expect this issue may also arise to varying extents in more complex data-deidentification approaches, including non-partitioning approaches. Query results, dimension reduction techniques, summary statistics and model training are still dependent on choice of schema.
- Importantly, subpopulations in the data may not be trivial to identify a priori. It is always valuable to evaluate with respect to basic demographic divisions, but distributionally differing subpopulations may also be defined by cultural, work, community, or lifestyle attributes that aren't captured by race or gender.
- We believe this topic requires further research.

Subpopulation Evaluation as Best Practice for Maintaining Equity

Topic 8: Existing best practices that are helpful for PETs adoption

For data deidentification, we recommend that utility for subpopulations be evaluated separately, in addition to overall utility, and that privacy and fairness be considered together when stakeholders are deciding on a data release strategy. We recommend this for all data deidentification approaches, including recent techniques (differential privacy, synthetic data) and techniques in common use currently (subsampling and cell suppression).

The previously referenced HLG-MOS Synthetic Data Test-drive website contains a large directory of utility evaluations for deidentified data. Additionally, in the coming months the SDNist library is expecting to release a set of benchmark data and evaluation tools designed to support detailed exploration of algorithm performance on diverse, real world populations. Updates will be posted here: <https://github.com/usnistgov/SDNist>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Kotevska, Olivera

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

RFI Response: Advancing Privacy-Enhancing Technologies (35250)

Attention:

Jeri Hessman

pets-rfi@nitrd.gov

Submitted by:

Olivera Kotevska
Research Scientist
Oak Ridge National Laboratory
1 Bethel Valley Rd
Oak Ridge, TN 37831-6013
PO Box 2008, MS6013

Introduction

This white paper summarizes my feedback regarding the “Advancing Privacy-Enhancing Technologies” in response to the subject request for information.

Dr. Olivera Kotevska received her Ph.D. from the University of Grenoble Alpes in France and was an international guest researcher at the National Institute of Standards and Technologies (NIST) Gaithersburg, MD, during her doctorate. She started working at Oak Ridge National Laboratory (ORNL) in 2019 and is currently a research scientist. Her research is in the area of privacy algorithms for complex systems. She received internal ORNL funding for her research of privacy algorithms for streaming data and edge computing and additionally works on the projects for privacy protection of machine learning models. She has a few publications on these topics in reputable IEEE conferences ^[1-3], DOE white paper ^[4-5], and DOE open-source license ^[6].

Specific Recommendations

The current strategic vision does an excellent job of identifying and describing issues that must be addressed by the U.S. in the development of privacy enabled technologies. The descriptions and issues identified remain accurate and timely. As the vision development progress, it would be valuable to identify specific recommendations, policies, and

investments that will be important in sustaining U.S. leadership in privacy preservation. The following paragraphs summarize these recommendations in topical areas.

- **U.S. Leadership.** In the last few years, privacy has been identified as an integral part of today's and future technologies. As the technology progresses toward interconnectivity ^[7], voice-activated systems, automated and autonomous intelligent decisions, data becomes crucial for their operation. This explosion of data is increasing rapidly and exponentially, collecting sensitive information from healthcare to critical infrastructure. In this direction, the EU developed the first privacy regulation called GDPR ^[8] to protect users from unauthorized data collection, sharing and publishing. They even had a few court cases regarding data privacy with top tech companies to enforce its use. However, the U.S. has the CCPA ^[9], HIPAA ^[10], ECPA ^[11], and COPPA ^[12] created by congress and signed into law, but these are for a specific state, a specific type of application, or target group. Many international companies follow the GDPR standard, and it is not clear that the U.S. has the leadership in this position internationally. Privacy must be part of any data sharing, collection, and transfer to protect sensitive information from potential intruders.

The plan should provide information regarding U.S. leadership in privacy preservation and regulations applicable on national and international levels. The development of privacy regulations that protect users and entities (e.g., power grid companies) should be developed and act as an example for other countries (e.g., EU and China) so U.S. can establish strongly their leadership position.

- **National Security Issues.** There have been cases where major U.S. entities had cybersecurity and data breach attacks. Some of them are U.S. Electric Grid'17, U.S. Federal Aviation Administration'15, U.S. Healthcare Network, SolarWinds supply chain data breach'20, First American Financial Corporation data breach'19, and many more. Because of these cases, the importance of privacy preservation technologies was addressed by the recent report published by the National Academy of Engineering ^[2]. For example, the electric grid is quite vulnerable to cyber-security attacks, and privacy preservation algorithms can provide an additional layer of data protection. Future applications (e.g., self-driving cars, transportation, energy automation, remote surgery, and patient monitoring) transfer raw unprotected data over the Internet. Privacy technology can protect sensitive data as close to the source as possible before it gets transferred to the cloud.

The plan should recommend national and international norms for privacy preservation in critical infrastructure, economic situations, and healthcare. The privacy preservation technologies should be incorporated into the existing solutions and act as an option for the desired level of privacy protection.

- **Regulations.** As mentioned in the U.S. leadership section, some state and federal regulations exist in the U.S. The private sector in the U.S. lacks transparency in its technical specification if they follow any regulations and their response to the attacks on the Internet of Things (IoT) devices ^[13-14]. While in the E.U., regulations are mandatory, and this is clear, so the tech companies that operate in the E.U. need to follow them. This information creates trustworthiness by delivering this information to the users.

The plan should include the creation of a law to address the mandatory privacy protection regulations across domains. Regulation could be two-fold: government policies giving the general direction in handling privacy and data protection or self-regulation to ensure that industries adopt best practices in cybersecurity and data minimization.

Research Investments. The U.S. government agencies (e.g., NSF, DARPA, NIH) invested in the scientific foundations of privacy preservation. Particularly in the area of fundamental development of privacy algorithms, databases, and healthcare. However, most of the research has been performed by universities and partly by the national laboratories and industry. There are a few ways to advance future privacy-enhancing technologies (PET).

- Taking the algorithms in the research community and making them in practical application.
- Development of software solution for mapping the privacy regulations (*mentioned in the Regulation section*) into rules and integrating them into the existing application platforms. These include:
 - Utilize analytics to automated policy configuration and management.
 - Explaining current access policies over data.
 - Explaining the impact of changes to access policy.
- Future research directions in privacy
 - Privacy for obfuscating data sources in images and other data types.
 - Better explanations and visualizations of privacy preservation.
 - Development of better privacy measurement and utility so that exploratory analytics is truly possible.
 - Enable meaningful control over data and mechanism of deciding how much privacy the user can have (*mentioned in the Industry and Application part*).
 - Development of mechanisms for checking if current data sharing matches historical records across the same user and other users and learning an optimal data sharing policy.
 - Incorporating privacy into decision-making, machine, statistical algorithms, and fairness analysis.

- Development of trustworthiness in privacy algorithms is needed as it will encourage the users to share sensitive information and use the new technology.
- Development of lightweight solutions that are robust to more powerful adversaries.
- Development of solutions resilient to privacy attacks, especially under distributed settings.
- Research opportunities in science
 - Development of privacy layer for supercomputing resources.
 - Development of various surrogate datasets for testing privacy advances the development of robust privacy algorithms.
 - Development of distributed machine learning approaches (e.g., federated learning) with privacy capabilities to encourage data sharing between different entities. National laboratories have a challenge in accessing diverse datasets, so a solution that can guarantee the protection of sensitive information will encourage data sharing with the industry.
- The next generation of privacy-protecting systems.
 - As we progress into new computational architectures, e.g., quantum Internet and quantum computing, a neuromorphic adaptation of the privacy algorithms is needed.
 - Incorporate diverse privacy models and incorporate a socio-cognitive approach to improve privacy.
- **Industry and Applications.** Industry plays a crucial role in developing privacy-enabled technologies and making these capabilities available for use. In a few cases, the industry has performed fundamental research, e.g., Apple, Google, and Samsung. However, it is unclear if and what type of privacy preservation is applied by reading the specifications. They are long, unclear, and written in a language most people cannot understand. Most people do not read them, and very rarely is there an option for the user to choose their privacy preservation preferences. The relationship between industry and government in the U.S. is not clearly defined.

The plan should address this issue explicitly and explore how to coordinate federal research investments with industry investments. Policies could be three-fold, i.e., a legal code prepared by lawyers and interpreted by courts, human-readable code easily understandable by consumers, and machine-readable code embedded in the IoT software. Data protection should be built into the IoT solution from the outset throughout the development life cycle as part of the principle of ‘privacy by design.’ Create awareness of the level of exposure associated with connectivity through interaction with IoT and ensure consumers are afforded opportunities to access and control their data.

- **Education.** A recent study showed that people have a minimal understanding of privacy preservation and complicated manuals, and programs make them give up easily.

The plan should create educational programs, materials, and delivery methods in high-school, universities, and companies for privacy awareness and introduce a basic understanding of existing regulations and how to choose products with privacy features.

The short paragraphs above are intended to present ideas for expanded or added content to the revised national Privacy-Enhancing Technologies plan. If the Council or Subcommittee would like additional information or has questions on any of the topics above, please contact Dr. Olivera Kotevska.

References

- [1]. Alfonso, S., Danfeng (Daphne) Yao, Kotevska, O., Measurement of Local Differential Privacy Techniques for IoT-based Streaming Data. In 2021, *18th IEEE Conference on Privacy, Security, and Trust (PST)* (pp. 1-20). IEEE. Acceptance rate: 24%.
- [2]. Kotevska, O., Alamudin, F., Stanley, C., Optimal Balance of Privacy and Utility with Differential Privacy Deep Learning Frameworks. In *2021 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 425-430). IEEE. Acceptance rate for regular paper: 16%.
- [3]. Peralta-Peterson, M., Kotevska, O., Effectiveness of Privacy Techniques in Smart Metering Systems. In *2021 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 675-678). IEEE. Acceptance rate for short paper: 24%.
- [4]. O. Kotevska, C. Stanley, R. Michael, B. Kay, A. Sarwate, R. Kannan, G. Tourassi. Challenges with Sensitive Data in Distributed Graph. *Published to DOE ASCR Workshop on Cybersecurity and Privacy for Scientific Computing Ecosystems.* Link: <https://custom.cvent.com/DCBD4ADAAD004096B1E4AD96F3C8049E/files/event/f9a6d32df34c43e5980b54f79ad848bc/738372bbcd246ac9a9c00a177c075ac.pdf>
- [5]. J.R. Michael, C. Stanley, R. Adamson, O. Kotevska, Addressing the Limitations to Distributed Learning Containing Sensitive Data. *Published to DOE ASCR Workshop on Cybersecurity and Privacy for Scientific Computing Ecosystems.* Link: <https://custom.cvent.com/DCBD4ADAAD004096B1E4AD96F3C8049E/files/event/f9a6d32df34c43e5980b54f79ad848bc/2e42c4c5dbe74b578e47188036244a10.pdf>
- [6]. Kotevska, Olivera. *Privacy Algorithms for Streaming Data.* Computer Software. <https://github.com/ORNLStreamingPrivacy/Algorithms.git>. Web.
- [7]. National Academy of Engineering 2019. *Privacy and Security in the 21st Century: Who Knows and Who Controls? Proceedings of a Forum.* Washington, DC: The National Academies Press. <https://doi.org/10.17226/25575>.
- [8]. General Data Protection Regularization (GDPR). Link: <https://gdpr-info.eu>. Last visited: 06.29.2022. The Fourth Industrial Revolution. Link: <https://www.weforum.org/about/the-fourth-industrial-revolution-by-klaus-schwab>. Last visited: 06.29.2022.
- [9]. California Consumer Privacy Act (CCPA). Link: <https://oag.ca.gov/privacy/ccpa>. Last visited: 06.29.2022.

- [10]. Health Insurance Portability and Accountability Act (HIPAA). Link: <https://www.hhs.gov/hipaa/index.html>. Last visited: 06.29.2022.
- [11]. Electronic Communications Privacy Act (ECPA). Link: <https://epic.org/ecpa/>
- [12]. Children's Online Privacy Protection Rule (COPPA). Link: <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa>. Last visited: 06.29.2022.
- [13]. IoT Devices Most Vulnerable to Internal Security Threats. Link: <https://www.iotworldtoday.com/2022/02/22/iot-devices-most-vulnerable-to-internal-security-threats/>. Last visited: 06.29.2022.
- [14]. Top 6 Hackable Medical IoT Devices. Link: <https://www.criticalinsight.com/resources/news/article/top-6-hackable-medical-iot-devices>. Last visited: 06.29.2022.
- [15]. Apple Differential Privacy Team. (2017, December). Learning with Privacy at Scale.
- [16]. Erlingsson, Ú., Pihur, V., & Korolova, A. (2014, November). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (pp. 1054-1067).
- [17]. Nguyễn, T. T., Xiao, X., Yang, Y., Hui, S. C., Shin, H., & Shin, J. (2016). Collecting and analyzing data from smart device users with local differential privacy. *arXiv preprint arXiv:1606.05053*.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Kulshrestha, Anunay; Mayer, Jonathan; and Scheffler, Sarah

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

July 8, 2022

Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, DC 20504

**Comment of Princeton University Researchers on
Advancing Privacy-Enhancing Technologies**

Thank you for the opportunity to inform the forthcoming national strategy on privacy-preserving data sharing and analytics. We are academic researchers at Princeton University who study information privacy and security using interdisciplinary methods. Our scholarship has examined data practices that affect individual privacy, the contours and efficacy of privacy law, and the design and implementation of new privacy-enhancing technologies.

In this comment, we briefly respond to two prompts in the request for information, drawing on our recent research experiences.

7. Risks related to PETs adoption: Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks.

Privacy-enhancing technologies can provide unprecedented capabilities, as the RFI rightly notes. But, in some instances, those capabilities can risk causing serious harms—calling into question whether the technology merits deployment.

In a recent publication, we constructed a novel privacy-preserving protocol for identifying harmful imagery in end-to-end encrypted services.¹ The conclusion of our

¹ Anunay Kulshrestha & Jonathan Mayer, *Identifying Harmful Media in End-to-End Encrypted Communication: Efficient Private Membership Computation*, Usenix Security Symposium (2021).

work was that, while the technology was viable, it also created a danger of misuse for surveillance and censorship by foreign governments.² Shortly before we were scheduled to present the publication, Apple announced a very similar privacy-preserving protocol for identifying harmful imagery stored with its iCloud Photos service.³ A global backlash followed—including from experts in PETs—because of the possibility that the system could backfire and *undermine* rather than *enhance* privacy.⁴ Concerns also emphasized possible threats to free speech and risks for vulnerable users. In response, Apple decided to pause its deployment of the PET.

We take two lessons from this experience. First, where a PET creates a capability that did not exist before, it is important to carefully examine the societal context for and consequences of that capability. Risks of misuse, exploitation, inaccuracy, or inequity may outweigh the upsides of the PET. These risks depend not only on the current implementation of the system, but also on the future motivations and capabilities of persons, organizations, and governments that could influence the system going forward. Communicating with the public about these risks may be extraordinarily difficult, and quantifying these risks may be effectively impossible.

Second, PETs can be more difficult to supervise and audit than conventional system designs, because they (intentionally) restrict the information that parties can obtain. Unfortunately, PETs research usually starts and ends with providing a new capability in a privacy-preserving way. There is an urgent need for complementary research on how to achieve transparency and accountability for PETs, while maintaining the strong privacy guarantees that make them so promising. In our own upcoming research, for example, we are developing new designs for public transparency in privacy-preserving hash matching systems.⁵

² Jonathan Mayer & Anunay Kulshrestha, *We Built a System Like Apple's to Flag Child Sexual Abuse Material—and Concluded the Tech Was Dangerous*, Wash. Post (Aug. 19, 2021).

³ Abhishek Bhowmick et al., *The Apple PSI System* (July 29, 2021).

⁴ See Hal Abelson et al., *Bugs in Our Pockets: The Risks of Client-Side Scanning* (Oct. 15, 2021); Sarah Scheffler & Jonathan Mayer, *Content Moderation in End-to-End Encryption* (in submission).

⁵ Sarah Scheffler et al., *Public Verification for Private Hash Matching: Challenges, Policy Responses, and Protocols* (in submission).

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs: Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs.

The federal government has an opportunity to lead by example. By adopting PETs, the government can offer new forms of public transparency and more rigorous statistics than were previously possible. The Census Bureau's implementation of differential privacy stands out as a particularly remarkable example of what the federal government can achieve by adopting PETs and how government use of PETs can advance the entire field.

Transparency about the data that the government holds about individuals, and the privacy implications of that data, is another promising application of PETs. In recent research, we examined how secure multiparty computation could provide a viable path forward for the decade-long challenge of estimating how Foreign Intelligence Surveillance Act collection affects ordinary Americans.⁶ While our work focuses on just one type of data (intercepted communications) and just one type of government component (elements of the Intelligence Community), the same concept could bring transparency to personal data holdings across the federal government.

With careful design, PETs could enable the best of both worlds: maintaining the necessary confidentiality and accuracy of important government systems, while also significantly improving their transparency and accountability.

* * *

Thank you again for the opportunity to provide input to the federal government's strategic plan for information integrity research and development. We would be glad to provide additional detail or discussion as would be helpful to OSTP.

⁶ Anunay Kulshrestha & Jonathan Mayer, *Estimating Incidental Collection in Foreign Intelligence Surveillance: Large-Scale Multiparty Private Set Intersection with Union and Sum*, Usenix Security (2022).

Sincerely,⁷

Anunay Kulshrestha

Graduate Researcher, Center for Information Technology Policy, Princeton University

Jonathan Mayer

Assistant Professor of Computer Science and Public Affairs, Princeton University

Sarah Scheffler

Research Fellow, Center for Information Technology Policy, Princeton University

⁷ We offer this comment as individual academic researchers.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

LeapYear Technologies

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>

Responding party: LeapYear Technologies, Inc (<https://leapyear.io>)

Entity type: Industry / software provider

Respondent: Head of Marketing, Garrett Long

1. Specific research opportunities to advance PETs:

The federal government has a unique opportunity to provide the public with a consolidated summary of the current state of PETs usage / testing in the public sector today. In certain scenarios it's well publicized that a PET is used (for example [Census](#)), but there are many other situations where PETs may be in use or trial that are hard to find out (for example this [DARPA](#) program, it's hard to understand what happened and the result) By providing the public and legislators a "state of play" it would help determine the actual state of PETs usage, highlight success and failures, and potentially identify areas to focus on for advancement.

A potential useful framework to report on these federal projects would include:

- The agency(ies) involved
- A description of the privacy problem to be solved, including definition of threat models and attack vectors
- A review of the type of data involved
- The PET(s) used or considered
- Time of the project (start, duration)
- A summary of the result (success, failure, other)

A second opportunity would be for the federal government to assemble a canonical guide to what PETs are, the maturity of a given PET, what problems they are well suited for, and when a PET is not useful. This type of guide can be used to help agencies and the public streamline consideration and selection of PET(s). This type of guide can also include when multiple PETs may be useful in certain contexts.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

The responses in this section will have to do with applications specifically for differential privacy.

In general, differential privacy solutions are well suited for applications that meet certain criteria:

- The data to be analyzed is sensitive (deemed so by regulation, law, contract or best practice)
- The data is structured / tabular
- The use case can tolerate a small and explainable amount of randomization in an analysis
- The analyst cannot (or should not) be trusted to receive access to row level data, but has legitimate purposes to query the data to reveal and use statistical properties contained therein

Sectors of application

Our work in the financial sector reveals many compelling use cases for differential privacy on market data. Market data can include consumer transactions (such as credit cards or bank accounts) and institutional data (such as equity transactions, fixed income pricing and over the counter traded derivatives). Generally, the financial sector is expected to, by regulation, provide appropriate liquidity, risk and best execution services for the market – ensuring a fair playing field for institutional and retail investors alike. Differential privacy has been proven in the private sector to have many useful applications under these considerations including:

- Helping asset managers receive best pricing from their banks
- Helping broker-dealers optimize demand forecasting (and by proxy) trading costs
- Helping protect citizens' financial behaviors while still enabling useful analysis of broader market trends
- Bringing new sources of information to derive investment strategies
- Eliminating unnecessary (and often mis-understood) leakage of private information

These same types of data analysis are routinely considered and executed by the financial regulatory bodies (including SEC and FINRA). Recent trends to require banks to hand over detailed transaction data sets (the consolidated audit tape) have created significant tension and public debate on how to secure this data and [keep it private](#). Differential privacy software can be used to improve the public/private partnership in this area. Federal agencies with differentially private access to data sets can achieve many of the stated goals of oversight,

without the need to stipulate massive data collection and storage which puts US institutions and citizens' privacy at risk.

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs:

[HIPAA](#) – today, HIPAA defines the mechanisms through which healthcare data can be modified (de-identified) and therefore deemed to protect the privacy of citizens whom are included in a data set. Unfortunately, the realities of information and computing power that is available today (due to the rise of digital systems) largely invalidates the assumptions under which the original regulation was written. Updating HIPAA (as has [already been proposed](#)) to require a more formal definition of privacy will help advance the adoption of PETs to protect the healthcare information of every US citizen.

SEC/FINRA – the SEC and FINRA enforce regulations on the financial services sector. Regulations are in place to ensure optimal market risk, liquidity and provide best execution to institutional and retail investors alike. These regulations often act in a way that reduces the sharing of data, either across an institution or with the larger ecosystem. Our work in the capital markets also reveals a mix of ad hoc methods, applied inconsistently across market participants, that attempt to deal with privacy issues. These methods often trap important signal in data sets, leading to assumptions that firms are safer than they really are and leaking privacy. The SEC and FINRA can motivate an industry standard on the way particular market data is protected and used leveraging PETs.

Treasury - The IRS collects sensitive demographic and income information on every citizen. The IRS also has the ability, and sometimes the requirement, to share and exchange this information with other federal, state, local and public agencies. But sharing this data or creating information products based on this data has inherent privacy risks. The IRS has already acknowledged the existence of shortcomings in the way they analyze data, and [has explored using differential privacy](#) to improve both data value and citizen privacy. Treasury can expand and formalize the use cases, risks and positive outcomes from the adoption of PETs for financial information. Regulations should be adopted in this particular set of use cases that would require the use of PETs to ensure citizens income and personal information is rigorously protected while still allowing the important and necessary work of the IRS and its partner agencies to continue.

5. *Specific laws that could be used, modified, or introduced to advance PETs:*

As above, an update to HIPAA.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Mastercard

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



July 8, 2022

By Electronic Mail to: PETS-RFI@nitrd.gov

Re: RFI Response: Privacy-Enhancing Technologies

Respondent: Caroline Louveaux, Chief Privacy Officer at Mastercard (Respondent Type: Industry)

Mastercard welcomes the opportunity to provide comments in response to the White House Office of Science and Technology Policy’s (“OSTP”) **request for information (“RFI”) on Advancing Privacy-Enhancing Technologies**. As a technology company in the financial services space, we are exploring a variety of privacy-enhancing technologies (“PETs”) in line with the data practices outlined in our [Global Data Responsibility Imperative](#).¹ In our comments, we provide background on Mastercard’s work in the PET space and outline some of the unique challenges that organizations face when deploying PETs. Our comments will provide examples of areas that could benefit from enhanced PET adoption and recommendations that we believe would advance the development and adoption of PETs in a way that offers high levels of protection for individuals, while also enabling data-driven innovation. In particular, we encourage OSTP and other regulators to promote PET adoption, to clarify that a dataset subject to appropriate PETs is not subject to data protection laws while still holding companies accountable for their use of PETs, and to actively monitor the field in the near term as technology continues to evolve. For ease of reference, we note the corresponding RFI topic numbers to which each section of our response pertains.

Exploring Opportunities for PETs (RFI topics 1, 3, 10)

At Mastercard, we have dedicated resources across a variety of teams to explore PETs including synthetic data, differential privacy, and transformation techniques with risk scoring for our data insights capabilities. Our research has demonstrated that while each solution has its own pros and cons and appropriate use cases, all of these technologies have a positive contribution to make to the field of privacy-compliant data analytics. We strongly support adoption of these tools to enhance individual privacy while enabling data innovation - both to power new products and services and help solve some of society's most pressing challenges.

Our exploration of these technologies is driven by a strong privacy culture combined with our technology mission. Bringing these two elements together has resulted in the development of a suite of new products and services. Our Data & Services team, for example, drives efficiency and value for our customers (from small businesses to large enterprises) and empowers them to solve problems with powerful analytics tools. As another example, our Cyber &

¹ Our Global Data Responsibility Imperative is a vision of how Mastercard believes data can be leveraged in a responsible and secure way to drive relevant benefits for individuals and society. It posits that individuals hold certain data rights and explains why data responsibility is a business and moral imperative. It proposes a model framework for how organizations could approach personal information and lays out principles to guide organizations' data practices, as well as characterizing the critical role data has in driving inclusive economic growth across the globe.

Intelligence teams help prevent fraud, financial crime, and cyber-attacks with sophisticated data-based tools that can spot individual and systemic level risks without compromising any personal information.

We are also committed to the use of data analytics to aid in humanitarianism, inclusive growth, pressing societal issues such as climate change, and global health crises. We work across industry and with private and public partners alike on projects for the public good. Mastercard's [Center for Inclusive Growth](#) has collaborated with governments, universities, and non-profits on a number of projects to responsibly deliver data-driven insights that enable a better understanding of inclusive growth around the world. For example, [insights based on Mastercard's anonymized and aggregated spending data](#) have helped with challenges faced by low-income communities in Chicago by offering a much more detailed, nearly real-time glimpse into not just what communities spend on, but where they spend, aiding public- and private-sector entities to ensure investments meet local needs and lower the cost of living. Our data insights have also been valuable following natural disasters and in recovering from the pandemic's impact on the global economy. We also recognize that synthetic data tools can support the advancement of inclusive growth and humanitarian objectives.

Other potential PET use cases include training machine learning models with synthetic data, which minimizes privacy risks while maximizing responsible innovation and healthy competition. PETs also enable new solutions to various problems through privacy protecting data combinations, where parties can share insights gained from multiple datasets without having to share the underlying data. For example, in the financial services industry the wide generation of synthetic data could help coalitions of companies and regulators to identify systemic risks or threats.

All of these uses of data will benefit from the evolution of tools and techniques in the PET space, provided that regulators and companies partner to support their development, acceptance, testing, and continued improvement. We believe that regulators should strongly endorse PETs, and clearly articulate that when the risk of re-identification of a dataset is sufficiently remote, the use of PETs should exclude the applicability of data protection law to a given dataset.

Regulatory ambiguity can have negative real-world consequences for innovation and social impact by disincentivizing companies and discouraging market entry. As examples, in markets where regulatory ambiguity increases risk and the cost of compliance for businesses, products that aid small businesses may take longer to launch, or tools for inclusive growth that require data to make decisions may be prioritized in other markets where risks and barriers to entry are lower. Certainty around privacy regulation, and the use of PETs in particular, will be fundamental in the coming years for innovation efficiency.

Reviewing the Challenges of Deploying PETs (RFI topics 2, 4, 5, 7, 9)

To fully appreciate the need for regulator endorsement of PETs, it is important to understand that organizations are naturally hesitant to enter a new field without a clear understanding of the technology, cost, and anticipated regulatory environment. The adoption of PETs within a large organization will require significant capital investment, potentially measured in the millions of dollars depending on the size and scale of a company.

At a minimum, deploying a PET will require financial resources to purchase or license the requisite software and hardware, and investments in personnel with a variety of skillsets. A successful deployment of a *single* PET may require software engineers, hardware engineers, corporate security specialists, privacy lawyers, privacy engineers,

project managers, data analysts, data modelers, database administrators, data strategy experts, and “red teams” to test, simply to make a PET *available* within an organization’s environments – before the PET is even applied responsibly to a product, service, or dataset. Testing and tuning a variety of different PETs further increases the demand on resources.

The complex and fragmented data protection regulatory landscape, both globally and within the United States, creates headwinds against PET adoption. Varied anonymization and/or deidentification thresholds under applicable state and federal laws create challenges for responsible industry players who seek to lead the way in unlocking the benefits of PETs. As we describe below, clear regulatory endorsement of PET development, acceptance, testing, and continued improvement will help mitigate some of these challenges.

Providing Regulator Endorsement to Drive Industry Adoption of PETs (RFI topics 5, 6, 8, 9, 10)

While some companies will take a pioneering role and invest in the PET space, the majority of them will take a wait-and-see approach which will slow innovation. To minimize this risk and remain competitive on the global stage, regulators should make clear that data subjected to appropriate PETs will fall outside the scope of data protection law, provided that the risk of linkage or re-identification in the dataset is within reasonable risk thresholds supported by research.

In addition, regulators should provide reasonable flexibility in the selection of various technologies until the field matures. Although a dominant technology is likely to emerge, we anticipate that forms of synthetic data, differential privacy, and transformation techniques with risk scoring will each play a role depending on use case (homomorphic encryption and secure multiparty computation will also add to the PET space depending on the use case, such as data sharing). For example, there are multiple techniques for generating synthetic data and different risks with each approach; while no single technique has yet been identified as the best model, regulators should continue to monitor the research and provide appropriate guidance on synthetic data as this field develops.² Anonymization by an independent third party may also be a solution or requirement in some cases. Existing technologies that were once computationally intensive will become a reality with newer computing resources. In each case, however, regulators can and should rely on reasonable re-identification risk thresholds to ensure that a technology maintains privacy, while endorsing the PET space in general.

We also highlight that use cases vary between companies and industries, and different use cases will have varying risks of re-identification and different “release models” (internal use only, external release of a dataset, or external release in aggregate form, as examples). In the financial services space, companies often find significant value in producing reporting that is aggregated, e.g., at the city level or, in some cases, a neighbourhood level with sufficient population sizes. The re-identification risk in such a use case is lower than what may be required when publicly releasing or allowing third parties to carry out analytics on granular data. Similarly, the risk levels with financial

² With regard to synthetic data in particular, it can be difficult to identify the age of a synthetic dataset, so markers should be added to the data to make this easy. In addition, only the original creator of a synthetic dataset will know how statistically close to the original dataset it is. If synthetic data is to be sold to or re-used by a third party, there should be a standard approach to describing the veracity of the synthetic dataset, without disclosing information that could reduce the privacy protections. These are areas where best practices are not yet defined but should be appropriately standardized.

transaction data may differ from health data or social media data that, by the nature of data elements included in a dataset, may be more identifiable or entail different privacy interests for individuals.

Nevertheless, as consumer-supporting industries seek to meet consumer needs in a variety of ways, a sector-specific approach to data would create uncertainty and inconsistency in how industries apply PETs for data analytics. For those reasons, companies should have internal flexibility to select risk thresholds for re-identification across a spectrum of identifiability and privacy risk, within reasonable constraints supported by research. Regulators could consider issuing grants to fund academic research or conduct meta-analysis in these areas. We also see that academic publications, patent applications, and commercial research are beginning to emerge that demonstrate the application of mathematical discipline to determining the risk of re-identification. Such research could inform the flexibility and guardrails given by regulators to companies using applicable technologies.

To drive adoption, regulators should also make clear that the act of applying PETs to identifiable information does *not* require additional or separate consent of an individual and falls outside of consent frameworks imposed by data protection law. PETs are intended to preserve and enhance privacy, and companies should be incentivized by removing uncertainty around legal requirements or withdrawing additional barriers. Similarly, it should be clear that companies may outsource PET functionality to vendors, as vendors are likely to have advanced capabilities backed by research and publications, and advanced technology.

Finally, due to the privacy-preserving nature of data processed by PETs and the inability to re-identify a dataset, it should be made clear that companies do not need to destroy an original dataset in order to comply with data protection laws, although it should be kept separate and under strict access controls. Companies may have separate uses and a separate legal basis for the personal data contained in an original dataset, such as to enable a primary business purpose or facilitate a transaction at the direction of a consumer, while data processed by PETs could be used for aggregated analytics or “data for good” projects.

Holding Organizations Accountable While Supporting Innovation of PETs (RFI topics 2, 7, 8, 10)

In line with privacy principles around accountability, and in exchange for regulator endorsement of PETs, regulators should hold companies accountable for their use of PETs. As with other technologies and uses of data, companies must deploy technical and organizational measures to demonstrate that their use of PETs in each context is appropriate. This could include documentation on why a PET was selected for a particular use case, the data elements involved, the type and nature of the processing and analytics, the outputs and levels of aggregation, the release model, and the selected risk threshold (or results of a motivated intruder analysis). These measures would be combined with records of processing, data protection impact assessments, and other normal recordkeeping as part of a robust privacy program.

Regulators should also encourage vendors in the PET space to consider the “explainability” of their PETs from the earliest stages of development, just as privacy by design requires engagement of privacy principles from day one of product development. Vendors in this space could build, for example, dashboards or reporting capabilities with appropriate explanations of how a tool was deployed, how risks of re-identification or singling out were minimized, and which risk thresholds were used based on the intended use case, release model, and aggregation level of the output. Regulators might also consider developing an industry code of conduct that provides guidance on how

companies can promote their use of PETs to consumers in a user-friendly and accurate manner, based on the controls and technical and organizational measures that they have in place.

Companies benefitting from the use of PETs should also be required to demonstrate they have made appropriate investments in their controls and auditing, which may include the retention of privacy counsel, privacy engineers, and auditors with expertise in the area.

Keeping Pace with Technological Innovation and Developments in the Law (RFI topics 1, 4)

Finally, we encourage regulators to stay actively involved in the conversation surrounding PETs, especially in the next few years as rapid advancements in the technology are expected. Academic and commercial research may shift thinking toward one technology or demonstrate that certain use cases are lower or higher risk than others. Continued regulator engagement in this area will encourage the rapid pace of innovation with a focus on privacy of the individual.

We are available to answer questions that OSTP may have regarding our comments, and we would welcome an opportunity to further discuss this topic with you at your best convenience.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Meta

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



July 8, 2022

Via Electronic Mail

Office of Science and Technology Policy
1650 Pennsylvania Avenue
Washington, DC 20504

RE: Request for Information on Advancing Privacy-Enhancing Technologies

Thank you for the opportunity to respond to this Request for Information. Meta is investing heavily in privacy-enhancing technologies (“PETs”), and we enthusiastically support a national strategy to advance the adoption of these technologies.

We have been investing in PETs for several years across our organization, and based on that experience we share the Government’s optimism that PETs—if widely adopted—will enhance data’s benefits while preserving people’s privacy. Meta’s Data for Good program, for example, uses differential privacy to assist global responses to COVID-19 by providing public health experts and researchers with Movement Range Maps.¹ These maps visualize aggregated, county-level changes in population mobility rates, which, for example, allow our partners to assess how lockdowns impact income inequality and to steer public health messaging. Differential privacy has also been critical to the work of Meta’s Facebook Open Research & Transparency (“FORT”) team, which supports independent research into social media’s impact on society.²

PETs also hold potential to advance civil rights and increase equity for marginalized groups, as the Request points out. At Meta, we are exploring secure multiparty

¹ See A. Hergađdelen et al., *Protecting Privacy in Facebook Mobility Data During the COVID-19 Response*, Meta Research (June 2, 2020), <https://research.facebook.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/>. For more information on PETs in similar initiatives, see E. Finkle, *How Facebook Protects Privacy in Our COVID-19 Initiatives*, Meta Newsroom (Apr. 29, 2021), <https://about.fb.com/news/2021/04/how-facebook-protects-privacy-in-our-covid-19-initiatives>.

² See FORT, <https://fort.fb.com/>. As one example, the URL Shares dataset includes differentially private counts of how many people viewed, clicked, liked, commented, shared, or reacted to URLs on Facebook. G. King & N. Persily, *Unprecedented Facebook URLs Dataset Now Available for Academic Research Through Social Science One*, Social Science One (Feb. 13, 2020), <https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one>; see also Meta, FORT Researcher Datasets, <https://fort.fb.com/researcher-datasets>.

computation (“SMPC”) as a tool to help measure whether people’s experiences with our products differ across racial and ethnic groups—without revealing individuals’ sensitive information to Meta.³

Meta’s investment in PETs also highlights their promise for driving the next generation of personalized ad experiences. Businesses—especially small ones—depend on personalized advertising to reach relevant audiences on limited budgets. Personalization does not have to be at odds with privacy, and PETs will help reinforce why. For instance, Meta is exploring SMPC to help advertisers understand how ads perform without disclosing to advertisers or Meta who sees an ad and who makes a purchase.⁴ We have open sourced this framework so that anyone in the industry can develop similar products,⁵ and we are working with industry partners like Mozilla to advance the use of PETs in personalization.⁶

Despite these promising applications of PETs, industry adoption has been limited. We agree with the Request that a host of challenges has contributed to this outcome. There are technical limitations and tradeoffs to many PETs, which must be both accounted for in policy and addressed with greater funding for research and development. There is also a lack of awareness around PETs and their benefits. Finally, entities using PETs face great regulatory uncertainty around how privacy concepts like deidentification are scoped, and what that means for how PETs and other protections can help meet obligations.

We elaborate on these and other challenges—and their potential solutions—in our responses to the Request’s topics below. The fulcrum of our responses is that PETs must be an essential part of *all* conversations around data and privacy. The work of the Fast Track Action Committee will be invaluable for achieving this, and we hope that our responses will serve as a springboard for future government collaboration with Meta.

1. Specific research opportunities to advance PETs

As our response to Topic 2 details, PETs are rapidly evolving, and some PETs are relatively new and less mature. Although newer PETs hold great promise, they face technical

³ See R. L. Austin, *Race Data Measurement and Meta’s Commitment to Fair and Inclusive Products*, Meta Newsroom (Nov. 18, 2021), <https://about.fb.com/news/2021/11/inclusive-products-through-race-data-measurement/>.

⁴ See Meta, *What Are Privacy-Enhancing Technologies and How Will They Apply to Ads?*, Meta Newsroom (Aug. 11, 2021), <https://about.fb.com/news/2021/08/privacy-enhancing-technologies-and-ads/> (also noting how on-device learning and differential privacy could enhance ads privacy).

⁵ See G. Mudd, *Privacy-Enhancing Technologies and Building for the Future*, Meta Business News (Aug. 11, 2021), <https://www.facebook.com/business/news/building-for-the-future>.

⁶ See M. Thomson, *Privacy Preserving Attribution for Advertising*, The Mozilla Blog (Feb. 8, 2022), <https://blog.mozilla.org/en/mozilla/privacy-preserving-attribution-for-advertising/>.

limitations and challenges. Additional federal funding for technical research programs to address these limitations would be immensely beneficial. Research could also help illuminate ways of strengthening privacy in artificial intelligence, such as assessing the vulnerability of models to data extraction attacks, which Meta is actively exploring.⁷ This Administration has already recognized the importance of research, and we applaud efforts like its collaboration with the United Kingdom on prize challenges to advance PETs tackling financial crime.⁸ We encourage further research opportunities and funding.

Regardless of the specific contours of research opportunities, they should strive to promote collaboration between a variety of stakeholders to the greatest extent possible. PETs will be advanced most effectively and responsibly if the Government, industry, academia, and civil society partner to make progress toward shared goals.

2. Specific technical aspects or limitations of PETs

Some PETs, such as at-rest encryption techniques, have existed in some form for several decades. Others are much newer, and this relative lack of maturity carries important implications for what PETs can and cannot achieve.⁹ There might be limitations on the particular computational problems to which a less mature PET could be applied. Partial- and somewhat-homomorphic encryption, for instance, are limited to additive and/or multiplicative functions. Less mature PETs might also be more computationally intensive, potentially limiting their ability to be used at scale. Fully-homomorphic encryption and SMPC, for example, both demand significant computing resources.

Another, related limitation is the costs that applying PETs often incurs. As referenced earlier, some PETs demand particularly high computing resources. Applying PETs might

⁷ Data extraction attacks are attempts by adversaries to discern whether a datapoint was in an AI model's training data. Meta recently simulated adversarial attacks to assess the privacy of its SEER computer vision model. See P. Goyal et al., *SEER 10B: Better, Fairer, Computer Vision Through Self-Supervised Learning on Diverse Datasets*, Meta AI Research (Feb. 28, 2022), <https://ai.facebook.com/blog/seer-10b-better-fairer-computer-vision-through-self-supervised-learning-training-on-diverse-datasets/>.

⁸ See The White House, *U.S. and U.K. Governments Collaborate on Prize Challenges To Accelerate Development and Adoption of Privacy-Enhancing Technologies*, Briefing Room (June 13, 2022), <https://www.whitehouse.gov/ostp/news-updates/2022/06/13/u-s-and-uk-governments-collaborate-on-prize-challenges-to-accelerate-development-and-adoption-of-privacy-enhancing-technologies/>.

⁹ For two explanations of some technical limitations of PETs, see The Royal Society, *Protecting Privacy in Practice: The Current Use, Development, and Limits of Privacy Enhancing Technologies in Data Analysis* (2019), available at <https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf>; Centre for Data Ethics and Innovation, *PETs Adoption Guide* (2021), <https://cdeiuk.github.io/pets-adoption-guide/>.

also require reconfiguring how data is stored, itself a very costly endeavor. And there are the human resources required to implement technically complex PETs. These costs vary by situation, but they carry implications for what is practicable. As a result, some applications of PETs might be possible for only the largest, most well-resourced companies. Even the Government had to expend significant effort to apply differential privacy to the 2020 Census data, changing system designs multiple times.¹⁰

In addition to these technical limitations, there are also technical tradeoffs. First, there is a fundamental tradeoff with utility. Detailed data is valuable and has enabled wide swaths of data-driven ecosystems that benefit people and society. But, in some instances, applying PETs can reduce the informational value of data and, as a result, limit these benefits. Navigating this tradeoff requires PETs adopters “to have a clear idea of what information or value they are trying to protect, and they need to determine the potential benefits and costs of different PETs.”¹¹ In some cases, further research and development could make this tradeoff less severe; technical advances could permit marginal increases in privacy resulting from the use of a PET to carry lower marginal costs to utility.

Another technical tradeoff—one the Request highlights—sometimes exists between applying PETs and achieving equity or fairness. For one, applying PETs can make analyses of data less accurate for smaller, historically marginalized communities represented in it.¹² Applying PETs may also be in tension with developing fairness-aware machine learning, which may require having access to sensitive information about individuals to measure and correct bias.¹³ That said, there may be instances—exemplified by our research on SMPC¹⁴—in which PETs can actually enable privacy-protective insights into equity.

These technical limitations and tradeoffs will have to be carefully considered by all stakeholders in conversations around privacy, from technical conversations about the

¹⁰ See M. Hawes, *Implementing Differential Privacy: Seven Lessons from the 2020 United States Census*, 2 Harv. Data Sci. Rev. (2020), <https://hdr.mitpress.mit.edu/pub/dgg03vo6/release/4>.

¹¹ The Royal Society, *supra* note 9, at 29.

¹² This tradeoff was recently discussed in the context of applying differential privacy to U.S. Census data. See H. L. Wang, *For the U.S. Census, Keeping Your Data Anonymous and Useful Is a Tricky Balance*, NPR (Aug. 2, 2021), <https://www.npr.org/2021/05/19/993247101/for-the-u-s-census-keeping-your-data-anonymous-and-useful-is-a-tricky-balance>.

¹³ See, e.g., Federal Reserve Bank of San Francisco, *Privacy Enhancing Technologies: Categories, Use Cases, and Considerations* (2021) 16, available at <https://www.frbsf.org/economic-research/wp-content/uploads/sites/4/Privacy-Enhancing-Technologies-Categories-Use-Cases-and-Considerations.pdf> (“[T]esting for representativeness in a dataset may be important. . . , but that may not be possible if only a portion of the data can be viewed.”); R. Cummings et al., *On the Compatibility of Privacy and Fairness*, Proceedings of the 27th Conference on User Modeling, Adaptation and Personalization 309 (2019), <https://dl.acm.org/doi/abs/10.1145/3314183.3323847>.

¹⁴ See Austin, *supra* note 3.

direction of research to policy conversations about properly protecting data. In doing so, stakeholders should appreciate that the exact contours of limitations and tradeoffs will be highly dependent on the specific use cases and datasets to which PETs are applied.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

PETs have the potential to be transformative across sectors and applications. Two stand out as particularly amenable to further innovation: data sharing with researchers, and advertising. PETs have great potential—as the Request notes—to enable sharing data with researchers to drive insights and innovation in a host of fields. We discussed previously how Meta’s Data for Good and FORT teams leverage PETs to this end, and we encourage industry, academia, and federal agencies to do the same.¹⁵ At the same time, parties should be mindful that any guidelines or frameworks used for sharing data will need to be flexible, accounting for limitations that PETs might impose in certain contexts on the representativeness of data and the generalizability of research findings.¹⁶

PETs will also be critical for building the next generation of personalized experiences and advertising. Consumers and businesses benefit from advertising relevant to their interests and needs, and industry and policymakers are increasingly working to develop even greater privacy protections. As discussed, Meta is exploring how a wide range of PETs—including SMPC, on-device learning, and differential privacy—can advance this goal.¹⁷ These opportunities should not be overlooked, and we are eager to collaborate with policymakers, industry peers, and other stakeholders to realize PETs’ full potential.

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs

Regulatory uncertainty is one of the most pressing challenges facing the development and adoption of PETs. The White House has cited research from the intergovernmental Financial Action Task Force finding “that uncertainty about the regulatory implications of

¹⁵ See *supra* notes 1-2.

¹⁶ Cf. European Digital Media Observatory, Report of the European Digital Media Observatory’s Working Group on Platform-to-Researcher Data Access 11 (2022), *available at* <https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf> (highlighting challenges that differential privacy might pose to research outcomes and replicability in some instances).

¹⁷ See *supra* notes 5-6.

using these technologies is a significant barrier to adoption.”¹⁸ And the Federal Reserve Bank of San Francisco has emphasized that PETs are often extremely costly and thus must be properly incentivized.¹⁹ Regulatory uncertainty does the opposite.

The regulatory landscape in the United States is rapidly evolving at both the state and federal levels, and the Federal Trade Commission has stated its intent to commence a privacy-focused rulemaking.²⁰ As this landscape takes shape, we encourage regulators to consider the kinds of ambiguity and uncertainty that exist in current data protection schemes. Novel approaches should avoid these pitfalls, and incentivize innovation while protecting privacy. Without clear guidance for what legal obligations exist and how PETs can help comply with them, investing in PETs is a risky enterprise. We outline some proposals below, recognizing that implementation will vary depending on the legal regime being introduced or modified.

One fundamental uncertainty is the lack of agreement on key concepts relating to identifiability. Many data protection schemes impose requirements on “personal data,” carving out data that have been “deidentified” or “anonymized.” But schemes vary widely in how they define “personal data,” and they use “deidentification” and “anonymization” in different, often overlapping or contradictory ways.

This discord stems from a deeper framework misalignment in many policy schemes - treating the risk of identifiability as a binary instead of a spectrum. Data have either been absolutely anonymized—reducing the risk of identifying an individual to zero—or not, and legal obligations are pinned to this binary. In practice, though, achieving absolute anonymization is often difficult, technically complex, and resource intensive.

But it can be possible to render data *effectively* anonymized—to reduce the risk of identification to a non-zero, but sufficiently negligible level. Data protection schemes should move toward this conception of effective anonymization, which the United Kingdom’s Information Commissioner’s Office has embraced.²¹ In doing so, they should recognize not only that identifiability is a spectrum, but also that it is not a purely technical question. PETs can reduce the risk of identification, but so can other

¹⁸ The White House, *supra* note 8. See also Financial Action Task Force, Stocktake on Data Pooling, Collaborative Analytics and Data Protection 32-33 (2021), available at <https://www.fatf-gafi.org/media/fatf/documents/Stocktake-Datapooling-Collaborative-Analytics.pdf>.

¹⁹ See Federal Reserve Bank of San Francisco, *supra* note 13, at 6.

²⁰ See FTC Trade Regulation Rule on Commercial Surveillance, RIN 3084-AB69 (2022), <https://www.reginfo.gov/public/do/eAgendaViewRule?pubId=202204&RIN=3084-AB69>.

²¹ See Information Commissioner’s Office, Chapter 2: How Do We Ensure Anonymisation Is Effective? (2021), available at <https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf>.

protections, such as access controls, policies, procedures, and legal mechanisms. In some cases, effective anonymization might be achieved by using PETs and these other protections in tandem. In other words, although PETs may not always be a silver bullet, they are an important part of a suite of privacy-protective tools, methods, and practices.

This complexity and nuance around identifiability means that data protection schemes pegged to effective anonymization will have to be flexible. Prescriptive, one-size-fits-all tests for determining effective anonymization will not work. Rather, policies should provide a clear and consistent set of factors to balance when assessing risk, recognizing that assessing risk is highly use case-specific. Policies could help provide adopters certainty by incorporating concepts of reasonableness and industry best practices. In other words, policies should derisk beneficial uses of data by ensuring that good faith, considered efforts to deidentify data with PETs do not face undue legal burdens.

Policies should also consider how to incentivize privacy-enhancing data practices that are beneficial even if they do not achieve effective anonymization. For one, there might be contexts in which—given the nature of the data, the parties involved, and other factors—pseudonymized data are particularly valuable. Some schemes already exempt pseudonymized data from certain obligations.²² There might also be ancillary privacy benefits of PETs other than reducing the risk of identification that are worth promoting. For example, some PETs, such as SMPC, could reduce the risk that a party to a computation can make inferences about individuals—an outcome that is valuable even if it does not relate to identifiability per se. These ancillary positive effects could be incentivized through tools like safe harbors and presumptions of good faith.

Finally, policies should explicitly permit data processing done in service of beneficial innovation. For example, processing should be allowed to deidentify data and to test the risk of identification.²³ Entities should also be permitted to process data for the purpose of measuring or increasing fairness, an end that the Request rightfully identifies as valuable and possibly in tension with strict, monolithic conceptions of privacy.

Exactly how these suggestions could be implemented in regulatory regimes will depend on their details and issuing authorities. But we encourage all agencies in the Federal Government to help bring greater nuance to the topic of deidentification. This could emerge in rules governing nongovernmental entities, those around interagency data

²² See, e.g., Virginia Consumer Data Protection Act § 59.1-577(D).

²³ The Federal Reserve Bank of San Francisco has highlighted the challenge of legally testing deidentification as a source of uncertainty. See Federal Reserve Bank of San Francisco, *supra* note 13, at 5 (“[P]rohibitions against the re-identification of data make the testing of the strength of de-identification techniques difficult.”).

sharing, or—ideally—both. The Government has an opportunity to set examples for other regulators—not just in the United States, but around the world.

5. Specific laws that could be used, modified, or introduced to advance PETs

The principles laid out in our response to Topic 4 apply equally here. Just as the regulatory landscape in the United States is very dynamic, so is the federal statutory landscape. As Congress considers privacy-related bills, we encourage all Government stakeholders to evaluate how federal law can recognize the complexity of deidentification and the tradeoffs involved. A comprehensive federal privacy law that draws on our proposals above has the potential to incentivize innovation and reap the full value of data while protecting privacy and advancing PETs. We also encourage stakeholders to consider the ways in which the use and adoption of PETs may drive new and different thinking around how data is regulated in any proposed privacy legislation.

Outside of substantive privacy laws, Meta also encourages statutes that enable further research and development into PETs. Laws like the draft “Promoting Digital Privacy Technologies Act,”²⁴ for example, could help drive efforts needed to address some of the technical limitations and tradeoffs of PETs discussed in our response to Topic 2.

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

Our responses to the previous topics underscore the importance of consensus building on what are very complex issues without clear answers. The Federal Government can play an invaluable role in building this consensus through two additional mechanisms: developing technical standards and guidance, and fostering regulatory co-design.

In our response to Topic 4, we describe how regulatory uncertainty stems from a lack of agreement on foundational concepts like how to define and assess risk. We also suggest that, if policies shift their focus from absolute anonymization to effective anonymization, assessing effective anonymization should take into account industry best practices. Both of these challenges could benefit from Government leadership. Agencies like the National Institute of Standards and Technology could put forward standards and guidance addressing exactly these foundational questions. These materials could include, for example, a taxonomy that provides consistent vocabulary for talking about PETs and what they accomplish. In doing so, this vocabulary should allow stakeholders to understand how different PETs mitigate different discrete, people-centered privacy risks.

²⁴ See Promoting Digital Privacy Technologies Act, H.R. 847, 117th Cong. (2022).

Standards and guidance could also sketch a risk framework that treats identifiability as a spectrum and recognizes the multifaceted nature of risks and associated tradeoffs.

The Government could also create avenues for stakeholders to be intimately involved in co-designing regulations. Workshops building on the insights gained from this Request—on topics like effective anonymization mentioned above—would be valuable. Opportunities could also encompass tools like policy prototyping and regulatory sandboxing, which are useful in problem spaces—like deidentification—that are technically complex and not amenable to bright-line rules. At Meta, we are harnessing the potential of co-design through our Open Loop program.²⁵ Open Loop is helping policymakers in Europe, Latin America, and Asia craft artificial intelligence regulations, testing prototype policies in real-world conditions. We recently kicked off our largest prototyping program to date, with more than forty AI startups providing qualitative and quantitative feedback on provisions of the European Union’s AI Act.

In these efforts, collaborating with a wide range of stakeholders will be key. Involving industry from different technical sectors will ensure that potential standards or policies are practicable, and that relevant use cases of PETs are not overlooked. The Government should also consult academic and civil society stakeholders representing marginalized and underrepresented groups, who have valuable perspectives on advancing equity. Finally, partnering with other governments—state, local, and foreign—will help prevent regulatory fragmentation. Data sharing and analysis are global activities that demand global consensus to fully incentivize privacy innovations and further their adoption.

9. Existing barriers, not covered above, to PETs adoption

Widespread embrace of PETs is currently hindered by a lack of public trust, understanding, and acceptance of PETs as a critical tool for protecting privacy. Without this trust and acceptance, it is harder for companies, particularly ones without significant resources, to justify the costs of implementing PETs. One challenge to building this trust is the difficulty of explaining PETs’ benefits, in part due to their technical complexity. Many PETs are highly advanced technologies that make complex modifications to data or computations. It is difficult to explain these nuances to non-technical privacy stakeholders using intuitive language. For example, explaining how on-device learning works requires communicating the concept of “aggregated model summaries.” That concept challenges even those with university-level degrees and high digital skills.

²⁵ See Open Loop, <https://openloop.org/>; N. Andrade, *Introducing Open Loop, a Global Program Bridging Tech and Policy Innovation*, Meta AI Blog (Jan. 11, 2021), <https://ai.facebook.com/blog/introducing-open-loop-a-global-program-bridging-tech-and-policy-innovation/>.

Another barrier to explaining PETs' benefits is that—even assuming simpler ways of communicating how they work—it is difficult to connect *what* a PET does in a technical sense to *how* it mitigates discrete privacy risks in real-world scenarios. Research by TTC Labs, a cross-industry effort we founded to create innovative design solutions, supports this. TTC Labs found that explanations of a PET are more beneficial when they focus on the PET's impacts on people's personal experiences than when they detail its inner workings.²⁶ Providing these more people-centered explanations might be easier for PETs like differential privacy, whose benefits are better understood, but harder for PETs like SMPC, that strengthen privacy in more nuanced ways.

Meta encourages further research into best practices for explaining PETs and their benefits, and we would welcome the opportunity to collaborate. The Government could also play a pivotal role in fostering a shared stakeholder understanding of PETs' benefits that would in turn bolster trust in PETs and prompt wider PETs acceptance and adoption. This could be achieved by, among other things, the taxonomy we suggest in Topic 6.

Conclusion

Meta is grateful for the opportunity to comment on this Request. PETs hold great potential for ushering in transformative, beneficial ways of using and sharing data while preserving privacy. But there are significant barriers. Research is required to advance the technical capabilities of PETs. There need to be more concerted efforts to explain the benefits of PETs to all relevant stakeholders. And data protection policies should move away from rigid conceptions of privacy that are incompatible with both the nuances of privacy risks and the complexities of how PETs mitigate them.

We are eager to lend our expertise in developing, applying, and explaining PETs to these efforts, and we welcome further conversations and collaborations. If you have any questions, please contact Brad Weltman.

²⁶ See L. Kaplan, *How Can Companies Help People Understand Privacy-Enhancing Technologies Like On-Device Learning?*, TTC Labs Research (2021), <https://www.ttclabs.net/research/how-to-help-people-understand-privacy-enhancing-technologies>.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

MITRE Corporation

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

**©2022 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release.
Distribution unlimited. Case Number 21-01760-26**



Response of The MITRE Corporation to the OSTP RFI on Advancing Privacy-Enhancing Technologies

July 8, 2022

For additional information about this response, please contact:

Duane Blackburn
Center for Data-Driven Policy
The MITRE Corporation
7596 Colshire Drive
McLean, VA 22102-7539

<<This page is intentionally blank.>>

About MITRE

MITRE is a not-for-profit company that works in the public interest to tackle difficult problems that challenge the safety, stability, security, and well-being of our nation. We operate multiple federally funded research and development centers (FFRDCs); participate in public-private partnerships across national security and civilian agency missions; and maintain an independent technology research program in areas such as artificial intelligence, intuitive data science, quantum information science, health informatics, policy and economic expertise, trustworthy autonomy, cyber threat sharing, and cyber resilience. MITRE's 9,000-plus employees work in the public interest to solve problems for a safer world, with scientific integrity being fundamental to our existence. We are prohibited from lobbying, do not develop or sell products, have no owners or shareholders, and do not compete with industry. Our multidisciplinary teams (including engineers, scientists, data analysts, organizational change specialists, policy professionals, and more) are thus free to dig into problems from all angles, with no political or commercial pressures to influence our decision-making, technical findings, or policy recommendations.

MITRE has extensive privacy experience supporting federal, state, local, and international government agencies. MITRE's demonstrated privacy capabilities include conducting research, development, test and evaluation (RDT&E) activities that help government agencies better manage privacy risk, meet privacy compliance requirements, and strategically address privacy policy and technology challenges. RDT&E activities include investigating and reviewing privacy-enhancing technologies (PETs) and shaping privacy best practices to maximize the value of new and emerging technologies.

Additionally, MITRE recently established the Center for Data Privacy and Protection (CDP2) to better streamline the demands on the institution's privacy capabilities and corporate compliance efforts. The mission of CDP2 is to build privacy considerations into business operations and engagements by implementing privacy policies that reduce risk and foster trust, accountability, and transparency. The establishment of CDP2 further illustrates MITRE's commitment and value to privacy and data and protection.

Introduction and Overarching Recommendations

Protecting sensitive data is more involved than simply removing personal information from datasets. Modern PETs offer the potential to protect sensitive data while also helping government agencies achieve their mission goals. PETs represent one set of a series of tools that can be used to protect data and minimize legal, privacy, and ethical risks. To ensure proper understanding and use of PETs, MITRE recommends the following three overarching activities:

1. Conduct an independent review and analysis of existing PET products and services. Test and evaluate how well PETs perform in different scenarios, identify the technical expertise required to implement and maintain PETs, document the potential risks and rewards, estimate financial cost, and determine which solutions can be adopted and implemented in the near term.

2. Ensure the use of project management and systems engineering best practices. Systems designed via a “solution looking for a problem” approach rarely succeed and are not recommended. The first step instead should be to thoroughly define the problem/use case and then design an appropriate solution, which could include PETs as a component.
3. Conduct pilot use cases and document the benefits, limitations, successes, and areas for improvement.

Responses to Selected Questions Posed in the RFI

1. *Specific research opportunities to advance PETs:* Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.

Advancing PETs into proper application requires late-stage research centered on specific use cases, which can often be overlooked while developing overarching research strategies.

Recommended use cases for this portion of a research strategy include:

Social Security Number (SSN) and personal information anonymization – Customer and personnel information systems frequently contain large collections of sensitive personal information, such as SSNs, bank account numbers, and vaccination records. Data processors frequently use manual processes to anonymize or mask sensitive data. PETs may speed up the process by automating anonymization and allowing the underlying, non-sensitive data to be used for intended purposes.

Zero Trust Architecture (ZTA) – Executive Order 14028 on Improving the Nation’s Cybersecurity instructs federal agencies to adopt ZTA. PETs using homomorphic encryption, multiparty computation, or zero-knowledge proofs may help ZTA technologies to perform better at protecting confidentiality, integrity, and accessibility of data.

Digital assets – Executive Order 14067 on Ensuring Responsible Development of Digital Assets addresses privacy and data security throughout the Order. Securing and protecting data is critical to the stability and trustworthiness of any digital assets ecosystem. PETs promise to play a key role in digital assets privacy and security protections.

Public data – Public data from social media sites such as Twitter and Facebook may serve as an early warning indicator for federal, state, and local first responders. This might include information about fires, floods, and tornados or missing persons alerts. PETs may allow government first responders to use public data in a privacy-preserving manner, if sufficiently consistent with social norms.

Synthetic data – Artificial intelligence synthetic data generators evaluate real-world data and then generate statistically accurate synthetic datasets that mimic real-world data. This allows for accurate data analytics without disclosing personal information and identities. Use cases may include census data, taxpayer filings, healthcare records, and immigration trends.

2. Specific technical aspects or limitations of PETs: Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

PETs are most effectively deployed based on a holistic view of a use case, the environment in which they are embedded, and the nature of the relevant PETs. Absent this kind of broad systems approach, PETs may enable ethically and/or societally problematic use cases, alleviating surface concerns while simultaneously undermining more fundamental privacy norms. From a systems engineering standpoint, PETs are not Band-Aids that can be simply dropped onto system designs to render those that are privacy problematic less so. PETs should be considered one set of tools in a larger toolbox of privacy and risk management tools and strategies, to be applied as appropriate in an integrated fashion throughout the systems engineering life cycle.

PETs provide a structured approach to protecting data. However, there are a wide range of subjective requirements that must be addressed. These include legal authorities to collect and use the data for specific use cases; guarding against known and unknown biases such as age, race, and gender discrimination; and ethical considerations. Objective PETs solutions do not always address subjective risks. They often require human subject matter experts to analyze the risks and develop and implement appropriate protections in conjunction with PETs.

This is particularly true for PETs grounded in cryptography and/or theoretical computer science, which offer certain kinds of mathematical guarantees. How such guarantees relate to actual privacy requirements and objectives is not necessarily straightforward, and the work PETs do (or don't do, as the case may be) must be properly situated within the larger socio-technical system. Mathematical guarantees have little intrinsic value outside of their disciplinary contexts; their value is a function of the real-world requirements they support and the conditions under which they hold.

Promoting trust is a core privacy principle, and transparency is critical to promoting trust. The transparency process should work to inform underserved and marginalized groups that do not have time or resources to read privacy notices, privacy impact assessments, and system of records notices that their personal information is protected using privacy preserving data sharing and analytics technologies.

Finally, most PETs products and services lack benchmarks and metrics. MITRE recommends an independent entity conduct test and evaluation benchmarking and propose standards and metrics that allow government agencies to examine their options.

3. *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:* Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of Executive Order 14058: Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government.

Key Areas

Government

Consumer protection – Each year the Federal Trade Commission and state consumer protection agencies receive millions of identity theft and fraud reports. PETs may help agencies process the records in a privacy-preserving manner that also identifies patterns leading to perpetrators.

Tax records – The IRS reported \$2.3 billion in tax fraud for fiscal year 2020. PETs may help identify instances of taxpayer fraud while also preserving the privacy of law-abiding taxpayers.

Homeland Security, law enforcement, and national security records and data – PETs can help ensure that data has been collected lawfully, is being used and maintained in accordance with regulatory and policy requirements, and is shared in a privacy-respecting manner.

Private Industry

Banking, financial, and payment systems, and tax records – PETs allow industry members to exchange data in a secure and privacy enhanced way, as well as comply with state, national, and international data protection regulations. PETs may also help identify potential financial criminal activities such as money laundering and payments for illicit goods.

Healthcare records and data – Data is critical to quality healthcare, medical research, and artificial intelligence/machine learning research and development. Protecting patient data is also critical and required by statutes and regulations. PETs have the potential to enhance privacy protections beyond the existing statutory and regulatory requirements such as the Health Insurance Portability and Accountability Act (HIPAA). In turn, this would open the door to more advanced medical research and development.

Insurance industry data – Automobiles generate “telematics” data that records information such as mileage, fuel, geolocation, speed, and engine diagnostics. The auto insurance industry could use this data to improve automobile and driver

safety, reduce accidents, and set more accurate premiums. However, privacy regulations limit the use of this data. PETs may offer a solution that resolves those limitations.

Marketing data – Big data and social media companies generate revenues based on advertising and marketing data, matching their users to specific products and services. This involves extensive collection of personally identifiable information (PII) that, when combined, could lead to disclosures of sensitive information. PETs may allow users and social media companies to enhance protections of personal information.

Statistical Organizations

Census and statistical bureau/organization data – PETs such as differential privacy can be used to inject “noise” into datasets in a manner that sufficiently preserves the accuracy and privacy of the underlying data.

Trustworthy artificial intelligence (AI) and machine learning (ML) – Trustworthy AI involves building a series of elements and protections into AI/ML algorithms and models. Elements include accuracy, explainability, privacy, security, and mitigation of differential performance. PETs may provide a pathway to achieving some trustworthy AI goals.

PETs were originally developed to protect individual privacy. However, many of the government and industry use cases noted above involve organizational data that may not specifically contain PII. PETs may be adapted to protect sensitive organizational data.

5. Specific laws that could be used, modified, or introduced to advance PETs: Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.

Safe Harbors assume that a technical privacy solution can resolve all privacy risks and should therefore allow the organization implementing PETs to escape any responsibility or liability pertaining to a privacy breach. The challenge with this approach is that PETs are one part of a multifaceted solution. PETs are an objective approach to privacy risks. But there are subjective approaches and analyses that also need to be conducted to understand the full scope of risks and mitigation strategies. Moreover, the protections afforded by PETs are not themselves absolute, and some residual risk will usually remain.

As stated in the response to Question 2, PETs provide an objective approach to protecting data. However, there is a wide range of subjective requirements that must be addressed. These include legal authorities to collect and use the data for specific use cases; guarding against known and unknown biases such as age, race, and gender discrimination; and ethical considerations. Objective PETs solutions rarely address subjective risks. They often require human subject

matter experts to analyze the risks and develop and implement appropriate protections in conjunction with PETs.

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs: This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

There can be considerable confusion both on the part of potential PETs adopters and on the part of PETs developers that hinders effective design and use.

Potential PETs adopters often struggle to understand the relevant technical and operational characteristics of particular PETs, while PETs developers are often unclear about the characteristics and exigencies of real-world use cases. One way of addressing the first issue is development of standard design patterns for distinct types of PETs, especially those that are cryptographically based, while an approach to the second issue is the development of structured use case specifications.

Design patterns are structured solution templates for addressing recurring problems and have a long history in software development. They are highly adaptable, including with respect to the amount of technical detail. Appropriately configured design patterns that, among other facets, convey trust relationships and processing states could help potential PETs adopters better understand the key operational characteristics of different types of PETs. This would enable more accurate assessments of their applicability to specific use cases. While design patterns could facilitate better understanding of PETs functionality on the part of potential adopters, structured use case specifications could facilitate better understanding on the part of PETs developers of the types of problems for which solutions are sought. Such documentation would also benefit potential PETs adopters, as it would force them to articulate problems with sufficient granularity to enable meaningful analysis of the applicability of different types of PETs.

7. Risks related to PETs adoption: Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

For reasons articulated in the response to Questions 2 and 5, sufficiently expansive risk analysis becomes more, rather than less, necessary for appropriate PETs deployment.

This applies to both the analytical methods and the risk models employed. This should leverage methodologies, where appropriate, beyond the typical privacy impact assessment, such as System Theoretic Process Analysis for Privacy.^{1,2} More specialized forms of assessment should also be considered where appropriate. For example, MITRE has developed a Supplemental Technology Assessment methodology for a federal agency that is specifically intended to assess the privacy implications of using prosaic technologies in unusual ways or under atypical circumstances. These methodologies, in turn, must entail the use of sufficiently rich risk models that go beyond the standard ones revolving around Fair Information Practice Principles, such as Solove's taxonomy of privacy problems,³ as well as synthetic consequences.⁴

Privacy risk models, though, need to expand beyond consequences (which most of them focus on) to model threats and vulnerabilities as well. Contextual integrity⁵ is one way of conceptualizing privacy vulnerabilities. MITRE is currently developing a Privacy Attack Taxonomy that will provide a standard structure for mapping privacy attacks that can be used to model privacy threats.

Privacy threats are currently not well understood, and privacy threat modeling is not actively included in risk management processes within many organizations. It is important to effectively assess privacy threats and use privacy threat information as input for PETs selection. Otherwise, organizations may not select the privacy-enhancing technologies that are appropriate for their environment. MITRE's Privacy Attack Taxonomy will provide a standard structure for mapping privacy attacks that can be used to model privacy threats and facilitate privacy risk management, including PETs selection. Security risk modeling typically focuses on confidentiality-based threats to information about individuals (e.g., data breaches). However, the Privacy Attack Taxonomy will enable identification of threats beyond those typically addressed in security risk modeling (e.g., threats related to consent, notice, and inappropriate use, sharing, or retention of information about individuals). This expansion in focus will enable consideration of a broader set of PETs for potential implementation.

Risk-appropriate PETs deployments may be undermined by poor implementation. This is particularly the case with PETs based on cryptography. As discussed in the response to Question 6, standardized PET descriptions, such as PETs-specific design patterns, can help guard against this, as well as against the application of particular PETs to use cases for which they are ill suited. Poor implementation of the right solution or selecting a misaligned solution in the first place may lead to greater problems.

De-identification is an aspect of PETs in which organizations are often challenged with selecting the appropriate methodology and properly implementing it. De-identification reduces the ability

¹ S. Shapiro. Privacy Risk Analysis Based on System Control Structures: Adapting System-Theoretic Process Analysis for Privacy Engineering. 2016. IEEE, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7527748>. Last accessed July 1, 2022.

² R.J. Cronk. Strategic Privacy by Design, 2nd edition. 2022. International Association of Privacy Professionals, <https://iapp.org/resources/article/strategic-privacy-by-design/>. Last accessed July 1, 2022.

³ D. Solove. Understanding Privacy. 2010. Harvard University Press, <https://www.hup.harvard.edu/catalog.php?isbn=9780674035072>. Last accessed July 1, 2022.

⁴ S. Shapiro. Deriving and Using Synthetic Consequences for Privacy Risk Modeling. In ICT Systems Security and Privacy Protection. 2022. Springer, <https://link.springer.com/book/10.1007/978-3-031-06975-8>. Last accessed July 1, 2022.

⁵ H. Nissenbaum. Privacy in Context: Technology, Policy, and the Integrity of Social Life. 2009. Stanford University Press, <https://www.sup.org/books/title/?id=8862>. Last accessed July 1, 2022.

to associate information with an identifiable individual, thereby supporting data privacy and security. However, de-identification is typically used in an all-or-nothing fashion, acting as the sole privacy risk control for a dataset. In principle, though, de-identification should be usable as one of a set of privacy risk controls. For example, MITRE is developing a Data De-Identification Process Architecture that appropriately guides the application of de-identification as a privacy risk control by aligning the extent of de-identification with utility requirements (i.e., intended or projected uses) via quantitative models, assessing residual privacy risk, and indicating additional controls to mitigate the residual risk. ***This approach is distinct from typical approaches that either prioritize addressing risk to enable dataset release or default to maximal security protection of minimally de-identified data.***

Privacy, ethics, and civil liberties risks are normally addressed on a use case-by-use case basis. Existing privacy regulations focus on protecting personal information. However, data analytics generates different risks. How will the use of PETs be integrated with existing, subjective privacy, ethics, and civil liberties reviews involving religion, ethnicity, gender, age, and disabilities? As noted earlier, legal, privacy, civil liberties, and ethics subject matter experts will still be needed to identify and mitigate these risks.

8. Existing best practices that are helpful for PETs adoption: Information about U.S. policies that are currently helping facilitate adoption as well as best practices that facilitate responsible adoption. This includes existing policies that support adoption, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This also includes information about where and when PETs can be situated within tiered access frameworks for accessing restricted data, ranging from publicly accessible to fully restricted data.

There are several existing privacy best practices and frameworks that can be used to facilitate PETs adoption. These include:

1. **NIST SP 800-53, Rev. 5, Security and Privacy Controls for Information Systems and Organizations**⁶ – Provides a catalog of security and privacy controls for information systems and organizations to protect organizational operations and assets, individuals, and other organizations from a diverse set of threats and risks, including hostile attacks, human errors, foreign intelligence entities, and privacy risks. The controls are flexible and customizable, and are implemented as part of an organization-wide process to manage risk. Consideration should be given in the next revision to including additional PETs-related controls beyond the current ones.
2. **NIST Privacy Framework**⁷ – Voluntary tool intended to help organizations identify and manage privacy risk to build innovative products and services while protecting individuals' privacy.

⁶ Security and Privacy Controls for Information Systems and Organizations. 2020. National Institute of Standards and Technology, <https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final>. Last accessed July 6, 2022.

⁷ Privacy Framework. 2022. National Institute of Standards and Technology, <https://www.nist.gov/privacy-framework>. Last accessed July 6, 2022.

3. **Fair Information Practice Principles**⁸ – Widely accepted as a general framework for privacy requirements that is reflected in numerous privacy statutes and regulations in the U.S. and internationally. The principles serve as the basis for analyzing privacy risks and determining appropriate mitigation strategies. In particular, the principle of Data Minimization calls for organizations to collect only personal information directly relevant and necessary to accomplish the specified purpose and retain data only for as long as is necessary.
4. **Privacy by design** – Originated from PETs development and implementation, incorporating privacy principles into system and business process development and operation.
5. **Privacy engineering** – Supports the operationalization of privacy by design by applying systems engineering principles and approaches to the development of socio-technical systems. MITRE’s Privacy Engineering Framework provides high-level guidance regarding fundamental privacy engineering activities, including how to map them to different types of life cycles (e.g., agile).
6. **MITRE Privacy Maturity Model**⁹ – Framework for developing, implementing, maintaining, and evaluating privacy programs within organizations.
7. **MITRE Supplemental Technology Assessment** – Enhanced method of identifying privacy risks and mitigation strategies to minimize risks and maximize rewards in specific contexts, beyond what traditional privacy impact assessments normally identify.
8. **MITRE ATT&CK Framework**¹⁰ – Curated knowledge base that tracks cyber adversary tactics and techniques, many of which frequently impact the confidentiality of PII and sensitive data.

MITRE has extensive experience with these and other privacy best practices and frameworks. This working knowledge has been applied to support federal, state, local, and international government agencies’ adoption and implementation of privacy and security policies and procedures. MITRE has also supported the testing, evaluation, and implementation of PETs at government agencies with privacy best practices incorporated.

9. *Existing barriers, not covered above, to PETs adoption:* Information about technical, sociotechnical, usability, and socioeconomic barriers that have inhibited wider adoption of PETs, such as a lack of public trust. This includes recommendations on how such barriers could be overcome. Responses that focus on increasing equity for underserved or marginalized groups are especially welcome.

⁸ The Fair Information Practice Principles: Framework for Privacy Policy at the Department of Homeland Security. 2008. Department of Homeland Security, <https://www.dhs.gov/sites/default/files/publications/privacy-policy-guidance-memorandum-2008-01.pdf>.

⁹ Privacy Maturity Model, Version 1. 2019. MITRE, <https://www.mitre.org/sites/default/files/publications/pr-19-3384-privacy-maturity-model.pdf>. (Note that Version 2 has been completed and will soon be published to www.mitre.org/privacy.)

¹⁰ MITRE ATT&CK. 2022. MITRE, <https://attack.mitre.org/>. Last accessed July 6, 2022.

Privacy and security overlap in various areas. However, there are unique aspects to privacy that are not addressed by security, particularly regarding notice, consent, individual participation, and collection and use limitation. Privacy and security are mutually supportive, and privacy and security teams should work closely together to protect information about individuals. Better integration between privacy, cybersecurity, and systems/technology development and acquisition is needed to successfully implement PETs.

Many organizations do not have mechanisms in place whereby these different areas can regularly engage. PETs implementation should include formal mechanisms that allow engagement across different PETs stakeholders in an organization. For example, a PETs advisory board can be used that is composed of representatives from areas such as privacy, security, legal, information technology, and data management so that inputs regarding PETs selection and implementation are provided from all relevant stakeholders.

More education regarding PETs, privacy engineering, and technical aspects of privacy is needed. Privacy professionals have historically been more focused on legal, regulatory, and compliance issues, and do not typically have the technical skills needed to manage privacy risks regarding the use of technology. Individuals working in PETs stakeholder areas besides privacy frequently do not have the right level of knowledge of privacy needed for engagement regarding the use of technology that handles information about individuals.

Organizations considering use of PETs should assess privacy workforce needs and identify privacy-related skillset gaps. NIST is currently leading development of a NIST Privacy Workforce Taxonomy, which will contain task, knowledge, and skill statements that are aligned with the NIST Privacy Framework and the National Initiative for Cybersecurity Education Workforce Framework for Cybersecurity. Considerations regarding PETs selection and implementation should be included in the NIST Privacy Workforce Taxonomy. Privacy training should be enhanced to include PETs selection and implementation considerations, and privacy certifications, such as the Certified Information Privacy Technologist certification available from the International Association of Privacy Professionals, should include knowledge of PETs areas as a requirement.

Conclusion

Privacy-enhancing technologies, in conjunction with other privacy and security risk mitigation methodologies, have the potential to substantially enhance PII and sensitive data protections, reduce privacy and security risks, and allow authorized users access to data in a secure manner. MITRE recommends an independent study and review of existing PETs be conducted. The review should include PETs not grounded in mathematical formalisms. Improved capabilities for detecting PII, tracking data flows, specifying and enforcing policies, and measuring privacy risk posture, to note just a few examples, are as important from a utility standpoint as those that leverage cryptographic protocols.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

MOSTLY AI Inc

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

MOSTLY AI Response
to the
Request for Information on
Advancing Privacy-Enhancing Technologies
from the
SCIENCE AND TECHNOLOGY POLICY OFFICE

Alexandra Ebert

Chief Trust Officer

MOSTLY AI Inc.

Type: Industry

Dr. Michael Platzer

Chief Strategy Officer & Co-Founder

MOSTLY AI Inc.

Type: Industry

Response of MOSTLY AI¹ to OSTP RFI on Advancing Privacy-Enhancing Technologies

8th of July, 2022

Introduction

MOSTLY AI Inc. ("MOSTLY AI") welcomes the opportunity to the OSTP's RFI on advancing privacy-enhancing technologies (PETs). As the global leader in **structured synthetic data**, we are dedicated to enabling an open data ecosystem where access to high-quality, diverse, granular-level data can be democratized while privacy remains protected. Synthetic data allows to distill the insights from existing data in a fully automated manner, and makes these insights accessible by generating statistically representative, highly realistic, yet completely new data samples at scale. With that, our synthetic data platform enables already today public and private sector organizations across the world to safely innovate and collaborate on top of large-scale data assets towards building a smarter and safer future.

And it's the data that serves as the **lingua franca** of the digital era. It is the source for learning and exploration by machines and by humans alike. In particular, it is data at the granular level, that represents subjects and events, that can be easily understood, interpreted, and reasoned upon by people of all backgrounds. For that reason, we see that synthetic data serves a unique need among the group of emerging PETs, as it allows the involvement of much broader communities and stakeholders in the process of building and validating the algorithms that will shape our society going forward.

We would also like to use the opportunity to bring attention to a recently published report by the European Commission on synthetic data², that concludes:

*"Synthetic data can become the unifying bridge between policy support and computational models, by unlocking the potential of data hidden in silos; thus **becoming the key enabler of artificial intelligence**. [...] More important than focusing on how to synthesize data is what can we achieve with the new data available at scale, how to convince data owners to unleash their coveted data to the broadest audience, and how to accommodate this **massive new ability** into the policy formulation and assessment."*

¹ <https://www.mostly.ai/>

² <https://publications.jrc.ec.europa.eu/repository/handle/JRC128595>

Response to the RFI

1. Specific research opportunities to advance PETs

Aside from research on the technical feasibility and validity of emerging PETs, we see a need to research

- 1) their anticipated impact on society and economy – analyzed from a non-technical perspective
- 2) the interplay of the various PETs with each other - PETs are oftentimes analyzed in isolation, but they can very well complement each other
- 3) how to encourage or even mandate open data sharing practices for the benefits of society

2. Specific technical aspects or limitations of PETs

We consider encryption-based mechanism to provide security, but not necessarily zero-trust privacy guarantees. While these can help to provision data safely to machines to perform pre-defined computation, they fall short in enabling broad data sharing with humans.

We consider aggregation- and query-based systems to be capable of satisfying privacy requirements, but they again do not allow the broad sharing of granular-level data with humans.

We consider synthetic data to take a unique position, that comes at the cost of carefully reduced accuracy, but can then provide unrestricted access to granular-level data. To cite the aforementioned report by the European Commission on the subject:

“Correctly performed synthesis introduces controllable and well described distortion of the original data, which is just a small price to pay for the availability of highly granular privacy unburdened data. [...] Among the privacy-preservation technique studies analysed (differential privacy, data perturbation, homomorphic encryption, secure private computing infrastructure), data synthesis gave the best price (effort)/cost ratio.”

A current limitation of synthetic data is still the lack of commonly accepted standards and benchmarks to empirically assess the accuracy and privacy of various approaches. Not all synthetic data approaches are necessarily accurate. Not all approaches are automatically private. The involved machine learning algorithms can suffer from underfitting as well as from overfitting to the original data, if done incorrectly. First research initiatives exist³, but more standardization is needed.

Another limitation of synthetic data exists with respect to scaling the required compute for very large data assets.

³ <https://www.frontiersin.org/articles/10.3389/fdata.2021.679939/full>

3. Specific sectors, applications or types of analysis that would particularly benefit from the adoption of PETs

We consider healthcare⁴ as well as the public sector to have the biggest potential for positive impact for society, if their data assets could be shared more openly in a truly privacy-preserving manner. Beyond that, we consider the financial service and the telecommunication industries to be segments that gather the largest pools of insightful behavioral data, thus posing an enormous economically opportunity, if such data can be safely utilized. But any sector, that deals with personal data at scale (retail, education, recreation, mobile services, etc.), is expected to benefit, if their insights become accessible in a privacy-safe manner.

It's important to emphasize that PETs not only allow safe data access across organizations, or across borders, but also foster data sharing within an organization. Data access is a crucial factor to accelerate innovation and thus to strengthen the competitive advantage of US entities, while safeguarding the privacy of all citizens.

4. Specific regulations or authorities that could be used, modified or introduced to advance PETs

PET adoption across industries can be greatly accelerated by authorities leading by example, and proactively taking them in use for protecting citizen data, while also publicly and transparently communicating about these taken initiatives.

5. Specific laws that could be used, modified or introduced to advance PETs

GDPR and its absolute and strict definition of anonymous data can serve as guidance for introducing a national level privacy law, which would certainly help the adoption of PETs. The more aligned these regulations are, the closer the collaboration of the western world in the digital era will be.

Aside from privacy protection, it is important to recognize the need for data to validate and assess algorithms that impact individual's lives. Ideally, upcoming laws shall consider requirements to put external parties into a position to stress-test algorithms for fairness, by being mandated to share not only model access, but also access to representative (synthetic) data samples at scale.

⁴ See eg Humana's synthetic data exchange <https://developers.humana.com/syntheticdata>

6. Specific mechanisms, not covered above, that could be used, modified or introduced to advance PETs

Competitions and benchmarks are essential to advance PETs. These shall be open for submissions, both from open-source as well as from proprietary solution providers, in order to encourage broad competition and investments. However, any solution shall be openly accessible for stress testing by a broad audience, as e.g., done by MOSTLY AI with their publicly available free version of their synthetic data platform⁵.

7. Risks related to PETs adoption

The unique strength of the US economy builds upon diversity, creativity, and open collaboration. We consider it thus as crucial to understand the non-technical implications of the various PETs in detail. Particularly with respect to whether they allow for broad, diverse communities to directly benefit from accessing data.

We see it as a potential risk, if PETs result in strengthen the already existing data monopolies of few organizations, that can afford to use PETs to exclusively gather even more data. Data is information, data is knowledge, and thus data shall be accessible to as many people as possible.

8. Existing best practices that are helpful for PETs adoption

We recommend a crawl, walk, run approach. We recommend to start small, and gradually and swiftly build up experience and expertise with respect to the already existing PETs, rather than investing in big initiatives with risky return in a far distant future.

⁵ <https://mostly.ai/>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Mozilla

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



July 8, 2022

Office of Science and Technology Policy
Eisenhower Executive Office Building
725 17th Street NW
Washington, D.C.
United States of America

Via Email to: PETS-RFI@nitrd.gov

Subject: “RFI Response: Privacy-Enhancing Technologies”

About Mozilla	2
Mozilla’s Thinking on Privacy Preserving Advertising	2
Opportunities & Challenges for Multi-Party Computation Solutions	3
Need for Complementary Regulatory Solutions	4
Importance of Standards Bodies	5
Mozilla’s Previous Work	6

—

Mozilla is submitting this response, focused on digital advertising, to the Office of Science and Technology Policy’s (OSTP) Request for Information¹ to inform the development of a national strategy on privacy preserving data sharing and analytics, along with associated policy initiatives. Regulators and technology companies together have an opportunity to improve the privacy properties of online advertising—an industry that has not seen privacy improvement in many years. Mozilla has worked on shaping the ongoing evolution of privacy preserving advertising for the past decade, both in our products² as well as in our external

¹ Request for Information on Advancing Privacy-Enhancing Technologies, Science and Technology Policy Office. 6 July, 2022. Available at: <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>

² Firefox rolls out Total Cookie Protection by default to all users worldwide, Mozilla Blog. June 14, 2022. Available at: <https://blog.mozilla.org/products/firefox/firefox-rolls-out-total-cookie-protection-by-default-to-all-users-worldwide/>

engagement with regulators³, industry⁴, and standards bodies.⁵ We welcome this opportunity to briefly share our thinking on this critical issue, with the intent of placing our extensive work so far on the record and paving the way for further conversations.

Specifically, our response covers the following items from the scopes listed in the RFI request “2. Specific technical aspects or limitations of PETs”; “3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs”; and “10. Other information that is relevant to the adoption of PETs”.

About Mozilla

Mozilla is the maker of the open-source Firefox web browser, the Pocket “read-it-later” application and other products and services that collectively are used by hundreds of millions of individuals around the world. Mozilla is also a global community of contributors and developers who work together to keep the internet open and accessible for all. As a mission-driven technology company and a non-profit foundation, we are dedicated to putting people in control of their online experience, and creating an internet that is open and accessible to all. To fulfill this mission, we are constantly investing in the security of our products, the privacy of our users and in advancing the movement to build a healthier internet.

Mozilla’s Thinking on Privacy Preserving Advertising

Mozilla has often outlined⁶ that the current state of the web is not sustainable, particularly in the context of how online advertising works today. It is a hostile place for user privacy, and is effectively an arms race⁷ between browser anti-tracking technologies and trackers. It’s opaque by design, rife with fraud, and does not serve the vast majority of those which depend on it - from publishers, to advertisers, and of course, the people who use the open web. It is also important to note that many critical aspects of internet architecture were not created with privacy in mind. However, the growing importance of privacy has led to experts at standards

³ Despite improvements, delaying the deprecation of third party cookies in the CMA’s Privacy Sandbox commitments will hold back privacy on the open web, Mozilla Blog. December 17, 2021. Available at: <https://blog.mozilla.org/netpolicy/2021/12/17/privacy-sandbox-cma-dec2021/>

⁴ Analysis of Google’s Privacy Budget Proposal, Eric Rescorla. October 1, 2022. Available at: <https://blog.mozilla.org/mozilla/google-privacy-budget-analysis/>

⁵ Interoperable Private Attribution (IPA) at the Private Advertising Technology Community Group (PATCG) of the World Wide Web Consortium. 6 January, 2022. Available at: <https://github.com/patcg/private-measurement/issues/9>

⁶ Mozilla responds to the UK CMA consultation on Google’s commitments on the Chrome Privacy Sandbox, Marshall Erwin. July 9, 2021. Available at: <https://blog.mozilla.org/mozilla/uk-cma-google-commitments-chrome-privacy-sandbox/>

⁷ Narayanan, A., 2018. The web tracking arms race: Past, present, and future. In Enigma 2018 (Enigma 2018). Available at: <https://www.usenix.org/conference/enigma2018/presentation/narayanan>

development organizations (SDOs) committing to plug the holes; this can be seen with the increasing focus on privacy at the World Wide Web Consortium (W3C).

At the same time, we also believe there's nothing inherently wrong with digital advertising. It supports a large section of services provided on the web and we believe it is here to stay. However, the ways in which advertising is conducted today - through pervasive tracking, serial privacy violations, market consolidation and lack of transparency - is not working and causes more harm than good.

At Mozilla, we have been working for years to drive the industry in a better direction, away from pervasive and opaque web tracking. We've done so by limiting⁸ the use of Third-party cookies, developing more privacy preserving ways to measure user interactions online,⁹ and working on advancing privacy preserving advertising.¹⁰ We believe the web can do better and are already doing work on the technical side to eliminate third party tracking, and engaging with the Private Advertising Technology Community Group (PATCG) at the W3C and other industry fora.

Opportunities & Challenges for Multi-Party Computation Solutions

In addition to our extensive work analyzing various PETs proposals put forward by industry, Mozilla has done considerable work exploring, developing, and implementing one class of PETs solutions - those involving *Multi-Party Computation (MPC)*. This includes developing, deploying, and testing Prio, a privacy preserving data collection technique developed at Stanford, within the Firefox browser. More recently, in partnership with Meta, we developed and proposed Interoperable Private Attribution (IPA), an MPC solution that would allow advertisers to measure and understand the effectiveness of their ad campaigns without having to track and collect extensive user level data.

The basic premise of these MPC solutions is that, using various advanced cryptographic techniques, data is shared across multiple servers in a way that avoids allowing any single entity — websites, browser makers, or advertisers — to learn about user behavior. Those servers can then work together to produce aggregate results that cannot be linked to individual users. For example, Mozilla has run a trial using 'Prio' to better understand what websites were being blocked in Firefox by privacy features.¹¹ This would have previously required directly collecting data about users' browsing activity. To know what was being

⁸ Firefox 86 Introduces Total Cookie Protection, Tim Huang, Johann Hofmann and Arthur Edelstein. February 23 2021. Available at: <https://blog.mozilla.org/security/2021/02/23/total-cookie-protection/>

⁹ Next steps in privacy-preserving Telemetry with Prio, Steven Englehardt. June 6, 2019. Available at: <https://blog.mozilla.org/security/2019/06/06/next-steps-in-privacy-preserving-telemetry-with-prio/>

¹⁰ Building a more privacy preserving ads-based ecosystem, Mozilla blog. Available at: <https://blog.mozilla.org/mozilla/building-a-more-privacy-preserving-ads-based-ecosystem/>

¹¹ Kogan, D., & Corrigan-Gibbs, H. (2021). Private blocklist lookups with checklist. In the 30th USENIX Security Symposium (USENIX Security 21) (pp. 875-892). Available at: <https://rwc.iacr.org/2020/slides/Gibbs.pdf> and <https://www.usenix.org/system/files/sec21-kogan.pdf>

blocked, we would have needed to know what websites users were trying to visit. As a privacy respecting company, that is data we do not want to collect. Instead, the Prio trial allowed us to collect aggregate information about websites without ever learning about individuals' browsing activity.

While MPC solutions have specific use-cases in browser measurement and advertising that have motivated Mozilla's work on this topic, the example above demonstrates the potential power of these tools. The internet today is having a profound, sometimes harmful impact on people's lives. But to really understand that impact and do data driven policy making today requires collecting highly sensitive data about people's online activities, even if we aren't interested in any individual's behavior. MPC provides a means to collect this data and understand aggregate activity without compromising people's privacy.

There are, however, a number of hurdles for this technology to ensure that it can both protect privacy and allow a diversity of parties to benefit from better aggregate insights. First, transparency and strong governance are critical to the success of MPC solutions and more work must be done to develop governance models for those that run pieces of these systems. One shortcoming of these approaches is that the different parties in the system might collude to break the privacy properties of the design and to gain access to individual level data. The relationship between those parties must be set up in such a way that they have strong incentives, structural limitations, and bylaws that prevent such collusion.

Second, it is important that a diversity of parties be able to offer MPC services and the federal government should explore ways to create and support that diverse ecosystem of MPC services. These services are complex to design and operate, requiring deep technical expertise and resources that can be found within a small number of major tech companies. Indeed, Apple, Google, and Meta have been on the forefront of developing this technology and have done very promising work on this topic. But a privacy-preserving world in which only the most dominant players have the means to offer MPC solutions would be a mixed blessing. Moreover, dominance of this space by major tech companies, which are not known for their transparency and are already known to abuse their market positions to disfavor competitors, would likely exacerbate the governance challenges mentioned above.

Finally, it is important that these services be available to diverse public and privacy sector parties. Again, a privacy-preserving world in which only the most dominant players are able to take advantage of these services, while researchers, academics, and watchdogs groups that could benefit from these services are left out in the cold, could reinforce the have/have not data divide that has allowed a small number of parties to gain significant insight into public life on the Internet while leaving outside groups starving for such insights.

Need for Complementary Regulatory Solutions

Apart from technical solutions, we believe in the need for better regulation and creating more partnerships with the ecosystem within the diverse range of actors. Without a better

regulatory framework, a sufficient incentive won't exist to move towards more privacy preserving techniques. And even with some of those techniques in place, various types of harm will persist that require regulatory intervention. We do not think that technical work alone will not solve the problem of the dependence on data, and the risks and harms that this causes, which include:

- Disinformation (elections, politics, etc.)
- Discrimination (race, age, gender, etc.)
- Societal Manipulation (vaccines, etc.)
- Privacy Violations (leaks, breaches, etc.)

Our vision is a web which empowers individuals to make informed choices without their privacy and security being compromised. While there is no denying behavioral advertising is the underlying business model of the web today, it does not mean that it cannot be reformed to minimize its societal harms. In order for this ideal to be realized, we think that:

- its impact on society should be understood better (research);
- Users—including publishers, advertisers, and oversight bodies—should be able to understand how or why they are being targeted (transparency);
- targeting should happen with minimal data collection (privacy preserving advertising);
- regulators should penalize violations of privacy and competition laws (enforcement);
- regulatory solutions should be explored which limit targeting and pave the way for better technical solutions (regulation)

In an ideal state, a combination of new research, technical solutions, increased public awareness and effective regulatory enforcement would reform behavioral advertising for the future of the web.

Importance of Standards Bodies

Open SDOs should play an essential role in the development and deployment of privacy preserving technologies for online advertising. Many of the underlying standards that form the bedrock of today's internet have been transparently debated and collaboratively developed at SDOs by relevant stakeholders through formal processes. For example, Transport Layer Security (TLS) is a foundational security protocol that was developed at the Internet Engineering Task Force (IETF). Cascading Style Sheets (CSS) is a cornerstone technology to develop webpages and was developed at the W3C.

These seminal standards were not developed overnight or by a single company; rather, multiple stakeholders, including different browsers and browser engines, worked

collaboratively and transparently over years in open SDOs. This approach, sometimes colloquially referred to as “rough consensus and running code” prioritizes real world deployment and validation and helps ensure critical technologies for the internet also serve the collective interest. This also allows third party vendors (service providers) and browsers to operate across diverse offerings of the same or similar services, and create consistently clear experience expectations for consumers, which is especially vital for online advertising.

Mozilla has always believed that a vibrant and open internet depends on fair conditions, open standards, and opportunities for a diversity of market participants to participate. We have substantial experience contributing to open SDOs over the last two decades, having played a key role in the development and implementation of critical standards such as TLS 1.3 and HTTP/3 as well as industry wide initiatives like Let’s Encrypt. It is through open standards that we believe the internet can remain decentralized, open, and interoperable.

Global standards development is a voluntary consensus-driven process. For this reason, stakeholder commitment to final specifications and deployment on specific timelines is vital for an open web. The commitment and deployment, particularly by large stakeholders, is necessary to materialize theory developed in SDOs into practical applications used widely across products in a particular industry. Markets can be distorted, and consumers impacted, either due to the absence of commitment to the final standard itself *and/or* not respecting the agreed upon timelines to deploy or deprecate relevant technologies. We think that this interplay of process and consensus is vital for privacy preserving technologies around advertising to develop in a sustainable manner rather than being unilaterally imposed upon the ecosystem.

The direct implementation of web features by dominant players often leads to them becoming de facto web standards, often well before the relevant SDO has formally adopted the underlying standard itself. This could put many current industry proposals in this space at risk of becoming de facto industry benchmarks merely via their unilateral implementation by a few players, leaving publishers, and other browsers without choice or say in the matter. Taking the relevant standards to the appropriate formal SDOs, on the other hand, will help ensure that their properties are validated rigorously while also allowing for relevant public visibility and engagement to stakeholders via open processes. It will also allow for other players to play the appropriate role in the development of standards themselves.

Mozilla’s Previous Work

The section below contains links to our previous work and regulatory submissions on the topic, which elucidate the thoughts we’ve mentioned above in greater detail.

Overview

- Building a more privacy preserving ads-based ecosystem¹²

¹² Building a more privacy preserving ads-based ecosystem, Mozilla Blog. May 28 2021. Available at: <https://blog.mozilla.org/mozilla/building-a-more-privacy-preserving-ads-based-ecosystem/>

- The future of ads and privacy¹³

Technical Analysis

- Privacy analysis of FLoC¹⁴
- Privacy analysis of SWAN.community and United ID 2.0¹⁵
- Analysis of Google's Privacy Budget Proposal¹⁶
- Privacy Preserving Attribution for Advertising¹⁷

Regulatory Submissions

- Mozilla responds to the UK CMA consultation on Google's commitments on the Chrome Privacy Sandbox¹⁸
- Despite improvements, delaying the deprecation of third party cookies in the CMA's Privacy Sandbox commitments will hold back privacy on the open web¹⁹
- Competition should not be weaponized to hobble privacy protections on the open web²⁰

Conclusion

We believe that there is a real opportunity now to improve the privacy properties of online advertising by drawing upon the internet's founding principles of transparency, public participation, and innovation to make progress. We hope these brief thoughts are helpful in the OSTP's efforts in understanding the space and looking forward to continued engagement on this issue with relevant agencies in the coming months.

Contact for Additional Information

Jenn Taylor Hodges, Head of US Public Policy, Mozilla Corporation -

Udbhav Tiwari, Senior Manager, Global Public Policy, Mozilla Corporation -

¹³ The future of ads and privacy, Eric Rescorla. May 28, 2021. Available at : <https://blog.mozilla.org/mozilla/the-future-of-ads-and-privacy/>

¹⁴ Privacy analysis of FLoC, Eric Rescorla. June 10, 2021. Available at: <https://blog.mozilla.org/privacy-security/privacy-analysis-of-floc/>

¹⁵ Privacy analysis of SWAN.community and United ID 2.0, Mozilla Blog. August 4, 2021. Available at: <https://blog.mozilla.org/mozilla/swan-uid2-privacy/>

¹⁶ Analysis of Google's Privacy Budget Proposal, Eric Rescorla. October 1, 2021. Available at: <https://blog.mozilla.org/mozilla/google-privacy-budget-analysis/>

¹⁷ Privacy Preserving Attribution for Advertising, Martin Thomson. February 8, 2022. Available at: <https://blog.mozilla.org/mozilla/privacy-preserving-attribution-for-advertising/>

¹⁸ Mozilla responds to the UK CMA consultation on Google's commitments on the Chrome Privacy Sandbox, Marshall Erwin. July 9, 2021. Available at: <https://blog.mozilla.org/mozilla/uk-cma-google-commitments-chrome-privacy-sandbox/>

¹⁹ Despite improvements, delaying the deprecation of third party cookies in the CMA's Privacy Sandbox commitments will hold back privacy on the open web, Mozilla Blog. December 17 2021. Available at: <https://blog.mozilla.org/netpolicy/2021/12/17/privacy-sandbox-cma-dec2021/>

²⁰ Competition should not be weaponized to hobble privacy protections on the open web, Udbhav Tiwari. April 12, 2022. Available at: <https://blog.mozilla.org/netpolicy/2022/04/12/competition-should-not-be-weaponized-to-hobble-privacy-protections-on-the-open-web/>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

NowVertical Group Inc

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

NowVertical Group Inc. (NOW) is a big data technology company that helps businesses, government institutions, and large enterprises conduct vertically intelligent (VI) transformations through industry-specific software and services.

NOW possesses deep expertise in developing and bringing to commercial and government markets a flexible platform-agnostic software solution that transforms data to open standards and automates data processes.

Respondent Organization	NowVertical Group Inc.
Respondent Points of Contact	David Whitmire President, NOW Solutions <hr/> Farid Kassam President, NOW Origin <hr/>
Respondent Type	Industry

Our RFI Response addresses: *Specific technical aspects or limitations of PETs*

There are currently two primary strategies to conduct data collaboration between data owners, namely:

- Peer-to-Peer sharing
- Data Clean Rooms

Both these strategies have been around for years and have their advantages and disadvantages. However, with emerging privacy legislation and focus on personal data rights, organizations need to find new, scalable strategies to enable turn-key, privacy-safe data collaboration strategies that will satisfy the needs for sharing data without violating the privacy of individuals. Privacy-preserving computation (PPC) refers to strategies that enable data privacy to be protected while still being able to model and generate insights. **As part of this RFI, we will consider the existing strategies of data collaboration and PPC models and propose an architecture that we believe is scalable and efficient that will satisfy the needs for future privacy-safe data collaboration requirements.**

Peer-to-Peer Data Collaboration

Traditionally, peer-to-peer data sharing was the only way to perform analytics between multiple parties (and continues to be used today in second party data exchanges). It requires copies of raw data to be sent directly between collaborators with privacy restrictions managed by both parties independently and through legal agreements between parties. Typically, the goal is to “join” the shared data to an existing party’s data, and thus often requires the data to be row-level and in raw format (no encryption, etc.). Although encryption can be used to secure data in transit, encryption keys are often shared between parties to revert back to raw state or a common encryption method is used between both parties so the data can be joined (however, this still

enables either party to translate the encrypted data back to an original data point - and thus potentially an individual).

The benefits of this model are:

- Full control over what data is share to which partner

The primary down-falls to this model are:

- Engagements are delayed by lengthy legal/privacy reviews
- Raw data is used (or encrypted data can be reversed back to an individual)
- Lacks scalability (to add another party increases the privacy risks per project)
- Requires a heavy lift on behalf of the data teams to manage, join, and analyze the data
- Copies of the data can easily be made, shared, and exposed increasing liability risks

Data Clean Rooms (First Generation)

For the purposes of scale, commercial organizations began offering “data clean rooms” which were primarily created in order for partners to share and collaborate using their first party customer data. Since this application was the driving force behind data clean rooms, most of the early partners offering the solution had an underpinning identity graph that was used to 1) translate customer Personal Identifiable Information (PII) to a single common ID, 2) provide insights around the percentage of segment overlap between collaboration parties, 3) identify an ID set that in the overlap for media activation. This model required all participating and collaborating parties to share copies of their raw PII data with the data clean room providers, creating a hub-and-spoke model where the agency offering the service is the “hub” and all participating parties represent the spokes.

The early iterations of this model have challenges from a privacy perspective since raw PII and sensitive data is being shared into the centralized hub agency (and is thus exposed to the raw PII of all participating spoke parties). In addition, it creates a single point of failure since the entire model is controlled by the centralized hub organization, creating yet another walled garden of sorts.

The benefits of this model are:

- Scale enabling multi-party collaboration across different partners leveraging a common ID
- Privacy from “spoke” partners - since raw data was never shared to all other partners, the privacy risk does not increase with the number of data collaboration projects

The challenges of this model are:

- One agency owning access to all private data (single point of failure)
- Private data exposed to the centralized “hub” agency
- Control - insights and flexibility of data collaboration restricted by the centralized agency

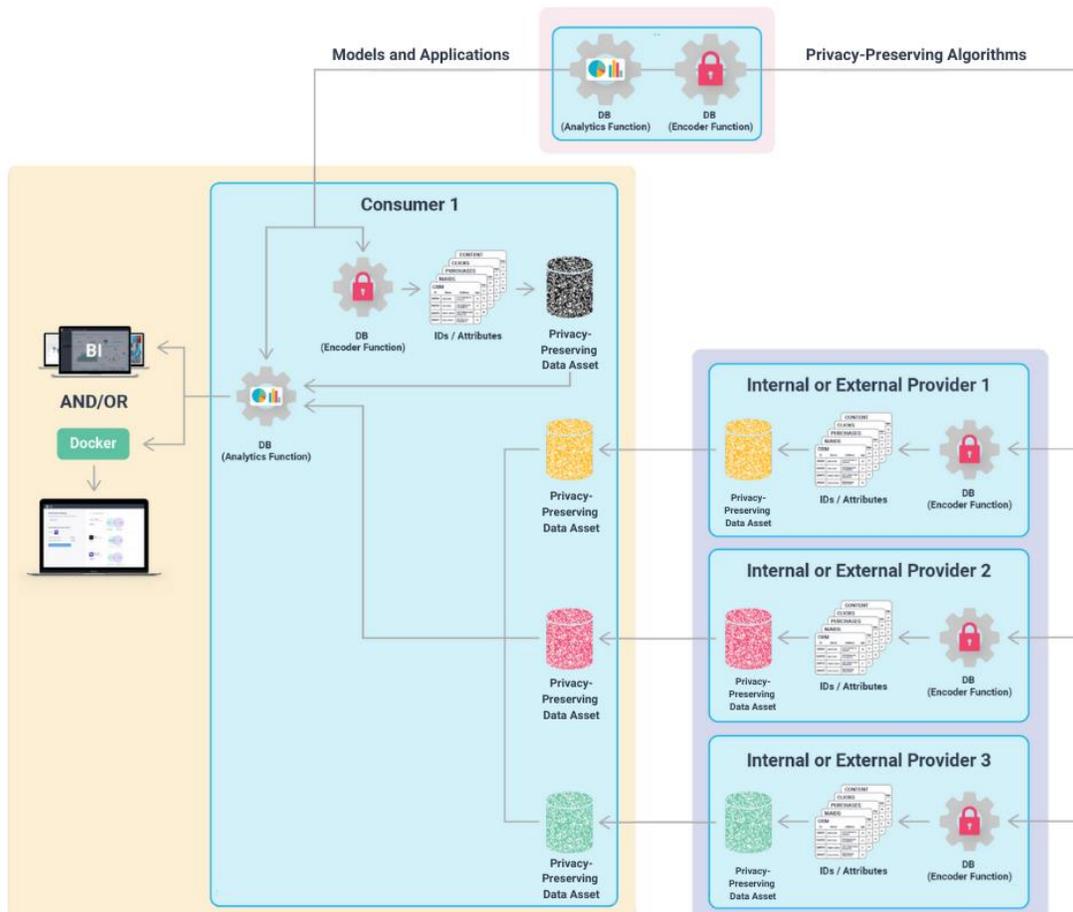
Data Clean Rooms (2nd Generation)

Over the last few years, a new generation of data clean rooms has emerged that leverages privacy-protecting strategies at the heart of the model. With this new model, private (PII) data of the “spoke” partners is shared to a “bunker” owned by the central agency where privacy-protecting algorithms are applied to the data to create a new data asset that can be used safely for analytics between “spoke” partners. This strategy reduces the privacy risk of the “at rest” data inside the bunkers since the original raw/private data is flushed after the privacy algorithm has been applied to the data set. This strategy has been proven to reduce legal and privacy due diligence, but still suffers from many of the challenges highlighted in the first-generation model.

Recommended Architecture for Privacy-Safe Multiparty Data Collaboration

The future of privacy-safe, multiparty data collaboration needs to leverage the benefits of the aforementioned models but change architecturally. This model needs:

- **A Distributed Model:** no one party “owns” the data, nor is there a centralized agency that manages all collaborations – each party owns their own data and has full control over what they share and with what party (similar to the peer-to-peer model)
- **Compatible Privacy-Preserving Data Assets:** no raw, private data should ever be shared. The model requires a unified privacy-preserving algorithm and strategy so that all participants use the same algorithm so that all privacy-safe data assets are compatible. This algorithm should be non-reversible and use privacy methods like noise injection to prevent privacy risks of “drilling down” to a single ID (and thus potentially exposing PII).
- **Secure Data Sharing:** copies of private data should not be shared; the architecture should enable secure data sharing without moving data. This simplifies data sharing and access revocation.
- **Modeling and Application Infrastructure:** beyond sharing the data, this model requires the ability for each participating party to access productized models and applications that can be applied to the privacy-safe data assets for insights and actionability.



A key to this architecture is where the privacy-preserving algorithms are being run. There are a host of well-known algorithms, each that often serve unique purposes, but in order for this future state to scale, there must be an agreement across all parties relating to which algorithm(s) is used.

We recommend that this algorithm(s) be set as an open standard to ensure compatibility across all collaborating parties.

This standardized algorithm(s) is applied to the data on the private servers of each participating party, creating a new privacy-safe data asset. All data sets are indexed/encoded/encrypted in the same way creating a foundation of compatible privacy-safe data assets. Regardless of which cloud or servers this new data asset has been created on, it next needs to be copied to a centralized platform or data warehouse for data sharing.

Once the data asset is on a centralized platform or data warehouse, it can be shared to other participating parties with no concerns about data privacy (since the new asset is non-reversible and anonymized). The data owner can choose who they share the data with and have full control to revoke that access if necessary.

With data sharing there is now a network of data assets accessible to all participants. The final feature of this future state is a common set of models and applications that can unlock aggregate insights and intel from the combined privacy-safe data assets.

We recommend that these models and applications can be user-contributed, but vetted (for security), authorized and managed as a library available to all participating parties.

These models and applications provide a means of creating pre-canned insights for specific needs (i.e., finding the percentage of overlap between two data sets). These models and applications also provide turn-key productized solutions so that data scientists are not needed to extract insights, and it can be done quickly and easily by citizen data scientists.

In summary, in order to create this distributed privacy-safe data collaboration model each participant must:

1. Create a privacy-safe data asset leveraging the agreed-upon algorithm.
2. Place a copy of this data asset on servers or a data warehouse that enables data sharing (without moving data).
3. Share this data asset with participating parties of its choosing (and likewise receive these data assets from others)
4. Leverage models and apps to compare, analyze and model multiple data assets in a secure environment.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

NTT Research

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

July 8, 2022

Ms. Stacy Murphy
Operations Manager
White House OSTP

Dear Ms. Stacy Murphy:

I appreciate the opportunity to respond to OSPT's RFI on Advanced PET.

NTT Research is NTT R&D's global presence outside of Japan with headquarter based in Bay Area of US. We expand a rich lineage of ground-breaking basic research heritage from Japan to the rest of the world. One of our key research topics in cryptography and information security area is Attribute Based Encryption (ABE) / Functional Encryption (FE). We would like to highlight the unique roles ABE/FE can play in many of the use cases for enhancing data privacy and security. We believe ABE/FE's characteristics are complementary to other PETs. Lastly, we also would like to provide you with our observations and feedback about the challenges and barriers in adopting PET.

If there is anything we can clarify for you further, please feel free to reach out directly to me or my staff member, Fang Wu

Sincerely,

Kazuhiro Gomi
President and CEO
NTT Research, Inc.

This page intentionally left blank



NTT Research, Inc.

940 Stewart Dr, Sunnyvale, CA 94085, USA

www.ntt-research.com

RESPONSE TO OSTP RFI ON ADVANCING PRIVACY-ENHANCING TECHNOLOGIES

July 8, 2022

Table of Contents

Executive Summary	1
1 Introduction.....	1
1.1 Attribute Based Encryption (ABE).....	1
1.2 Functional Encryption (FE).....	2
2 ABE Use Cases.....	2
2.1 Data Security and Data Privacy.....	3
2.1.1 Challenges of legacy systems	3
2.1.2 ABE to Address the Issues.....	4
2.2 Digital Identity and Verifiable Credential.....	6
2.2.1 Different Use Cases of Digital Identity and Verifiable Credential	6
2.2.2 ABE for Digital Identity and Verifiable Credential.....	7
2.2.3 ABE Value Proposition for Digital Identity	8
3 Applications and Use Cases for Functional Encryption (FE)	9
3.1 Secure Email Filtering.....	9
3.2 Surveillance.....	9
3.3 VPN Content Inspection	9
4 PET Adoption and Barriers	10
5 Conclusion	10

Table of Figures

Figure 1: legacy way of access control policy configurations.....	4
Figure 2 legacy data encryption - "all or nothing"	4
Figure 3 ABE access control policy for entire data lifecycle	5
Figure 4 ABE Encryption: large scale data breach avoiding	6
Figure 5 ABE for verifiable credential concept.....	7
Figure 6 biometric attribute(s) to add additional security	8

Executive Summary

NTT Research is NTT R&D's global presence outside of Japan with headquarter based in Bay Area of US. As an organization that has some elements of a public utility and some elements of a private company, NTT Group makes full use of various business resources and capabilities, including research and development, ICT infrastructure and personnel, to solve social issues by promoting digital transformation and CSR through collaborations with our partners. NTT Research expands a rich lineage of ground-breaking basic research heritage from Japan to the rest of the world. Our current focus research areas are:

- physics and informatics
- medical and health informatics
- cryptography and information security

One of our key research topics in cryptography and information security is Attribute Based Encryption (ABE)/Functional Encryption (FE). While ABE and FE have been recognized as part of the Privacy Enhancing Technologies, e.g., by NIST Privacy-Enhancing Cryptography (PEC) project, they have been overshadowed by technologies such as FHE (Full Homomorphic Encryption). We would like to highlight the unique roles ABE/FE can play in many of use cases for enhancing data privacy and security. We believe ABE/FE's characteristics are complementary to other PETs. Lastly, we also would like to provide our observations and feedback about the challenges and barriers in adopting PET. In fact, a number of these factors contribute to slow adoption of ABE even though the technology is mature and can provide benefits to enhance data privacy and data security.

1 Introduction

1.1 Attribute Based Encryption (ABE)

Legacy encryption technologies have been an “all-or-nothing” proposition. For any given encrypted dataset, or ciphertext, if you had the correct key to open it, you had access to the entire dataset. If you didn't have the appropriate key, you could access none of it.

In 2005, cryptographic researchers Brent Waters and Amit Sahai changed that dynamic with a paper that introduced the concepts of attribute-based encryption. The paper¹, titled [“Fuzzy Identity-based Encryption,”](#) introduced the idea that multiple keys could exist for a given ciphertext to enable different users to access different parts of the underlying dataset with attribute-based access control policy. This idea opens up new use cases and applications for cryptography in information security and data protection. Over the years, a lot of more research has been done. ABE algorithms have been proven to be secure and enhanced to be increasingly efficient. With current state of the computing technologies, ABE is ready to be deployed in production environment for many practical use cases.

Moreover, ABE can be upgrade to a quantum-resistant algorithm by swapping the underline cryptographic library to a post quantum cryptography algorithm. Moreover, this upgrade can be

transparent to applications. The NTT implementation demonstrates such transparency for easy migration to post quantum cryptography.

ABE provides the ability to implement Attribute-Based Access Control (ABAC) at cryptographic layer. ABAC is a powerful way to enforce data access policy for data privacy and data security. The concept has been around for a long time. NIS published a guideline in this areaⁱⁱ. With ABE, ABAC can be implemented in the way that is system independent because policy is embedded into encryption. Data sets can be encrypted with ABE to allow different data segments to have different ABAC access policies. Policy implemented with ABE supports “fine-grained” access control.

For example, only users tagged with an attribute of “financial team” can see financial related segment of the data set, and only users tagged with an attribute of “account support” can see customer contact information. Segment(s) of data sets in database implementation can be row(s), column(s), or even down to cell(s).

Embedding access control policy into encryption provides several benefits for data privacy and data security that will be demonstrated in our use case discussions later in this response.

1.2 Functional Encryption (FE)

Functional Encryption (FE) is a generalization of Attribute-Based Encryption (ABE)ⁱⁱⁱ. It allows a user who has a proper key:

- to compute a specific function on encrypted data, and
- to obtain the computing result (unencrypted)

while the original data remains encrypted/protected.

Use cases involving delegating certain actions based on outcome(s) of computing result(s) on encrypted data are perfect for FE. A data owner can provide specific key(s) to third party to allow this third party to compute certain function(s) without decrypting original data. Based on the result(s) of computing function(s), the third party can take proper action(s) without the need to get the data owner involved in the process.

Unlike ABE (even though ABE can be viewed as a specialization of FE), full functionality of FE is not yet ready for production. More research in this area is needed to realize its full potential.

2 ABE Use Cases

Until recently, most of the use cases identified for ABE are related to data protection. Digital Identity and Verifiable Credential (VC) emerge as new applications. ABE provides flexibility to support a wide range digital credential verification use cases from “full knowledge proof” to “Zero Knowledge Proof”.

2.1 Data Security and Data Privacy

As organizations increasingly deploy AI/ML capabilities and adopt data driven decision making processes, data becomes one of the most valuable assets. On the other hand, data related to personal privacy needs to be properly handled and protected. Moreover, today's hostile cybersecurity environment demands organizations to eliminate every single potential vulnerability for data breach. All these factors ask for better data protection for privacy and security. There are several limitations with legacy ways of protecting data. ABE can be used to bridge these gaps.

ABE provides data privacy enhancements and protection to sensitive data in database, data lake, lake house, data warehouse, blockchain or entire data asset of an organization. The key benefit is the ability to embed access control policy into encryption, so that consistent access control can be enforced starting from the point the data is created throughout the entire life cycle regardless of where it gets moved or stored. The end results are:

- empower data owners to have full control of their data through entire life cycle with consistent access control policy
- allow organizations to have better data governance
- reduce policy configuration overhead
- minimize large scale data breach

2.1.1 Challenges of legacy systems

2.1.1.1 System Based Access Policy Limitations

Many organizations deploy multiple systems for data storage/process (see “ Figure 1: legacy way of access control policy configurations”). Data is created at the data source and flows through multiple systems to downstream, e.g., data lake for analytics/ML/AL purpose. Implementing access control policy in a consistent way for every system is a challenge task. Often time, configurations for all systems are required to achieve such goal. Multi-system configurations come with overhead that consumes resources and drives up cost.

Moreover, system-based access control policy can only be applied and enforced to data residing inside of given system(s). Once the data is sent out, downloaded, leaked, or hacked out of protection boundary, there is no access control policy enforcement anymore.

Lastly, because system-based access control needs system specific knowledge to make necessary configuration(s), most of data owners do not have the ability to perform such task(s) unless another layer of system being deployed for policy building. These configurations most likely rely on members of system admin personnel. As a result, system admin personnel might have access to sensitive data that they are not supposed to access to if proper processes and/or system capabilities are not in place - data privacy and data security might be compromised.

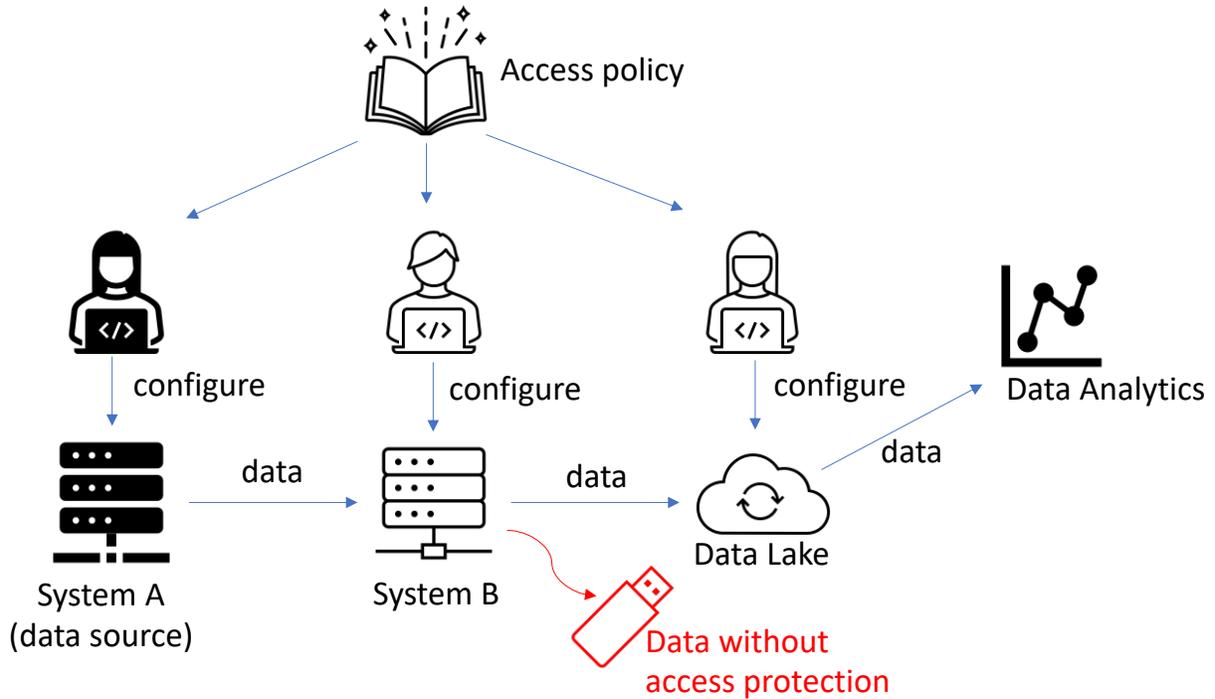


Figure 1: legacy way of access control policy configurations

2.1.1.2 Legacy Data Encryption Limitations

To protect sensitive data, legacy encryption is often utilized. However, while legacy encryption technologies such as AES provide high degree of security for data protection, there is a single decryption key being used by all users in daily operation. Once the decryption key is hacked with the encrypted data, all data will be exposed (see Figure 2 legacy data encryption - "all or nothing"). "All or nothing" access is too vulnerable for large scale data set such as customer database.

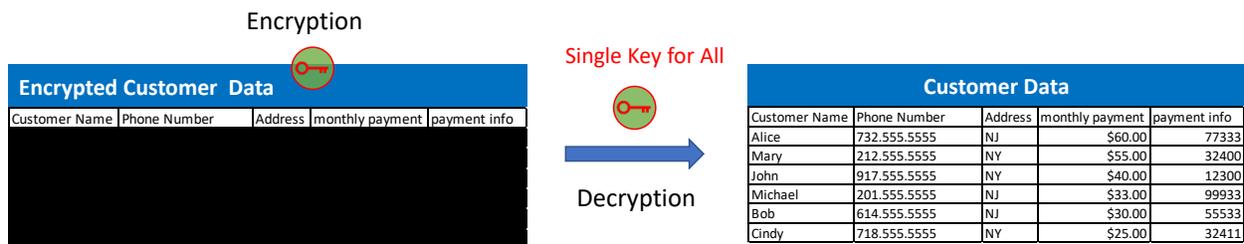


Figure 2 legacy data encryption - "all or nothing"

2.1.2 ABE to Address the Issues

ABE encryption can be applied when data is created or as close to the point that the data is created (see "Figure 3 ABE access control policy for entire data lifecycle"). This provides several benefits in terms of data protection and large-scale data breach avoiding.

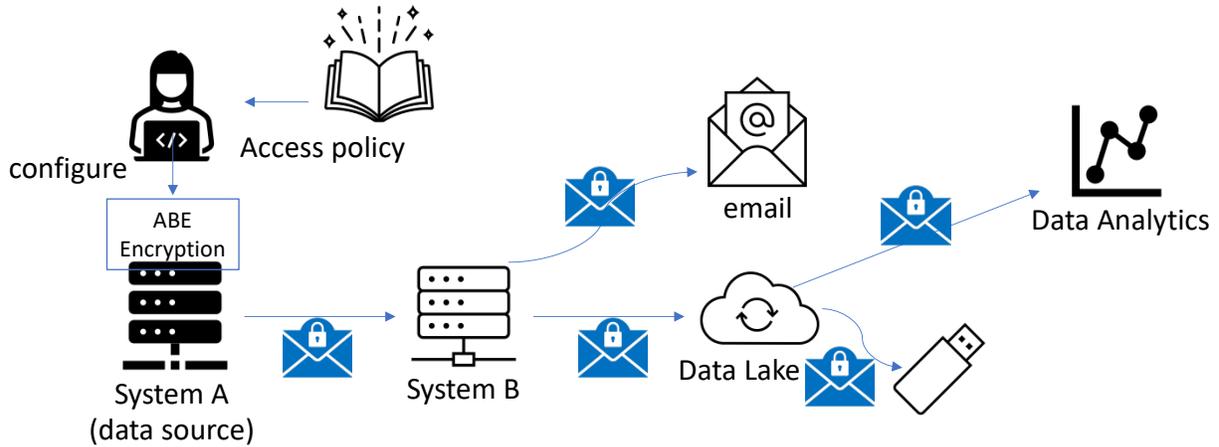


Figure 3 ABE access control policy for entire data lifecycle

2.1.2.1 Data Layer Access Control

Because data access policy of ABE is embedded in the data layer, this eliminates the need to configure system level access control policy on a system-by-system basis. With that, cost saving and resource efficiency can be achieved. It also provides the benefit of consistent policy flowing down with data from upstream to downstream, even for off-stream, e.g., download. Lastly, with proper policy builder deployed in conjunction with ABE, data owners can configure their own access control policy from upstream without worrying about any element of the policy being tempered or compromised when data flows down from one system to others.

2.1.2.2 Large Scale Data Breach Avoiding

2.1.2.2.1 Needed Basis Decryption Keys Instead of “Single Key for All”

ABE provides the ability to generate different user secret keys for different users with data access policy on a needed basis (see “ Figure 4 ABE Encryption: large scale data breach avoiding”). Only data segment(s) users have right to access can be decrypted instead of “single key for all”. Users use their own keys instead of a single decryption key used by all users in daily operation. As such, as long as there is proper protection for Key Generation System (KGS), possibility of large-scale data breach is decreased.

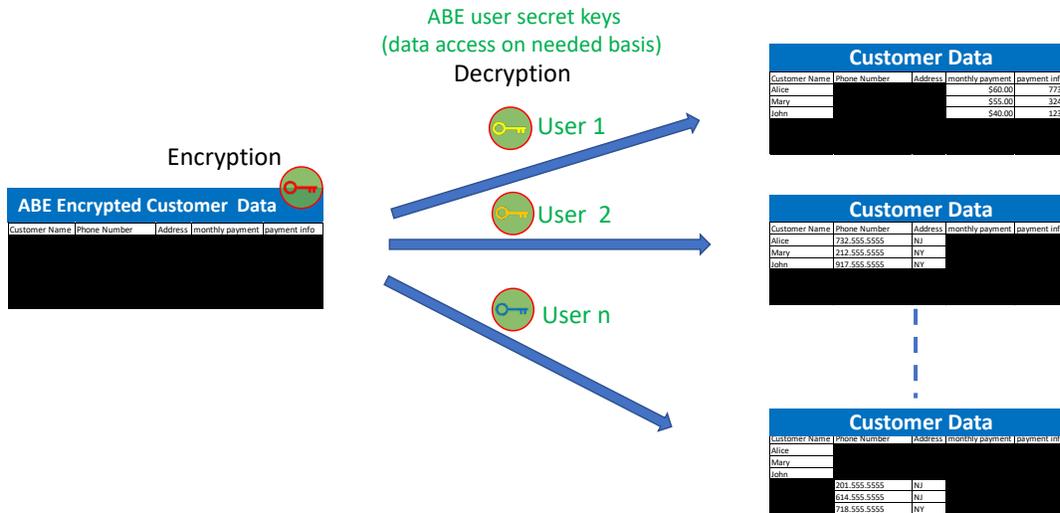


Figure 4 ABE Encryption: large scale data breach avoiding

2.1.2.2.2 Additional Layer of Security – Environmental Conditions

Additional security measures can also be implemented through environment/conditional attributes supported by ABE to ensure decryption keys only function in certain conditions, e.g., time of the date, specific location(s), network(s), device(s), and such. Proper access policy with these conditional attributes makes it harder for hackers.

2.2 Digital Identity and Verifiable Credential

Recent research expands ABE use cases from data protection to Digital Identity and Verifiable Credential (VC).^{iv}

2.2.1 Different Use Cases of Digital Identity and Verifiable Credential

Different use cases might have different requirements for identity/credential verifications. Individuals to be verified can be completely anonymous for some use cases. For other use cases, they are required to provide a set of detailed identity attribute(s), and these attribute(s) need to be verified. There are also use cases falling in between:

- Individuals need to be verified of detailed information of identity and personal information (“full knowledge proof”):
 - Banking, e.g., to open an account
 - Government agencies, e.g., airport security check
- Individuals can be completely anonymous (“zero knowledge proof”):
 - senior discount, youth discount, Low-income discount
 - Resident based discount: transportation, theme park ticket/admission, etc.
- Individuals need to be identified, but other detailed identity attribute(s) might not be required; only threshold verification for certain attribute(s) is required (“partial knowledge proof”):
 - Age restricted purchasing: alcohol, tobacco, recreation drug, lottery tickets, etc.
 - Age restricted admission: casino, gym, etc.
 - Income threshold qualification for rental applications

ABE can be used to support all these use cases with added value in comparing to the current ways of operations as well as other potential technologies.

2.2.2 ABE for Digital Identity and Verifiable Credential

A high-level architecture is depicted in “ Figure 5 ABE for verifiable credential concept”:

- Identity Authority or Trusted Service Providers: This is an entity to maintain/to collect and to proof identity data; it also performs user authentication and ABE key generation.
- Person: S/he registers with Identity Authority to allow her/him to obtain digital identity in the form of ABE user key
- ABE user key represents digital identity containing identity attributes such as ID #, name, etc.
- Identity Verifiers: These are business or agencies that need to verify individual’s identity for different use cases such as age restricted purchase, banking activities and such.
- Verification Steps:
 1. Person gets authenticated from Identity Authority and obtains digital ID, i.e., ABE user key containing person’s identity attributes, e.g., passport #, name, citizenship, DoB (Day of Birth), etc.
 2. Identity Verifier use ABE (public parameter of ABE) to encrypt a “secret value” as “challenge” to send to Person. The access policy of this ABE encryption corresponds to verification requirements of the use case, e.g.,
 - age threshold for supermarket senior discount
 - specific person’s name, passport # and such for international flight boarding process
 3. If the person can use her or his ABE user key to decrypt and return the “secret value”, verification is complete with a positive result. If not, the verification fails.

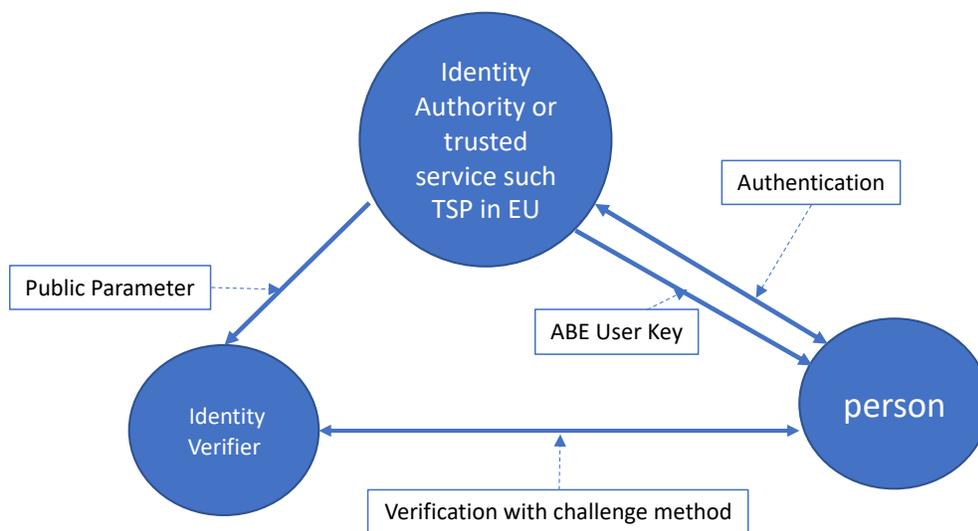


Figure 5 ABE for verifiable credential concept

ABE provides the ability to use biometric attribute(s) to tie digital identity to individuals in a secure way for critical use cases:

- Biometric attributes can be range from a simple hash value of an individual’s of ID photo to attribute(s) for more advanced technologies, e.g., three-dimensional facial recognition and iris recognition. The implementation can evolve along with technology innovation. This provides additional layer of security for preventing stolen keys being used for critical use cases.

Please see “Figure 6 biometric attribute(s) to add additional security” for illustration purpose:

- Person’s user secret key contains hash value of ID photo as one of the identity attributes. Person provides the ID photo to airport along with other required identity info electronically.
- Airport uses the ID photo to inspect if it matches Person either by human inspection or facial recognition.
- If there is a match, Airport calculates hash value of the ID photo and uses it along with other needed identity attributes to form access policy for “challenge” for Person.
- If Person’s ABE user key contains all proper attributes, e.g., ID photo hash value as well as other required attributes such as passport #, etc., Person can decrypt the “secret value” and return it to the airport inspection agent to complete the verification process.

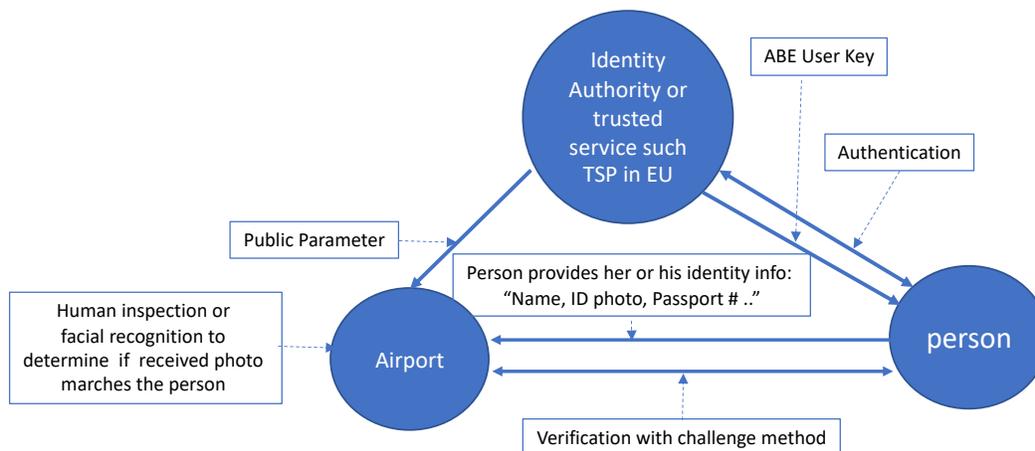


Figure 6 biometric attribute(s) to add additional security

2.2.3 ABE Value Proposition for Digital Identity

ABE identity verification provides value for all use cases:

- “Full knowledge proof”: Detailed credential verification:
 - All required attributes from an individual are verifiable through crypto technology with extremely low tampering possibility.
- “Zero knowledge proof”: no detailed identity information is shared with verifiers while a given requirement can be verified, e.g., age threshold:
 - For the persons being verified: no worry about their identity information being stolen – everything is secured without privacy exposure.

- For the verifiers (typically business entities): no need to touch PPI (Personal Private Information) data, it eliminates any liability as well as system requirements for handling PPI data.
- In comparing to the traditional physical ID, it eliminates the possibility of fake IDs being used as well as potential identity theft
- In comparing to software/system-based implementations, ABE provides much more secure and flexible ways in an end-to-end cryptographical fashion:
 - Attributes of user secret keys are generated by authorities or trust services, and they cannot be tampered for entire key cycle – from generation, to utilization, to expiration.
- Distributive verifications with only key generation relying on centralized system(s) – highly reliable architecture
- ABE provides multiple layers of added security protection:
 - the ability to tie digital ID to specific person by using biometric attribute(s)
 - the ability to tie digital ID to specific hardware/device with other environmental attributes
- Crypto technology of ABE provides a future proof evolution path in many ways, e.g., biometric attribute(s) can evolve along with technology innovation

3 Applications and Use Cases for Functional Encryption (FE)

Functional Encryption (FE) can be used for applications where it's useful to reveal just specific result(s) of the computing while original data remains to be encrypted/protected.

3.1 Secure Email Filtering

With FE encrypted email, several actions a user can delegate to email server to do without getting email content decrypted:

- To determine email message is spam, but without revealing the contents of the message
- Similarly, it could be used to determine whether an encrypted email is from important person(s), and thus flagged as “important,” but again without revealing the contents

3.2 Surveillance

FE could be used to provide privacy preservation to individuals while a specific key to allow law enforcement to examine a series of surveillance photos to determine whether a particular person is included in an image – without revealing other contents of the images.

3.3 VPN Content Inspection

The [Google Transparency Report^v](#) shows the percentage of encrypted web traffic has increased from less than 50% in early 2015 to about 90% or better today for all major platforms except Linux, which sits at 77%. This has presented huge concern for organizations from cyber security point of view. To detect any threats in encrypted traffic, organizations often use Transport Layer Security (TLS) inspection. TLS inspection decrypts TLS traffic, enabling security tools to look for malware and other threats, then re-encrypt the session. This, of course, presents new risks because the traffic is in the clear, unencrypted form during the inspection process - user privacy is compromised.

In the future, FE can be used for VPN, and specific computing key(s) can be used by security tool(s) for such inspection while VPN content remains to be encrypted/protected – no privacy compromise.

4 PET Adoption and Barriers

While a lot of research work and “proof of concept” trials have been done for different PETs, large scale adoption has not been widely deployed in actual business. There are number of reasons. Here are our observations:

- The common perception in the industry is that cryptographic technology needs to be part of the NIST standard portfolio to be considered secure. While there is an on-going PEC project within NIST, hardly any PET is part of NIST standard yet.
- In fact, data privacy regulations such as HIPAA do not provide clear requirements for data protection technology, and they are sometime conflicting. We recently conducted a study in this area. Such study is time consuming and very costly.
- To avoid such confusion, consistent and more clarity of regulations will certainly help. Moreover, regulations should provide a clear path to allow continued technological innovation.
- We all agree that it would be beneficial to have a common reference to guide business/government agencies about specific type of PET that is ready for production. If NIST standard is the one, current NIST standard process might not serve the needs with multi-year timeline. Furthermore, resource constraint might put things into queue before such technology can be put on the starting line of standardization.
- Lastly, it would be good to recognize the fact that different PET might be at different maturity stage. Based on maturity stage, different actions need to be prioritized to speed up research, adoption, and deployment. For example,
 - for technologies such as ABE that are mature and ready for production deployment, the urgent action is to speed up any needed standardization or deployment guideline as well as to provide clarity about its ability to support regulatory compliance.
 - For technologies that require more research and enhancements, proper government funding and policy will accelerate the speed to achieve maturity.

5 Conclusion

NTT Research appreciates OSTP taking a leadership position in advancing PET. We are honor to provide you with our inputs. We welcome any follow-up opportunity to support this initiative.

References

-
- ⁱ Amit Sahai and Brent Waters – “Fuzzy Identity-Based Encryption”:
<https://eprint.iacr.org/2004/086>
- ⁱⁱ NIST Special Publication 800-162 – “Guide to Attribute Based Access Control (ABAC) Definition and Considerations”: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-162.pdf>
- ⁱⁱⁱ Boneh, Dan; Amit Sahai; Brent Waters – “Functional Encryption: Definitions and Challenges”:
<https://eprint.iacr.org/2010/543.pdf>
- ^{iv} Giovanni Bartolomeo – “Attribute-Based Encryption for Access Control in Cloud Ecosystems”:
[https://www.techrxiv.org/articles/preprint/Attribute-Based Encryption for Access Control in Cloud Ecosystems/13546871](https://www.techrxiv.org/articles/preprint/Attribute-Based%20Encryption%20for%20Access%20Control%20in%20Cloud%20Ecosystems/13546871)
- ^v Google Transparency Report - <https://transparencyreport.google.com/https/overview?hl=e>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

OpenMined

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



RFI Response: Privacy-Enhancing Technologies

Office of Science and Technology Policy
White House
1650 Pennsylvania Avenue NW
Washington, DC 20502

Re: Request for Information on Advancing Privacy-Enhancing Technologies (87 FR 35250)

We appreciate the opportunity to respond to this Request for Information on Advancing Privacy-Enhancing Technologies.

[OpenMined](#) is a nonprofit organization that develops leading free and open-source software allowing data scientists and A.I. or machine learning practitioners to use advanced privacy enhancing technologies (PETs). We [educate](#) thousands of professionals about PETs; organize a technical and professional community; and deploy boundary-pushing implementations of PETs with institutions including Twitter and the UN PET Lab. The best place to learn more about our work is our [blog](#).

Our mission is to unlock the stunning potential of non-public data to advance social good, while simultaneously helping engineers, institutions, and society navigate the complex hazards presented by these technologies. Neither of these aims, we believe, can be effectively pursued without the other.

- 1. PETs are infrastructure. Network effects and technical lock-in make it critical for the government to use and support free and open-source PETs standards. (Questions 1, 2, 5, 7, 8)**

Infrastructure and the public interest

PETs are more than useful tools. They are vital infrastructure for tomorrow's digital networks. Therefore we believe the most important way the government can ensure that PETs work in the public's interest is to **favor free and open-source PET standards through policy, resources, and official action.**

When the modern internet coalesced in the early 1990s, the work of farsighted policymakers helped ensure that it was built on a foundation of free and open source software, such as the

TCP protocol from the earlier ARPANET. This stroke of leadership was critical to bringing about a true, open “world wide web”. Otherwise, the internet would have emerged in the form of walled gardens dominated by rent-seeking private actors, like AOL and the major telecommunications companies. The dynamic startup culture of the 1990s and 2000s might not have emerged. Or, the revolution of an open internet might have started elsewhere, with the United States failing to lead it.

Network effects, technical lock-in, and competition

In the coming years, PETs will increasingly supplement legacy data transmission protocols with more precise and targeted methods of information sharing. This will represent a tectonic shift in network architecture, comparable to the emergence of the early internet.

As in the 1990s, network effects will create technical and economic pressure toward the consolidation of a *single* interconnected network (or at most a few), through which information may be shared using common software protocols. This means that whatever PET-enabled networks seize the first-mover advantages could quickly achieve global scale and lasting strategic dominance.

Let us walk through a hypothetical example of how this could happen. Suppose numerous important holders of sensitive medical data, like hospitals, decide to make their combined data available to researchers through advanced privacy preserving methods. To this end, they set up access and permissions to use a very large combined dataset. Their initiative is successful: it dramatically accelerates researchers’ progress, without disturbing upstream privacy, confidentiality, or intellectual property interests.

This could result in that network growing exponentially. Why? Since data is more valuable when combined with complementary data, a leading PET-enhanced data sharing network would be able to pay the market’s best prices for additional contributions of information, compounding its advantage. The network’s utility would likely expand beyond its initial ambit into other fields with no obvious limit. Access to it would become indispensable not only to medical researchers, but to practitioners of a wide variety of A.I. and machine learning disciplines, and the businesses that rely on their work (i.e., most large businesses).

In this way, the maturation of PETs could have extremely negative consequences and lasting consequences regarding competition. A dominant network dependent on proprietary PET standards would give private actors irresistible opportunities to extract rents and impose switching costs. Exclusive reliance on free and open source PET protocols is therefore a necessary (albeit not sufficient) requirement to protect the public and the economy from foreseeable harm.

The government should use its substantial leverage to send the industry down the path of free and open-source PETs standards, as described below.

Supporting open-source PETs

The government should support free and open source PETs through (a) policy, (b) resources, and (c) official action to speed adoption. By throwing its considerable weight behind these efforts, the government can help ensure that the dominant PET-enabled networks that emerge in coming years will not rely on proprietary standards.

Regarding *policy* and *resourcing*, we broadly support the **Promoting Digital Privacy Technologies Act** (H.R. 847 and S. 224) and the related **DATA Privacy Act** (H.R. 5807 and S.3065), both of which would accelerate the development (and therefore adoption) of PETs. H.R. 847 contains language at Section 3(7) directing the National Science Foundation (NSF) to support “development of freely available privacy enhancing technology software libraries, platforms, and applications”. We applaud this clear emphasis. **This and similar legislation should be complemented by a firm operational understanding of PETs’ competition implications, by decision-makers across all Federal agencies.** To this end we believe that efficient and focused education initiatives directed toward key Federal officials is warranted. OpenMined would be happy to comment further on this.

The **Secure Research Data Network Act** (S. 4200) is also an encouraging initiative. This bill would create responsible Federal data stewards helping researchers benefit from Federal data. Such stewards could help shape best practices, increasing the chances that important PETs networks are operated in the public interest. At Section 3(D)(2), the proposal provides that such networks must be built with free and open source software. This provision is of vital importance for reasons we have articulated.

Regarding *official actions to speed adoption*, much can be undertaken in advance of or in addition to legislation. For example it is already possible for any and all agencies to host dedicated federated data networks providing safe access to the information they steward. Early examples include OpenMined’s present work on pilot initiatives with the [U.S. Census Bureau](#). In addition to replicating and expanding projects of this kind, the government should support academic work in any discipline that demonstrates the utility of the newly-accessible data. For example, NSF grants could favor applications that express an intention to explore or leverage such data.

2. PETs serve democracy by enabling transparency into important networks. (Questions 1, 2, 3, 4, 9)

PETs can unlock maximum transparency into technical systems that have strategic or public importance – such as those run by governments or large technology companies – while respecting privacy, security, or intellectual property concerns.

This is one of PETs’ most important implications for justice, economic fairness, and the future of democratic institutions.

Examples of systems that PETs can make more transparent include:

- social media companies' algorithms, which may fuel radicalism, polarization, or foreign political influence
- new digital networks using decentralized technologies like blockchains, which may generate complex, subtle, and hard-to-foresee distortions of social, political, and economic life
- enforcement patterns at regulatory or law enforcement agencies, which may exhibit complex unfairness along racial, gender, or other protected qualities

Today, only the parties that own and operate these systems have the information necessary to understand and identify their complex public implications. Yet, analyzing and understanding these implications is at best outside the expertise, and at worst against the incentives of such parties. It is therefore critical to develop techniques allowing third parties like researchers, civil society organizations, and government regulators, to study and understand the patterns of activity in these networks.

PETs are our sharpest tools for advancing the frontiers of transparency while respecting the privacy, security, and intellectual property concerns that weigh in favor of opacity. Thus, investing in both fundamental PETs techniques and ambitious applied projects is necessary if our technical systems are to operate in accordance with open and democratic values.

Supporting ambitious demonstrations of techniques that can address algorithmic bias and related problems

OpenMined is conducting a [pilot with Twitter](#) that advances and demonstrates PETs' transparency-unlocking possibilities, and could serve as a precedent for other social media companies.

The project uses a large, 1-billion-tweet private dataset from Twitter, which researchers could not access through the ordinary Twitter API. It creates a synthetic version of that data, and makes it available to researchers to perform computation on it. Researchers can see the results of their computations using an adversarial differential privacy system. This system tracks privacy budgets on the level of individual Twitter users, not as an aggregate dataset. This enables researchers to obtain much more detailed information than conventional differential privacy deployments, while still ensuring that they cannot reconstruct individual information that would have privacy implications.

The project is experimental. It remains to be seen, for example, how useful researchers' findings will be concerning Twitter's algorithms' influence on political discourse. However, if successful, it will shift industry norms in terms of how much transparency platforms should provide into their algorithms. **Platforms offering researchers less access than the standards set by this and future programs will not be able to claim the fig leaves of privacy, security, or trade secrets.**

Applied projects like this are perhaps the most important activity that the government can support to nudge the technology industry towards better transparency practices.

Regulating decentralized networks with PET-enabled safe havens

In the short term, most of the transparency-related opportunities for PETs concern building windows into centralized networks. However, the government should also be aware of the medium-term relevance of PETs to the regulation of decentralized networks.

Today, the largest blockchain networks are anonymous, but public. This permits a [surprisingly high degree](#) of traceability for regulators. However, some less-popular decentralized networks are truly resistant to tracing, and therefore potentially problematic for law enforcement. Such hard-to-trace networks can be expected to proliferate and/or grow in sophistication. It is not yet clear what strategy the government can or should take in this area.

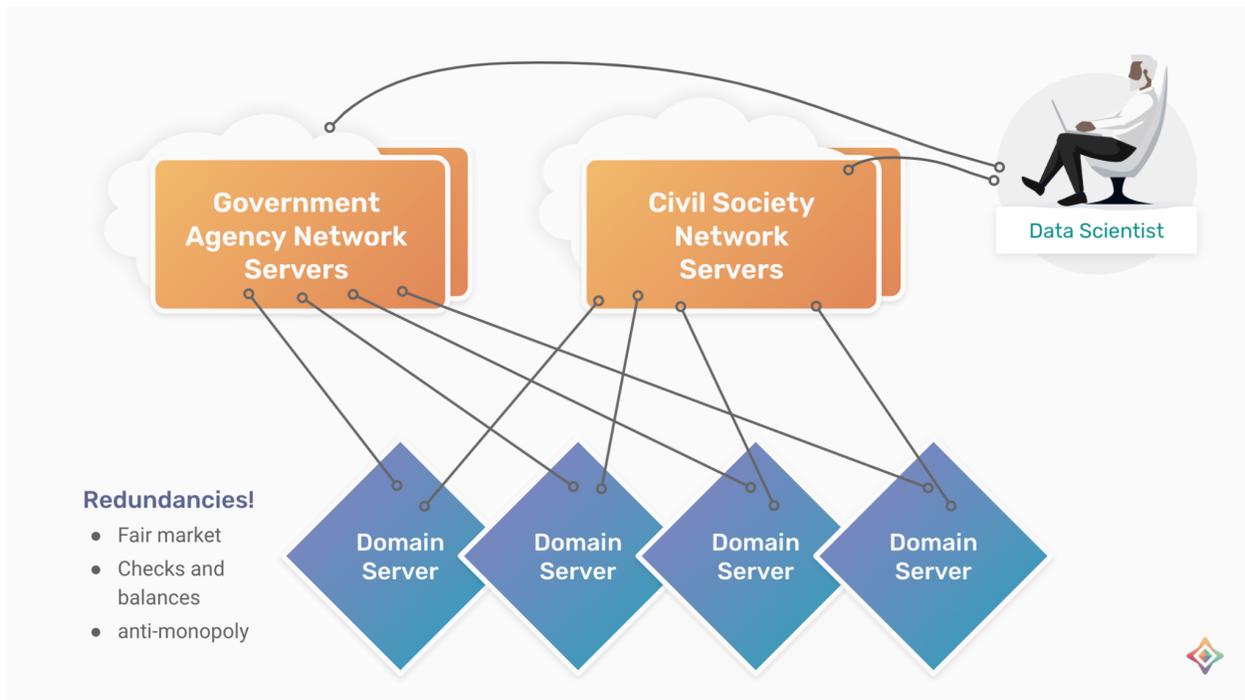
PETs create interesting, although still speculative possibilities. First, it is important to acknowledge there are legitimate and potentially beneficial applications for hard-to-trace decentralized networks, which should be weighed against their propensity for misuse. For example, consider communities wishing to share resources in oppressive regimes; or legitimate activists seeking to coordinate mutual aid networks without exposing themselves to social or professional retribution. The government might permit hard-to-trace networks to operate for such legitimate purposes by crafting PET-enabled safe-harbor rules. For example, otherwise-untraceable networks could be licensed if their architecture enabled transparency sufficient to rule out illicit uses while preserving essential privacy features.

Helping public institutions to develop best practices for using PETs

Government agencies and regulators themselves often harm the public through structural bias. Letting researchers study internal data from agencies such as the IRS, Census, or DOJ could uncover such bias and improve the government's ability to serve the public. The **Secure Research Data Network Act** (S. 4200) would be a welcome step forward in this regard.

OpenMined has partnered with Census, IRS, and the UN PET Lab to explore the utility of PETs for agency use. Census and IRS manage sensitive data of huge public value, which our systems help them open to new uses. Similarly, our work with the UN PET Lab lets trade authorities of competing nations reconcile complex data on imports and exports, helping them identify accounting discrepancies (and accelerate trade), without needing to share sensitive data with non-trusted counterparts.

Going a step further, our systems enable unique forms of collaboration between government and civil society. Responsible civil society organizations can be directly incorporated into the networks through which researchers access sensitive data; this serves as a check and balance on the decisions of agency network administrators.



Analogously, OpenMined’s AdAstra project has established a federated learning network across private data servers located at over 100 universities and research institutions. This is an unprecedented effort to unify disparate internal review processes for accessing research; while also building a system of unhindered privacy-preserving access so that researchers can work from exponentially larger datasets.

Establishing PETs in the public and nonprofit sector has long-term strategic and economic implications. Agencies, more than private companies, are genuinely accountable to the public interest. This means that, with good leadership, their incentives are better aligned toward developing responsible uses of PETs. Moreover, dovetailing with our argument in Section 1, PET networks operated by government agencies could grow exponentially and become strategically important.

It is much better for the stewards of strategic networks to be agencies, which are accountable to the U.S. public, than private companies accountable to shareholders and potentially foreign interests.

3. Encouraging bottom-up and shared data governance. (Questions 5, 10)

The architecture of data sharing will change fundamentally in the coming years. Optimistically, this shift will involve the establishment of a new class of innovative, fiduciary-style institutions that manage data on behalf of interest-aligned consumers, businesses, and others. Broadly similar versions of this idea give different names to such hypothetical institutions, including [data coalitions](#), [data trusts](#), data cooperatives, [mediators of individual data](#), [middleware](#), and [more](#).

These evolving ideas hold medium-term promise. Trustworthy institutions could, for example, help individual consumers aggregate their interests and mitigate the take-it-or-leave-it dynamics that characterize their present interactions with dominant platforms. PETs are an important part of this vision: they will enable such institutions to exert much more meaningful control over data.

Requirements that digital service providers not monopolize their customers' data – similar to those imposed on IoT providers by Chapters II-III of the [EU Data Act](#) – would help support the emergence of such PET-enabled data governance, by preventing data from being locked in to dominant platforms.

Conclusion

The government should view PETs as basic infrastructure. Supporting free and open-source standards reduces the danger of harmful monopolies or rent-extraction that could harm the public and hinder U.S. competitiveness.

PETs have an important role to play in restoring and maintaining the health of democracy by enabling secure transparency into important institutions, from government agencies and technology companies to decentralized networks.

PETs may also unlock possibilities for promising forms of shared data governance that could empower consumers and others.

The government can and should support positive progress by supporting legislation, encouraging research, and actively incubating more and better-funded experiments. Decision-makers across federal agencies should be educated about PETs' strategic considerations. To accelerate adoption, federal funds should focus on the demand side in addition to the supply side, e.g., training and supporting researchers whose work makes use of public data made available through PETs.

OpenMined looks forward to actively assisting this work. Please do not hesitate to contact us – we would be delighted to continue the conversation.

Respectfully Submitted,

Andrew Trask

Leader, OpenMined

Matt Prewitt

Advisor, OpenMined

President, RadicalxChange Foundation

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Palantir

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

ORGANIZATION INFORMATION

Organization Name: Palantir Technologies Inc. (“Palantir”)

Organization Type: Industry

Organization Description: Palantir builds innovative software that helps organizations make sense of disparate data at scale. Palantir’s software incorporates robust security, privacy, and data protection capabilities at every level of its platforms, allowing organizations to individually secure each piece of integrated data, set granular access controls, and maintain context-enriched audit logs of user activity. These features, baked in from the beginning of development, help ensure responsible, proportionate, and legitimate interactions with sensitive information. Organizations in national security, defense, law enforcement, health, and finance trust Palantir to safeguard, manage, and analyze their most important data assets.

We built privacy controls, data protection, and governance into our platforms from the start. Our implementation of privacy-enhancing technologies (PETs) across our platforms has helped our organization build trust with the institutions whose work relies on safekeeping information and protecting the data of their constituents.

INTRODUCTION

Data sharing and analytics is a critical component to solving many of the nation’s most pressing problems: from critical emergency response (e.g., natural disasters, pandemics) and supply chain bottlenecks to carbon footprint tracking and clean energy development. While these projects vary in scope and impact, they all begin with the same fundamental question: is the relevant data available? If so, what is the quality of this data, in terms of completeness, accuracy, and timeliness? From that foundation, organizations may logically turn to questions about which PETs they might most effectively deploy to protect their most valuable assets. However, without a solid data foundation, novel PETs can be a distraction or, worse yet, fundamentally mislead, alter, or otherwise hide data quality issues and impede operational outcomes and decisions.

Data quality is a central concern because it is the foundation upon which all successful data and analytics projects are built. The application of PETs is also contingent upon data quality. A single bad dataset has the potential to compromise an entire data-driven initiative, triggering consumer distrust and rendering the effort moot. Worse, it could get into the hands of users and be misinterpreted or actioned upon.

Sometimes these lapses in data quality are simply a result of having the wrong source data (e.g., an incorrect mapping table), sometimes the data just represents a different perspective of correctness (e.g., different definitions of metrics between business groups), or sometimes it is due to mistakes in the data preparation (e.g., incorrect data cleaning logic).

When users are blind to assumptions or issues in their data, the decisions made upon that flawed data can propagate quickly throughout an organization, with successive levels of users left none the wiser. This is why data quality must be paired with robust data transparency: all users, with relevant roles and permissions, should have the full context necessary to appropriately use and trust their data.

Coupled with a strong foundation of quality data, governance and technology are equal enablers of effective PET implementation. Rigorous privacy protections coded into governance policy is socio-technical and must factor in the institutional workings as much as the pure data/technology considerations.

RESPONSES TO QUESTIONS

1. Specific research opportunities to advance PETs

Privacy-enhancing technologies (PETs) are best understood not as standalone tools, but as instruments within more complete systems. Integrated together with other PETs as well as other technical products or organizational governance procedures, PETs produce a contextually configurable, holistic data governance arsenal. Whilst innovative approaches such as homomorphic encryption may herald great promise as privacy ‘cure-alls’ in specific circumstances, this should not distract from the tremendous effects that can already be generated by combining a comprehensive data foundation with powerful and flexible data governance tools such as granular access controls and sophisticated data minimization regimes.

Accelerating the adoption of individual and novel PETs should therefore factor in three critical systems-level considerations:

- **Integration:** As with any software product, proper technical integration is key. Integration overhead, execution risk, and ongoing maintenance requirements can be minimized by relying on commercially available software wherever possible, rather than developing one-off integrations.
- **Governance:** Technology-driven governance is a powerful tool to promote best practices and the enforcement of PETs. For example, privacy-enhancing governance policies can and should be effectively applied, enforced, and monitored in integrated software platforms. Technology-driven governance can automate difficult or time-consuming actions, such as the correct application of policies and security controls to specific users, that might otherwise hamper the adoption of PETs.
- **Orchestration:** Software-based orchestration acts as a force multiplier for PET adoption. Instead of relying on a patchwork solution backed by complex, manual processes and policies, software can reduce total cost of ownership (TCO) and remediate the burdensome technical sophistication required by end users to implement or use PETs. In this way, commercially available software is a critical component in the adoption of PrivacyOps, which integrates teams within an automation-driven common framework that enables communication and collaboration for most important practices of privacy compliance.

Relying on a diversity of interoperable fundamentals rather than narrow, “magic bullet” technologies will be advantageous in large or complex processing environments which feature a range of heterogeneous data sources and workflows. Not all of these may be amenable to a single PET; instead, an interoperable, layered “fundamentals” approach is necessary. This approach is akin to best practices in related fields like cyber security, in which a defense-in-depth strategy adds intentional redundancy and strengthens protections against varied and sometimes unanticipated failure-modes.

Given the interconnected nature of PETs implementation and governance, future educational and training programs would benefit from the inclusion of privacy-focused governance topics. Such programs should focus on the human aspect of governance policy and implementation, as well as how technology can expedite and more consistently enforce privacy-enhancing tools, techniques, and best practices.

Like the current state of PETs, there are many ways to expand and advance research in support of PETs. These methodologies and approaches suffer from many of the same problems facing academic, medical, or other research: progress is slow, collaboration is limited, and the process is fraught with barriers. PETs research would benefit from software platforms that can enable secure, collaborative, and reproducible research, which would enhance outcomes including the development and improvement of existing PETs.

2. Specific technical aspects or limitations of PETs

PETs are only effective if the underlying data foundation is sound. Organizations need a comprehensive overview of their processing operations, including the ability to check their data for quality, accuracy, and

representativeness. In the absence of this, organizations will struggle to configure and apply PETs effectively and sustainably. This is true regardless of the specific platform architecture that a data governance team is responsible for overseeing, whether data is held in a single data lake, a federated system, or various siloed systems.

The use of PETs can be useful in connection with data protection requirements, e.g., granular access controls that enable controllers to limit processing of personal data or prompts that ask users to confirm the purpose of processing prior to accessing sensitive data. These approaches can fail when implemented either on a poorly-understood data foundation, or as standalone, “magic bullet” solutions that fail to bridge the gap from academic innovation to real world impact due to the factors described below:

- **PETs’ technical sophistication:** The technical sophistication of PETs yields a number of challenges for organizations attempting to implement and enforce them across their enterprise:
 - Novel PETs are often difficult to implement and utilization without an underlying orchestration infrastructure.
 - Effective PETs governance may require semi-automation or augmentative controls over the processes and policies supported by some PETs.
- **Configuring solutions marketed as one-size-fits-all:** Such solutions often claim “out of the box” functionality that must actually be contextually situated and adapted to the particular processing and threat circumstances (e.g., the specifics of particular ontologies, data, use cases, attack vectors).
- **Lack of interoperability and the inadequacies of point solutions:** Operators are often left to navigate whether the specific PET interoperates with the other governance or oversight processes on which the system depends.
- **High set-up costs:** The set-up costs required to establish PETs may be prohibitive for certain organizations that would otherwise like to deploy PETs. The expertise to deploy PETs may also come at a high cost as the required skillsets are highly specialized and in limited supply.
- **Lack of scalability and prohibitive computational requirements:** Organizations may find that the additional costs accrued in the transition from controlled environment testing to real-world operations on large scale enterprise platforms are prohibitive.
- **Adaptability:** Organization leaders will need to assess whether their proposed PET solutions are applicable to anticipated future changes in processing operations on dynamic, changing environments. PETs that are highly context dependent and brittle should be avoided due to the up-front costs and minimal long-term viability and impact.
- **Lack of communicability and comprehensibility:** PETs should be interoperable and promote communication across the range of relevant stakeholders in the privacy-promoting space (e.g., operations, IT, legal). Effective PETs can be integrated within a large enterprise solution that can be tailored for the variety of relevant stakeholders and users to promote, rather than hinder, communication and comprehensibility.
- **Misguided understanding:** Resistance to PETs adoption may also stem from the belief that PETs can impede and introduce friction to the process of innovation. This view should be challenged: in our experience, this perception is misguided. In fact, the absence of PETs can undermine consumer and citizen trust in personal data processing in both the public and private sectors. The lack of trust and consensus can serve as a more significant impediment to innovation.

The limiting conditions of PETs adoption outlined above may be mitigated with an approach to innovation that focuses not just on the narrow purpose impacts of specific products in standalone, controlled environments, but also aims to address the full ecosystem and lifecycle of data management in complex real-world systems. In this more holistic setting, privacy risks may be better addressed through a combination of several inter-related and reinforcing technological safeguards including but not limited to:

- **Access Permissions:** Ensuring that users only have access to precise subsets of data necessary for their responsibilities.

- **Action Permissions:** Restricting permissions to conduct potentially sensitive actions, such as importing, exporting, transferring, or combining data to those users who absolutely need to do so.
- **Marking Data:** Persistently tagging sensitive datasets to clearly indicate their sensitivity, and to restrict actions such as joining them with datasets bearing other markings that may be risky in combination.
- **Obfuscation by Default:** Making data encrypted and unreadable by default. Users must enter an acceptable justification in order to decrypt necessary subsets of the data.
- **Auditing:** Empowering oversight bodies to check and verify compliance with data governance policies around de-identified data, and that no spurious, malicious, or risky actions are undertaken.
- **“Inferring” Sensitive Data:** Running background checks to infer sensitive data across the system, automatically flagging and locking down sensitive data uploaded accidentally or de-identified insufficiently.
- **Testing and Validation:** Providing the ability to do validations and “battle-test” anonymized data before it is shared more widely within the system or exported for external use.
- **Data Lineage:** Leveraging lineage tracking to understand how data is flowing within the system: which users have access to what level of identifiable data, and for what purposes at different stages.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

Broadly, PETs can benefit a broad range of sectors, applications, and analyses by promoting consumer confidence and public trust. The most effective communications regarding relatively complex and niche areas like emerging PETs must directly and simply address people’s basic concerns about the use of their data.

This means understanding what people’s concerns are, explaining how a technology addresses these concerns, providing an approachable overview of how the tech works, what the outcomes are, and why these outcomes are desirable. This discussion of outcomes is integral to the broader adoption of PETs because it demonstrates their value beyond the privacy community. Namely, PETs are critically important because they lead to protections in contexts that carry the most risk or harms, particularly in the realm of public health and medical research.

Given the potential complexity of this subject, concise, easily understood, easy-to-consume communications are important: illustrations and video demonstrations can help in this regard. These should be posted somewhere prominent and easily available, and where relevant, should be displayed or shared by projects leveraging these PETs. This must be part of an ongoing cycle—trust in any one technology is contingent on a broader sense that the developer behind the technology is trustworthy, which requires demonstrating a continuous investment in thoughtfulness and a responsible approach to technology. These communications will be particularly effective if they are developed or released in collaboration with an already highly trusted individual or institution.

For example, scientific and health/medical researchers have demonstrated the value of leveraging a centralized research platform to manage data and its use in a secure and traceable environment. At the National Institutes of Health (NIH), data scientists, clinicians, and researchers use commercial software for organization-wide data, aggregation, harmonization, sharing, and knowledge management; data quality improvements; the use of imaging and genomics data for patient care; informatics workflows; AI/ML model management and performance monitoring; and analysis of high-throughput screening data. Organizations benefit from using commercial software because it is:

- **Quickly Deployable:** Commercial software platforms often require minimal configuration out-of-the-box, allowing rapid deployment of the platform to meet NIH needs in days or weeks instead of months or years.

- **Highly Scalable:** Commercial software is typically backed by the latest in dynamic scaling technologies, allowing it to scale in lockstep with user, compute or data size—preventing latency issues and workflow restrictions.
- **Easily Maintainable and Highly Secure:** Under the software as a service (SaaS) continuous delivery model, cloud-based software instances receive frequent updates and security patches. This software is often highly (re)configurable to reflect changing needs—and doesn't require large-scale reinvestment for its sustainability.

The benefit of leveraging commercial software to implement and enforce PETs is demonstrated by NIH's National COVID Cohort Collaborative (N3C). At the beginning of the COVID-19 pandemic, the National Center for Advancing Translational Sciences (NCATS) and the Center for Data to Health (CD2H) team required the ability to securely and rapidly integrate, harmonize, and make available clinical data relating to the COVID-19 outbreak in the United States. With a diverse and complex clinical data landscape, NCATS required a flexible data infrastructure that could integrate previously siloed data sources (e.g., Electronic Health Records (EHR)) as well as open source data (e.g., Social Determinants of Health data, US demographic data, etc.). Prior to the creation of N3C, technical barriers to data ingestion, harmonization, and secure sharing capabilities slowed critical collaboration. Furthermore, many entities had access to partial views of the current situation from their own limited EHR; no single entity had access to a comprehensive view of COVID-19 patient data needed for scientific research.

To enable scientific research of COVID-19 patient data, NCATS and CD2H partnered with Palantir to configure its commercially available software solution to support the N3C data enclave. N3C provides a secure, national resource of EHR and related data that can serve as a foundation and approach for future multi-stakeholder, multi-system medical research efforts. N3C is highly secure and governable, national research data infrastructure. N3C provides an expanding data and research asset comprised of harmonized data from 5.5 million COVID-positive patients and 815 million visits across 74 different provider sites. The research enclave enables team science among a community of more than 2,600 researchers from 280 institutions; a collaborative approach resulting in the publication of more than 37 peer-reviewed research projects in the world's leading medical journals.

As an open and interoperable data infrastructure, N3C enables researchers to overcome the complexities of COVID-19 data heterogeneity and sensitivity. Through open APIs and out-of-the-box connectors, N3C allows researchers to share and leverage lab and other health data for COVID-19-related research in open and interoperable formats. By using open source languages, researchers can also publish their work from N3C to GitHub. N3C impact includes:

- **Collaborative Science and Research.** To date, there are approximately 400 research projects in N3C run by more than 2,600 researchers. This includes projects to identify the efficacy of repurposed drugs for COVID-19, investigate how social determinants of health (SDoH) affect real-world outcomes for COVID-19 patients, and characterize the pharmacoepidemiology of COVID-19. These projects and many others are made uniquely possible by N3C.
- **Privacy Preserving Record Linkage (PPRL).** The commercial software underlying N3C can integrate directly with PPRL PETs to securely link de-identified data across data sources. For example, N3C has used PPRL to link clinical data from EHR records with additional information such as images (TCIA), viral variants information (NCBI), mortality data, and Medicare claims. This provides numerous benefits: Analysis of x-ray images provides deeper insight into the impact of the disease on the lungs, variant data enables understanding of the clinical differences caused by different variants, mortality data promotes an accurate picture of patient outcomes, and claims data provides a complete picture of the medicines a patient is being prescribed. This enables a richer understanding of the disease while preserving the privacy of patients.
- **N3C External Dashboard.** With open APIs, N3C populates an external dashboard with data suitable for public consumption and displays that data in an accessible web portal. In this way, N3C fosters data transparency with the public to promote understanding and spur innovation and

scientific discovery.

- **Interoperability with Observational Health Data Sciences and Informatics (OHDSI) Ecosystem.** With JSON definitions that are used by OHDSI, researchers can input concept sets and cohort definitions that they defined in the ATLAS tool into N3C and then in turn export those created in N3C to other OHDSI environments.

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs

PETs can be useful tools for engendering trust in data privacy considerations required to support collaborative consortia and multi-party deployments/partnerships. However, additional technical and governance frameworks should be established to ensure risk averse institutions are confident in the efficacy of PETs and the underlying infrastructure for collaborative or other data-sharing environments. Fundamental features should include:

- **Purpose-based Access Controls:** As discussed in more detail in response to Question 6, purpose-based access controls play a critical role in fundamental trust in PETs as well as environments where PETs should be applied to protect data and promote participation.
- **Collaboration functions, to include branching and version control:** A shared environment should include clear provenance and traceability, including the ability to branch data and code as well as maintain clear version control.
- **Data integration and compounding knowledge:** Participating organizations know that they will benefit from large-scale, PET-protected data and access to compounding knowledge.
- **User friendly environment:** To ensure the broadest possible participation and utilization of PETs within a shared environment, PETs should be applicable in an intuitive manner. Any shared capability should include user-friendly security, compliance, and audit features for platform administrators.

Additionally, clarifications to regulatory ambiguities would greatly benefit the PETs community. For example, OSTP/NITRD could facilitate agency-specific or broader initiatives to define the preferred, context-specific technical means and operating procedures for the deletion of sensitive data¹. Information blocking poses another opportunity for meaningful clarification.

As OSTP/NITRD is aware, information blocking is a practice by a health IT developer, health information network, health information exchange, or health care provider that is “likely to interfere with access, exchange, or use of electronic health information (EHI),” except in exceptions as identified by HHS. “EHI,” however, is defined as “electronic protected health information (ePHI) to the extent that it would be included in a designated record set, regardless of whether the group of records are used or maintained by or for a covered entity.” To qualify as PHI, such information must:

1. Identify (or reasonably could be used to identify) an individual;
2. Relate to past, present, or future physical or mental health conditions of an individual, the provision of health care to an individual, or payment for care, and;
3. Be maintained or transmitted in any form of media.

This definition presents a challenge for the use of data as it relates to some PETs, since PETs may be used to fundamentally change the nature of data to remove its PHI characteristics. As such, some PETs could render information blocking rules inapplicable. In a PET environment, the obligation for entities to share EHI could evaporate, reversing the important progress the ONC Cures Act Final Rule has (and will continue) to propel. We recommend expanding EHR Conditions of Certification (CoC) to adapt to the implementation of PETs, in whatever technical capacity OSTP/NITRD sees fit within the current

¹ An explanation of our approach to “Designing for Deletion” is below:
<https://blog.palantir.com/designing-for-deletion-palantir-explained-6-adfe25fda810>

regulatory landscape for information sharing/blocking, which does not adequately account for the adoption and use of PETs.

In addition, the future of data analytics and PETs would benefit from expansion of information blocking regulations to cover sharing information that is not EHI. For example, it is unclear how an organization would train a machine learning model using federated learning on data accessible only through Fast Healthcare Interoperability Resources (FHIR)/US Core Application Programming Interfaces (APIs). API standards specified in 45 CFR § 170.215 were simply not designed with PETs in mind. Sharing broader data sets, even when not entirely necessary given the use case, can vastly increase the impact of analysis, and the power of that analysis to improve programmatic and organizational effectiveness.

5. Specific laws that could be used, modified, or introduced to advance PETs

Due to the high up-front cost of PET implementation, the public's tenuous perception of organizations' ability to protect privacy, and the limits of industry self-regulating, the legislative branch would be well-advised to enact policies to incentivize the implementation of PETs across key industries. Broadly, this could include:

- Establishing funding for PET prototyping and subsidies for the implementation of PETs in the next LHHS appropriations bill.
 - OSTP/NITRD could consider a budgetary authority request to fund an incentive program that encourages PET adoption. Such an incentive would not only speed the adoption of PETs, but it would advance many principles of data exchange outlined elsewhere in the U.S. Government's data policy work, for example in the health context with the Department of Health and Human Services (HHS) Trusted Exchange Framework and Common Agreement (TEFCA):
 - **Safe Harbors:** The adoption of PETs by entities liable for inappropriate information disclosure and/or use should be incentivized through the expansion of existing safe harbor regulations (such as HIPAA Safe Harbor) and the establishment of new safe harbor regulations to cover the use of PETs.
 - **Standardization:** Ensures that sensitive data will be subject to the same, and highest, level of protection that does not vary based on the sophistication of the agency establishing the PET.
 - **Openness and Transparency:** Offers clear standards to the community as to the types of security measures adopted within the PET.
 - **Cooperation and Non-Discrimination:** Establishes standard rules around access to PETs, ensuring organizations of all sizes can benefit and removing the barrier of establishing (or independently assessing) the strength of an independent PET.
 - **Privacy, Security, and Safety:** Improves access to anonymized data for analytical/machine learning purposes to occur within those secure data enclaves.
 - **Equity:** Levels the playing field by enabling organizations of various sizes and technical skillsets to securely access and analyze the data, which can also lead to important health equity data insights and improvements.
 - **Public Health:** Accelerates the opportunities for data analysis to address health disparities and public health issues.
- The inclusion of PET implementation as a condition of maintaining certification for health information technology under 42 U.S.C. § 300jj-11(c)(5). This could be included, for example, as part of a new iteration of the Cures Act (21st Century Cures Act/H.R. 34).

While information blocking regulations help unlock access to electronic health information (EHI), PETs can enable access to health data without the disclosure of protected health information (PHI)/personally identifiable information (PII), thereby enhancing patient privacy protections. As set forth in 45 C.F.R. Part 171, the Information Blocking Final Rule prohibits Actors from undertaking any practice likely to

interfere with, prevent, or materially discourage access to, exchange of, or use of EHI. Under these regulations:

- There is no actual obligation to share data once it has been de-identified (i.e., once it is no longer EHI).
- There is an obligation to share the data when it is EHI (i.e., when it is still identifiable) as and when appropriate (per relevant statutes and regulations).
- The API standards specified in 45 CFR § 170.215 were not designed, nor are they being implemented, with PETs in mind. For example, it is unclear how organizations could train an ML model using federated learning on data accessible via FHIR/US Core APIs. Instead, organizations are incentivized to share EHI (which often incurs liability for covered entities (CEs)) even when unnecessary given the applicability and availability of PETs.

Given the potential of PETs to mitigate concerns related to EHI, PHI, and PII, PETs would benefit from changes to the existing body of legislation, regulation, and policy by:

- Expanding the EHR Chain of Custody (CoC) to cover PETs implementation.
- Expanding existing Information Blocking regulations to the sharing of non-EHI information. The scope of non-EHI information will vary depending on the specific PET and should be established clearly in regulations and guidance.

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

While much effort and attention has recently been directed towards de-centralized tools for addressing privacy risks, it is important to acknowledge that, for both technical and institutional reasons, this class of PETs has fundamental limitations. There is and will remain a critical role for centralized information and analytics environments for the foreseeable future. Centralized software suites, therefore, can be a powerful tool to increase the efficacy of data sharing and analytics to benefit individuals and society. Software can be used to effectively implement policy, enable partnerships, or create collaborative research workspaces. For example, OSTP/NITRD may recommend reliance on templated data sharing agreements outlining risk mitigation and mutual responsibilities. Risk mitigation and mutual responsibilities can be applied to the mechanism of data sharing itself through the Governance and technical infrastructure by leveraging a central data sharing platform to track provenance and use. Trust in these programs (e.g., a data sharing agreement or partnership between public and private organizations), and the tools backing them, requires capabilities to enforce security measures and provide provenance. One such mechanism is the configuration and implementation of purpose-based access controls. Controlling data, and tracking its use across a platform, is a powerful mechanism to build trust and enable better outcomes.

This creates an acute challenge for data governance teams. Tracking who has access to what information and why, across thousands of datasets and thousands of users, quickly becomes exceptionally complex.

This challenge grows exponentially with organizational scale. As the number of access requests grows, so too does the number of potential failures. Auditing decisions to grant access is difficult if the access requests were made by telephone, email, or even in person. Data governance teams often rely on more technical colleagues to grant access to the data itself, and this can make it hard for the data governance team to check whether their decision has been appropriately enforced.

To mitigate these challenges as data scale and use grows, organizations require a close integration between data governance process and access control system. Purpose-based access controls aim to:

- Introduce structure and clarity to data access decisions.
- Capture missing context and make it available to the people who need it.
- Build intuitive tooling for non-technical data governance teams to enforce rules.

Instead of applying for access to an individual data set, a potential user applies for access to a purpose. The purpose is set by data governance teams to contain data specifically scoped to help the user meet their goal—no more, no less. Every user must apply to a purpose, and they only have access to the data that’s been assigned to that purpose.

Data governance teams must record a rationale for their decision at the same time they grant a user access to data. Likewise, data owners must record a rationale when they approve the use of a data set for a purpose. Recording these justifications prompts both sides to continually consider the necessity and proportionality of their decisions. The output of that assessment can be captured in commercial software, making it available to data governance teams for review.

At any point, an auditor can understand not just who has access to what data, but also why they were given access—with all the context that went into that decision.

7. Risks related to PETs adoption

Perhaps the greatest risk to PETs adoption is inevitable disappointment and distrust that is bound to arise when PETs are touted as silver-bullet solutions to privacy challenges, but fail to deliver fully or as promised. Part of this risk is a consequence of privacy literature and technology landscape littered with artifacts of a seemingly binary categorization of data: anonymized and non-anonymized (or raw) data. The underlying presumption being that anonymized data, by being stripped of designated identifiers, has been sufficiently cleansed so as to wholly eliminate the risk of direct (personal) attribution. The problem with this binary notion is that data can almost never be fully anonymized (there will always be some residual re-attribution risk in the face of sufficiently motivated and well-resourced adversary) and the degree to which one pursues complete anonymization often will come at the price of diminished utility of the data.

We therefore recommend in discussions of PETs a terminological movement away from the use of “anonymization” (and even slightly more nuanced concepts like “pseudonymization”) and instead suggest a focus on “de-identification” as a concept that captures the spectral rather than binary nature of privacy risks. Focusing on de-identification enables a more clear-eyed view of a range of techniques and spectrum of attendant impacts in reducing (but not necessarily eliminating) re-identification risks. In this framing, the following class of tools and techniques might be better conceptualized and characterized as methods for minimizing re-identification risk to varying degrees and with corresponding benefits and drawbacks:

- **Generalization:** Reducing the granularity of information (e.g., converting Date of Birth to Age or Age Range).
- **Aggregation:** Grouping data about individuals together and continuing analysis at the aggregate level.
- **Obfuscation:** Hiding or disguising identifying data to unauthorized parties, perhaps by masking or encryption.
- **Dynamic Minimization:** Showing only parts of the data depending on the needs or role of the user.
- **Synthetic Data:** Producing artificial data that replicates important underlying trends in the original data.

8. Existing best practices that are helpful for PETs adoption

Certain questions can be applied to guide the adoption of relevant, PETs implementation best practices. These questions help organizations establish PETs beyond a multitude of various de-identification techniques, most of which are subject to a tradeoff between optimizing data utility and minimizing re-identification risk. To make data less identifiable, organizations should consider:

- **How sensitive is the data?** There are many ways data could be sensitive: it could contain information on protected characteristics such as health, gender, or ethnicity; or it may be in some

other sense intimate, personal, or confidential. A related question to ask is “What would the potential harm to these individuals be if this information were re-identified?”

- **How easy is it to re-identify the data?** To answer this question, consider how unique the individual data point is, i.e., to how many individuals it could apply. The fewer people to whom it can apply, the higher the risk of re-identification.
- **What happens if it’s joined with other data?** Consider the other data in your system, now and in the future. Could that data, joined with the de-identified data, result in re-identification? How likely is such a join to occur (in the system or if the data is published elsewhere)? What protections are in place to guard against it?

Organizations must also consider their internal accountability, oversight, and governance structures, and how these structural artifacts could either increase or decrease the likelihood of re-identification. Namely, organizations should consider impacts across four categories:

- **Users:** How many users will have access to this data? How will this change over time? Risk grows with each user who gains access. This is particularly important if users might be motivated to try re-identifying the data, perhaps to learn about public figures or people they know.
- **Permissions:** How much data can users access? What other data can they access (i.e., outside of the platform where they access the de-identified data), and could this be combined with the de-identified data? Do these users have permissions that would allow them to import, export, or transfer the data in unanticipated ways?
- **Policies:** Are there clear data governance policies in place, and how well does the average user understand them? Does the platform enforce these policies? Can data governance teams monitor and measure compliance?
- **Metadata:** Are datasets within the platform clearly labelled and described, so that data governance and operational users can quickly understand their sensitivity, intended use, and the applicable policy protections?

9. Existing barriers, not covered above, to PETs adoption

The greatest challenge to the deployment of PETs comes when attempting to bridge the gap from developing purely academic innovations to deploying these innovations at scale in complex, changing, expansive real world processing platforms. Additional hurdles to implementing PETs at this step include:

- **Technical Skillset Shortages:** Some of the most critical and fundamental PETs are underutilized because they require unique technical skills, ensuring that those individuals tasked with implementing PETs and best practices are almost always under-resourced. One such example is the practice of audit log analysis: even with tools and automations to support the analysis, this can be a difficult and time consuming task. Specific challenges include inconsistent log formats, decentralized logs, and an expert knowledge requirement to perform change log analysis. Broadly, this gap in PET literacy or familiarity demonstrates how so often the individuals tasked with implementing and enforcing PETs are over-worked or under-resourced.
- **Corollary Staffing Shortages:** The implementation and enforcement of PETs often falls to oversight/governance teams. PETs introduce a paradox whereby many of the most effective PETs work best when they augment human review/oversight, but without such oversight teams in the first place, it is hard to implement PET-driven protections. This demonstrates the socio-technical nature of data governance and privacy, where both technical and human controls are required to uphold data governance effectively and in context.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Peisert, Sean

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Privacy-Preserving Data Sharing for Science and Public Policy

Dr. Sean Peisert
Senior Scientist,
Lawrence Berkeley National Laboratory

July 14, 2022

1 Introduction

Data useful to science, public policy, healthcare, and other vital functions in the national interest is not shared as much as it should or could be, particularly when that data contains sensitivities of some kind. We advocate the use of hardware *trusted execution environments (TEEs)* and *differential privacy* as means to significantly change approaches to and trust relationships involved in secure data management [1]. There are many reasons why data may not be shared, including laws and regulations related to personal privacy or national security, or because data is considered a proprietary trade secret. Examples of this include electronic health records, containing protected health information (PHI); IP addresses or data representing the locations or movements of individuals, containing personally identifiable information (PII); the properties of chemicals or materials, and more. Two drivers for this reluctance to share, which are duals of each other, are concerns of data owners about the risks of sharing sensitive data, and concerns of providers of computing systems about the risks of hosting such data. As barriers to data sharing are imposed, data-driven results are hindered, because data isn't made available and used in ways that maximize its value.

And yet, as emphasized widely in scientific communities [2, 3], by the National Academies, and via the U.S. Government's initiatives for "responsible liberation of Federal data," finding ways to make sensitive data available is vital for advancing scientific discovery, public policy, and other important functions. When data isn't shared, certain research may be prevented entirely, be significantly more costly, take much longer, or might simply not be as accurate because it is based on smaller, potentially more biased datasets.

Computing systems used for data analysis today include institutional computing resources and commercial clouds, and, for certain functions, supercomputers such as those present in high-performance computing (HPC) centers sponsored by U.S. Department of Energy's Office of Science and the U.S. National Science Foundation. Not all data analysis is large, but at the largest scale, it can be characterized by massive datasets and distributed, international efforts to analyze that data. However, when sensitive data is used, computing options available are much more limited in computing scale and access [4].

2 Limitations of Current Privacy Approaches

2.1 Current Secure Computing Environments

Today, where remote access to sensitive data is permitted at all, significant technical and procedural constraints may be put in place, such as instituting ingress / egress "airlocks," requiring "two-person" rules to move software and data in or out, and requiring the use "remote desktop" systems. Architectures like this are becoming more and more common as means for computing involving sensitive data [4]. However, even with these security protections, traditional enclaves still require implicitly trusting system administrators and anyone with physical access to the system containing the sensitive data, thereby increasing the risk to and liability of an institution for accepting responsibility for hosting data. This security limitation can

significantly weaken the trust relationships involved in sharing data, particularly when groups are large and distributed. These concerns can be partially mitigated by requiring data analysts to be physically present in a facility owned by the data provider in order to access data. However, in all these cases, analysis is hindered for communities – such as scientific communities — whose abilities and tools are optimized for working in open, collaborative, and distributed environments. Further, consider the pandemic in which a requirement of physical presence in a particular facility for analysis would be a significant public health risk at various times.

2.2 Reducing Data Sensitivity Using “Anonymization” Techniques

Sometimes attempts are made to avoid security requirements by making data less sensitive by applying “anonymization” processes in which data is masked or made more general. Examples of this approach remove distinctive elements from datasets such as birthdates, geographical locations, or IP network addresses. Indeed, removing 18 specific identifiers from electronic health records satisfies the HIPAA Privacy Rule’s “Safe Harbor” provisions to provide legal de-identification. However, on a technical level, these techniques have repeatedly been shown to fail to preserve privacy, typically by merging external information containing identifiable information with quasi-identifiers in the dataset to re-identify “anonymized” records [5]. Therefore, de-identification doesn’t necessarily address the risk and trust issues involved in data sharing because re-identification attacks can still result in significant embarrassment, if not legal sanctions. In addition, the same masking used in these processes also removes data that is critical to the analysis [5]. Consider public health research for which the last two digits of a zip code, or the two least significant figures of a geographic coordinate are vital to tracking viral spread.

3 Confidential Computing

Hardware TEEs can form the basis for platforms that provide strong security benefits while maintaining computational performance [6]. TEEs are portions of certain modern microprocessors that enforce strong separation from other processes on the CPU, and some can even encrypt memory and computation. Common commercial TEEs today include ARM’s TrustZone, Intel’s Secure Guard Extensions (SGX), and AMD’s Secure Encrypted Virtualization (SEV). All three vendors take extremely different approaches and have extremely different strengths, weaknesses, use cases, and threat models.

TEEs can be used to maintain or even increase security over traditional enclaves, at minimal cost to performance in comparison to computing over plaintext. TEEs can isolate computation, preventing even system administrators of the machine in which the computation is running from observing the computation or data being used or generated in the computation, including even from certain “physical attacks” against the computing system. They can implement similar functionality as software-based homomorphic and multiparty computation [7] approaches, but without the usability issues and with dramatically smaller performance penalties.

The use of TEEs to protect against untrustworthy data centers is not a novel idea, as seen by the creation of the Linux Foundation’s Confidential Computing Consortium [8] and Google’s recent “Move to Secure the Cloud From Itself.” [9]. Google has comparing the importance of the use of TEEs in its cloud platform to the invention of email [10]. However, TEEs have not yet seen broad interest and adoption in data analysis, although they are now present in Amazon, Google, and Microsoft’s cloud computing platforms.

The approach we envision is to leverage TEEs when data processing environments are out of the direct control of the data owner, such as in third-party (including DOE or NSF) HPC facilities or commercial cloud environments, in order to prevent exposure of sensitive data to other users of those systems or even the administrators of those systems. Data providers can specify the configuration of the system, even if they are not directly the hosts of the computing environment, to specify access control policies, a permitted list of software or analyses that can be performed, and output policies to prevent data exfiltration by the user. The notion of being able to leverage community HPC and cloud environments also enables the use

of data from multiple providers simultaneously while protecting the raw data from all simultaneously, each potentially with their own distinct policies.

Researchers at the Berkeley Lab and UC Davis empirically evaluated Intel SGX and AMD SEV TEEs for their performance under typical HPC workloads including both traditional modeling and simulation benchmarks, ML/AI benchmarks, and real-world ML/AI applications. Our results [11] show that AMD’s SEV generally imposes minimal performance degradation for single-node computation and represents a performant solution for high-performance computing — including large-scale data applications — with lower ratios of communication to computation. Importantly, the major commercial clouds, as well as modern HPC centers, such as the the DOE’s NERSC-9, contain AMD processors that support the SEV TEE, and thus it is our hope that our results will provide some of the evidence needed to justify the use of TEEs in large-scale, data-driven computing.

4 Research Opportunities to Advance Privacy-Enhancing Technologies

4.1 Trusted Execution Environments

Although numerous commercial TEEs exist, no TEEs yet exist in processors other than CPUs, such as in GPUs and accelerators, although NVIDIA has indicated plans to expand TEEs to GPUs, and Google has indicated plans to expand TEEs to GPUs, TPUs, and FPGAs. There are also issues with low-latency communication between TEEs, and also the cost of virtualization, that must be addressed to enable secure data analysis and machine learning scale [11]. In addition, promising RISC-V efforts such as Keystone [12] exist that carry both the promise of broadening the scope of processors that contain TEEs, while also being open source and possible to formally verify. However, RISC-V based TEEs have not yet been developed that target algorithms that center around data, such as large graph workloads and machine learning. Most likely, an entirely new TEE architecture tailored for scientific computing and data analysis applications will be needed, which is a focus of Berkeley Lab’s efforts in this space [13, 14].

4.2 Differential Privacy

Output policies are another area that deserve investigation. While TEEs protect against untrusted computing providers, and can provide certain measures of protection from malicious users, output policies determine what data is returned to the user. Differential privacy [15] is a particularly interesting approach to providing strong privacy protection of data output. Differential privacy is a statistical technique that can guarantee the bounds on the amount of information about a dataset that can be leaked to a data analyst as a result of a query or computation by adding “noise” and enforcing a “privacy budget” that bounds information leakage. It is now a mainstream solution, with production use by Apple, Google, and the U.S. Census Bureau, the existence of several open source distributions, and successful application to a diverse range of data types. However, differential privacy is not appropriate everywhere, and applying it is currently challenging, requiring a high degree of expertise and effort. Thus, differential privacy is highly useful today, albeit in a limited set of situations for datasets that have sufficiently wide use to justify the time and expense required. Work is needed to advance the usability of differential privacy so it can more easily be broadly leveraged. Energy data and mobility studies are two areas that Berkeley Lab has demonstrated success in applications of differential privacy [16, 17].

5 Summary and Next Steps

In contrast to traditional secure enclaves, TEEs enable sensitive data to be leveraged without having to trust system administrators and computing providers. However, while the application of TEEs have now been widely heralded in cloud environments, they have not advanced to be performant for large-scale data

analysis, despite the significant concerns frequently expressed by both data providers and computing facilities about hosting sensitive data. But while improvements are needed to truly harness TEEs for large-scale data applications, the current generation of TEEs is here, those TEEs are available, and until we start making use of them in scientific computing, data is not shared as much as it should or could be by leveraging TEEs to address the trust issues underlying current limits on data sharing.

What is missing is a connection to the particular infrastructure used in large-scale data-driven computing, including I/O subsystems, custom workflows, highly specialized instruments, community data repositories, and so on. Therefore what is needed is a conversation between processor manufacturers, system vendors, and cloud and scientific computing operators regarding enabling the TEE functionality already present in the AMD EPYC processors — and presumably in other, future processors — into scientific computing environments, while simultaneously developing and preparing for the next generation of TEEs. However, the path forward is not solely technical. It requires the community to build infrastructure around TEE technology and integrate that infrastructure into scientific computing facilities and workflows, and into the mindset of operators of such facilities.

Acknowledgements

This response is being provided provided by Dr. Sean Peisert, a Senior Scientist at Lawrence Berkeley National Laboratory, a Federally Funded Research and Development Center (FFRDC) operated by the University of California for the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author.

References

- [1] Sean Peisert. Trustworthy Scientific Computing. *Communications of the ACM (CACM)*, 64(5):18–21, May 2021.
- [2] Justine S. Hastings, Mark Howison, Ted Lawless, John Ucles, and Preston White. Unlocking Data to Improve Public Policy. *Communications of the ACM*, 62(10):48–53, September 2019.
- [3] Jane Macfarlane. When Apps Rule the Road: The Proliferation of Navigation Apps is Causing Traffic Chaos. It’s Time to Restore Order. *IEEE Spectrum*, 56(10):22–27, 2019.
- [4] Sean Peisert. An Examination and Survey of Data Confidentiality Issues and Solutions in Academic Research Computing. Trusted CI Report — <https://escholarship.org/uc/item/7cz7m1ws>, September 2020.
- [5] Arvind Narayanan and Edward W. Felten. No Silver Bullet: De-identification Still Doesn’t Work. <http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf>, July 9, 2014.
- [6] Mark Russinovich, Manuel Costa, Cédric Fournet, David Chisnall, Antoine Delignat-Lavaud, Sylvan Clebsch, Kapil Vaswani, and Vikas Bhatia. Toward Confidential Cloud Computing: Extending Hardware-Enforced Cryptographic Protection to Data While in Use. *Queue*, 19(1):49–76, February 2021.
- [7] Joseph I Choi and Kevin RB Butler. Secure Multiparty Computation and Trusted Hardware: Examining Adoption Challenges and Opportunities. *Security and Communication Networks*, 2019(1368905), 2019.
- [8] Fahmida Y. Rashid. The Rise of Confidential Computing. *IEEE Spectrum*, 57(6):8–9, 2020.
- [9] Lily Hay Newman. Google Moves to Secure the Cloud From Itself. *Wired*, July 14, 2020.

- [10] Sunil Potti and Eyal Manor. Expanding Google Cloud’s Confidential Computing portfolio. <https://cloud.google.com/blog/products/identity-security/expanding-google-clouds-confidential-computing-portfolio>, September 8, 2020.
- [11] Ayaz Akram, Anna Giannakou, Venkatesh Akella, Jason Lowe-Power, and Sean Peisert. Performance Analysis of Scientific Computing Workloads on General Purpose TEEs. In *Proceedings of the 35th IEEE International Parallel & Distributed Processing Symposium*, 2021.
- [12] Dayeol Lee, David Kohlbrenner, Shweta Shinde, Krste Asanović, and Dawn Song. Keystone: An Open Framework for Architecting Trusted Execution Environments. In *Proceedings of the Fifteenth European Conference on Computer Systems (EuroSys)*, 2020.
- [13] Ayaz Akram, Venkatesh Akella, Sean Peisert, and Jason Lowe-Power. Enabling Design Space Exploration for RISC-V Secure Compute Environments. In *Proceedings of the Fifth Workshop on Computer Architecture Research with RISC-V (CARRV), (co-located with ISCA 2021)*, June 17, 2021.
- [14] Ayaz Akram, Venkatesh Akella, Sean Peisert, and Jason Lowe-Power. Simulating Trusted Execution Environments in gem5 (extended abstract). In *Proceedings of the Workshop on Modeling & Simulation of Systems and Applications (ModSim)*, October 6–8, 2021.
- [15] Cynthia Dwork. Differential Privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, part II (ICALP)*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12, July 2006.
- [16] Nikhil Ravi, Anna Scaglione, Sachin Kadam, Reinhard Gentz, Sean Peisert, Brent Lunghino, Emmanuel Levijarvi, and Aram Shumavon. Differentially Private K-means Clustering Applied to Meter Data Analysis and Synthesis. *IEEE Transactions on Smart Grid*, 2022.
- [17] Ammar Haydari, Michael Zhang, Chen-Nee Chuah, Jane Macfarlane, and Sean Peisert. Adaptive Differential Privacy Mechanism for Aggregated Mobility Dataset. arXiv preprint 2112.08487, 2021.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Prifina Inc

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



**Response to the notice of request for information on
advancing privacy-enhancing technologies
(87 FR 35250, Doc. Nr. 2022-12432)**

To

*Office of Science and
Technology Policy (OSTP)*

July 8, 2022

On behalf of

Prifina, Inc.

Address

1250 31st Avenue
San Francisco, CA
94122 USA

Email

policy@prifina.com

To whom it may concern,

Prifina Inc. is pleased to have the opportunity to respond to the request for information on advancing privacy-enhancing technologies. We are grateful for making it possible for various interested parties to express their views on this highly significant and timely area of innovation.

Prifina appreciates the work that the Office of Science and Technology is doing. The privacy-enhancing technologies have been developed for quite a while, and we are reaching the tipping point when such PETs will be available not only for legacy technology giants but also to benefit individual consumers.

The lack of a uniform data and privacy framework on the federal level is only one factor that contributes to the lack of certainty about how companies handle user-generated data. In our response, we propose that privacy-enhancing technologies in general, and the user-held data model in particular, provide a viable long-term solution to ensuring that user-generated data is utilized responsibly and ethically. Without major regulatory interventions, the user-held data model helps achieve many policy objectives - individual autonomy, data minimization, lawful and transparent data usage by third parties, data portability, and interoperability. We also propose steps that the executive branch could take to advance the adoption of PETs.

We admire that the Office of Science and Technology Policy has taken a firm stance to advance privacy-enhancing technologies and protect consumers' data privacy rights while maintaining the fair and efficient functioning of the digital market.

Should you have any questions, please do not hesitate to contact us.

Sincerely yours,

Markus Lampinen, Jouko Ahvenainen and Paul Jurcys

About Prifina

Prifina is a VC-backed company building a new data architecture where individuals own and control their data ("user-held data model"). This type of data architecture enables new use cases and personal applications to be designed and built on top of user-held data.

Prifina's mission is to create an environment where individual users can get daily value from their personal data and where developers can build applications that help generate value from such data. We believe that personal and user-generated data can help individuals live healthier and happier lives and that personal data has a long-term value to individuals.

The starting premise of Prifina's user-held data model is the ability of each individual to collect their data from various data sources (wearable devices such as smartwatches or smart rings, online accounts, paper documents, etc.) into their "personal data clouds." Every personal data cloud is supported by a dynamic software layer that cleans and organizes the data format and makes data efficiently utilizable by apps. By default, only the user can access data in the personal data cloud; third parties cannot access any data unless the user gives prior express authorization. Prifina's user-held data model is user-centric: the user has exclusive and ultimate agency and control over the data held in the personal data cloud. Furthermore, in the user-held data environment, individuals are also legal owners of their user-held data (i.e., the data in each user's personal data cloud).

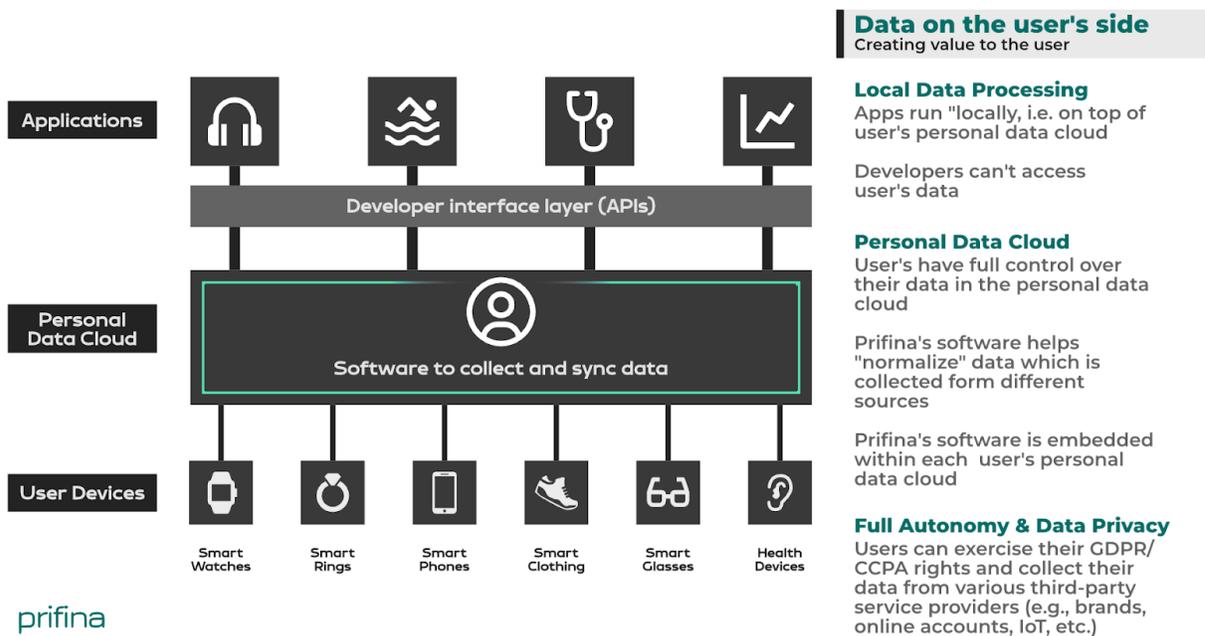
The "user-held data model" opens new opportunities for generating value from personal and user-generated data. Prifina is developing an intelligent data layer that helps normalize the data that is collected in users' personal data clouds. Using Prifina's resources and tools, developers can easily add new data sources and build new applications that run on top of user-held data (i.e., apps run locally in each user's personal data cloud). As a result, the value from user-held data is captured on the user's side.

The user-held data model has two important implications: First, users can better understand the depth and breadth of their data and have full ownership and control over it. Secondly, the user-held data model separates data from the applications.

This user-centric, user-held data approach is in line with the general principles of data privacy laws: that data is being used only with the user's prior consent, data minimization (here, the service provider does not have to hold any data on its own servers), transparency, purpose limitation (that data is used only for clearly defined purpose), data security and data portability.

The user-held data model opens new perspectives concerning the portability of personal and user-generated data. Rather than data being "ported" from service provider A to service providers B and C, service providers come to every user through new applications that run in users' personal data clouds. This means that service providers can better understand their potential customers and avoid huge risks associated with holding customer data on their own servers.

The user-held data model offers compelling technological architecture and multi-stakeholder incentives to realize the Commission's vision to build a new data ecosystem based on human-centric data values. This data model inspires people to think about "activating" data to unlock the value from data for individuals and developers/businesses and open the gates to building user-centric data apps for "smart citizens." Furthermore, the user-centric data model will likely become one of the possible technological solutions for utilizing user-generated data for research and reaching sustainability goals.



1. Specific research opportunities to advance PETs

The user-held data model offers opportunities to make more user-centric data services. It means that the service can include richer data on an individual, combine data from many sources and then provide tailored services for the individual. Adopting such an approach to data requires new models to analyze data and train Machine Learning models. The user-held data model offers many new research opportunities, from technology to data science and business models.

Building on a user-specific data set, there is the opportunity to research individual baselines and establish different ways of preventing and handling deviations. For example, establishing a baseline for an individual's sleep patterns can help prevent deviations such as sleep disruptions, circadian rhythm disruptions, handling shift work, and many other situations in everyday life based on personalized data sets.

However, centralized (or "siloed") data models prevail today. Such centralized data models contain data sets that are very rich on an *aggregate* level, but narrow as it comes to individuals. In practice, centralized data models mean that while there may be many individual data points about the specific question, such as sleep data, the other factors, such as contextual data, behavior, activities, interests, etc. are lost from that aggregated data set. This is so because the privacy concerns and the size limitations of a centralized data set constrain what is possible to collect.

An individualized dataset, such as the one proposed by Prifina, unlocks this limitation because the individual's dataset can be incredibly rich and deep.

2. Specific technical aspects or limitations of PETs

The user-held data model is very different from the traditional big data approach. The user-held data model means changes and improvements for the technology, data models, analytics, and applications. For example:

1. Data storage and cloud solutions must enable individuals to store their own data and run applications so that the data never leaves the user's environment.
2. Analytics is based on rich data on an individual rather than sparse data from many individuals. This requires new types of algorithms and models to analyze data and teach machine learning models (e.g., federated learning type models).

3. If an individual wants to share some data, new cryptographic models are needed to limit the use of the data only for the purposes it was shared.
4. In many cases, bringing the processing near an individual would be better. This requires an edge computing type of technology and models to store and process the data.

From a technical perspective, the main challenges in building the new user-centric data ecosystem relate to the fact that (i) data is locked in silos and that there and (ii) there are no common/unified data layers. These two factors are among the most challenging obstacles that need to be overcome to facilitate the flow of data.

Further references:

- **Paulius Jurcys, et al., 'My Data, My Terms: A Proposal for Personal Data Use Licenses'**, *Harvard Journal of Law and Technology Digest* (2020), available at: <https://jolt.law.harvard.edu/digest/my-data-my-terms>

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

Health and well-being are becoming more important all the time for individuals and societies. Digital healthcare and wellbeing services offer significant opportunities to improve personal health and well-being and make healthcare more cost-effective. However, privacy and data protection are crucial with health data and health services. Responsible use of personal health data and wellness data has become even more critical after the Supreme Court's ruling in *Dobbs*.

The user-held data model with personal applications that can be run on a person's own data offers significant new opportunities to improve health and wellbeing services. At the same time, it is possible to get more data from individuals, for example, wearable data that already now has information on heart rate, exercising, and sleeping. New sensors enable measuring e.g., blood pressure, blood glucose, lactate, or inflammations. With the user-held data model, individuals can also share data with healthcare providers, making it easy for individuals to use those services they wish and always share necessary data.

The user-held data model can be adopted in various other sectors and verticals: IoT, geo-location and mobility, financial, insurance, smart cities, GovTech, etc. The key objective here is putting the individual in the middle of digital interactions, empowering them to *opt-in* to various services (rather than placing a hefty burden

on individuals to opt-out), and, most importantly, making sure that the value from user-generated data is captured on the user side.

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs

The impact of currently existing laws and regulations on the federal and state levels is relatively limited. Practically, individuals have very little control over how organizations access and use individuals' personal and user-generated data. While the adoption of the federal privacy act is pending, government agencies could take steps to adopt "soft" measures that could facilitate a more transparent and responsible use of personal and user-generated data (see section 6 below).

Clarification of data access/ownership rights.

From a legal and technical perspective, it is important to ascertain that individuals have unlimited access to the data about themselves regardless of where it is held.

The Use of Biometric Data.

The emergence of virtual work and play environments (including "metaverses") raises many questions about the utility of user-generated and biometric data. The user-held data model could prove helpful because it allows individuals to control their avatar data. This, of course, raises various issues related to data portability and interoperability.

The use of non-clinical data in health care.

As health-monitoring sensors become widespread, various questions emerge about the utility of such user-generated wearable data in health care. Currently, wearable device makers collect vast amounts of user-generated data, which is locked in the walled gardens of such devices. In the future, it would be desirable if such data could be unlocked and made available to individual users who generate such data. Homomorphic encryption could prove to be especially useful in this context.

Further references:

- **Paulius Jurcys et al., 'Ownership of User-Held Data: Why Property Law is the Right Approach', *Harvard Journal of Law and Technology Digest* (2021), available at: jolt.law.harvard.edu/digest/ownership-of-user-held-data-why-property-law-is-the-right-approach**

- **Jouko Ahvenainen, 'The metaverse is coming for your biometric and health data'** available at: medium.com/prifina/the-metaverse-is-coming-for-your-biometric-and-health-data-1d185a93519c
- **Marcelo Corrales Compagnucci, et al., 'Homomorphic Encryption: The 'Holy Grail' for Big Data Analytics and Legal Compliance in the Pharmaceutical and Healthcare Sector?', 3(4) European Pharmaceutical Law Review (2019) available at: <https://eplr.lexxion.eu/article/EPLR/2019/4/5>**

5. Specific laws that could be used, modified, or introduced to advance PETs

Federal privacy law.

It would be most desirable if the US Congress could adopt a federal privacy act that is based on user-centric principles (data minimization, data portability, data interoperability, transparency, consent, and user-held data model). Such a uniform law would significantly curtail compliance costs and put the US on equal footing with other countries and regions (EU).

Solving cross-border data transfers with the user-held data model.

The user-held data model could be incredibly impactful in solving cross-border data transfer problems.

Learning from the EU.

In the past few years, the EU has adopted several regulations that aim to unlock data from silos and address some of the most complex issues concerning data use and AI. The US could consider learning from the EU and aligning the US regulatory landscape with the EU.

Further references:

- **Paulius Jurcys, et al., 'The Future of International Data Transfers: Managing New Legal Risk with a 'User-Held' Data Model' *The Computer Law and Security Review* (2022, forthcoming), available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4010356**

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

Focusing on User-Centric Solutions. It is difficult to get privacy to work appropriately only by restricting the use of data by businesses and creating more regulations for technology. As emphasized above, data is also y valuable for individuals. The real long-term solutions should be based on models that empower individuals to manage and utilize their own data. In the past decades, companies have built successful businesses on data. As a result, five data giants dominate the digital space. However, solutions for individuals to make their life better and healthier based on personal data are still in a very early phase. This area offers many opportunities for innovations, economic growth, and improving the lives of individuals.

The best way to support this development is to offer individuals more rights and tools to access such user-generated data, create frameworks that help them own such data, and benefit from services and applications to utilize such user-held data.

Possible actions for government agencies:

- Create incentives for direct-to-consumer sellers and service providers to adopt privacy labels explaining how they use their customer data;
- Promote user interface and user experience frameworks that are based on **opt-in** (rather than opt-out) approaches;
- Create **incentives** for data market participants to develop solutions that facilitate access to data. This could include API protocols, **open-source standards** for data portability, and data interoperability;
- Create programs to facilitate collaboration between academic organizations, think tanks, and market participants to conduct studies relying on **behavioral economics**. The end result could be to figure out areas where nudges could be useful to improve user experience, enhance transparency and nudge people to control better how their data is used;
- Consider areas where the government introduces **regulatory sandboxes** to validate the use of PETs.

References:

- **Paulius Jurcys et al., 'My Data, My Terms: A Proposal for Personal Data Use Licenses'**, *Harvard Journal of Law and Technology Digest* (2020), available at: <https://jolt.law.harvard.edu/digest/my-data-my-terms>

7. Risks related to PETs adoption

One of the dimensions of unintended consequences relates to rule-making: adopting rules in rapidly changing technological environments is challenging. The regulator needs to balance the necessary freedom for developing new privacy-enhancing technologies and complex compliance requirements that could restrict innovation. Imposing specific rules on the use of data could likely lead to the concentration of data in the hands of several organizations. The end result could be that individuals will not be able to get better services based on their own user-generated data.

Further references:

- **Mark Fenwick and Paulius Jurcys, 'From Cyborgs to Quantified Selves: Augmenting Privacy Rights with User-Centric Technology and Design'**, 13(1) JIPITEC (2022) available at: <https://www.jipitec.eu/issues/jipitec-13-1-2022/5512>

8. Existing best practices that are helpful for PETs adoption

Wearables can be a testament to the opportunity for more direct-to-individual applications. Some wearables companies (such as Ōura) could be an excellent example of how their products and applications could improve individual lives, personal health, and well-being through data. At Prifina, we aim to create an environment where anyone can build applications on top of user-held data and bring value closer to the user.

Further references:

- **Jouko Ahvenainen, 'People are interested in protecting personal data, although some deny it'** available at: medium.com/prifina/people-are-interested-in-protecting-personal-data-although-some-deny-it-fc35977a712f

- **Jouko Ahvenainen, 'Today's wearables are a bit like 1980s computers and that's not great'**, available at: medium.com/prifina/todays-wearables-are-a-bit-like-1980s-computers-and-that-s-not-great-771afd58d2ca

9. Existing barriers, not covered above, to PETs adoption

In the current centralized/siloed data ecosystem, companies rely on product-centric business models; there is a lack of incentives to create user-centric services, i.e., services that offer value on a more individualized level. Currently, enterprises collect and generate the most value from personal data for their own corporate interests. Turning this around to benefit individuals more requires a collective change of mindset and incentives. It could be further incentivized through policy changes, making it easier for individuals to gather this data and shifting the reputational risk from building data monopolies for companies.

Further references:

- **Jouko Ahvenainen, 'Who Protects Whom In The Battle For Personal Data And Privacy? You?'** available at: medium.com/prifina/who-protects-whom-in-the-battle-for-personal-data-and-privacy-you-c267b5a48200
- **Jouko Ahvenainen, 'Turn it around — could businesses sell data to consumers?'** available at: medium.com/prifina/turn-it-around-could-businesses-sell-data-to-consumers-705382cccc1

10. Other information that is relevant to the adoption of PETs

The Prifina team has published a series of articles about the benefits of adopting the user-held model. Shorter articles are available at www.medium.com/prifina.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Privacy Analytics

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

*Safe Data Enablement for Health Services & Research through Privacy-Enhancing Data Sharing and Analytics: A Spectrum of Perspectives**Luk Arbuckle, Chief Methodologist and Privacy Officer**Jordan Collins, Data Privacy Solutions Business Leader*

We would like to thank the Office of Science and Technology Policy for this opportunity to provide an industry perspective and help inform the development of a national strategy on privacy-enhancing data sharing and analytics, along with associated policy initiatives (as published in the Federal Register on 06/09/2022 and available online at [federalregister.gov/d/2022-12432](https://www.federalregister.gov/d/2022-12432), and on [govinfo.gov](https://www.govinfo.gov)). We believe that public consultation is an important step in producing an informed strategy that is more likely to produce solutions that are scalable and proportionate to the intended uses cases and needs.

Abstract

Since 2007, Privacy Analytics has been providing services and software in privacy-enhancing data sharing and analytics for organisations in the consumer and healthcare industries. We are particularly interested in safe data enablement for health services & research. From our 15 years of experience working in this space, we agree that there is a tremendous opportunity to use data safely and responsibly to the benefit of people and society. We are pleased that the request for information acknowledges the need to consider operational challenges and incentives to adoption.

In this response, we provide an overview of the landscape of tools considered for privacy-enhancing data sharing and analytics with a view towards the needs and perspectives of different stakeholders involved in health services improvement and health research for the full lifecycle of data. While there are many such tools, we focus our attention on the challenges of integration and interoperability, especially for complex health data and analytical pipelines, and the impact on end users driving to improve health outcomes. We believe that, in order to drive the adoption of safe, useful, and timely data and analytics at scale, there is an increasing need for the integration and deployment of suits of tools that are interoperable and complimentary.

Introduction

Health data is often described as some of the most sensitive since it deals with the intimate details of a person's body and mind. The information itself exists before it is measured, inferred, or assessed through various machines, tests, or questions to patients, eventually recorded in a form that may be used in data sharing or analytics. Once captured and collected, health data is continuously updated, transformed, harmonized and restructured to meet the myriad needs of extracting the greatest insights from advanced statistical methods that are themselves continuously updated and improved on throughout the data lifecycle.

The importance of data sharing and analytics to drive evidence-based decision making has been highlighted by data science consortia and researchers alike. In the highly competitive and innovative fields of health research and treatment development, solutions for data sharing and analytics must contend with data that include a large number of variables, have spatial and

temporal dependence, and inconsistent or missing data (eg, due to non-response bias in surveys or complex data collection practices and linking challenges). Even static data is refactored and reharmonized to suit the various acrobatics of statistical analysis.

Take the pharmaceutical industry as an example of the complexities and challenges of working with health data. Besides requirements to maintain copies of data that support drug approvals, the data submission guidelines include thousands of variables needed to understand chemical and health interactions in the drug development process.(1) Pharmaceutical companies collect data in many different formats (eg, approximately 40% is still based on electronic capture from forms), with many steps before delivering to different functional groups in the drug development process, including thousands of mapping rules for billions of data points. Each company will have their own internal standards and constantly evolving schema that result in continuous data integration and harmonization.

It takes over a decade to develop a new medicine, at a cost in the billions of dollars with only a 0.01% chance of success for compounds in preclinical research.(2) Trial participants expect the effective use of data about them, including reuse, provided safeguards are in place. The safe reuse of trial data also reduces the burden on participants by making the best use of data already collected, accelerating research discoveries by improving access to data and analytics. Coupled with the regulatory recognition of research benefits to improve health outcomes, there is clearly a general consensus of the need and importance to make the best use of data and analytics in clinical research.(3)

In this response, we introduce a data lifecycle view and stakeholder perspectives to set up a baseline for discussion of expectations and needs. We then provide a categorization of tools and how they align with stakeholder perspectives before introducing a spectrum of tools that incorporate some of the main considerations that were established. Finally, we provide a brief view into those tools as we have seen them used in practice, culminating in a revised view of the data lifecycle with the various tools that may best fit the needs of users in that stage of the lifecycle. Our hope is that this landscaping exercise will drive a conversation so that more practical and applied methods get the attention they deserve.

Data Lifecycle

For the purposes of privacy enhancing data sharing and analytics, for sensitive data in particular, we consider a data lifecycle focused on those activities where there is a particular need for privacy considerations. Throughout this lifecycle, the goal of producing safe, useful, and timely data and analytics, often at scale, should be kept in mind so that proposed solutions are practical and more likely to be adopted in practice. The subject of legislative or regulatory incentives are out of scope for our review (eg, see the “Failure” of PETs in (4)).

We begin with a simple 5 step data lifecycle with general needs for thinking about privacy enhancing data sharing and analytics, which we will revisit after introducing a spectrum of perspectives and technologies. The intent here is to lay out where opportunities may exist for privacy technology that serve practical needs, without consideration for legislative or regulatory incentives. This list is only a summary and by no means exhaustive.

Get data

- Why and what to collect serves to define purpose and minimize the capture and collection to what is truly necessary at the very outset.

- Needs and wants are evaluated through strategic thinking to ensure the right breadth is included for future opportunities (including a full lifecycle view).
- Capture and collection is designed to ensure the right concentration of data will deliver the sharing and analytics that provide the desired strategic intent.

Link data

- Separation anxiety, in the form of stakeholder concerns with linking, is acknowledged while focusing on data capture or collection with strategic benefit (ie, demonstrating value).
- Identify use cases so that linking of data across capture and collection points have a defined purpose and are appropriately minimized (ie, demonstrating means).
- Match data subjects or insights, introducing opportunities to disassociate from the analytics (eg, prior distributions, transfer learning), based on actual needs.

(Re)Use data

- Isolation anxiety, in the form of stakeholder concerns with data enablement, is acknowledged while focusing on use and reuse with strategic benefit (ie, demonstrating value).
- Build trust from users, in the use and reuse of data (ie, primary and secondary purposes) through the principles of trustworthy data sharing and analytics.
- Drive adoption in the trustworthy use and reuse of data as standard practice and with suitable privacy defaults for data enablement.

Share data

- Win stakeholder trust in sharing useful and timely data and insights with demonstrable benefit to current and future data subjects (eg, reducing costs, collaboration).
- Safe design and delivery through trustworthy methods that safeguard what is shared and how, for the intended purposes.
- Share insights, wherever possible, as collaborators in understanding human health and improving health outcomes (ie, evidence based).

Open data

- Citizen data as a government commitment towards transparency, innovation, and accountability (eg, G8 Open Data Charter)
- Data transparency as a regulatory commitment to democratizing data, analytics, and insights for all stakeholders (eg, European Medicines Agency Policy 0070)
- Data trust as an asset management framework with independent stewards acting on behalf, and for the benefit, of a broader group stakeholders.

Stakeholder Perspectives

Having considered a data lifecycle for privacy enhancing data sharing and analytics, we will now turn our attention to different stakeholder types to understand needs and expectations. The 4 stakeholder types we consider represent an evolving set of perspectives, and the interplay between them will be important to explore the spectrum of tools we consider in the next section.

End user. We start by assuming the end user of data, responsible for making sense of data and performing analytics, believes information assets and technologies should be available to support

their functional role, and that they should be trusted. The end user is, after all, a professional with a job to do, a job that often requires specialized training and is respected in our modern world of evidence-based decision making. Consider the biostatistician, epidemiologist, or a health scientist as standard examples of end users that are pursuing improvements in efficiencies and health outcomes. The interests of the end user include:

- Access to data that is timely and useful to meet their needs and objectives
- Data clean-up and preparation so that they can perform their analyses
- Access to suitable, and likely familiar and preferred, analysis tools
- Flexibility to develop their own custom or tailored analysis methods

Risk-based IT Security Officer. Information assets and technologies should be adequately protected, and in any modern environment there will be a role to ensure this takes place. This can be thought of as “trust, but verify” what the end user is able to do and is doing (ie, authorized uses only). To make this scalable, guardrails are set up in advance to ensure the end user can do their job but in a safe enough way. Risk-based implies that the tolerance around security controls will be commensurate with the sensitivity of the information and expectations of what the end users can and should be doing. The interests of the risk-based IT security officer include:

- Reducing information security risks through mitigating controls and audit logs
- Selectively restricting access to what data is needed, by whom, and only when they need it
- Preventing the loss of data, whether it be accidental or intentional

Privacy Officer. Whereas a security perspective will aim protect all organizational information assets, privacy has emerged as an additional consideration focused on societal rules of behavior, decorum, and civility.⁽⁵⁾ Security plays an integral role, and the privacy perspective brings attention to how people represented in data feel regarding the uses of identifiable information about them. Ethics will often be a consideration, with the privacy officer concentrated on the applicable legal frameworks that attempt to codify these societal norms and values into an established set of principles for the responsible sharing and use of sensitive data. When privacy is also evaluated in terms of human rights, it is done in the context in which privacy issues or protections are being considered, and with regards to all other human rights, including economic, social, and cultural rights (ie, it is contextual). The interests of the privacy officer can therefore be summarized as:

- Reducing privacy risks from working with personal information and codified data
- Respecting legally established data subject rights based on jurisdiction
- Norms for trustworthy data processing, including governance and privacy technology

Zero-trust IT Security Officer. The concept of zero-trust, or trustless, security is a relatively recent IT architecture design principle in which a breach is always assumed and every request for data is verified as though it's coming from outside (ie, never trust, and always verify). This perspective has also been introduced as a technical solution to privacy, although it doesn't really address responsible uses of data. The interests of the zero-trust IT security officer include:

- Eliminating all information security risks with always on protection
- Risks from everywhere, and from everyone, assuming every interaction is a threat
- An end-to-end protection strategy with provable guarantees

As was mentioned, these represent an evolving set of perspectives that should, in theory at least, build on one another. However, we need to go back to the beginning and ask what end users think of all these added perspectives, and how they could limit their ability to achieve their goals of

delivering insights that improve efficiencies and health outcomes. While aiming to deliver trustworthy systems that address potential privacy concerns, we need to ensure that the ultimate goal of producing safe, useful, and timely data is still possible in the eyes of those professionals that derive value from working with sensitive data. Eg, do they believe their needs will be met? Will they be forced to change how they work? Is it even realistic to ask them to change how they work in practice?

Recall that the end users in health sciences include the preprocessing, or data engineering, required to prepare data for analytical uses across the data lifecycle, as well as providing the flexibility to develop novel and tailored analyses that achieve the greatest insights. This will inform the type of privacy enhancing solutions that are suitable for safe and responsible data sharing and analytics.

Spectrum of Tools

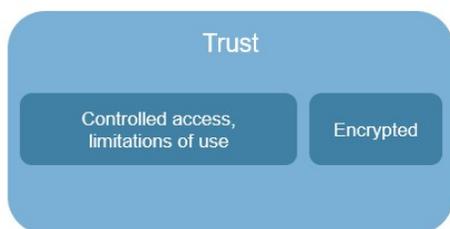
Before we consider a spectrum of privacy-enhancing technologies, we will first look at categories in which they may be considered. This will allow us to consider the different stakeholder perspectives that were introduced, and how they may perceive these tools that are intended to address their needs.

Trust Based

Traditionally a risk-based approach has been used, as described from the perspective of the risk-based IT security officer. For sensitive data, this has met the needs of the end users because it usually allows them to do any preprocessing that they require prior to doing their statistical analysis, and they can get access to the analytical software tools they need for analysis (including the most recent algorithms, and customization for more advanced modelling) From the end user's perspective, this is better termed a trust-based approach. We can therefore summarize this category of tools as:

- Trust (but verify) end users to do their job ethically and responsibly
- Provide access to authorized users/roles to minimize exposure
- Encrypt data in transit and at rest to avoid breaches

→ Confidentiality, in the form protecting information from unauthorized access, is the primary set of tools considered in this category.

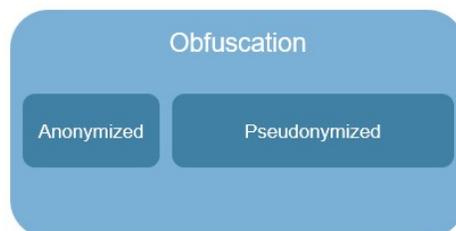


Obfuscation Based

While the trust-based tools serve an important role, primarily in protecting confidentiality, the sensitive data that is made available to end users may nonetheless raise privacy concerns. The well-established concept of data minimization can reduce some of these concerns, or more advanced methods may be used to remove the personal from data, especially for purposes other than what the data was originally collected for, known as secondary purposes. Legislation to

ensure non-identifiable information is used for secondary purposes is becoming increasingly common, although the HIPAA Privacy Rule has required this for large classes of health information since 2003.(6) We can summarize this category of tools as:

- Obfuscate to hide personal information, or confuse what the information actually contains
 - Pseudonymize to hide pieces of information with a coded replacement
 - Anonymize or de-identify to hide and confuse to create non-identifiable information
- Disassociability, in the form of removing the personal from data, based on the use case and needs to ensure the appropriate degree of useful data is still available



The use of trust-based and obfuscation-based tools is well established in health services & research because they typically meet the broad set of needs for complex data and complex data pipelines that involve advanced statistical analysis by experts.(7) Privacy concerns nonetheless exist and have resulted in another category of emerging tools.

Limited to Zero Trust

Here we begin to combine the concepts of confidentiality and obfuscation into a single category of tools. We can start with limiting trust so that the input data to statistical analyses is protected.

From a security engineering perspective, and maintaining the confidentiality of inputs, we find tools that:

- Hide input data entirely from the end user or anyone else (eg, intruders)
- Hide the computations performed on data, for the purposes of statistical analysis, entirely from the end user or anyone else (eg, intruders)

While the input data and computations are entirely hidden through means of encryption (in a broad sense), in our experience there are simple ways to unpack and reveal the underlying input data unless additional measures are introduced to prevent this from happening. The most sophisticated approach would be a database reconstruction attack, although there are much easier ways to do this when a person has access to any computational function in the applicable software library that implements such approaches.

From a privacy engineering perspective, and disassociating the input data, we find tools that:

- Obfuscate input data to hide and confuse the information, rendering it non-identifiable in the circumstances of use (eg, anonymize or de-identify)
- Provide access to the outputs of statistical analyses only, which are less granular by definition (ie, statistics are summaries)

The key advantage of using these approaches is that the data feeding into an analytics pipeline is in this case non-identifiable. That can mean the difference between having access to the input data in the first place, such as for secondary purposes in which the use of personal information may not be permitted. To ensure the most statistically useful input data is available, however, reasonable assurance that the input data is non-identifiable will in part rely on access to statistical analyses only. Concerns around the ability to unpack and reveal the underlying input data will still exist, although this is less of a concern since the input data is at least partly minimized.

For these reasons, we can also extend to limiting trust so that the inputs and outputs to statistical analyses are protected.

From a security engineering perspective, and maintaining the confidentiality of input data and statistical outputs, we find tools that:

- Do all of the previous, ie, hide input data and computations entirely
- Limit operations to a set of protocols that are believed to avoid reconstructions of input data

From a privacy engineering perspective, and disassociating the input data and statistical outputs, we find tools that:

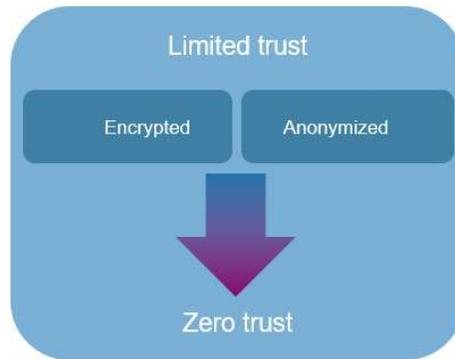
- Do all of the previous, ie, obfuscate input data to hide and confuse, and provide access to statistical outputs only
- Limit output disclosure by "checking" what the statistical outputs are and deciding whether or not they can be provided (using manual or automated means)

Both these security and privacy engineering perspectives, requiring limits imposed on operations that can be performed or checking statistical outputs, can be manual, daunting, and seemingly ad hoc no matter how well thought out they may be.⁽⁸⁾ Which is why another class of tools are emerging.

Recall the zero-trust IT security officer assumes there is always a breach, and systems should be designed assuming they will be in a constant state of breach. This perspective has evolved to include provable security, and attempts at provable privacy (although much more complicated to define because privacy is a societal good without clear boundaries).

From a security and privacy engineering perspective, we find tools that are emerging in an attempt to provide both confidentiality and disassociability to inputs and outputs with provable guarantees in one form or another. In practice, and especially for health data, we consider tools that:

- Satisfy a narrow set of use cases due to their complexity, computational performance, possible need of specialized hardware, and significant restrictions on end users

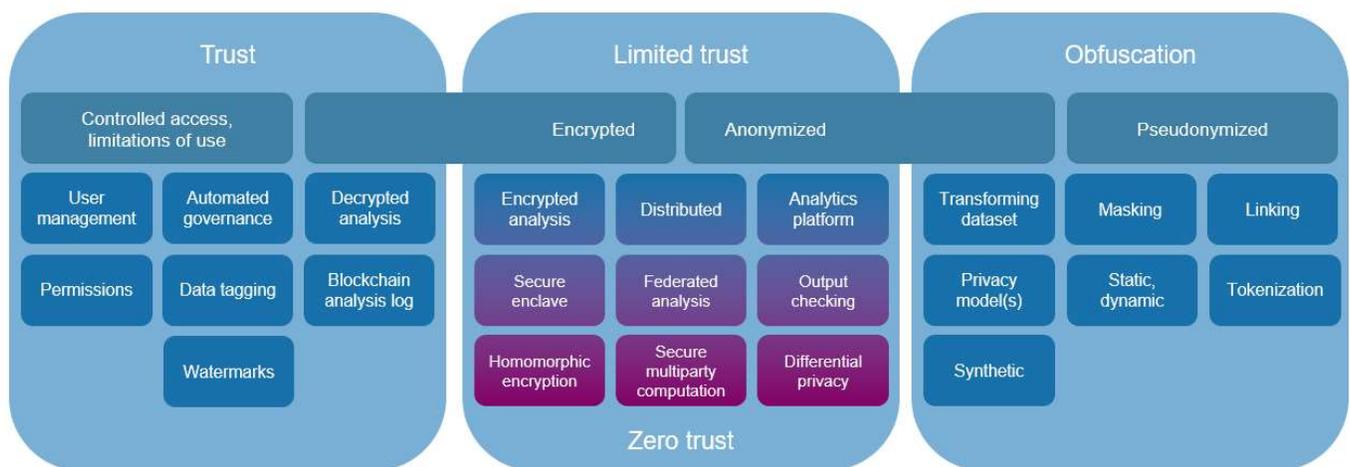


The health researcher, for example, would need to know in advance that they have clean, well formatted input data, with a limited set of predefined statistical analyses in order to use many of these tools. The reality of health data is that of complex structures, sparsity, and disparate formats that rarely line up across departments in the same organization let alone different sites. While study protocols are often defined in advance, for example to get ethics approval, they would rarely if ever define the exact statistical algorithms that will be used. In more sophisticated and established data pipelines, such as in drug development, even static data is refactored and reharmonized to suit the various acrobatics of statistical analysis.

The truth is that a great deal of analysis goes into deriving meaningful statistical results from health data: evaluating various forms of bias, understanding error distributions, inspecting outliers, testing assumptions, refining algorithms and methods, etc. While there are good examples of zero-trust tools being used, in our experience they are often academic or pilots of limited use more broadly, and ill-suited to the realities of health service improvement and research intended to meaningfully improve health outcomes.

Full Spectrum

With the above framing and description of categories, we now provide a more detailed view of the full spectrum of tools (inspired by (9)). We provide this diagram without delving deeper into each tool or subcategory, in the hopes that it will at least motivate discussion and exploration.

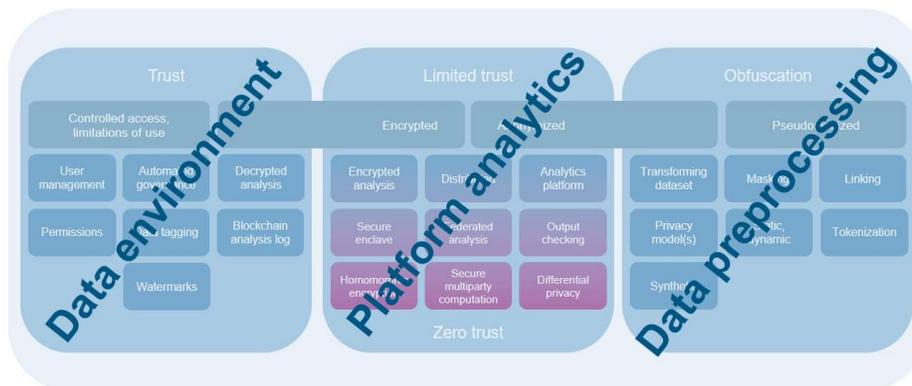


Tools in Practice

PETs and Data pipeline

Another way to think about these tools is in terms of their application areas in a data pipeline, ie, secure data environments (to preserve confidentiality), data preprocessing to reduce privacy risks (disassociation), and platform analytics to further protect through a combination of technical security and privacy controls. A view of privacy tech in the data pipeline, from sensitive information to insightful analytics, can be described as:

- *Data preprocessing*: privacy processing to disassociate or remove the personal from data, in preparation for statistical analyses, from pseudonymized to anonymized
- *Data environment*: protecting confidentiality of sensitive data in the environment in which the data will be used for statistical analyses, including the security posture, protecting against unauthorized access, and technical data governance (non-PETs related protections are out of scope for this discussion)
- *Platform analytics*: protecting confidentiality and possibly disassociating in an environment of limited trust, where access to the data by end users is through the platform only, with a minimal view into outliers, and minimal ability to clean or prep the data



PETs and Data Lifecycle

We can now revisit the data lifecycle we described in the introduction, and consider where different tools may be deployed from sensitive data to insightful analytics (a glossary is provided in the appendix that describes the different tools):

Get data

- Creation, collection by privacy model (anonymization)
- Pseudonymization
- Randomized response, local differential privacy
- Simulated data

Link data

- Federated statistics or AIML
- Secure linking and dataset anonymization
- Secure multiparty computation

(Re)Use data

- Analytics platform (output checking, global differential privacy)
- Anonymized, synthesized dataset
- Homomorphic encryption

Share data

- Anonymized, synthesized dataset
- Personal data store (ie, releasing control, allowing individual control)
- Secure enclave or environment

Open data

- Aggregation and statistical outputs
- Anonymized, synthesized dataset
- Redaction or summarization
- Simulated data

Conclusions

While we may have provided some sweeping generalizations to challenge certain ideas and perspectives, our goal is very much to drive a conversation so that more practical and applied methods get the attention they deserve. Our hope is that the adoption of tools that enable privacy-enhancing data sharing and analytics can be increased through an increased focus on real end users that need safe, useful, and timely data and analytics that can work with the complexity and scale of real health data and modern health challenges.

In our experience, it is the entire spectrum of privacy-enhancing tools that are needed for the safe enablement of data and analytics, depending on the needs of end users and the specific use cases being deployed. The more practical approach is therefore, in our opinion, a combination of tools rather than any one tool. While this may seem obvious to some, it bears mentioning so that more effort is put towards the integration and deployment of suits of tools that are interoperable and complimentary.

We wish to thank you again for this opportunity to provide our views on the operational challenges with, and development needs for, privacy-enhancing data sharing and analytics. We hope that you have found our feedback helpful and insightful towards developing a national strategy on this topic. We look forward to participating in future consultations, such as exploring more detailed views on the interplay between different tools that support privacy-enhancing data sharing and analytics.

References

1. Clinical Data Interchange Standards Consortium. Study Data Tabulation Model (SDTM) v2.0 [Internet]. Clinical Data Interchange Standards Consortium. 2021. Available from: <https://www.cdisc.org/standards/foundational/sdtm/sdtm-v2-0>
2. PhRMA. Biopharmaceutical Research & Development: The process behind new medicines [Internet]. PhRMA. Available from: http://phrma-docs.phrma.org/sites/default/files/pdf/rd_brochure_022307.pdf
3. Stephen Bamford, Sarah Lyons, Luk Arbuckle, Pierre Chetelat. Sharing Anonymized and Functionally Effective (SAFE) Data Standard for Safely Sharing Rich Clinical Trial Data. Applied Clinical Trials [Internet]. 2022; Available from: <https://www.appliedclinicaltrials.com/view/sharing-anonymized-and-functionally-effective-safe-data-standard-for-safely-sharing-rich-clinical-trial-data>
4. Office of the Privacy Commissioner of Canada. Privacy Enhancing Technologies: A Review of Tools and Techniques [Internet]. Gatineau, Canada; 2017. Available from: https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711/
5. Daniel J Solove. "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. San Diego Law Review. 2007;44:745.
6. Office for Civil Rights. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [Internet]. Washington, DC: Department of Health and Human Services; 2012 [cited 2021 Jul 7]. Available from: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#_edng
7. Arbuckle L, Muhammad Oneeb Rehman Mian. Engineering Risk-Based Anonymisation Solutions for Complex Data Environments. Journal of Data Protection & Privacy. 2020;3(3):334–43.
8. O'Keefe C, Chipperfield J. A Summary of Attack Methods and Confidentiality Protection Measures for Fully Automated Remote Analysis Systems. International Statistical Review. 2013;81(3):426–55.
9. Mobey Forum. The Digital Banking Blindspot: Emerging Privacy Enhancing Technologies and their Role in Privacy Risk Mitigation and Business Innovation [Internet]. 2021 p. 17. Available from: <https://mobeyforum.org/the-digital-banking-blindspot/>

Glossary

Anonymization: transformation, including synthesis, of data with inclusion of privacy model and controls

Differential privacy: noise addition to produce indistinguishable outputs up to a defined information limit

Federated analysis: combining the insights from the analysis of data assets without sharing the data itself

Homomorphic encryption: encrypted data that can be analyzed without decryption of the underlying data

Output checking: verifying disclosure risk of analysis results conducted on confidential data

Privacy model: syntactic evaluation of data threats or formal proof of information limits

Secure enclave: isolated execution environment to ensure integrity of applications and confidentiality of assets (aka trusted execution environment, or confidential computing)

Secure multiparty computation: combining the encrypted insights from the analysis of data assets without decrypting the underlying insights themselves

Simulated data: artificially generated data by a theoretical, representative model

Synthetic data: artificially generated data by a statistical or learning model trained on real data

Tokenization: secure process of substituting sensitive data elements with non-sensitive and secure data elements

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

PUBLIC Group International Ltd

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

**REQUEST FOR INFORMATION ON ADVANCING PRIVACY-ENHANCING
TECHNOLOGIES**

REFERENCE NO: Doc861397707

Document Citation: 87 FR 35250

Document Number: 2022-12432

Agency: Office of Science and Technology Policy

Respondent Name: PUBLIC Group International Ltd

Respondent Type: Industry (government-focused technology company)

Website: <https://www.public.io/>

Introduction

Starting life as an accelerator and investor, **PUBLIC** is now a leading European government-focused technology company working to reimagine and build digitally-enabled public services. Having delivered numerous programs to tackle thorny governmental data problems, PUBLIC approaches privacy-enhancing technologies (PETs) from a perspective of encouraging privacy-preserving data collaboration, championed by the public sector. In our convening role, PUBLIC has developed an ecosystem of 9 leading PETs providers.

Our work in PETs focuses on:

- **Leading discovery projects** to map sensitive datasets and explore PETs use cases
- **Designing and running innovation challenges** to 'crowd in' innovative PETs solutions for priority use cases from our PETs startup ecosystem
- **Building and scaling PETs solutions** through an Agile Digital Delivery function and Software engineering capability, supporting data governance and strategy through its Data Services team and partnering with PETs industry leaders to bring in the best-fit solution

PETs have a unique opportunity to unlock data sharing within and beyond the public sector. Building on the last decade of digitally-enabled service transformation, the next phase of government data-innovation will leverage PETs to enable inter-organisational data collaboration at scale across the public and private sector.

Given our public sector focus, our response to this RFI concentrates on learnings and recommendations that the US federal and local government can take to harness privacy-preserving data sharing and analytics responsibly to benefit individuals and society.

Q1: Specific research opportunities to advance PETs

Benchmarking and standards

PETs promise to remove the uncertainty from data privacy and information security with mathematically-provable improvements to confidentiality. However, despite market growth across multiple sectors, including telecommunications, and predictions of faster-than-average technical advances,¹ PETs adoption will be limited if the technical definitions of privacy, confidentiality, and information security are not precisely matched to regulatory requirements.

Regulatory benchmarking: PET solution providers and buyers would benefit not just from advice, but also prescriptive guidance from the Federal Trade Commission and other regulatory authorities. The first step, however, is additional research and development to establish baselines on how existing configurations of PET solutions satisfy technical definitions of confidentiality, such as [k-anonymity standards](#), as well as federal data classification standards.

Encryption and trust-setting standards: Multiple types of encryption can be layered or applied at each stage of the data and analytics pipeline to increase protection, not only when data is 'at-rest' or 'in-transit', but also when it is 'in-use', thereby reducing risk and improving security

¹ According to MIT research from [Singh et al, 2021](#), the technical improvement rate of the PETs domain (707G06F) is 178.1% p.a. compared to a typical technical improvement rate of less than 25% p.a. See [search portal](#).

posture. Typical encryption methods (ie. AES256 encryption) are applied to data 'at rest', whereas homomorphic encryption is applied to the target dataset to enable encryption 'in-use', empowering computations and analytics to be performed without ever decrypting the underlying data. While technically possible to combine these methods today, it is slow and impracticable. Investing in research, and subsequently engaging standards-setting bodies, could accelerate the development of solutions which use this highest standard of protection.

R&D into hardware acceleration

A major barrier to scaling PETs for public sector use cases is improving the performance of computationally-intensive data processing. For example, performing computations on Fully Homomorphic Encrypted (FHE) data is 0.5-19 million times slower, depending on the computing operation, than the same computations on plaintext data ([Sidorov et al, 2022](#)).

Hardware acceleration refers to the use of specifically-designed hardware to improve the performance of computations which, when run on software on a general CPU, have increased latency, decreased throughput, and increased energy consumption. Hardware acceleration is a promising solution to help PETs achieve scale appropriate for running secure computations on billions of records. However, research and development investments specifically targeted at this problem are needed to understand appropriate configurations and commercialize solutions.

Combining PETs to solve real-world problems

A number of think tanks and regulatory authorities have issued recent guidance on using PETs for public sector,² which typically assesses the array of technical methods and diagnostic criteria to match a single technology to a use case. Given the nascency of the PET provider market, vendors have also largely to date developed solutions utilizing a single technical method. However, each PET technology has different strengths and drawbacks (see Q2), which demands a more nuanced, blended approach.

Solving real-world problems requires the integration of technical and non-technical solutions to protect privacy, provide acceptable levels of information security, and facilitate data utilization. The Office of Science and Technology Policy (OSTP) could commission research on combinations of PETs appropriate to different use cases – for instance, using differential privacy to add privacy controls to a data asset, and then placing this data into a Trusted Execution Environment. Testing multiple PETs and the appropriate configuration of confidentiality, utility, and performance, can help establish the baseline approach for PET providers to solve a 'whole problem' for clients.

Q2: Specific technical aspects or limitations of PETs

PETs represent a range of emerging technologies, at varying levels of technical maturity and market penetration. They do not simply replace or augment existing privacy, information security, or data protection management tools, but instead enable new data sharing capabilities. As a result, they need to earn trust. According to the Royal Society's '[Protecting privacy in practice](#)' [report](#), for any given use case, "the choice of a PET requires considering forms of control or oversight in any given system, and what attacks that system might be vulnerable to."

² Recent guidance includes the Royal Society [report on the state of PET use in data analysis](#), CDEI [Adoption guide to PETs](#), and The Rise of Privacy Tech (TROPT) [Defining the Privacy Tech Landscape Whitepaper](#)

Chief Information Security Officers (CISOs) need to be onboard; their trust needs to be earned. Currently, federated learning, data protection, and anonymisation have security and functionality drawbacks which would not hold up against existing [NIST cybersecurity standards](#); however, these can be addressed with combinations of other tools, such as homomorphic encryption.

The core limitation of fully homomorphic encryption and multi-party computation today is that, while they can compute any function and guarantee security, they are typically inefficient off-the-shelf, beyond basic computations. Future technology development and hardware acceleration can mitigate these limitations. For instance, one homomorphic encryption company in our start-up network, [Duality](#), works with [Ice Lake](#), Intel’s scalable server processor, to mitigate computational load with hardware.

Each PET lends itself to a different approach to security and access controls– i.e., which users are given access to data, in which circumstances, and how – depending on the use case and computational technique. Some may fit agency requirements better than others. Existing cybersecurity standards, such as ISO27001, are a strong-baseline, but do not fully address security implications of PET technologies which are less understood or well-trusted. The mathematical precision of different PET methods create an opportunity to clarify the definitions and benefits of these more traditional approaches which are more intuitively understandable, but for the same reason, are less reliable at guaranteeing privacy protection. For more details, see Figure 1 below.

Technology <small>(not exhaustive)</small>	What is it?	Advantages	Limitations
Trusted Execution Environment	Processing environment isolated from a computer’s main processor - code/data held within a TEE cannot be accessed from the main processor	<ul style="list-style-type: none"> Permits outsourced computation where data re-identification is not a concern, but trust is low Able to handle complex big data processing 	<ul style="list-style-type: none"> Use cases where output privacy is paramount because it is vulnerable to repeated query attacks
Differential Privacy	A formal definition of privacy requiring that the output of any statistical analysis reveals no information specific to an individual in the dataset	<ul style="list-style-type: none"> Permits data collaboration with low or intermediate trust, where accuracy of analysis does not need to be perfect 	<ul style="list-style-type: none"> High data utility requirements Use cases where repeated querying will be the norm
Homomorphic Encryption	A cryptographic technique which allows for encryption of data in process rather than just in-transit or at rest	<ul style="list-style-type: none"> Where trust is low or nonexistent, empowers collaboration and computations without ever decrypting 	<ul style="list-style-type: none"> Requires extra compute resources and performance is generally 10-1,000x slower than computing in the clear
Federated Learning	A paradigm for executing a computation against decentralised data	<ul style="list-style-type: none"> Multi-party analytics where trust is low or data transfer risk is high Commonly used with DP Training models with small datasets 	<ul style="list-style-type: none"> Computationally intensive models “Messy” data use cases

Figure 1

Given the wide variety of use cases, both buyers and suppliers require clarity on what is needed for provisioning models of access to data and analytics with PETs, data schemas which accommodate various PET techniques, and data infrastructure requirements. These process elements sit apart from the PETs approaches in Figure 1, but are vital to successful deployment. For this reason, we recommend that early adopters of PETs within federal agencies begin with a robust [discovery sprint](#), according to the U.S. Digital Service (USDS). This can help illustrate the ideal configuration for a given use case, and when repeated across use cases, represent best practice for more standardized PET configuration.

Q3: Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

Based on PUBLIC's experience in cross-governmental data sharing, we have identified four high-value application areas for PETs in a public sector context, with potential secondary benefits for the private sector:

1. **Cross-departmental data access** represents use cases where two or more departments have a desire to access the others' data, but there are barriers to access due to data sensitivity or interpretation of controllership.
2. **'Centralisation' of disparate datasets** represents use cases where both public and private sector entities would benefit from a 'central' repository of relevant datasets to solve or monitor a specific challenge. In these cases, the barriers to centralisation typically result from differences in data infrastructure, IP or competitive concerns around data transfer, and/or other regulatory concerns.
3. **Risk assessment & monitoring of sensitive populations** represents use cases where sensitive data is used as a predictor of adverse outcomes to deliver intervention services or model potential economic/planning impact. In these cases, the barriers to prior implementation of these models typically include the ability for private contractors to access the data to run models or other data access challenges.
4. **Unlocking scalability for sensitive data assets** represents use cases where testing or access to sensitive data assets in the Open Data context or for researchers/developers collaborating with the government has been limited due to the sensitive nature of the data.

PUBLIC's initial research has highlighted four policy areas that represent early adopter opportunities for PETs, which we believe should be flagged in the OSTP's forthcoming national strategy. These have been identified through a combination of the specific demands for data sharing within these sectors we have witnessed in advisory work across these sectors, and nascent demand signals of key players' willingness to experiment and adopt these technologies:

Healthcare

PETs can help enable innovation in the health and social care system. In the United Kingdom, the National Health System (NHS) is developing a new 'Federated Data Platform' which includes dedicated acquisition of PET-based solutions as part of an initiative to enable better utilization of data and "unlock the value of NHS data." PETs can help link between datasets controlled both by public health authorities as well as health and care systems in order to drive population-level insights. They can also be used to increase the privacy and security of existing data flows which currently rely only on contractual protections. In the United States, where healthcare is more decentralized than the UK, public health agencies are well positioned to anchor the industry with use cases and set overall standards for harmonizing with, including HIPAA reform.

From collaboration with our PETs startup network and engagement with the NHS, we have identified a few use case themes:

- Research and development for therapeutics, devices and health service delivery
- Population health management, particularly by enabling greater collaboration between insurers, health services and clinicians

- Coordinating care effectively between social workers, carers and health services, particularly towards the end of life, while complying with HIPAA laws

Case study

Gradient0 - University Hospital Basel

Through its GovStart accelerator program, PUBLIC helps start-ups scale their work in all areas of the public sector by equipping them with the knowledge and network needed to succeed. One of the members of the current cohort of start-ups, [Gradient0](#), implements DQ0, a software platform for privacy-preserving analytics and machine learning, at the University Hospital Basel, to enable internal and external researchers conduct medical studies without compromising data protection and data ownership. The clinical data is both highly sensitive and very valuable, which is why the clinical partner attaches great importance to data sovereignty. DQ0 provides a zero-trust solution where the clinic retains full control over the data at all times.

With complete audit trails, enhanced built-in policy mechanisms and specifically implemented PETs like Differential Privacy, the Basel clinic can work on their study with internal and external partners safely and securely.

National Statistics and Open Data

Building on the foundations of Tumult Labs' [U.S. Census' implementation of a differential privacy program for sensitive data releases](#), national and subnational data publication programs can leverage tools such as differential privacy to lower risk of disclosure. By increasing output privacy (the level of confidentiality achieved in the 'output' of a computation) through adding 'noise' to a source dataset, programs could publish more detailed data otherwise routinely published in aggregate form, or 'locked' behind business intelligence tools, on official government websites.

These tools could also be routinely applied to state and local open data programs. Industrially scaling the use of PETs in open data programs could unlock value previously inaccessible with a binary public-private view of data taken by the government open data community.

Online Safety

Through extensive work with Ofcom and the UK's Department for Digital, Culture, Media and Sport (DCMS), PUBLIC has identified opportunities for PETs to benefit policymakers and the wider online safety ecosystem. The core data problem centers on a lack of access to high-quality sensitive online harms data from online platforms, due to barriers around data protection and loss of commercial intellectual property. There are major risks around re-identification of a data subject from processing of publicly indexed data, which could lead to re-victimisation and legal challenges.

Two priority high-level use cases for PETs have emerged from our project work:

1. Facilitating access to online harms data for researchers and online safety technology providers ('Safety Tech') to develop solutions to tackle priority online harms and facilitate safer online experiences, including building, training and testing AI/ML content moderation
2. Facilitating access to online harms data for an online safety regulator to audit compliance with emerging online safety legislation (e.g., for Ofcom in the UK/California Department of Justice)

Case study

Delivering a Trusted Research Environment (TRE) to tackle child grooming online

In 2021, a consortium including [PUBLIC](#) created a TRE, a secure environment where data access can be controlled, to combat child grooming online. Many online platforms do not have sophisticated moderation mechanisms for child grooming, therefore this activity can go undetected while Safety Tech providers often lack the data to build models to help address this problem. The TRE received positive feedback from multiple Safety Tech providers as a necessary and useful tool to build more effective models.

A TRE sandbox offers the optimal combination of utility to Safety Tech companies and data privacy. It would also allow government departments and law enforcement to control data access and prevent the extraction or manipulation of sensitive data.

Defense, Intelligence & Policing

There are a number of PETs-addressable data problems faced by public bodies responsible for defense, intelligence and policing, driven by the nature of data with potential importance for national security.

PETs are particularly well-suited to use cases centered around:

1. **Sharing near real-time threat intelligence data** between public bodies across international boundaries and organizational jurisdictions in order to coordinate responses to sophisticated, cross-border threats e.g. cyber attacks, money laundering (including crypto-based transactions) and military or policing intelligence.
2. **Facilitating access to secure datasets** between public bodies and non-governmental partners to reduce the barriers of security classification to enable procurement and collaboration.

Q6: Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

'Data challenges'

'Data challenges' effectively 'crowd-source' particular data sharing pain points and opportunities, while introducing the topic of PETs to wider government audiences. The [Civil Service Data Challenge](#) was launched by the UK Government in 2021. All civil servants were asked to identify key datasets officials needed to better fulfill their government roles and propose which datasets needed to be linked. The best ideas were selected by an expert government panel and provided with funding and institutional backing.

The key strength of 'data challenges' in this case was the bottom-up crowdsourcing from civil servants at all levels who had domain or departmental expertise and direct experience of data pain points. A selection criterion was that solutions must be replicable and scalable with potential applications across government. Canada has now launched its own [Public Service Data Challenge](#), while the UK Government has launched the Data Challenge for the second year in a row.

For the OSTP, a ‘Data Challenge’ approach could create a valuable longlist of potential data sharing initiatives, highlight potential PET pilots to address data sharing challenges across the federal government and build momentum for PETs adoption through increased awareness.

Innovation challenge programs

Innovation challenge programs are agile and bespoke programs, aimed at supporting governments to find innovative tech solutions capable of solving tightly defined policy and delivery challenges.

On the supply side, a challenge program is an opportunity for innovators to gain traction, enabling them to find the right ‘front doors’ for their technology. On the demand side, a challenge program allows the public sector to get a wide view of the market through a rapid market engagement exercise. Through assessing new technologies and engaging with tech providers, regulators, and end users, policymakers can become better informed of the latest trends, developments, challenges and concerns around PETs.

Case study

Safety Tech Challenge Fund (STCF)

In 2021, PUBLIC designed and managed the [Safety Tech Challenge Fund](#), a UK Government-backed grant fund program that brought together new interventions in Safety Tech with funding to develop proof of concepts to keep children safe in end-to-end encrypted environments whilst upholding user privacy.

The program stimulated the creation of 5 innovative proof of concepts over the course of just five months. It also encouraged cross-sectoral collaboration and knowledge sharing among the online safety community. Through this program, Government has fostered broader links with the burgeoning Safety Tech sector to address key issues in online safety while upholding user privacy.

Q7: Risks related to PETs adoption

The relative nascency of the privacy-enhancing technology market, the limited number of case studies outside of early-adopting market subsectors and the technical complexity of the solutions present certain risks to PETs adoption. Through experience delivering a Trusted Research Environment for the UK government’s [Online Safety Data Initiative](#) and our public sector research, PUBLIC has identified a number of risks and mitigations for government customers.

Risk Category	Risk Description	Recommended Mitigation(s)
Technical implementation	There is a risk that a supplier does not support privacy-preserving AI/ML or lacks the off-the-shelf integrations to deploy across the data stack.	Ensure that this is a threshold requirement for all technical suppliers at procurement.

	There is a risk that limited experience of PETs project delivery impacts the quality of the technical implementation, increasing the probability of major technical defects or privacy breaches.	Ensure the project engineering team are knowledgeable in the products and services of the chosen PETs provider. Conduct penetration testing of the solution to identify and mitigate vulnerabilities.
Change Management	There is a risk that users fail to adopt the PETs solution and secure benefits post-implementation due to poor communication and a lack of trust.	Design training and internal/external comms into solution delivery as part of a comprehensive target operating model in order to establish the purpose of the PETs solution.
	There is a risk that the customer lacks the necessary resources to operate the solution, as PETs providers do not typically provide services, but simply the infrastructure.	Identify the necessary roles to operate and maintain the PETs service, performing a gap analysis against existing internal skillsets and resources.
Procurement	There is a risk that PETs suppliers fail to meet their stated product performance levels due to potential technical immaturity on the supplier side and/or lack of understanding of what 'good looks like' on the customer side.	Perform an upfront market assessment to assess technical, strategic, operational and commercial fit. Depending on data access, infrastructure and budget, there may be scope to use test datasets to assess AI/ML performance.
	There is a risk that piecemeal adoption of PETs 'point solutions' rather than a privacy-by-design approach fails to address the data privacy problem.	During the discovery phase, perform a cost-benefit analysis of various solutions, including integrated or embedded solutions, that factors in residual risk.
Equality and human rights	There is a risk that uneven distribution of benefits from PETs could create equality and human rights issues and citizen complaints.	Perform a equality and human rights impact assessment at the project outset. Perform monitoring and evaluation as part of a pilot rollout before scaling the solution.

Q8: Existing best practices that are helpful for PETs adoption

Barriers to PETs adoption largely stem from a lack of trust. The PETs ecosystem currently suffers from limited standardization across technical and regulatory definitions and requirements for privacy, confidentiality and information security, which prevents clients from being assured of compliance. PUBLIC has identified two key areas of best practice to encourage PETs adoption:

Decreasing risk and severity of potential issues

The dominance of a single-technology, single-supplier approach is a barrier to responsible integration and adoption of PETs. PETs should rather be seen as part of a privacy tech 'stack' in which application layers network with and reinforce one another to provide a fully secure solution.

For example, federated learning requires all involved parties to pre-determine the exact analyses that will be performed on datasets, meaning that a third party could analyze updates to the model to make inferences about the underlying data, putting it at risk. Homomorphic encryption adds an additional layer of security, encrypting both data and results, meaning that nothing can be inferred about the model or data. By adding secure multiparty computation, the models used to analyze the data are built on encrypted data, and all parties must agree to access any results.

Approaches of this kind are often found in [fraud prevention](#) where a bank's fraud officers can collaborate, pooling data and training models on encrypted data to predict which types of fraud are most common in their country, or which flagged accounts are likely to repeat suspicious activity.

Increasing organizational trust

Open source repositories (such as Github and Kaggle) can help to solve trust issues around the use of PETs. This allows security and intelligence customers to examine and experiment with solutions in order to ensure they are 'quantum-safe', meaning their algorithms are resistant to attacks by both classical and quantum computers.

- [OpenDP](#): is an open-source initiative started by Harvard University, building a suite of tools for enabling privacy-protective analysis of sensitive personal data; the target use cases are to enable government, industry and academic institutions to safely share sensitive data to support scientifically oriented research and exploration.
- [Microsoft SEAL](#): Microsoft SEAL provides open-source encryption libraries that allow computations to be performed directly on encrypted data. This replaces typical cloud storage solutions where only unencrypted data can be analyzed, rather enabling cloud services with both encrypted storage and computation capabilities, whilst still guaranteeing customer data will never be exposed in unencrypted form.

The development of standards and accreditation by governmental and regulatory bodies is likely to catalyze adoption and new applications. This is especially important where public-private regulatory data sharing is involved, such as in healthcare. In 2022, the ICO invited health sector organizations to participate in PETs workshops to upskill the sector and inform [ICO regulatory guidance](#), which is currently at consultation stage.

Defining appropriate organizational infrastructure and ownership around data and privacy is critical to responsible PETs adoption. As PETs providers are onboarded, specific teams with mandates around data policies, system audits and risk management should work with suppliers to communicate a strong understanding of an organization's network architecture, including their firewall boundary, to design and deploy their solution. This will ensure that data feeds, algorithms and outputs comply with data governance and are designed around security boundaries.

Q9: Existing barriers, not covered above, to PETs adoption

Knowledge and cultural barriers

The first fundamental barrier to PETs adoption is a lack of knowledge. Many data protection officers and budget holders may be unaware of the existence of PETs and the benefits that they may provide to the data enterprise. Commercial teams may be uncertain on how to design a requirement to encourage innovative PETs responses, entrenching business as usual.

The [structured transparency framework](#) has the goal of making PETs accessible, building a bridge between technical and non-technical communities. The practical application of this framework and its underlying concepts (e.g., input privacy, output privacy and their verification) should form a core element of a general learning course on PETs and effective data sharing in the public sector, which could then be shared with industry.

There is a risk that partial knowledge of the PETs landscape may entrench poor outcomes for underserved or marginalized groups. A lack of customer or supplier delivery experience in PETs projects may lead to unfair impact on more vulnerable data subjects, possibly without their knowledge. Therefore, it is critical that the government conducts an [equality impact assessment](#) on public sector PETs programs and encourages this as best practice in industry.

Organizational culture is also a major barrier to PETs adoption and new applications. Effective, scaled adoption requires large-scale organizational change and board-level recognition of the risk-reward trade-off of deploying PETs. In order to influence cultural change and accelerate adoption, PUBLIC recommends appointing a single public sector organization with the budget and remit to play a promotion and educational role. This ideally rests with a regulator or a data protection authority, such as the Federal Trade Commission (FTC), or the OSTP.

Ethical barriers

From recent strategic advisory work for the UK's [Equality and Human Rights Commission](#) (EHRC), PUBLIC recognizes the potential opportunities and barriers from an equality and human rights point of view, including data privacy.

The Bureau of Democracy, Human Rights, and Labor could be well positioned to evaluate new PETs-based digital services, partnering with a small group of public sector service providers to understand how they adopt PETs, monitor the potential risks at each stage of the software development lifecycle and then work with them over time to ensure they are adhering to equality and human rights standards. Transparency reporting on this evaluation would result in greater confidence in PETs-driven public services and provide the evidence base to communicate real-world benefits to citizens.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

R3

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



***R3 Response to the Office of Science and Technology Policy Request for Information on
Advancing Privacy-Enhancing Technologies
July 2022***

R3 welcomes the opportunity to respond to the Office of Science and Technology Policy's (OSTP) Request for Information (RFI) on advancing privacy-enhancing technologies (PETs).

We are encouraged by OSTP's launch of the interagency Fast Track Action Committee within the National Science and Technology Council to develop a vision and national strategy for privacy-preserving data sharing and analytics.

We support OSTP's stated goal of "putting forth a vision for responsibly harnessing privacy-preserving data sharing and analytics to benefit individuals and society." We also agree with a recent OSTP blog post titled, "Advancing a Vision for Privacy-Enhancing Technologies," in which the authors conclude "we believe PETs can help enable progress in these key areas, from improving healthcare and combatting financial crime, to responding to natural disasters and driving next-generation digital technologies with equity in mind and in a manner that upholds our democratic values."¹

Government and institutions have the power to effect positive, transformational change by exploiting and combining the data they hold. Such information can be used to help policymakers address important social issues such as social and health inequalities, human trafficking, and climate change. However, many of these benefits have not been realized due to the equally important need to preserve the privacy of individual citizens. PETs represent an opportunity to realize the myriad of benefits on offer from wider data sharing while simultaneously guaranteeing the rights and privacy of individuals.

It is for this reason R3 believes that the future of computing is confidential and that PETs, particularly confidential computing mechanisms built on Trusted Execution Environments (TEEs), are key to enabling this future. To that end, we stand ready to engage with U.S. policymakers as they work constructively on this critical topic and to assist in the development of a national strategy on privacy-preserving data sharing and analytics. We are keen to serve as a resource for OSTP and the Fast Track Action Committee as they consider actions to bolster responsible and effective PET adoption in the United States.

Below we outline our approach to many of the PET-related issues raised in this RFI. We would be delighted to discuss our response in more detail at any time.

Introducing R3

R3 is a New York-based enterprise software company that enables direct, digital collaboration in regulated industries where trust is critical. Multi-party solutions developed on our platforms, Corda and Conclave, harness the "Power of 3" – R3's trust technology, connected networks, and regulated markets expertise – to drive market innovation and improve processes in banking, capital markets, global trade, insurance, and other sectors.

As the only company to deliver both a private, secure, and scalable distributed ledger technology (DLT) platform designed for regulated markets, and a confidential computing platform, R3 empowers

¹ OSTP's blog post on PETs can be accessed here: <https://www.whitehouse.gov/ostp/news-updates/2022/06/28/advancing-a-vision-for-privacy-enhancing-technologies/>



institutions to realize the full potential of trusted direct collaboration in the new digital economy. We maintain one of the largest DLT production ecosystems in the world, connecting over 400 institutions in both the private and public sectors.

Introducing R3's Trust Technologies: Conclave and Corda

Corda is R3's DLT software, which is used throughout regulated financial industries to record, manage and execute institutions' financial transactions in perfect synchrony with their peers. Corda is unique in the blockchain space by offering an open core model alongside privacy, settlement finality, and scalability.

Developed in collaboration with our ecosystem, Corda is transforming entire industries by digitalizing the processes and systems that firms rely on to connect and transact with each other. The Corda ecosystem is the largest in the world with more than 400 institutions deploying and building on Corda Enterprise and Corda. Our customers and partners have access to a network of leading systems integrators, cloud providers, technology firms, software vendors, corporates, and banks.

Building on the success of Corda, and in reaction to the increasing digitization of personal data, we developed Conclave, a software development kit and suite of complementary cloud services for the rapid development of privacy-first applications using hardware TEEs.

Conclave Core utilizes confidential computing and Intel Software Guard Extensions (SGX) technologies to empower businesses to develop applications that can analyze sensitive data from multiple parties – all without compromising on confidentiality. It also allows firms to securely aggregate their datasets to solve shared business problems for their customers across markets and, crucially, without revealing the actual data to anyone.

A wide variety of institutions can be transformed by building on Conclave, including financial services firms, analytics firms, companies leveraging artificial intelligence (AI) and machine learning (ML), data service providers, exchanges, fraud detection and compliance software firms, corporates, and regulators.

Conclave offers a range of benefits for customers, including access to previously inaccessible client data to discover new insights, and the ability to:

- Deliver collaborative data services to grow and differentiate their businesses;
- Monetize data-driven insights without sacrificing an individual's privacy or an institution's processes;
- Reduce the risk of third-party misuse of customer data;
- Simplify building of a privacy-preserving application to collect and process confidential data from customers;
- Benefit from cutting-edge technologies such as confidential computing to secure data at the hardware level; and
- Provide high-value analytics to customers by building apps on a platform that assures multi-party data privacy.

R3's Confidential Computing Observations

The rise and proliferation of the Internet into all facets of life, a process which began in the 2000s and has rapidly accelerated during the COVID-19 pandemic and ensuing pivot to hybrid work, has delivered many benefits but also significant data privacy risks.



Businesses are facing new challenges in this changing world in which more confidential data is being stored and shared online than ever before. Many organizations seek to address potential risks by simply not sharing confidential data with their ecosystem partners in order to prevent tampering or leaks. This results in unintended economic costs and prevents organizations from realizing the benefits of data-driven insights.

Confidential computing, a type of PET that is built on hardware TEEs and encrypts data while it is still in the processing stage, can enable organizations to securely aggregate their datasets to solve shared problems for customers across markets, without revealing raw confidential and personal data to anyone.

A recent report by the Everest Group forecast the confidential computing Total Accessible Market (TAM) to grow to \$54 billion by 2026, a striking total that illustrates the sheer scope and importance of this emerging technology.² U.S. policymakers should work with industry and other stakeholders to foster innovation as well as consumer protection and trust in these new technologies.

Working with OSTP

R3 believes that the work of OSTP is essential to fostering responsible and robust adoption of PETs. We have long recognized the value of working with policymakers and regulators to support their efforts relating to emerging technologies and look forward to further assisting as this important work progresses.

If you have any questions relating to our submission, please contact:

Michael Greco, Government Relations Policy Manager:
Isabelle Corbett, Global Head of Government Relations:

Responses to Selected Questions

Below are R3's responses to selected questions that are most relevant to our perspective as a technology provider.

- 1. *Specific research opportunities to advance PETs:* Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.**

As OSTP explores the broad topic of PETs and formulates a national strategy, we encourage it to include hardware TEEs. We believe that TEEs are certainly among the most promising technologies in this space for a variety of reasons and can accelerate the development and adoption of PETs.

TEEs are a hardware-based solution that allow programs to operate inside secure enclaves which are isolated from the rest of the computer on which they run. They also ensure that data and code can be processed without the computer's owner gaining access to the raw data. As a result, programs are resistant to physical and software attacks by the owner or operator of a computer, as well as by outside

² Everest Groups' paper, "Confidential Computing – the Next Frontier in Data Security," can be accessed here: https://confidentialcomputing.io/wp-content/uploads/sites/85/2021/10/Everest_Group_-_Confidential_Computing_-_The_Next_Frontier_in_Data_Security_-_2021-10-19.pdf

parties. In this example, organizations that share data with partners by utilizing TEEs can be assured that the raw data will not be shared and cannot be tampered with.

In our experience, TEEs are an effective, efficient, and flexible confidential computing technique, while being both highly scalable and production ready. TEEs also can run all code inside its secure enclave – an improvement from many other PETs which limit the number of use cases that can be addressed.

Additionally, R3 urges the U.S. government to bolster its research and development activities to support fundamental research on PETs, with a focus on the relationship between PETs and AI and ML technologies. Specifically, we encourage OSTP to coordinate and spearhead research initiatives exploring how both AI and ML can be leveraged in combination with TEEs. TEEs do have the capability to leverage data-driven technologies like AI and ML while preserving data privacy, but more research and development would benefit the overall adoption of PETs.

We also recommend that the National Artificial Intelligence Initiative, which is administered by OSTP, consider TEEs when evaluating PET's interplay with AI and ML technologies. ML models in particular benefit from large, diverse training data sets, and their emergence as a mainstream technology has revealed the importance of multiple firms being able to share data securely. For ML models to function effectively, it often requires trust in a data aggregator. In sensitive applications, such as healthcare and financial services, the potential leak of trade secrets and risk of privacy breaches can prevent data sharing.

2. *Specific technical aspects or limitations of PETs:* Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

[No response]

3. *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:* Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of [Executive Order 14058](#): Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government.



Confidential computing is broadly applicable across a wide range of sectors and industries but will likely offer the most significant near-term benefits to sectors and applications where data are extremely sensitive, such as the financial services, healthcare, and public sectors. Each of these sectors process large quantities of personally identifiable information (PII) and other sensitive data.

As an example of confidential computing's benefits for a sector where the processed data are extremely sensitive, consider the Hope for Justice charity, a nonprofit that collects data on human trafficking. The data gathered by Hope for Justice is based on highly sensitive, individual cases that often contain PII. Before the advent of confidential computing, analysis of the charity's data may very well have been impossible because any leaks in the original data could threaten the precarious position of those who provide it. One solution has been to apply blockchain along with confidential computing, allowing the charity to create aggregated analytics about human trafficking trends without exposing the underlying data and people's personal information. In one country, these technologies pinpointed a 35-mile radius from which 95 percent all domestic human trafficking reports originated. This information provided the relevant authorities the intelligence they needed to locate and stop the traffickers.³

Additionally, we note that a range of data collaboration use cases can be unlocked with this technology. For example, R3's Conclave enables collaboration with its focus on developer and user experience, as well as its support of many coding languages which unlocks large developer communities. Users of Conclave can use high-level coding languages such as Java, Kotlin, Python, and JavaScript to develop hardware-secured services and privacy-first applications.

We also note that R3's Conclave offers both code integrity and data integrity. Importantly, this functionality enables us to deliver technical proof through remote attestation reports in which code is executed without being altered.

R3's Conclave can open several business use cases for software vendors to process previously inaccessible data and build high-value analytics solutions. With cutting-edge PET, the Conclave platform allows businesses to capitalize on Market Data Aggregation, Fraud Detection, Private Order Matching, and Analytics Solutions. We explore these four challenge areas in greater detail below:

- **Challenge 1: Market Data Aggregation:** Firms need to know how big their market is in order to calculate their own market share and forecasts; however, they do not feel comfortable sharing proprietary sales data such as new and recurring license revenue with a third-party provider.
 - **Solution:** Conclave enables firms to securely upload revenue and other proprietary sales data to a service that calculates relevant statistics, all without allowing the operator of the system to see the raw data, saving time and costly development resources.
- **Challenge 2: Fraud Detection:** Fraud that is spread across multiple institutions often remains undetected as there has been no way to securely pool multiple private data sets from industry participants to identify and prevent fraud. This challenge impacts both the private and public sectors. There are limits to the types of services that fraud detection software vendors can provide as there are data privacy regulations and concerns with sharing confidential customer data with a third party.

³ "Blockchain is restructuring ESG." Alisa DiCaprio, Chis Ford, Yingying Zhang.
<https://www.r3.com/blog/blockchain-is-restructuring-esg/>

- **Solution:** Conclave enables fraud detection software firms to access these private data sets by providing assurances to their customers that their data will not be viewed by the provider or the provider’s other customers, just processed. As a result, these software firms can build new solutions that aggregate data from multiple firms in a trustworthy manner, while reducing false positives and detecting new fraud. In the public sector, federal departments and agencies that distribute benefits to the public could utilize this technology to detect fraud and error by comparing other government data sets without needing to examine individual records that could contain confidential data.
- **Challenge 3: Private Order Matching:** Traders often want to do price discovery and order matching privately, and want to make sure that no one – not even the market operators – can misuse their private transaction details.
 - **Solution:** Conclave enables firms to build private order matching solutions that give assurances that private transaction data will not be viewed or misused. As a result, software firms can build solutions that protect market participants’ data, giving traders and regulators confidence that they are trading in a fair and transparent marketplace.
- **Challenge 4: Analytics Solutions:** Companies want to benefit from insights that Analytics and AI firms can provide, but do not feel comfortable sending confidential data outside of their organization, or having it pooled with their peers.
 - **Solution:** Conclave enables Analytics and AI firms to deliver solutions that provide transparency into how data is processed and assurances that customer data is protected. This enables end users to benefit from insights derived from analytics solutions without having to compromise on data privacy. Software firms can either build new solutions or develop on existing analytics solutions and run them across new private data sets.⁴

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs: Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking authorities under the Office of Management and Budget, the Federal Trade Commission, and financial regulatory bodies, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (e.g., Other Transaction Authority) to procure PETs for development.

We encourage the U.S. government’s hiring authorities to recruit federal employees with the necessary expertise to advance effective PET technical standards. A robust federal workforce with strong technical skills and knowledge would benefit the entire PET sector. It is important for federal agencies and departments to develop in-house expertise that enables robust and advanced PET projects and initiatives that bolster innovation while applying the necessary controls to protect consumers. Recently, the Office of Personnel Management (OPM) outlined a detailed skills-based hiring guide for federal agencies, which could serve as a template for recruitment of federal employees with PET expertise.⁵

⁴ More information on R3’s Conclave and the problems it can solve can be accessed here: https://www.r3.com/wp-content/uploads/2020/10/Conclave_Platform_FS_R3_2020.pdf

⁵ OPM’s Skills-Based Hiring Guidance can be accessed here: <https://www.opm.gov/news/releases/2022/05/release-opm-releases-skills-based-hiring-guidance/>

R3 also encourages the Office of Management and Budget, Federal Trade Commission (FTC), and other federal agencies that issue privacy-related rulemakings to consider benefits offered by confidential computing as they explore and issue regulations and other guidance to ensure the responsible sharing of data.

Overall, the PET sector would benefit from the development of voluntary consensus standards and best practices that provide useful and actionable guidance. In certain cases, clear regulations that effectively balance consumer protection while promoting innovation and adoption may be necessary and ultimately useful. Improving consumer protection and fostering innovation are complementary goals – increased consumer protection can bolster public trust which in turn will lead to increased adoption of PETs.

5. *Specific laws that could be used, modified, or introduced to advance PETs:* Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.

As of July 2022, five U.S. states (California, Colorado, Virginia, Utah, and Connecticut) have passed their own respective comprehensive consumer data privacy laws. This growing patchwork of state-level privacy laws complicates compliance for businesses and may ultimately harm consumer protection efforts. U.S. consumers and businesses utilizing PETs would both benefit from clear and comprehensive regulation.

We also encourage OSTP to continue to promote cross-agency collaboration with NIST, the National Science Foundation, the Federal Trade Commission, and other relevant federal agencies and departments. A harmonized whole-of-government approach to PETs would benefit both the sector and consumers. We also encourage the U.S. government to continue its collaboration and cooperation with foreign governments, including the United Kingdom, on prize challenges and related initiatives to foster PET adoption.

6. *Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:* This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

R3 believes that all the mechanisms referenced in the above question (open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and territorial governments) have the potential to advance PETs.

Open-source software is particularly important because it accelerates trust in the technology and removes barriers to adoption via a variety of features including internal audits.

Additionally, we believe there is a clear need for increased public awareness and understanding of PETs. The U.S. government can help promote greater awareness of PETs, including confidential computing, and the opportunities they present through prize challenges and collaborations. We also believe that technical standards, if developed in a collaborative and inclusive process, can benefit the PET industry. NIST has a particularly strong record of developing voluntary, consensus-based standards that promote innovation while applying the necessary controls to encourage consumer protection.

We appreciate and support the U.S. and UK joint Prize Challenges to Accelerate Development and Adoption of Privacy-Enhancing Technologies. We support the prize challenges' feature of enabling innovators to engage with regulators, including the UK's Financial Conduct Authority and the U.S. Financial Crimes Enforcement Network. Collaboration between industry and regulators, along with increased public awareness, are key measures to maximizing the potential of these technologies, and we appreciate the U.S. and UK governments' willingness to address both factors.

Finally, as U.S. policymakers evaluate federal standards that can be used, modified, or introduced to address PETs, we encourage NIST to update its Privacy Framework and Cybersecurity Framework to keep pace with recent PET developments.

7. Risks related to PETs adoption: Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

[No response.]

8. Existing best practices that are helpful for PETs adoption: Information about U.S. policies that are currently helping facilitate adoption as well as best practices that facilitate responsible adoption. This includes existing policies that support adoption, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This also includes information about where and when PETs can be situated within tiered access frameworks for accessing restricted data, ranging from publicly accessible to fully restricted data.

We encourage the U.S. government to collaborate with industry and other stakeholders to develop a shared taxonomy and terminology of PETs, following a public comment period. A common, well-designed taxonomy could cultivate trust in the design, development, use, and governance of PETs while raising public awareness and providing greater clarity. NIST's Taxonomy and Terminology of Adversarial Machine Learning⁶ is a useful template that U.S. institutions could look to when considering this taxonomy.

Additionally, we believe the U.S. government should work with a broad and inclusive range of stakeholders to develop guidance and, in certain cases, clear rules governing how regulated entities use PETs. The PET industry would benefit from clear rules of the road that govern PET usage for specific data classifications, particularly when PII or other confidential data are shared.

⁶ NIST's Taxonomy and Terminology of Adversarial Machine Learning can be accessed here: <https://csrc.nist.gov/publications/detail/nistir/8269/draft>



9. Existing barriers, not covered above, to PETs adoption: Information about technical, sociotechnical, usability, and socioeconomic barriers that have inhibited wider adoption of PETs, such as a lack of public trust. This includes recommendations on how such barriers could be overcome. Responses that focus on increasing equity for underserved or marginalized groups are especially welcome.

As discussed in our response to Q8, a shared taxonomy that clearly defines relevant terms would foster wider adoption of PETs. Such a taxonomy could provide the clarity needed to accelerate PET market growth.

Additionally, as also discussed in our response to Q8, the U.S. government can help promote prize challenges and other educational initiatives to increase public trust in and adoption of PETs. We encourage OSTP and the broader U.S. government to evaluate ways to build trust among underserved and disadvantaged Americans in an equitable and inclusive way.

10. Other information that is relevant to the adoption of PETs: Information that is relevant to the adoption of PETs that does not fit into any of the topics enumerated above.

We encourage OSTP to refer to our confidential computing white paper, “Conclave: An Introduction,” written by R3’s Chief Technology Officer, Richard Gendal Brown.⁷

⁷ R3’s Conclave white paper can be accessed here: <https://www.conclave.net/wp-content/uploads/2022/01/ConclaveIntroductoryWhitepaper.pdf>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Restore the Fourth

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Restore the Fourth Comment on Privacy-Enhancing Technologies (PETs)

The Office of Science and Technology Policy (OSTP)—on behalf of the Fast Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics of the Subcommittee on Networking and Information Technology Research and Development (NITRD) of the National Science and Technology Council, the National Artificial Intelligence Initiative Office, and the NITRD National Coordination Office—recently requested public comment to help inform development of a national strategy on privacy-preserving data sharing and analytics, along with associated policy initiatives.

Restore the Fourth is a 501(c)(4) nonprofit organization committed to strengthening the United States' notion of privacy as enshrined in the Fourth Amendment and other applicable statutes and case law. While our focus is on curbing mass government surveillance, any and all correlation of data and data analytics has surveillance implications, whether directly by the government or nefarious actors. This particular comment will focus on privacy-enhancing technologies for medical, financial, and location data.

There are several broad principles that apply to collection and application of this data:

- Data, and thus surveillance applicability, begins at the point of collection
- Not all data is alike in terms of its risk profile
- Once data is in the hands of entities outside of its original scope, it can be bought, sold, and exploited completely free of privacy protections accorded to the original research

Data, and thus surveillance applicability, begins at the point of collection

In the modern Internet-connected information landscape, every US resident has multiple devices on their person and in their homes that are capable of collecting private information beyond the scope of its design and primary implementation. In the case of medical and financial data, the usual devices are Internet-connected smartphones and personal computers, as well as implantable and wearable medical devices that collect real-time biological information. Many of these datasets are geolocated to varying degrees of precision, and thus have embedded location data with its own privacy implications.

Financial data is particularly thorny from a data privacy and security perspective because of its relative centralization; while ATMs and credit card processing operate closest to the consumer, Americans already are governed by opaque algorithms derived from central processing of their transactions: the credit rating system. Modern advances in payment processing, from electronic payments (PayPal, Stripe, Plaid, Spreedly) to gamification/incentivization of preferred spending patterns (store rewards cards, digital healthcare platforms) have enabled financial transactions to originate from websites and regions more and more remote from this centralization.

Regardless of its scope and content, the data generated is stored onboard, *somewhere*, before transmission to other systems or over the Internet. Thus, common data security and privacy protection technologies, like “access control, data anonymization, data encryption, differential privacy protection, digital watermarking...identity authentication” [1] should be performed on-device wherever possible. In addition, the particular use case of machine learning algorithms being applied to large datasets allow for the possibility of federated learning, in which training algorithms are executed “across [decentralized] edge devices (e.g., individual mobile phones) or servers hosting different local samples (e.g., data owned by different samples). Data samples are not shared or [centralized] and only the trained models are communicated, which might improve data security and privacy of patient data” [2]. This of course does not limit the scope of surveillance that can be collected by law enforcement officers (LEOs) at the device level, many techniques of which carefully skirt US privacy law via overbroad and intentionally backdoored legal regimes, but it can reduce the ability of other nefarious actors to collect and act on this information.

Contacting research participants is also a day-to-day activity in research with this kind of data; unfortunately, the risks of email makes it trivial to de-anonymize research communications, even without a 1:1 match to exact data points. Implementing robust, scalable pseudonymous remailers allows for one more layer of data privacy; e-commerce sites like Craigslist have implemented this already with a minimal tech footprint.

Not all data is alike in terms of its risk profile

One of the most exciting prospects in current medical research is the ability to examine large swathes of disparate data algorithmically and extract useful patterns. While there is ultimately human oversight of the analysis of the data itself, the techniques and the access to the data vary depending on budget, centralization of resources, and the composition and behavior of researchers and supervisors. The federal government, despite its best efforts, is not a monolith; project-by-project myopia (and thus duplication of resources) virtually ensures that data privacy, and PETs by extension, vary widely in existence, implementation, and maturity.

Access control technology in particular has become a hotspot of current research, but by being targeted mostly at the operating system level, ignores the differential risk associated with different kinds of medical data. While role-based access control (RBAC) is a fundamental, if antiquated, tenet of identity access and management (IAM) practices across governmental organizations, more care must be taken to closely examine the risk of exploitation by the *kind* of data, and not just the access to data itself.

Some researchers have expanded RBAC through mathematical methods to something they call Risk-Adaptive Access Control; in this case, risk is quantified as the “deviation degree between users’ access to medical information and their work tasks...[the] greater the deviation degree is, the greater the risk”. By calculating the information entropy of users accessing medical information, taking that entropy as an input data set for advanced data processing (K-means

clustering) and seeing what patterns emerge with respect to a defined baseline risk, supervisors and administrators can “dynamically access control policies based on users’ access conditions” [1].

While participant data itself can be exploited by any number of entities, the analytical techniques themselves are at risk as well. A recent paper by researchers at UC Berkeley, MIT, and the Institute for Advanced Study (IAS) demonstrates the ability to plant undetectable backdoors in machine learning classifiers [3]—incidentally, a class of techniques used broadly in analysis of Big Data of this nature. Backdoors of this nature can be used for any purpose, from altering research results to enabling de-anonymization of research protected via differential privacy techniques.

Once data is in the hands of entities outside of its original scope, it can be bought, sold, and exploited completely free of privacy protections accorded to the original research

The biggest data privacy risk of medical, financial, and location data is the ability for it to be correlated to other large datasets, and used to re-identify participants via this correlation. Datasets, while large in size, are ultimately portable, and their ability to exploit their participants via this type of correlation is ultimately limited by computing power and time.

In an era of data leaks, third-party data brokers, and overt intragovernmental information sharing, surveillance and exploitation is always on the table no matter who is doing the watching. The long tail of the 2020 Minneapolis protests and the armed insurrection on January 6 has also shown that there is always a law enforcement and thus profiling use case for re-identifying of this data. It is never fully possible to eliminate this possibility, given that this data is often freely exchanged on the black market, but one way to reduce legally-sanctioned scope creep of the data’s applicability is to ensure that use of the data is governed by licenses that require disclosure to, and agreement by, the original researchers for activities that re-identify it. In addition, the US government should adopt standard data retention policies for types of data, particularly those that are re-identifiable (whether at present or by undisclosed surveillance and analytical capabilities).

In conclusion, the expansion of Big Data capabilities of the US government comes with numerous open questions and pitfalls given by its applicability to surveillance, but extending back to data privacy and their corresponding PETs. Restore the Fourth as an organization focuses on the Fourth Amendment as a cornerstone of individual privacy, but actual policy that protects and enshrines this idea depends not just on accompanying statutes and case law, but specific principles that are obeyed in implementing PETs at any level of data collection and data sharing.

References

- [1] - Rong Jiang, Shanshan Han, Mingyue Shi, Tilei Gao, Xusheng Zhao, "Healthcare Big Data Privacy Protection Model Based on Risk-Adaptive Access Control", Security and Communication Networks, vol. 2022, Article ID 3086516, 12 pages, 2022. <https://doi.org/10.1155/2022/3086516>
- [2] - Yadi Zhou, Fei Wang, Jian Tang, Ruth Nussinov, Feixiong Cheng, Artificial intelligence in COVID-19 drug repurposing, The Lancet Digital Health, Volume 2, Issue 12, 2020, Pages e667-e676, ISSN 2589-7500, [https://doi.org/10.1016/S2589-7500\(20\)30192-8](https://doi.org/10.1016/S2589-7500(20)30192-8).
- [3] - <https://arxiv.org/abs/2204.06974>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Richardson, Douglas

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to the Office of Science and Technology Policy (OSTP) for
Information on Advancing Privacy-Enhancing Technologies

Submitted by
Douglas Richardson, PhD
Center for Geographic Analysis
Harvard University
July 8, 2022

Innovative and Interactive Geospatial Virtual Data Enclave (GVDE) Technologies:

*Robust and reliable privacy-enhancing infrastructure for
accessing, sharing, and analyzing confidential geospatial data and locational privacy*

INTRODUCTION

The generation, analysis, and protection of geospatial data is now at the frontier of many governmental, scientific, and private sector domains. Several contemporary trends are driving new privacy concerns: massive quantities of data streaming, data warehouses from global positioning system (GPS)-enabled devices and sensors, location-aware technologies, advances in web services and cyberinfrastructure, and new geoprocessing tools for analyzing, exploring, and visualizing large and multi-scale spatiotemporal data sets (Richardson, 2013).

These trends increase opportunities for exciting new use and integration of data sets to create multi-disciplinary and data-intensive collaborations. However, the unique confidentiality characteristics of geospatial and locational data present special challenges to such collaborations by governmental agencies and the public. Individuals are often identifiable when geospatial and locational data is presented in maps and other visualizations, or when combined with sensor data or other related geospatial data sets.

Our interactive GVDE technology identifies, integrates, and builds on **four interrelated components** required to create and implement a robust and reliable GVDE system for widespread use by governments, the public, and scientists conducting projects involving confidential geospatial data:

I. Develop the Interactive Geospatial Virtual Data Enclave and its Core Functions. Our GVDE technologies, funded by the National Science Foundation (NSF), address challenges for working with geospatial data and to improve the user experience with regards to geospatial data. This component of the technologies evaluates and integrates software tools and procedures (e.g., data management, GIS, analytics, modeling, spatial statistics, etc.) to enable governmental agencies using confidential geospatial data to a) *share*, b) *access*, c) *analyze*, d) *replicate*, and (e) *build on* projects **within** the GVDE.

II. Evaluate and Implement Masking and Encryption Capabilities for the GVDE. We have evaluated and implemented multiple geomasking methods, encryption, and other processes to enable government agencies to anonymize and then *export maps, analyses, and visualizations* derived from their analyses of confidential geospatial data from the GVDE (after review) for use in public dissemination via public reports, presentations, or publications. This component of the technologies has researched and tested numerous anonymization methods and related disclosure risks for specific types of geospatial data (point, line, polygon, raster, vector, etc.) including rapidly growing new sources of confidential geospatial data such as GPS trajectories, crowdsourced, and social media data used by government agencies. We are also extensively engaged in ongoing research on **Differential Privacy** methods. Our research team is collaborating closely with the Harvard *OpenDP Project*, which includes the Harvard Institute for Quantitative Social Science (IQSS), the Harvard Center for Geographic Analysis, and Microsoft.

III. Develop the GVDE User Credentialing System. Our research team has developed and implemented an innovative, robust, and reliable system to provide trained and with a durable digital identifier. This digital user passport creates an efficient mechanism for large numbers of credentialed government personnel to safely use the GVDE. This innovative system builds on research evaluating multiple access mechanisms at restricted data facilities around the world and improves on a decade-old on-line application systems for restricted data. This GVDE passport serves as a transferable and durable credential to allow governmental personnel to access restricted data at multiple agencies.

IV. Sustainability. To ensure the long-term sustainability of the GVDE technologies and its widespread usage by government agencies, researchers, and others, the GVDE is maintained as part of a portfolio of ongoing data management and stewardship services. To support the usage of the GVDE technologies for the broader governmental communities, we provide training, outreach and dissemination activities on the use of the GVDE technologies, data confidentiality ethics, credentialing requirements, and on policies and best practices.

Our GVDE team involves leading scientists with deep experience with privacy-enhancing technologies and confidential data protection, and is uniquely positioned to apply its expertise in data management and disclosure risk assessment to the development of the GVDE technologies for use by the broad range of users, including governmental agencies. The creation and use of geospatial data is becoming pervasive in many federal (and state) governmental agencies, but the unique privacy challenges of confidential geospatial data access, sharing, analysis, and safe dissemination are not yet fully understood by the public. To address these challenges, our scalable, robust, and reliable Geospatial Virtual Data Enclave (GVDE) privacy-enhancing technologies can easily be adapted for use by special needs of various governmental agencies, as needed.

Our GVDE technologies provide workable and sustainable solutions to key geospatial data confidentiality issues, in government agencies and the broader society. Projects and policies using geospatial data is transforming many governmental agencies. The ability to share and analyze these data securely advances public affairs broadly. The development of a new transferable user credentialing system also benefits governmental agencies, and significantly, at less well-resourced institutions, as their access to valuable data is no longer be limited by their agencies' ability to offer these services. Resolving practical issues related to access to confidential geospatial data benefits many other governmental agencies where progress is impeded due to limited data sharing. The creation of the GVDE technologies has created important new data infrastructure for governmental agencies to share data and to commit to data management plans that enable sharing confidential geospatial data.

RATIONALE

In order to leverage investments in geospatial data creation and analysis and to share data, governmental agencies put data into a trusted analytic repository where data can be safely accessed and analyzed by other agencies and the broader society. Our Geospatial Virtual Data Enclave (GVDE) technologies enable governmental agencies to build on prior data collected, safely, securely, and at low cost. Below we provide the rationale and need for our technologies address key issues and challenges of (a) data sharing; (b) creation of a robust and reliable GVDE technologies for widespread use by the governmental geospatial community; (c) implementation of an innovative credentialing system for accessing, sharing, and analyzing confidential geospatial data; and (d) disclosure risks and methods for protecting confidential data and geoprivacy.

Challenges to confidential geospatial virtual data sharing and analysis

In 2013, the White House issued an executive order establishing the Open Data Policy, which addressed data sharing practices. In response to the Open Data Policy and using NSF practices as a model, many federal agencies are now

requiring data management plans (Adler, 2015). Confidential data was more frequent and was associated with a wide range of stakeholders, including governmental agencies, researchers, and private sectors (Bishoff and Johnston, 2015). Data sharing is especially challenging for governmental agencies using geospatial data because of the risk of revealing both subject identities and precise locations when data is visualized as a map or linked to other datasets.

The opportunity to build on successful GVDE technologies

Our GVDE research teams developed an experimental, functional prototype GVDE that demonstrated proof of concept and was successfully tested for a small number of users. From 2014 to 2018 the experimental GVDE allowed authorized testers to access restricted and confidential geospatial data in a secure environment. In the prototype GVDE technologies, users in multiple locations connected to a virtual desktop to view and analyze data secured on a server. Within the virtual desktop users were able to access confidential data files as well as a wide range of statistical and other analysis software, including geospatial software such as ArcGIS and GeoDa. While they could view and analyze these data, they could not export them from the virtual desktop or download them to a local computer. Users were able to share a workspace, fostering collaboration; however individual projects were isolated from one another to maintain security. The prototype GVDE technologies design was able to replicate the performance and computing power of a standard desktop computer, so that users could perform the same geospatial analyses within the GVDE that they would have been able to do on their local personal computer, with minimal performance degradation.

A robust and reliable credentialing system for the GVDE

An integral component to the GVDE technologies is a credentialing system that establishes transferable digital identities for trained and trusted users to expedite their access to restricted geospatial data. The creation of a trusted user passport builds on the GVDE research team's prior work examining the standards and processes for accessing restricted data at repositories around the world. In a recent white paper, the GVDE research team proposed the establishment of a user passport (Levenstein, Tyler and Bleckman, 2018). By reducing the time and paperwork necessary to access restricted data, the passport helps to overcome the justifiable concerns of governmental agencies and the broader community about working in a restricted environment. While there is a tradeoff between ease of access and confidentiality protection, this project moves the frontier of that tradeoff outward to provide *both* greater access and greater protection so that governmental agencies can achieve desired standards of analysis.

In the current environment, restricted data are often available to users only after a lengthy and complicated application process. This process usually requires the interested users to address the following:

- *Detailed data request:* The user must specify the requested datasets, and in some cases, particular variables, and may include specification of data requested from the provider (both restricted and public use) and other data to be used in the analysis.
- *Research topic and plan:* The user is usually required to provide an analysis plan explaining why the restricted data are necessary to complete the study and the project.
- *Computing environment and data security plan:* Restricted data requests often require that the user describe a particular computing environment that the user or the user's agency/institution provides. A required data security plan specifies the rules, process, and location for accessing and analyzing data. The security plan must be reviewed and approved by the data custodian; in some cases, this includes physical on-site inspections.

This process is burdensome both for those who try to make data available and for users trying to use data. It creates opportunities for people to hoard data and refuse to share, under the guise of protecting confidentiality, or to claim

quite legitimately that it is simply too costly to share data safely. The GVDE credentialing system addresses these concerns. A system used to identify and credential users using the GVDE is essential and benefits governmental agencies in four primary ways.

First, to implement the digital passport, the GVDE technologies standardize the vocabulary used to describe potentially disclosive data and its degree of sensitivity. The GVDE research team's analysis of repository practices around the world found that the language used to describe levels of data restriction, confidentiality, and access methods differs significantly both between and within restricted data repositories. This language inconsistency confounds the challenge of developing a transferable digital user identity, therefore the GVDE research team standardizes the terminology used to describe the elements of restricted data security and access. Establishing a common set of terms and definitions allows different repositories to understand and integrate shared standards and technologies into their own processes. The GVDE research team harmonizes language characterizing disclosure risk associated with the geospatial data (distinguishing, for example, between small area estimates and trajectories of individuals) that undergird the standards necessary for a user to access data of different levels of risk.

Second, by creating a durable and transferable user ID that maintains a record of both responsible use and any prior breaches in handling confidential data, the user passport creates incentives that reduce risk and encourage data sharing. The passport reduces the risk of irresponsible user behavior, because there is a reputational consequence that affects future data access. The passport thus also increases the willingness of potential data providers to share because they can have more confidence that their data is protected.

Third, by establishing a common set of standards across restricted data custodians, the passport facilitates new, creative analyses of datasets of data held by multiple custodians; current inconsistencies in standards for access often make such analyses of multiple restricted datasets impossible.

Fourth, because all analysis takes place in the secure GVDE, the local computing environment is less critical. This is particularly advantageous for users from less resourced governmental agencies and institutions who do not have the technical staff or facilities to establish a secure local environment (which, for example, often requires a dedicated computer and locked office space).

Disclosure risks of confidential geospatial data

The GVDE research team develops and evaluates geomasking and encryption techniques, with particular emphasis on rapidly growing new geospatial big data sources including GPS trajectories, mobile GPS data with sensor inputs, and social media and crowdsourced data. We evaluate which types of geomasking and encryption tools are most appropriate for addressing multiple traditional types of geospatial data (e.g., point, line, polygon, vector, raster) as well as these new big data sources. The disclosure risks associated with these geospatial data are unique as these data can be highly identifiable when presented in maps, visualizations, or when combined with other related data (Richardson, 2015). To protect individual identities in confidential geospatial data sets, various methods and privacy protection metrics have been developed in the past two decades. These include several geomasking and geospatial encryption methods. They mainly seek to modify or hide the original location information in georeferenced data through adding statistical noise or including more data records when responding to spatial queries in order to render re-identification difficult. The GVDE research team investigated an expanded set of analytical tools and geomasking and encryption methods, and their applications to new confidential geospatial data sources and types. These techniques are useful in the analysis of confidential data and can also be used to anonymize results of analysis, maps, and other anonymized visualizations of confidential data within the GVDE so that they can be exported from the GVDE for use in presentations, publications, and other outlets for sharing results with scientists or public audiences.

METHODOLOGY AND ACTIVITIES

The GVDE research team identifies four key areas of work required to create and implement a robust and reliable fully scaled-up GVDE resource for widespread use by governmental agencies involving confidential geospatial data. These key areas are:

- I. Develop the Geospatial Virtual Data Enclave and its Core Functions
- II. Evaluate and Implement Masking and Encryption Capabilities for the GVDE
- III. Develop the GVDE User Credentialing System
- IV. Ensure Sustainability of the GVDE

Our approach to these key interrelated GVDE technologies and implementation components is discussed below.

I. Develop the Geospatial Virtual Data Enclave and its core functions

We evaluate and implement software tools and procedures (e.g., data management, GIS, analytics, modeling, spatial statistics, etc.) to enable users to a) *share*, b) *access*, c) *analyze*, (d) *replicate*, and e) *build on* confidential geospatial data within the GVDE. This part of the project tests the suitability, efficacy and efficiency of a set of analytical methods within the GVDE environment using different types of data and for different applications. This part ensures that the analytical tools made available to users in the GVDE enable them to perform analyses on a variety of data types and formats. Providing these analytical tools within the GVDE benefits the governmental agencies and the broader society, which may not have as much access or exposure to geospatial analytical tools—especially for less resourced governmental agencies and institutions. Specific examples of the datasets to be used for suitability testing are listed later in this section.

Enhancing the core capabilities of the GVDE technologies

To enhance the GVDE technologies prototype, we test and evaluate the system's efficiency, reliability, security, and the user experience. A diverse set of georeferenced datasets are used for testing in the GVDE.

- *User experience captures* the friendliness of the system to users. A system with excellent user experience is critical to successful adoption by users. The following four aspects of the user experience are evaluated: (a) GVDE set-up and login process; (b) system interface; (c) user control and input; and (d) display. Feedback is collected to ensure that the set-up and login process is as user-friendly as possible. The system interface is designed to replicate the standard Windows desktop environment that most users are familiar with. The system interface and user control and input can be affected by network lag if the user has a slow internet connection or is located at great distance from the host servers. The intention is to test the boundaries of these issues to minimize their effect on the user experience while maximizing system efficiency, reliability, and security.
- *System efficiency* is the computational and analytical performance of the GVDE. Comparison between the performance of a range of spatial analysis using the GVDE and local computers is conducted to compare analytical performance, operation time, and so on.
- *Software reliability* assesses whether the GVDE system is reliable. It addresses questions like: Are there any connection failures from remote desktops and how often do they happen?
- *System security* is the ability to secure geospatial data in the GVDE and minimize disclosure risk of confidential data. High system security is one of the most important features of the GVDE system, which must ensure the secure sharing of confidential data in a controlled and safe environment. System security must include data confidentiality, data accessibility, and data integrity and cover both access to the system

(ensuring only authorized users can access given data, and that unauthorized removal of data is prevented) as well as proper vetting of data that is authorized to be removed from the GVDE system. The GVDE system is designed to meet Federal Information Security Management Act (FISMA) Moderate standards in regard to system security. No data or analytic output can be removed from the GVDE system without undergoing formal statistical disclosure control and approval of sponsor agency and confidentiality officers.

There are two authorization steps during the login process. After initially logging in to the GVDE Account, where digital credentials from the proposed user credentialing system resides, the two-factor authentication requires a second authorization step; users need to provide a passcode dynamically generated by a pre-assigned electronic device (smartphone app or hardware token), or respond to a push notification sent to their smartphone app. A new passcode is generated every 60 seconds and is specific to each individual user. This process can detect and stop unauthorized access in the event of a compromised password, as both the password and electronic device are needed for login. Only users who successfully pass these two independent authorization steps can access the GVDE system. The GVDE system prevents users from uploading or downloading data or files to their local computing environment (e.g., their PC). External files can be added to a user's secure computing space by staff. Output files are made available to the user outside the GVDE only after disclosure review by a GVDE staff with expertise in confidentiality protection. Finally, to ensure that users disconnect or lock the GVDE account when he/she intends to leave the connected computer, the remote server automatically disconnects or locks the users' account when no activity is detected for a defined period in order to prevent unauthorized access to the GVDE system. Results of our testing confirm that these security technologies and processes ensure high system security of the GVDE.

Evaluation and implementation of GIS and analytical software tools for the GVDE

The GVDE research team evaluates and implements geographic information system (GIS), statistical and analytical tools for use within the GVDE. These tools are useful for analysis, sharing, and display of data within the GVDE, and also for visualizing data for geomasking and encryption methods so that output can be safely removed from the GVDE. Below is a brief description of some of the geospatial data management and analysis software tools which have been implemented and tested in the GVDE. We also monitor new and emerging techniques and methods for geospatial analysis and for confidential geospatial data protection and integrate them as they mature.

Social statistics: A suite of commonly used social statistical techniques is supported in the GVDE. These include multiple regression, principal component analysis, cluster analysis, factor analysis, discriminant analysis, contingency table analysis, general linear models, survival analysis, log-linear models, multi-level models, and structural equation models.

Spatial statistical techniques for area-based data: All major geospatial analytical techniques for area data are supported in the GVDE together with procedures to ensure that the aggregation level is adequate for geoprivacy protection. These techniques include various area-based measures of spatial association and spatial cluster analysis methods (e.g., Anselin's (1995) local indicator of spatial association (LISA), Moran's I, Geary's C, and Getis's Gi), and a suite of spatial regression models.

Geospatial methods for point-based and linear data: Many geospatial methods for point-based and linear data can generate results, in most cases, that do not reveal the original locations of the records in the dataset. These methods include geographically weighted regression, kriging, spatial point pattern analysis, spatial cluster analysis, kernel density estimation, and the K function. However, in some data sharing situations, users may need to see the original point locations during the analytical process (e.g., to visually assess the spatial distribution of the points). We develop and test a set of procedures for visualizing point locations while preventing disclosure of the identity of the subjects, masking the point locations and evaluating cartographic output to examine the effect of different parameters (e.g., bandwidth and impedance functions that model the effect of distance decay) on disclosure risks, and testing the value of these procedures.

Geographic Information Systems (GIS) tools for data management and analysis: The suitability of GIS data management tools, including ArcGIS and standalone open-source GIS, is evaluated and implemented in the GVDE. In the ArcGIS environment, each geospatial data protection method is implemented through scripting tools. Each method is implemented in the form of a Python script that uses the ArcGIS Python package (ArcPy). Through ArcPy, a script can access the geoprocessing environment of ArcGIS and can be easily distributed and reused without programming knowledge on any computational platforms with ArcGIS installed. In addition, the geospatial data protection methods is developed using standalone Python code based on open-source mathematics and GIS libraries, such as NumPy (doing mathematical calculations), SciPy (including data processing, optimization and statistics), Shapely (manipulating and analyzing geometric objects), and GDAL (processing vector and raster data formats). These standalone open-source tools run in the Python environment with the open-source library installed in the GVDE (users working in the GVDE cannot directly call on web-based software). The geoprocessing environment of ArcGIS also offers many user-friendly functions such as selection set support, validation of inputs, error messaging, and recording of history, as do some open-source GIS environments.

II. Evaluate and implement masking and encryption capabilities for the GVDE

Many users may wish to *export* the results of their analyses via maps or other visualizations from the GVDE, for use in publications or presentations. This component of the project examines appropriate anonymization methods and related disclosure risks for several types of geospatial data (point, line, polygon, raster, vector, etc.) as well as for rapidly growing new sources of confidential geospatial data such as GPS trajectories and geospatial data confidentiality issues pertaining to crowdsourced and social media data. The GVDE technologies have the ability to apply geomasking and encryption methods as well as other techniques to anonymize the data in order to protect the identities of human subjects when users need to extract maps or data analysis summaries and graphics from the GVDE.

Masking and Encryption Techniques

The GVDE research team evaluates multiple geomasking methods, encryption, and other processes to enable users to anonymize and then export *visualizations, maps, or analyses* derived from confidential geospatial data from the GVDE (after review) for use in publications or presentations. Below are some of the geomasking and encryption techniques that we evaluate for inclusion in the GVDE.

Geomasking - Geomasking techniques modify and hide the original location in georeferenced data by adding statistical noise to the original data (Kwan et al., 2004; Armstrong and Ruggles, 2005; Leitner and Curtis, 2006). By masking the locations in a data set, users may still use illustrations that include the locations of subjects' homes or workplaces in their maps or geovisualizations when publishing their results, while protecting their geoprivacy. Various geomasking methods have been developed to date (Armstrong et al., 1999; Kwan et al., 2004; Chen et al., 2008; Zimmerman et al., 2008; Zandbergen, 2014; Zhang et al., 2017), and are described below.

In *aggregation*, data may be grouped by areal units (areal aggregation) or multiple individual records can be assigned to one point-location (point aggregation). Further, aggregate patterns can be used to make it impossible to identify individual subjects (pattern aggregation, e.g., hot spot maps). (b) An *affine transformation* translates, contracts, or expands a point pattern. For instance, the scale of the point pattern may be altered so that relative positions and orientations between locations are maintained while the location pattern's relation to the study area is modified (re-scaling). Alternatively, all locations may be shifted a determined distance and direction from their original locations (shifting). (c) A *random perturbation mask* allows both the amount and direction of spatial displacement to vary between points, thus altering the relative locations and orientation of the points in a particular point (or location) pattern. For instance, each point may be randomly placed along some line feature, such as a circle defined by a center at the original point and a chosen radius (circular masks). The size of the perturbation circle

could be weighted by the population density at each point (weighted masks) in order to take into account its effect on the risk of disclosure (as lower population density in an area leads to higher disclosure risk). (d) A *donut mask* is similar to random perturbation within a circle, but in this method a smaller inner circle is also created inside a larger outer circle, creating the “donut,” and the perturbed location is then placed outside of this smaller circle but inside the larger one (Zhang et al., 2017). This method thus sets the minimum and maximum distance of random perturbation. (e) In *Gaussian displacement*, the direction of spatial displacement is random while the distance follows a Gaussian distribution. The dispersion of the distribution may vary based on other parameters of interest, such as local population density. (f) *Bimodal Gaussian displacement* is a variation on Gaussian displacement, but it uses a bimodal Gaussian distribution for the random distance function. In effect, this is similar to donut masking, but with a less uniform probability of spatial placement (Zandbergen, 2014).

(g) The *location swapping* method “replaces an original location with a masked location selected from all possible locations with similar geographic characteristics within a specified neighborhood” (Zhang et al., 2017). (h) An extension of this method is *location swapping with donut*, which uses the same method as location swapping, but like donut masking, a smaller internal circle within which points cannot be displaced is employed. The radius for creating the inner and outer circles can vary based on local population density.

Geomasking techniques may be applied either to the data before analysis or to the products (e.g., maps) after analysis. Most users prefer post-analysis masking. Since all masking procedures change the data in some way, pre-analysis masking may affect the results of georeferenced individual-level data analysis, obscuring important geographic patterns. The analysis of confidential individual-level geospatial data is important for understanding critical social and policy issues. Our GVDE technologies provide users with access to detailed geographic data *inside the GVDE* so that they can conduct their analyses using this preferred approach, and then offers tools for making their output (e.g., maps and analytical tables) safe before removing it from the GVDE.

Geospatial cryptography (Encryption) - Using cryptographic techniques to protect confidential geospatial data can be achieved in many ways including “transforming all data to a different space using cryptographic techniques so that they can be mapped back to spatial information only by the user” (Andrés et al., 2013). For example, Clarke (2016) employs forward and inverse algorithms to mask point-based data by switching the digits of data coordinates. Jacquez et al. (2017) have proposed that, using cryptographic techniques, it is possible to design and implement geospatially encrypted geographic information systems (GEGIS) to promote the sharing and spatial analysis of confidential data. For instance, one may conduct geospatial analysis in an encrypted space using original geographic coordinates and report the results without revealing individual locations. These and other basic geospatial cryptography techniques are evaluated for feasibility and appropriateness for the GVDE.

Evaluation of confidential geospatial data protection methods

The GVDE research team has evaluated the effectiveness of the confidential geospatial data protection methods by assessing how the disclosure risk and analytical utility of each analytical or visualization outcome has been altered. To do so, we conduct simulation experiments to generate estimates of disclosure risk and utility for various geomasking and encryption methods and analyze the trade-off between the quality of analytical results and preservation of confidentiality for sensitive geospatial data (e.g., comparing the results generated by masked and unmasked data). The GVDE team evaluates patterns of disclosure risk for maps after the application of a masking or encryption method to determine levels of confidentiality risks (Kwan et al., 2004; Zandbergen, 2014). In addition, raw and masked data undergo spatial statistical analysis (Haining, 1990; Bailey and Gatrell, 1995; Anselin, 2013; Chun and Griffith, 2013) to identify and compare geographic patterns that could be relevant to data confidentiality (e.g., spatial distribution and clustering patterns). Using a “baseline” model the GVDE research team compares statistical inferences generated from a subsequent “masked” model to see whether perturbed coordinates shift enough to result in different beta coefficients. Based on the evaluation results concerning the suitability, efficacy, and efficiency of each geomasking or encryption method (e.g., what types of geospatial data protection methods are suitable and effective for what types of data and analyses), guidelines in the form of a suitability matrix and

guidebook are prepared and disseminated to the GVDE users, including governmental agencies and the broader community.

III. Develop the GVDE user credentialing system

The development of an innovative credentialing system is an integral component of the GVDE technologies. The credentialing system provides a transferable digital identifier to trusted users of the GVDE, to ensure responsible data stewardship and the protection of confidential geospatial data. Our GVDE technologies build on a foundation of preliminary work by the research team to develop a standardized and broadly accepted system of user credentialing. That previous research identified significant discrepancies in rules and modes of access to confidential data, and even to the language and definitions to describe data confidentiality, modes of access, and user requirements. This GVDE project proposes standardized language and best practices regarding data confidentiality, user requirements, and modes of access for confidential geospatial data. It implements these criteria into a system of digital identities that control access to data in the GVDE, so that the system verifies, at login, that a particular user has the appropriate credentials for accessing the requested datasets. This system increases the willingness of potential geospatial data producers, including governmental agencies, to share data, because they can have confidence in its security, and the ability of users to more readily undertake creative analyses with the least possible risk to privacy and confidentiality.

The output from our GVDE technologies' component is threefold. The *first* is a recommended matrix system of user credentialing. These recommendations draw on interviews with organizations delivering geospatial data as well as survey, administrative, and other non-designed data. In order for a system of user credentialing to be successful, it must reflect the concerns of data providers and funders. In many cases, data providers distinguish citizens or residents of a particular country; "legitimate" users, journalists, and commercial entities; researchers at institutions with Institutional Review Boards; those subject to subpoenas or Freedom of Information Act requests; and those with the legal ability to submit to the requirements of the data custodian. The conditions under which users access data depend on the interaction between data and user characteristics. In addition to describing current practice and identifying areas of overlap and difference, the GVDE team identifies steps to move the field forward in terms of best practices and required elements for a system of standardized user credentials.

The *second* component is software and an interface that can be used to match user credentials (a passport) with dataset requirements (captured in a dataset-specific visa) in order to access the GVDE. When a user wishes to use a particular dataset, their information is compared with the requirements associated with those data. The GVDE team issues a "visa" when a user is approved to access a particular dataset. The GVDE technologies verify the passport and visa before providing the user access to the data in the secure computing space. In order for the system to work properly, the credentials necessary for the user, the user's institution, and the data themselves have to be clearly defined – which means that the interface has to be user-friendly and intuitive, and the software has to provide a bridge that performs "checks" for matches between the user's information and the data requirements. Our GVDE technologies may also provide a foundation for broader applications beyond the GVDE, which would allow user access to restricted datasets held at multiple repositories.

Thirdly, users need training in awareness regarding the availability of the user credential and in geospatial confidentiality protection itself. The GVDE research team offers training to users and other stakeholder groups, including governmental agencies, to build community understanding of the process and value of geospatial user credentialing. Users who successfully complete this training receive a digital badge identifying their knowledge in geospatial data stewardship. This badge is recognized by the user credentialing digital access system. This outreach and training ensure that the social and technical infrastructure changes described above make the process of accessing and using geospatial data with potential disclosure issues more conducive to data sharing.

IV. Ensuring sustainability of the GVDE

To ensure sustainability and to build the usage of the GVDE technologies, we undertake extensive training, outreach, and dissemination activities. The GVDE technologies are maintained as part of the larger portfolio of data management and stewardship services that the GVDE research team provides to the user community, including governmental agencies. As the GVDE is scalable, the GVDE system has the capacity to meet expanded demand. The most efficient way to expand capacity is to move to the cloud. This also requires careful attention to security issues, of course. The GVDE research team is now largely operating in a secure Amazon Web Service environment which provides the GVDE with a FISMA compliant and ATO (authority to operate) FedRAMP compliant platform; that is, it is approved for uses by DoD, IRS, and Census. The GVDE research team has a similar contract with Azure, the Microsoft cloud environment, so it is not tied to a single provider.

GVDE training programs

The GVDE research team is a leader in offering regular training to users regarding data management, handling, ethics, and analysis. The research team has conducted numerous training and workshops under grants from the NSF and the NIH and are highly experienced in developing training materials and conducting training directly related to these technologies.

The GVDE team also develops training and technology transfer modules for using the GVDE system, and on sharing and analyzing confidential geospatial data within the GVDE. These training modules support users in the use of the GVDE system (including user support and frequently asked questions), the GVDE credentialing process, articulating NSF data management plans, and ethics related to the handling and use of confidential geospatial data. All training activities are assessed based on the number and diversity of participants and trainee evaluations (e.g., anonymous feedback forms, group discussions, and post-event follow up). Evaluations cover both content understanding and retention, as well as the effectiveness of training materials and instruction. Trainees are asked to assess the quality of masking tools, usefulness of results, ease of use, and to comment on their general satisfaction with the GVDE system and information provided. Trainees are recruited from multiple sectors based on their needs for using geospatial data, and our selection process ensures diversity among our participants. The training materials and activities help to develop usage of the GVDE technologies.

Contact or Correspondence:

Douglas Richardson, PhD
Distinguished Researcher
Center for Geographic Analysis
Institute for Quantitative Social Science
Harvard University

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Stealth Software Technologies Inc

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



Steve Lu
CEO

Stealth Software Technologies, Inc.
12100 Wilshire Blvd. 8th Floor
Los Angeles, CA 90025

July 8, 2022

Submitted via email to PETS-RFI@nitrd.gov

Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, D.C. 20504

Re: Notice of request for information on Advancing Privacy-Enhancing Technologies

Dear Sir/Madam:

Stealth Software Technologies, Inc. (“Stealth”) appreciates the opportunity to comment on the Office of Science and Technology Policy’s request for information on Advancing Privacy-Enhancing Technologies (the “OSTP RFI”). Stealth is a small business co-founded by Dr. Rafail Ostrovsky, a Distinguished Professor of Computer Science and Mathematics at UCLA. Our team consists of experts in cryptography, computer science, and mathematics; experienced software engineers; and numerous consultants who are leading academic researchers. Stealth has performed on numerous Federal Government contracts over the past decade, for agencies and federal entities including DARPA, IARPA, ONR, and NIST.

Our comments below address several of the specific topics presented in the OSTP RFI, on a topic-by-topic basis.

Topic 1: Specific research opportunities to advance PETs

As a small selection of the many potential Federal research initiatives that could substantially accelerate and improve both the development and adoption of privacy-enhancing technologies (PETs), we identify some particularly productive opportunities for research programs and strategies to develop such programs:

- A primary barrier to adoption of PETs—particularly those based on modern cryptographic tools such as secure multi-party computation or homomorphic

encryption—is that, at present, even the *use* of such technologies to their full privacy-enhancing potential frequently requires highly specialized expertise and extensive training. A key contribution to overcoming this barrier would be the development of a *programming language and software stack for expressing privacy-preservation requirements* designed for users who are *not experts in cryptography*; such a programming language would enable users to obtain optimally privacy-preserving and application-appropriate cryptographic protocols while shifting the burden of cryptographic expertise to the developers of the language.

Toward this end, in 2019 the Intelligence Advanced Research Projects Activity (IARPA) created a broad agency announcement program titled “HECTOR” aimed at developing such a software stack. Both the HECTOR program itself and its source-selection process required a partnership between programming-language experts, systems experts, and cryptography experts. Although the four selected performer teams were first-rate and the first year of execution resulted in substantial accomplishments, the program was unfortunately canceled. We believe that resuming such an interdisciplinary research and development effort would substantially advance the field of PETs—particularly with a view toward the critical aspects of *usability* and *accessibility*.

- A frequently encountered limitation of PET tools—particularly those that provide *exact* computations with strong privacy guarantees through cryptographic protocols—is that their memory requirements are currently too great for many common real-world applications. Such limitations arise particularly often in highly computationally intensive applications to artificial intelligence and machine learning on large data sets, and a targeted research effort to enhance the performance of cryptographic PETs, such as secure multi-party computation and fully homomorphic encryption, on large data sets is very much needed.

A primary reason for this excessive memory consumption is that such PETs usually require computer programs to be converted into digital circuits for secure execution. Theoretical techniques enabling random access memory (RAM)-based secure computation, thereby avoiding the need for conversion to circuits, do exist; however, practical solutions for implementing these theoretical techniques and integrating them with existing data systems and infrastructure—such as large data-base management systems like Apache Spark, large training sets for artificial intelligence and machine learning applications, data analytics platforms, and data visualization formats—are missing. Again, research targeted toward such practical implementation of cryptographic theory are critical to advancing the usability and accessibility of PETs.

- The Defense Advanced Research Projects Agency (DARPA) has been instrumental in administering programs advancing fundamental and applied research in cryptography for the purpose of enhancing privacy. As evidence of its impact, leading PETs including work in secure multi-party computation such as Prio, Poplar, and Prio+; work in fully homomorphic encryption including PALISADE, and works in differential privacy that intersected with the 2020 decennial United States Census have all been funded in part by

various DARPA programs. Although research advancing privacy and confidentiality fits naturally within the Department of Defense in the interest of national security, it is clear that privacy considerations present common to diverse agencies across the Federal Government. Programs for PET research, tailored toward the activities and goals of specific Government entities, should be more widespread and funded by agencies beyond the Department of Defense. To this end, actively recruiting program managers with expertise in privacy and cryptography into research agencies in other departments is crucial to ensure that PETs advance further.

Topic 2: Specific technical aspects or limitations of PETs

Decision tools for PET selection

A common difficulty voiced by interested potential adopters of PETs is that they do not know where or how to begin in their efforts to select optimal technologies for their particular applications. Due to the concurrent emergence of various cryptographic, statistical, and hardware-based approaches to PETs in recent years, there has been significant confusion regarding which PET techniques to use for which purposes. This confusion is well founded—indeed, the various techniques and tools comprising the wider category of PETs are meaningfully distinct, and PETs are not fungible; depending on the specific application scenario or requirements of a given use case, some PETs may not even be meaningfully applicable.

This difficulty is of course not unique to PETs—all systems are to some extent sensitive to subtle differences in requirements—but many of the underlying issues have largely been adequately resolved in other technological areas, in which principled, requirements-based solution selection can be accomplished using many basic tools that are currently lacking in the field of PETs. These include accessible, well-designed, and widely available decision aids including decision trees and flowcharts, guidelines and best practices, pillars of success, knowledge bases and knowledge diagrams, and general systematizations of knowledge; such tools can be compounded with prior successful approaches to applications through the use of white papers, case studies, and roadmaps. The field of PETs would be advanced significantly through the development of such decision aids for the adopting public.

Limitations of particular PETs

Each particular PET approach has drawbacks, and understanding these drawbacks is critical to inform the optimal solution selection addressed above. Differential privacy, for example, relies inherently on the introduction of noise to data in order to ensure privacy, resulting in some likely amount of distortion of computational output. Accordingly, differential privacy is an inappropriate technology for applications in which an *exact* computation is needed, and users seeking privacy in such applications should turn to secure computation frameworks such as secure multi-party computation or homomorphic encryption. These latter approaches guarantee that inputs to a computation remain hidden, while the exact result—and only the result—is revealed to specified parties; it is frequently not appreciated by general audiences that such

cryptographic techniques do not involve “fuzzing” of data to guarantee privacy. Differential privacy is similarly ill-equipped to handle database queries asking whether there exists a database entry satisfying a specific condition, and secure computation frameworks are again more appropriate. Inversely, for applications in which privacy is compromised by inferences that can be drawn from the exact results of an analysis, differential privacy is a more appropriate tool.

Finally, we note certain significant technical differences between secure multi-party computation and homomorphic encryption. So-called *fully* homomorphic encryption—which is required to handle *arbitrary* computations—necessarily requires a conversion to digital circuits for secure computation, whereas, as discussed in our response to Topic 1 above, there exist techniques for executing secure multi-party computation using random access memory. Additionally, homomorphic encryption requires key-management mechanisms and trust in another party holding the decryption key, whereas secure multi-party computation essentially accomplishes decryption “within the system” through a protocol, at the cost of requiring interactivity and in some cases trust assumptions of non-collusion.

Finally, with respect to hardware-based PETs such as secure enclaves, although the application of such technologies to software attestation has been a major success, maintaining privacy of computations inside enclaves has been an elusive goal, and such technologies have routinely been compromised by novel attacks.

Interaction of PETs and blockchain

Blockchain technology is another cryptographic paradigm that has significantly increased in visibility in recent years; it enables an immutable recording of data at a large scale, applicable not only to data held by a single entity but also across multiple organizations and data fields. A combination of the privacy afforded by PETs with the auditability afforded by blockchain technologies, when applied to Government records, could potentially enable substantial advances in preventing, identifying, and mitigating fraud and abuse—it should be stressed that although auditability and privacy may initially appear to be competing or even mutually contradictory goals, the appropriate use of PETs that enable *computation* on private data can enable these goals to be realized simultaneously.

Topic 4: Specific regulations or authorities that could be used, modified, or introduced to advance PETs

In the near term, any new regulations or modifications to existing regulations mandating or otherwise controlling use of PETs should be carefully drafted to remain appropriately flexible and technology-neutral, in order to avoid prescribing the use of particular PETs before the technologies—and their optimal spheres of application—have been sufficiently standardized. Indeed, as addressed in our comments on Topic 6 below, given the relatively nascent and rapidly evolving nature of PETs as practically applicable technologies, rigid standardization itself may be premature at present. Moreover, until PETs become more accessible through widespread

adoption and awareness, strict regulatory mandates for their use in the private sector may be logistically or economically infeasible for many regulated entities. However, appropriately flexible initial measures short of highly specific regulatory mandates can further such adoption and awareness. In this direction, requirements in the Federal Acquisition Regulations requiring appropriate use of PETs by contractors in data management or in deliverables to the Government would be welcome, but—particularly prior to standardization—such requirements should be crafted to allow contractors to incorporate PETs feasibly and in a genuinely beneficial manner.

There may be significant opportunities to encourage PET adoption by issuing guidance clarifying the application of existing regulations to appropriate uses of PETs. Particularly in sectors covered by existing privacy-regulation regimes, disruption of established compliance frameworks and particularly uncertain risks of non-compliance can deter regulated entities from adopting novel technologies, which in many cases may actually provide highly effective technical measures for compliance with privacy regulations if applied appropriately.

Guidance from regulatory bodies can mitigate such deterrents, simultaneously advancing the adoption of PETs and facilitating effective compliance; it may be appropriate to issue guidance specifically devoted to addressing likely uses of PETs by regulated entities and the compliance status of such uses under applicable privacy and security rules.

In the context of financial regulation, the Federal Trade Commission (FTC) Privacy of Consumer Financial Information Rule¹ (“Privacy Rule”) and Standards for Safeguarding Customer Information Rule² (“Safeguards Rule”), covering financial institutions under the Gramm–Leach–Bliley Act, are prime examples of regulations that could benefit from such clarifying guidance. As a particular example, the Privacy Rule generally prohibits covered entities from disclosing customer account numbers to nonaffiliated third parties for marketing purposes, but specifically provides that “[a]n account number . . . does not include a number or code in an encrypted form, as long as [the covered entity does] not provide the recipient with a means to decode the number or code.”³ The phrase “encrypted form” is not defined directly in the Privacy Rule; although the Safeguards Rule’s definition of “encryption” could arguably apply to, for example, the use of account numbers in a secret-sharing-based secure multi-party computation protocol,⁴ there may be sufficient ambiguity in the language to deter a covered entity from considering such use—even if the entity were willing and able to undertake the combination of regulatory and technical analysis required to assess it. By contrast, the applicability of the Privacy Rule’s exception to an analogous computation involving account numbers using homomorphic encryption, with substantially similar privacy and security guarantees, would be considerably less ambiguous due solely to terminology. We offer no opinion in this comment regarding how the Privacy Rule or Safeguards Rule *should* apply to any specific application of any particular PET. We simply note

¹ 16 C.F.R. § 313.1 (2021).

² *Id.* § 314.1.

³ *Id.* §§ 313.12(a), (c)(1).

⁴ *See id.* § 314.2(f) (defining “encryption” for purposes of the Safeguards Rule as “the transformation of data into a form that results in a low probability of assigning meaning without the use of a protective process or key, consistent with current cryptographic standards and accompanied by appropriate safeguards for cryptographic key material”); *id.* § 313.3 (containing no definition of “encrypted” for purposes of the Privacy Rule).

that guidance from the FTC clarifying the application of these rules to foreseeable applications of a diverse variety of PETs would be welcome, particularly to the extent that it encourages beneficial uses of PETs.

In a potentially related vein, the President’s July 2021 Executive Order on Promoting Competition in the American Economy⁵ has encouraged the Consumer Financial Protection Bureau (CFPB) to consider rulemaking under section 1033 of the Dodd–Frank Wall Street Reform and Consumer Protection Act.⁶ Section 1033 directly addresses portability and sharing of consumer financial transaction data rather than computation involving such data, and we offer no opinion on the scope of the CFPB’s section 1033 rulemaking authority. However, to the extent that data-privacy considerations are implicated in any such rulemaking, they should be informed by an awareness of the potential applicability of PETs to sharing of and analytics and other computation on consumer financial data.

PETs may have beneficial applications to anti-money laundering (AML) and know-your-customer (KYC) activities, potentially allowing for both effective execution of AML and KYC policies and demonstration of compliance with such policies while preserving the privacy of sensitive data. Accordingly, rulemaking and enforcement bodies, notably the Financial Crimes Enforcement Network (FinCEN) and the Office of Foreign Assets Control (OFAC), should remain aware of the evolving state of the art in PETs and, if appropriate, issue guidance concerning such uses of PETs. As a specific example, PETs could potentially be relevant to FinCEN’s Customer Due Diligence Rule, particularly with respect to the requirement that regulated institutions implement and maintain customer due diligence procedures for “[u]nderstanding the nature and purpose of customer relationships for the purpose of developing a customer risk profile” and ongoing risk monitoring.⁷

In the realm of health care regulation, PETs-specific guidance from the Department of Health and Human Services (HHS) similar to that proposed above from the FTC would be welcome, in the context of the HHS Privacy of Individually Identifiable Health Information Rule (“Privacy Rule”) and Security Standards for the Protection of Electronic Protected Health Information Rule (“Security Rule”)⁸ under the Health Insurance Portability and Accountability Act (HIPAA). In particular, by facilitating joint computation involving protected health information (PHI) without requiring the direct sharing of such information, PETs can further the goals of the “minimum necessary standard” of the HIPAA Privacy Rule, which restricts the use and disclosure of PHI to that necessary to satisfy a particular purpose or carry out a given function.⁹ However, uncertainty regarding the application of the HIPAA Privacy and Security Rules to particular applications of PETs may deter beneficial adoption of privacy-preserving analytical tools for treatment and research. As one particular example, the Privacy Rule allows a covered entity to use or disclose a “limited data set,” defined to exclude certain direct identifiers of individuals, for

⁵ 86 Fed. Reg. 36,987 (July 14, 2021).

⁶ 12 U.S.C. § 5533 (2018).

⁷ 31 C.F.R. § 1020.210(b)(2)(v) (2021).

⁸ 45 C.F.R. §§ 164.102 *et seq.* (2021).

⁹ *See id.* §§ 164.502(b), 164.514(d).

research, public health, or health care operations subject to a suitable data use agreement with the recipient of the limited data set.¹⁰ Such direct identifiers may be particularly useful, perhaps even practically essential, to *link* data sets held by different covered entities; various PET tools and techniques can be used to achieve such linking without revealing any information about the underlying direct identifiers, significantly expanding the scope of available analytics for research and public health activities. However, it may be unclear whether such a use would be permissible under the regulations’ definition of “use.”¹¹

Topic 5: Specific laws that could be used, modified, or introduced to advance PETs

Although at present we have no specific proposals for new legislation that could be introduced, we support the existing bills introduced for the Promoting Digital Privacy Technologies Act,¹² the Digital Accountability and Transparency to Advance Privacy Act (“DATA Privacy Act”),¹³ the Algorithmic Accountability Act,¹⁴ and the National Secure Data Service Act¹⁵ in furtherance of a national strategy for advancing PETs.

Regarding data sharing—or more accurately, in this case, data *processing*—among international entities, we specifically note guidance from the European Data Protection Board explicitly recognizing secure multi-party computation as an “effective supplementary measure” under the European Union’s General Data Protection Regulation (GDPR) for data processing by independent processors located in distinct jurisdictions, as well as an effective technical privacy-by-design measure to ensure that data are not subject to unauthorized access.¹⁶

Topic 6: Specific additional mechanisms that could be used, modified, or introduced to advance PETs

The National Institute of Standards and Technology (NIST) should play a central role in advancing the development and adoption of PETs in the United States (and, indeed, globally), as recognized in some of the proposed legislation discussed in our comments on Topic 5. As noted in our response to Topic 4, however, NIST should be careful not to set standards recommending

¹⁰ See *id.* § 514(e).

¹¹ See *id.* § 160.103 (“Use means, with respect to individually identifiable health information, the sharing, employment, application, utilization, examination, or analysis of such information within an entity that maintains such information.”).

¹² H.R. 847, 117th Cong. (2021); S. 224, 117th Cong. (2021).

¹³ S. 3065, 117th Cong. (2021); H.R. 5807, 117th Cong. (2021).

¹⁴ H.R. 6580, 117th Cong. (2022); S. 3572, 117th Cong. (2022).

¹⁵ H.R. 3133, 117th Cong. (2021).

¹⁶ European Data Protection Board, Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data, Version 2.0, at 33–34 (June 18, 2021), https://edpb.europa.eu/system/files/2021-06/edpb_recommendations_202001vo.2.0_supplementarymeasurestransferstools_en.pdf; *id.* at 45.

the use of specific PETs—in the form of specific cryptographic protocols or specific algorithmic implementations of particular PET techniques, for example—prematurely. Early adoption of excessively specific standards can stifle innovation in the field of PETs by discouraging openness to new technologies that have developed outside the parameters of such standards; we emphasize again that all aspects of PETs are rapidly evolving, and prematurely articulated standards risk rapidly falling behind the state of the art or, worse still, impeding progress of the state of the art. This situation should be distinguished from that of cryptographic *primitives* underlying many PETs, such as encryption and digital signature schemes, which have been productively standardized; PETs in general are significantly more complex than these basic “building-block” technologies, and as such greater care is required in crafting standards for them.

As such, in the near term, NIST can take a progressively more formal approach to guiding the public regarding the development and use of PETs, beginning with comparatively informal reports and progressing to recommended best practices and guidelines before issuing formal standards. In keeping with our comments elsewhere in this response, we stress that such guidance from NIST would be particularly useful regarding the *optimal selection* of PET technologies for particular types of application, including clear, accessible explanations of the distinct privacy challenges addressed by distinct PETs.

In addition to such guidance and long-term standard-setting efforts from NIST, which can advance development and adoption of PETs at a “macroscopic” level, at a more “microscopic” level short-term prize challenges—spanning a few weeks to a few months—can generate momentum in developing PET-based solutions for specific areas of application and, additionally, raise awareness of PETs in targeted areas of the public or private sectors. Such challenges are particularly useful for bringing together *interdisciplinary* teams of domain experts in the application area and experts in PETs, both from academia and industry. As two examples of productive such prize challenges, we note the iDASH competition addressing privacy-preserving computation and analysis of biomedical data, in particular human genome data,¹⁷ and the United Kingdom Financial Conduct Authority’s 2019 Global AML and Financial Crime TechSprint.¹⁸

Topic 7: Risks related to PETs adoption

We have emphasized throughout this response distinct PETs are optimized for distinct use cases, and indeed a given PET may be entirely *inappropriate* for use in a particular use case. Without deliberate efforts to raise public awareness of not only the existence of PETs in general but also the appropriate use of specific PET tools and techniques, unprincipled adoption of PETs poses two significant risks: (1) an inappropriate PET *approach* (e.g., secure multi-party computation, zero-knowledge proof, homomorphic encryption, differential privacy, or federated learning) may

¹⁷ See iDASH PRIVACY & SECURITY WORKSHOP 2022 – SECURE GENOME ANALYSIS COMPETITION, <http://www.humangenomeprivacy.org> (last visited July 8, 2022).

¹⁸ See 2019 GLOBAL AML AND FINANCIAL CRIME TECHSPRINT, <https://www.fca.org.uk/events/techsprints/2019-global-aml-and-financial-crime-techsprint> (last visited July 8, 2022).

be applied to a particular use case based on a loose understanding that the technology is “privacy preserving”; and (2) even if an application-appropriate PET approach is selected, the user may lack the technical knowledge to understand the privacy guarantees and limitations of a particular *implementation* of that approach, and accordingly may be unable distinguish sound implementations from unsound ones. Our encouragement that NIST play a central role in public guidance regarding PETs is significantly motivated by the need to mitigate these risks.

A related risk is that PETs may be developed or applied without sufficient appreciation or understanding of either the domain of application or the technical aspects of the relevant PETs themselves, particularly as they relate to data privacy and security. Accordingly, it is critical that general policies and specific deployments of PETs be informed by *interdisciplinary* expertise—including, as appropriate, academic experts in the theory underlying PETs, industry practitioners versed in the practical implementation and deployment of PETs, stakeholders with well-informed perspectives on the policy and legal implications of a given use of PETs, and domain experts in the specific area of application. Indeed, given our limited resources as a small technology company focused on the design and development of PETs, it is for this reason that we have declined to opine, for example, on the optimal *substance* of regulations in our comments on Topic 4 above and have simply called for increased regulatory clarity.

Topic 9: Existing barriers to PETs adoption

A primary obstacle to the adoption of PETs is inertia due to the overhead—whether economic or logistical—of integrating novel technologies into existing practices and infrastructures. We believe that the benefits of PET adoption at a societal level—and in many cases at the level of individual adopters—outweigh these costs. But the relative technical sophistication required to understand and trust in the benefits of PETs, let alone to adopt them successfully, can skew the cost-benefit analysis away from the adoption of new technological solutions. The Government can help overcome this barrier by demonstrating the feasibility and efficacy of PETs through adoption in its own operations and by encouraging or requiring their use in appropriate situations.

A second barrier to realizing the full potential of PETs is a potentially narrow focus on privacy of *individuals* when considering privacy-preserving technologies, and an attendant focus on informed-consent frameworks and related privacy-assurance mechanisms. To be sure, considerations of individual privacy are vitally important; we simply caution against a narrow conception of privacy that potentially ignores many beneficial applications of PETs. Indeed, PETs can be applied to overcome data-privacy obstacles to collaboration that are based more on competitive or security concerns than on considerations of individual privacy; PETs can enable, for example, joint computation for activities ranging from collaborative risk analysis and mitigation between competitors to industry-wide auctions. Such applications may be less obviously within the ambit of “privacy” than protection of individuals’ personal data and accordingly may be neglected in discourse addressing PETs; policies for advancing PETs should be designed to realize the full range of benefits that such technologies can offer.

Thank you for the opportunity to inform the development of a national strategy for privacy-preserving data sharing and analytics and associated initiatives; we hope you find our comments useful.

Sincerely,

Steve Lu
CEO
Stealth Software Technologies, Inc.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Syntheticus Ltd

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

All comments are done by Mr. Aldo Lamberti, Founder & CEO of Syntheticus Ltd.
The contact information is:

- Syntheticus Ltd, Mr. Aldo Lamberti, Arnikastrasse 7, 8953 Dietikon, Zürich Switzerland
-

Syntheticus Ltd. is a Switzerland based company that empowers global organizations to democratize access to data, accelerate speed-to-market and mitigate privacy risks by enabling secure collaboration on synthetic data at scale through an enterprise-grade SaaS platform which results in business value and innovation. The Syntheticus Privacy-Enhancing Platform orchestrates the latest technologies such as 'AI-powered Synthetic Data', 'Differential Privacy' and 'Microsoft-based Trusted Execution Environment'.

1. *Specific research opportunities to advance PETs:* Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.
 - a. Better explainability of Deep Learning models ('Generative Adversarial Networks' to generate privacy-preserving synthetic data) to mitigate the "Black Box" assumptions from the users.
 - b. Research opportunities are certainly within the realm of Quantum Computing => how easy will it be to re-identify the original real-world data from the artificially generated synthetic data?

2. *Specific technical aspects or limitations of PETs*: Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.
 - a. Pseudonymized data is not free to use according to GDPR. Solely fully anonymized data or synthetic data are compliant with PII requirements and thus free to use ([Recital 26 GDPR](#)). According to our personal exchanges with the highest European authorities, the European Data Protection Supervisor (EDPS), even encryption-based technologies are seen as 'not free to use' given the high risk of losing the "key" and thus being able to decrypt the sensitive data.
 - b. Internal enterprise constraints are IT silos and legal/compliance which hinder Analytics and Software Dev Teams to get full access to the rights data for their daily jobs. Currently either (i) they don't get any access to sensitive information, (ii) they only get anonymized data with lower utility or (iii) they get access to full sensitive data with high privacy/security risks involved.
 - c. When generating differentially-private synthetic data there is a mathematical tradeoff between utility and privacy. The higher the chosen privacy level, the lower the utility of the synthetic data, and vice versa. The level of privacy can be adjusted by the customer through a parameter (epsilon = privacy budget).
 - d. The mitigation of bias within the underlying real-world data is an issue. The conditional generation of synthetic data can overcome specific biases and create data sets that are not biased anymore and thus can be used for further advanced analytics without harming minorities and its equity.

3. *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:* Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of [Executive Order 14058: Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government](#).

We are active in the following areas with specific applications to bring benefits to stakeholders:

- i. Financial Services Industry
 1. Fraud Detection, Anti-Money-Laundering
 - a. Challenge: Fraud patterns are different between countries, regions, businesses. It's hard for regulators or private enterprises to detect the anomalies in real-time, given the little amount of training data.
 - b. Solution: enterprises and regulators can generate differentially private synthetic data and securely collaborate on that data to better train their individual detection algorithms.
 2. Enterprise Risk Rating
 - a. Challenge: the underlying data is limited and thus the enterprise risk rating is lacking forward-looking scenarios.
 - b. Solution: by generating an augmented synthetic data set and creating artificial edge cases, those scenarios can be tested right away.

3. Lending Risks

- a. Challenge: When banks run their Machine Learning analytics to classify their customers or to predict default risks, there are inherent biases, such as from gender or ethnicity.
- b. Solution: With the help of synthetic data these biases can be equaled out by generating datasets that have parity between the biases => 50% men and 50% women in a dataset, while mitigating the privacy risks.

ii. Healthcare

1. One of the most sensitive data is our own health data, which in turn is also of utter interest to unlock value from it for our personal and societal well being
2. We are active in various Horizon Europe consortia around the topic of opening up access to health data for augmented insights while mitigating privacy risks through synthetic data. One of our partners is Prof. Chou from Stanford University.

iii. Telecommunications

1. TelCo providers are under constant pressure to digitalize and enhance their products and services. During their app development their engineers obviously need to test the apps before launching to customers. The access to real-world data is limited, and thus the testing is mediocre which results in customer churn => synthetic data is used instead of real-world data for a neatless and compliant app testing for improved apps and happy customers.
2. Synthetic data allows TelCo providers to interchange between each other insights about customer behavior or business logics to prosper the ecosystem while mitigating privacy risks.

4. *Specific regulations or authorities that could be used, modified, or introduced to advance PETs:* Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking

authorities under the Office of Management and Budget, the Federal Trade Commission, and **financial regulatory bodies**, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (e.g., Other Transaction Authority) to procure PETs for development.

- a. Office of Management and Budget
 - i. United States Digital Service
 - 1. Challenge: in order to improve federal websites and its applications, a thorough testing workflow needs to be in place. For software testers, internal or external (in foreign jurisdictions), access to test data is mission-critical. However, sharing sensitive test data should be avoided in order to avoid privacy breaches
 - 2. Solution: synthetic data is used as test data instead of real-world citizen data. Software testers can build better products and maintain them, while mitigating privacy risks.
- b. Federal Trade Commission
 - i. Bureau of Consumer Protection
 - 1. Challenge: private organizations are applying more and more advanced analytics (Machine/Deep Learning) to unlock insights and value from its collected consumer data. The used algorithms are private and won't be disclosed.
 - 2. Assumption: The Bureau of Consumer Protection has little overview over the used algorithms and whether they're benign or malign versus the consumers.
 - 3. Solution: The Bureau of Consumer Protection can install a synthetic data sandbox and then let the private algorithms perform on it to create a market benchmark and possibly detect outliers that potentially are working against the wills of the consumers.
 - ii. Office of Technology Research and Investigation
- c. SEC U.S. Securities and Exchange Commission
 - i. The Financial Conduct Authority (FCA) from the United Kingdom started with pilot projects on the exploration of synthetic data and its benefits for the financial sector one year ago.
 - ii. According to the FCA, the benefits are manifold and promise to unleash the synthetic data economy:

1. enable greater data sharing for the purposes of competition, without undermining data protection laws that are in place to protect consumers.
 2. alleviate the data access challenge by making synthetic financial data available to participating firms
 3. enable more effective testing and product development
- iii. Therefore, the FCA opened up a 'Call for Input' on specific questions around synthetic data to explore market attitudes towards synthetic data, and its potential for opening data sharing between firms, regulators and other public bodies. Syntheticus answered to this 'Call for Input' with their deep expertise in the field.
 - iv. The SEC can build upon the previous experiences of the FCA and explore the setup of synthetic data within the financial sector.

5. *Specific laws that could be used, modified, or introduced to advance PETs:* Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.

6. *Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:* This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and

territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

- a. The introduction of a standard on the quality of synthetic data is of utter importance. The questions from the public and private sphere are most of the time around the guarantee (both technical and legal) on how good synthetic data is. For that, we're in close contact with the highest European organization, the European Data Protection Supervisor (EDPS), to work on some initial thoughts about a potential standard.
- b. The 'Privacy Shield' with the EU is an on-going issue that needs to be reflected on.

7. *Risks related to PETs adoption:* Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

- a. The risk analysis of synthetic data is an ongoing research field, given the above mentioned trade-off between utility and privacy.
- b. We at Syntheticus implement quality assurance and validation steps after every synthetic data generation. However, there could still be the risk of re-identification of the original data through 3rd party data and dedicated inference attacks. Such a re-identification would cause tremendous damage to the individuals or business information present in the original real-world dataset.
- c. There is no 'silver bullet' PET that can deliver the highest utility without any privacy risk. It's a balance act and needs to be carefully taken into consideration for every individual use case. Anyhow, policy and governance measures need to be in place and also enforced, in order to mitigate the risks.

8. *Existing best practices that are helpful for PETs adoption:* Information about U.S. policies that are currently helping facilitate adoption as well as best practices that facilitate responsible adoption. This includes existing policies that support adoption, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This also includes information about where and when PETs can be situated within tiered access frameworks for accessing restricted data, ranging from publicly accessible to fully restricted data.

9. *Existing barriers, not covered above, to PETs adoption:* Information about technical, sociotechnical, usability, and socioeconomic barriers that have inhibited wider adoption of PETs, such as a lack of public trust. This includes recommendations on how such barriers could be overcome. Responses that focus on increasing equity for underserved or marginalized groups are especially welcome.
 - a. Gartner => Target market adoption for synthetic data is <5% for enterprises (Innovation Insight for Synthetic Data, Choudhary Farhan, Published 7 February 2022)
 - b. The privacy breaches throughout the last 10 years have been tremendous. But still, the enforcement of fines is lacking on a large scale. If there's no enforcement, privacy breaches remain a small risk in the overall enterprise calculations. The EDPS, with its leader Wojciech Wiewiórowski, is trying to counteract that topic.
 - c. Given the global privacy breaches, public trust has suffered. The trust needs to be re-established first - PET's will play a crucial role in this endeavor.
 - d. A global, or at least a more regional legal framework concerning privacy is needed. The EU, with the GDPR, takes a leading role in that. Other regions are taking similar approaches (China, India, Brazil). The USA instead is rather fragmented, having privacy regulations on state level

(California, New York, Massachusetts, etc), and also on specific sectors (HIPAA for health data or GLBA for financial data). This could be a recommendation to overcome barriers of technical and legal usability.

10. *Other information that is relevant to the adoption of PETs*: Information that is relevant to the adoption of PETs that does not fit into any of the topics enumerated above.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

The Ethical Tech Project

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



July 8th, 2022

To: Office of Science and Technology Policy

From: The Ethical Tech Project

Re: **RFI Response:** [Advancing Privacy-Enhancing Technologies](#)

To Whom It May Concern:

[The Ethical Tech Project](#) is an advocacy group that works across the worlds of business, academia, and government to build ethical data and privacy practices in organizations of all types and sizes.

Our vision is a future where technology supports human flourishing, where everyday consumers' rights are aligned with business practices such that data can build better products, services, and experiences while still respecting the dignity of consumers and citizens.

Privacy-Enhancing Technologies (PETs) are essential to achieving this vision and bridging the gap between regulations and standards set by policymakers, legal and compliance experts, engineers, product managers, and technologists on the ground building.

PETs will not see broad adoption unless regulators work in concert with technologists that build, maintain and own the systems responsible for enforcing the regulations. Whether it is a business, a government or state actor, or an NGO, the bottom line of the privacy problem is that someone somewhere is engineering a system around data that must have privacy-by-design. We believe that the only way to ensure this happens within the context of our market system is through the mechanism of voluntary Privacy Standards that achieve a critical level of adoption. We look forward to offering the details in the following responses to your RFI on Advancing Privacy-Enhancing Technologies.

Sincerely,

The Ethical Tech Project

Responses

1. Specific research opportunities to advance PETs

No comment.

2. Specific technical aspects or limitations of PETs

PETs do not exist in a silo: they must be part and parcel of a concerted architecture that includes established tech already broadly in use today in order to accomplish the specific demands of a privacy-by-demand system. At the Ethical Tech Project, we have identified **five requirements** for data users (businesses and other organizations that make use of individual data) to respect the permissions of data owners (individual citizens that are the originators of personal data).

Permission – the granular conditions for the use of every datum for every user and every purpose – is fundamental to privacy and the notions of Data Control (the opportunity for individual data owners to have the choice over what happens to the data that originates with them) and Data Dignity (the concept that data owners must have negotiation power over the use of their data, enabled by Data Control). The technical aspects of PETs must possess *all five* of these requirements in order to comprehensively respect permission:

- A. **Purpose** - Make purpose a ‘first-class citizen’ in the consideration set for processing data through systems to **declare, enforce, and audit** permissions.
- B. **Control** - Afford data owners the means to control their data through the **granting, revoking, and enforcing** of permissions and the ability to execute data **control operations**.
- C. **Recognition** - Make explicit the **recognition** and identification of all entities participating in the data transaction with associated **registration, verification, and revocation** procedures.
- D. **Transmission** - Support transmission of instructions and permissions from end to end across data supply chains through **subscription and broadcasting** procedures across the chain and **auditing and enforcement** procedures within each link.
- E. **Rectification** - Take remedial steps to **rectify** instances when permissions or instructions are not respected and **monitor and alert** for such instances.

How do these requirements translate into a technical architecture? At a technical level, privacy-by-design is in practice applied through a robust system across Apps, Services, and an organization’s data stack in order to protect the individual’s privacy rights (please refer to our “Privacy Stack” diagram in the Appendix). The components of the “Privacy Stack” exist in at the application and service levels, and comprise existing off-the-shelf solutions such as Identity Management that must work in concert. Our delineation of those components are:

- **Apps**
 - Data Risk Management
 - Responsive Regulatory Compliance
 - Automated Subject Rights Fulfillment
 - Consent and Preference Management
- **Services**
 - Policies for the Access and Use of Data
 - Identity and Entity Management
 - Granular Data Permissioning
 - Privacy CMS
 - Data Discovery and Classification
 - Access Control
 - Data Control
 - Integrations

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

No comment.

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs

No comment.

5. Specific laws that could be used, modified, or introduced to advance PETs

No comment.

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

We are advocates for standards, coupled with reference architectures, as a mechanism to make PETs easily adaptable. In other areas where data is sensitive or regulated, standards such as PCI-DSS, SOC 2, and others have achieved broad adoption and made it simple to understand how compliance may be reached. Even if achieving compliance is arduous, standards provide a template to snap to and give industry a space to innovate around a shared understanding and concrete expectations of what to build.

Privacy Standards are already on the market, but they face three challenges when it comes to adoption by engineers building products:

1. Deriving **technical specifications from otherwise non-technical requirements** is difficult - better standards must be flexible while still offering engineers a firm pattern to build around. Coupling standards with example reference architectures will support adoption and adherence from technologists.
2. Evaluation of various PETs on the market comes with a **high level of friction and cost of doing business** that current privacy standards do not come close to alleviating - a standard by engineers for engineers with the blessing of policy leaders would reduce cost of doing business by removing and fasttracking technology evaluations.
3. **Incentives are not currently aligned to motivate market forces** to drive the behavioral change that PETs aim for - PETs fundamentally are the tools necessary to advance respect for permission, Data Control, and Data Dignity, but as tools they will not be adopted without stronger incentives that a uniform standards-based ecosystem creates (see [here](#) for reference).

Surmounting these challenges is essential to advance the mechanism of standards and drive the adoption of PETs broadly.

7. Risks related to PETs adoption

As technologists seeking to bridge the gap between the work of law and policy and the world of those on the ground building products, we see an existential risk to PETs adoption are laws made in a silo without the input of engineers who then have to implement those laws. Laws must be accompanied by standards, frameworks, and patterns for engineers to quickly turn policies into action. These can be created by government, academia, nonprofits, advocacy groups, or for-profit enterprise, but the bullhorn of policymakers can help ensure awareness, adoption, and eventual success.

8. Existing best practices that are helpful for PETs adoption

Standards and reference architectures for engineers are essential to PETs adoption. Engineers are fundamentally problem solvers. When they see a problem, they seek a solution, and every engineer knows there are multiple ways to get the job done with accompanying trade-offs in terms of risk, time, and resource. Every engineer thus seeks out established patterns, frameworks, and solutions that have worked for others before them and can be adapted for the product they are building. We need to see this best practice - which has achieved broad adoption in areas such as security - applied to the privacy space. Therefore, there needs to be disseminated everywhere patterns, designs, reference architectures, and standards for how to implement PETs across different use cases and industries. A good place to start is the “privacy stack” architecture being generated by The Ethical Tech Project.

9. Existing barriers, not covered above, to PETs adoption

To detail our third challenge articulated in our response to #6, PETs adoption requires incentives. Incentives can be carrots, or they can be sticks. Privacy regulations such as GDPR or CPRA are sticks driving adoption. Conversely, there aren't any carrots incentivizing adoption at the same scale as privacy law regulation. This does not mean that positive incentives aren't possible: there are many instances of industry self-regulation that have used the positive incentives of a standard to reward those that adopt. A barrier today is that there is no positive incentive framework or standard for privacy that has achieved critical mass in terms of awareness and adoption. This is an opportunity for the government to step in and advocate for such a standard with positive incentives.

10. Other information that is relevant to the adoption of PETs

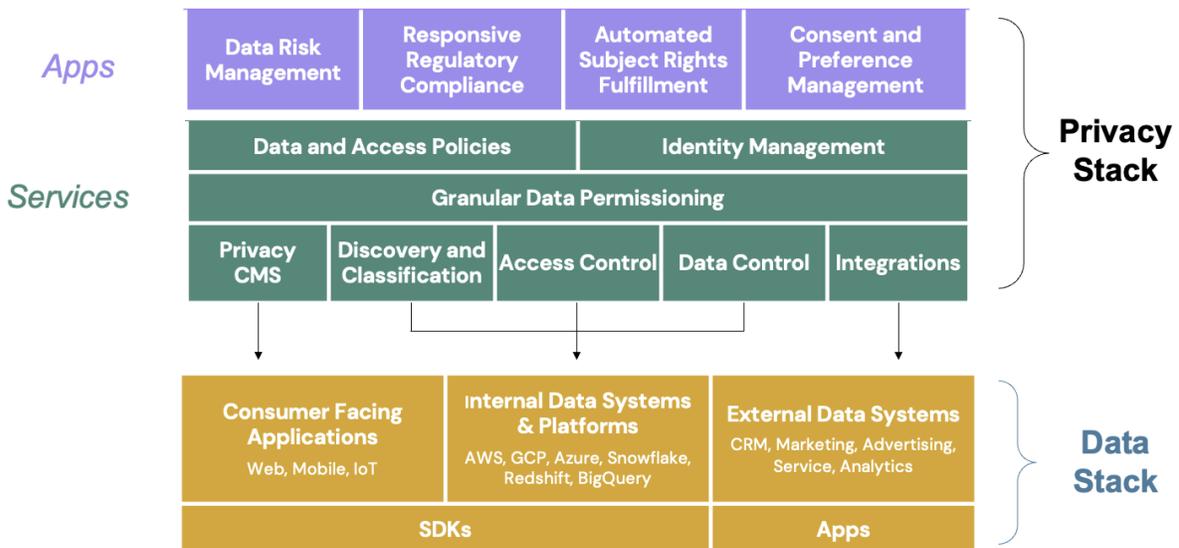
No comment.

Appendix

Privacy Stack Diagram

The Privacy Stack

Standard Architecture Diagram



Request for Information (RFI) on Advancing Privacy Enhancing Technologies

The Rise of Privacy Tech (TROPT)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

The Rise of Privacy Tech's (TROPT)

**Response to the
White House
Office of Science and Technology
Policy's (OSTP)**

**Request for Information on
Advancing Privacy Technologies**

July 8, 2022



AUTHORED BY THE RISE OF PRIVACY TECH (TROPT)

Lourdes M. Turrecha (Founder, The Rise of Privacy Tech)

With leading privacy and technology domain experts and thinkers, including select TROPT Innovators members and TROPT Privacy Tech Visionaries

Melanie Ensign (CEO, Discernible)

John Gevertz (Former Chief Privacy Officer/SVP, VISA and ADP)

Dr. K Royal (Chief Privacy Officer, Outschool, Inc.)

Dr. Mary Berk (Founder, Bluestreak Product Advisors)

Alexandra Ross (Senior Director, Senior Data Protection, Use & Ethics Counsel, Autodesk)

Gilbert Hill (Chief Strategy Officer, Pool Data)

Fatima Khan (Advisor, The Rise of Privacy Tech)

Debra Farber (Advisor, The Rise of Privacy Tech)

In consultation with other unnamed privacy and technology domain experts and thinkers

And researched by

Ece Gumusel (Researcher & Associate, The Rise of Privacy Tech)



TABLE OF CONTENTS

TABLE OF CONTENTS.....	2
EXECUTIVE SUMMARY OF THE RISE OF PRIVACY TECH’S NATIONAL PRIVACY TECH STRATEGY RECOMMENDATIONS	4
1. The White House national privacy tech strategy should standardize definitions.....	4
2. The White House national privacy tech strategy should broadly address privacy tech in general, not just PETs.....	6
3. The White House national privacy tech strategy should broadly address privacy technologies throughout the data lifecycle.....	8
4. The White House national privacy tech strategy should also broadly address privacy technologies throughout the technology development lifecycle.	8
5. The White House national privacy tech strategy should prioritize adoption incentives ahead of research initiatives at this time.....	10
a. Market funding towards privacy tech.....	10
b. An equal playing field for privacy-preserving business models.....	11
c. Economies of scale through incentives and disincentives	11
d. A privacy by design agenda in privacy law.....	12
6. The White House national privacy tech strategy should be technology neutral.....	12
7. Proposed vision for the White House national privacy tech strategy: a world where technology is designed and engineered with privacy in mind, to serve individuals and society, respecting and securing their rights to privacy, self-sovereignty, and self-determination.....	13
TROPT’S RESPONSES TO OSTP’S RFI	14
1. White House OSTP RFI on specific research opportunities to advance privacy tech.....	14
TROPT response to specific research opportunities to advance privacy tech.....	14
2. White House OSTP RFI on specific technical aspects or limitations of privacy tech.....	14
TROPT response to specific technical aspects or limitations of privacy tech.....	15
3. White House OSTP RFI on specific sectors, applications, or types of analysis that would particularly benefit from the adoption of privacy tech.....	15
TROPT response to specific sectors, applications, or types of analysis that would particularly benefit from the adoption of privacy tech	16
4. White House OSTP RFI on specific regulations or authorities that could be used, modified, or introduced to advance PETs.....	17
TROPT response to specific regulations or authorities that could be used, modified, or introduced to advance PETs:	17



5. White House OSTP RFI on specific laws that could be used, modified, or introduced to advance privacy tech..... 18
TROPT response to specific laws that could be used, modified, or introduced to advance privacy tech: 18

6. White House OSTP RFI on specific mechanisms, not covered above, that could be used, modified, or introduced to advance privacy tech 19
TROPT response to specific mechanisms, not covered above, that could be used, modified, or introduced to advance privacy tech..... 19

7. White House OSTP RFI on risks related to privacy tech adoption 20
TROPT response to risks related to privacy tech adoption 20

8. White House OSTP RFI on existing best practices that are helpful for privacy tech adoption..... 21
TROPT response to existing best practices that are helpful for privacy tech adoption..... 21

9. White House OSTP RFI on existing barriers, not covered above, to PETs adoption 21
TROPT response to existing barriers, not covered above, to PETs adoption..... 22

10. White House OSTP RFI on other information that is relevant to the adoption of PETs..... 23
TROPT response to other information that is relevant to the adoption of PETs..... 23

REFERENCES 23



EXECUTIVE SUMMARY OF THE RISE OF PRIVACY TECH'S NATIONAL PRIVACY TECH STRATEGY RECOMMENDATIONS

The Rise of Privacy Tech (TROPT) submits these recommendations to the White House Office of Science and Technology Policy's (OSTP) request for information (RFI)ⁱ to help inform the development of a national strategy on privacy tech.

TROPT is the only formal privacy tech community. TROPT community members control significant market power to build, invest in, design, and implement emerging privacy technologies. TROPT's mission is to fuel privacy tech and innovation. We do this by breaking down silos and bridging the gaps between privacy tech key players: founders of privacy tech startups, privacy tech investors, privacy domain experts who are privacy tech board advisors, privacy tech buyers and users, and other privacy professionals with the experience to integrate these technologies into corporate systems and consumer products. In this executive summary, we highlight seven (7) critical recommendations for the future of privacy tech and innovation that apply directly to OSTP's stated objectives for developing a national strategy on privacy tech.

1. The White House national privacy tech strategy should standardize definitions.

Words matter. To be effective in the development of a national strategy on privacy tech, there must be clarification and standardization of what mean by *privacy tech*, *privacy enhancing technologies (PETs)*, and related terminology. The lack of a universally accepted definitionⁱⁱ brings up questions of scope and allows organizations to misrepresent the purpose, use, and effectiveness of their privacy tech solutions.



There is currently no universally accepted definition of PETs. While PETs have come up in several failed bills and congressional hearing transcripts, it remains undefined under US law. Even in academia, researchers use different PETs definitions, or avoid defining the term altogether.

During the June 7, 2022 Fast Track Action Committee (FTAC) roundtable meeting, there was much debate among attendees about the use of appropriate terminology. The OSTP’s RFI acknowledges this tension as follows: “For the purposes of this RFI, privacy-enhancing, privacy-preserving, and privacy-protecting are used as equivalent terms.”ⁱⁱⁱ

At TROPT, we’ve kickstarted the work of defining *privacy tech* and exploring its relationship to *PETs* and related terminology in the foundational *TROPT Defining the Privacy Tech Landscape Whitepaper 2021*.^{iv} We did this because having a baseline understanding of privacy tech’s scope helps facilitate healthy debate and avoid talking past each other. Some may argue that definitions are mere semantics, but being clear about what we’re talking about has significant implications, for buyers purchasing privacy tech, investors investing money into privacy tech, and founders building privacy tech. A common starting point saves us valuable time and resources and enables us to instead focus on the greater goal of moving the needle on privacy and true privacy tech solutions. We offer this existing work as a starting point, not the end goal.

2. The White House national privacy tech strategy should broadly address privacy tech in general, not just PETs.

At TROPT, we focus broadly on *privacy tech* (versus solely on *PETs*) because privacy tech appropriately encapsulates the breadth of technical solutions needed to solve privacy problems.

We define *privacy tech* as technological solutions to privacy problems or privacy harms.^v

TROPT PRIVACY TECH DEFINITION

privacy tech

pri·va·cy·tech | \ 'prī-və-sē'tek

1 (noun) technological solutions to privacy problems.

2 (noun) emerging industry of tech companies building solutions to privacy problems, in an effort to address the mounting privacy technical debt.

SOURCE: TROPT DEFINING THE PRIVACY TECH LANDSCAPE WHITEPAPER 2021



In addition to refraining from focusing solely on PETs, we've described the relationship between privacy tech and PETs:

“PETs unfortunately do not have a universally accepted definition. That said, the OECD defines PETs as “a “wide range of technologies that help protect privacy,”^{vi} aiming to give the individual user (in a consumer context) or technology manager (in an organizational environment) the capability of controlling if, how much, or under what circumstances



personal information is disclosed. Therefore, based on this commonly used OECD PETs definition, PETs are privacy tech. But privacy tech tools are not necessarily always PETs because privacy tech solves for other privacy problems beyond control over personal information.

Another second widely cited definition of PETs is: a coherent system of technology measures that protect privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system.^{vii} Under this second common PETs definition, a technology that blocks online ad tracking would not be a PET because it interferes with system functionality, but it would certainly be privacy tech because it solves for a privacy problem: an individual's choice not to be tracked. In comparison, a technology that allowed ad personalization while giving users control over their personal information would be a PET under this second definition and, therefore also qualify as privacy tech because it both solves for a privacy problem (individual control over their personal information) without losing ad functionality.

Finally, S.224, the recently introduced Promoting Digital Privacy Technologies Act, defines PETs as “any software solution, technical processes, or other technological means of enhancing the privacy and confidentiality of an individual’s personal data in data or sets of data.”^{viii} S.224’s PETs definition also falls under privacy tech because any software solution, technical process, or other technological means that enhances the privacy and



confidentiality of an individual’s personal information in data or sets of data qualify as technological solutions to privacy problems.”^{ix}

3. The White House national privacy tech strategy should broadly address privacy technologies throughout the data lifecycle.

The White House OSTP’s RFI focuses on “privacy-preserving data sharing and analytics,”^x which could suggest that only privacy tech during the data *use* and *sharing* stages of the data lifecycle are in scope of the RFI and the resulting national policy.

Privacy protections must be bound to the data objectives themselves, so that organizations and consumers have reassurance that the appropriate privacy protections are enforced throughout the data lifecycle. By the time data is collected and reaches the data *use* and *sharing* stages of the data lifecycle, transparency, data minimization, and other privacy principles are often violated already.

As such, TROPT proposes that the national policy broadly address privacy tech throughout the data lifecycle, from collection, to and use and sharing, and up to storage, retention, and deletion.

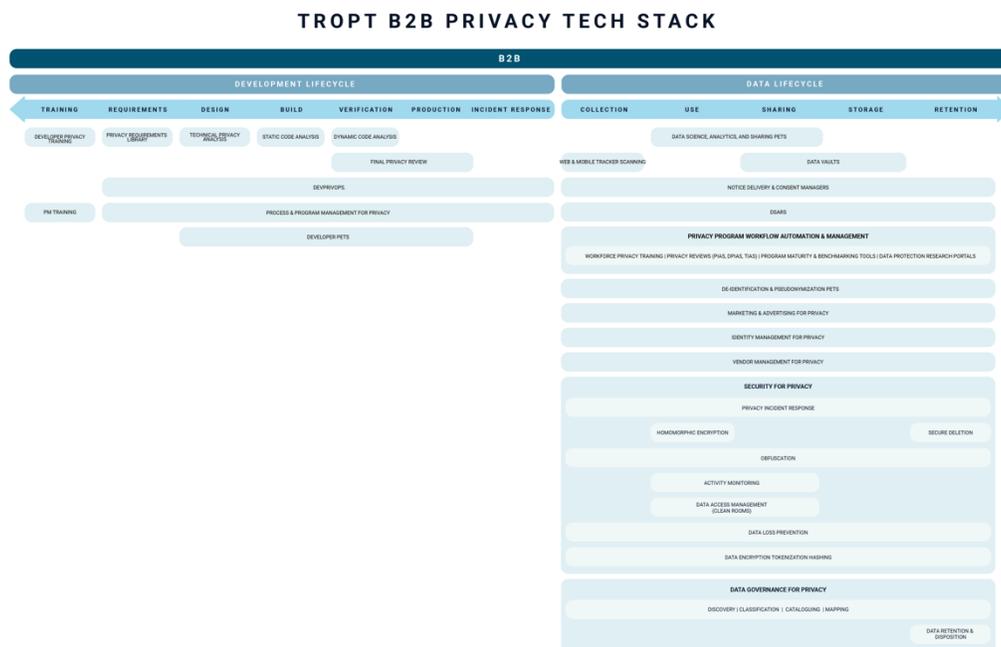
4. The White House national privacy tech strategy should also broadly address privacy technologies throughout the technology development lifecycle.

In addition to broadening the scope to privacy tech that solve privacy problems throughout the entire data lifecycle, we urge the White House to further broaden its scope by “shifting privacy left” to the technology development lifecycle. The technology development lifecycle is where many privacy decisions are made, even before data is collected, used, shared, stored, retained,

and deleted. Many privacy problems arise from the development of technologies without regard to privacy, resulting to decades worth of privacy technical debt.

At TROPT, we reported on the trend involving privacy tech tools during the development lifecycle: from developer privacy training, privacy code scanners, privacy requirements libraries, synthetic data for testing, etc.^{xi} These developers are equally—if not more—deserving of strategic highlight and support given they anticipate and solve privacy problems before they’re created. Development lifecycle privacy tech tools also help organizations adhere to global data protection rules requiring privacy by design and default.

Figure 1: TROPT B2B Privacy Tech Stack showing privacy technologies under both the data and development lifecycles.



SOURCE: TROPT DEFINING THE PRIVACY TECH LANDSCAPE WHITEPAPER 2021





5. The White House national privacy tech strategy should prioritize adoption incentives ahead of research initiatives at this time.

Academics, researchers, and technologists have been conducting PETs research since at least the 1980s. There is an existing wealth of PETs research from academia, industry, and the public sector. In academia, we have Carnegie Mellon’s CyLab Privacy & Security Institute^{xiii} and Harvard’s Privacy Tools Project^{xiii}. From industry, we have Apple’s Differential Privacy research^{xiv} and Google’s differential privacy library.^{xv} And in the public sector, the National Science and Technology Council (NSTC) has issued a report^{xvi} detailing a national privacy research strategy, and referencing the PETs research initiatives the federal government has engaged in.

While research is clearly important, it’s hardly enough to move us to the next stage: widespread privacy tech adoption. Research is not where we fall short on privacy tech. Instead, we need adoption incentives, including the following:

a. Market funding towards privacy tech

In addition to funding privacy tech research, we need to fund privacy tech solutions that are mature enough to go to market. The United States Government can fulfill this need. The same way that In-Q-Tel invested in important cybersecurity developments, the U.S. Government can invest in the emerging privacy tech industry.

The U.S. Government, through the Securities & Exchange Commission (SEC), could also incentivize investors to fund privacy tech—which aligns with the broader Environmental, Social,



and Governance (ESG) and impact investing trends in venture capital—or require them to be more transparent about their investments (or lack thereof) in privacy tech.

b. An equal playing field for privacy-preserving business models

Successful privacy tech adoption depends in part on appropriate business models. Several privacy tech companies that have made it to market have failed because they were competing against companies with anti-privacy business models, such as those rooted in the opaque monetization of personal data. Thus, the national strategy should also encourage an equal playing field for privacy-preserving business models, including those that include ownership-, value-, and profit-sharing with individuals.

c. Economies of scale through incentives and disincentives

To help privacy tech startups entering the market, the national privacy tech policy could play an external role in developing economies of scale. The national policy could pursue the introduction of tax breaks for startups with privacy tech products and/or to privacy tech users, similar to the tax incentives historically offered to climate tech or any other industry that the U.S. Government supports. The national policy could explore preferential treatment for privacy tech vendors in government contracts or introduce actual privacy tech requirements in government contracts through the Federal Acquisition Regulation (FARs).

In addition, the national policy could require or incentivize Big Tech players to accommodate PETs in the infrastructure they build and offer. Some Big Tech companies have already dipped their toes into PETs waters, largely due to brand incentives. But for those which haven't, perhaps



the national policy could pursue the introduction of disincentives, such as disqualification from government contracts and tax breaks.

d. A privacy by design agenda in privacy law

Beyond advancing privacy tech, the national policy should play a bigger role in mandating privacy design and engineering of technologies, in general. Leading privacy thinker, Professor Woodrow Hartzog, outlines this proposal in his book, *Privacy's Blueprint: The Battle to Control the Design of New Technologies*.^{xvii} Additionally, there is a wealth of proposals and ideas on privacy by design and privacy engineering from leading privacy technologists and academics, such as former Canadian Privacy Commissioner Dr. Ann Cavoukian,^{xviii} Dr. Lorrie Cranor,^{xix} Michelle Finneran-Dennedy,^{xx} and R. Jason Cronk,^{xxi} to name a few. Building on this solid foundation would provide a quicker path to success.

6. The White House national privacy tech strategy should be technology neutral.

Much has been said about the need for privacy tech—and specifically, PETs—in the context of artificial intelligence (AI) and machine learning (ML) technologies. We agree. That said, we urge the White House to develop its national privacy tech strategy in a technology neutral way, and not focus solely on AI and ML (or other specific) technologies.

We urge a technology neutral approach because privacy problems arise with every major technological advancement, from the invention of the portable camera and the proliferation of smart phones, to the advent of cloud computing, the Internet of Things (IoT), and now AI and ML. We are already seeing privacy problems arise with Web 3.0 technologies, as we build the next Internet age. With the benefit of privacy and technology history in mind, the national



privacy tech strategy should be technology neutral, transcending technological trends like AI and ML and applying to emerging technologies.

This means the national strategy needs to go beyond supporting only PETs used in the context of AI and ML and extend to many other privacy tech use cases.

7. Proposed vision for the White House national privacy tech strategy: a world where technology is designed and engineered with privacy in mind, to serve individuals and society, respecting and securing their rights to privacy, self-sovereignty, and self-determination.

At TROPT, we believe that privacy is a critical component in designing and building technology to serve people. Privacy design and engineering are prerequisites for product excellence, creating privacy value. Privacy tech and innovation offers market opportunities to those who can recognize the value of privacy, beyond compliance. We see a world where technology is designed and engineered with privacy in mind, to serve individuals and society, respecting and securing their rights to privacy, self-sovereignty, and self-determination.^{xxii} Mainstream privacy tech adoption can help get us there by providing solutions to a broad array of privacy problems that harm individuals and society and infringe on their rights to privacy, self-sovereignty, and self-determination.



TROPT'S RESPONSES TO OSTP'S RFI

Having laid out our seven (7) critical recommendations for a national privacy tech strategy in the previous section, we now turn to our specific responses to each of OSTP's ten (10) requests for information:

1. White House OSTP RFI on specific research opportunities to advance privacy tech

Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.

TROPT response to specific research opportunities to advance privacy tech

We recommend the following research opportunities to advance privacy tech:

- Carnegie Mellon University CyLab PETs research initiatives, under Dr. Lorrie Cranor's leadership^{xxiii}
- The numerous technical privacy research projects presented at USENIX's PEPR conference^{xxiv}
- The upcoming TROPT Privacy Tech Stack Whitepaper 2022

2. White House OSTP RFI on specific technical aspects or limitations of privacy tech

Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or



anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections or reduce the risks or costs of adopting PETs.

TROPT response to specific technical aspects or limitations of privacy tech

As privacy tech industry leaders and privacy practitioners in the tech industry, we've seen the following privacy tech limitations:

- Immaturity of privacy tech solutions
- Limited scalability of privacy tech solutions
- Commercial viability of privacy tech solutions
- Limited interoperability with existing systems
- Lack of privacy tech awareness and understanding among potential adopters
- False tradeoffs between usability and privacy

3. White House OSTP RFI on specific sectors, applications, or types of analysis that would particularly benefit from the adoption of privacy tech

Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to



the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of [Executive Order 14058](#): Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government.

TROPT response to specific sectors, applications, or types of analysis that would particularly benefit from the adoption of privacy tech

As privacy tech industry leaders and privacy practitioners in the tech industry, we've seen the following applications:

- Privacy code scanners to identify privacy problems in products before they're launched
- Privacy tech for secondary use, including in the context of data science and analytics (for marketing, trends, insights, etc.) especially in regulated industries like healthcare, finance, education
- Privacy tech for data retention and end-of-life
- Privacy tech (e.g., tokenization) for payment processing
- Use of synthetic data (instead of real personal data) during product testing
- Privacy tech allowing for the processing of personal data for cybersecurity threat and response analysis



- Privacy tech providing automated and accurate just-in-time transparency
- Privacy tech solutions in smart cities and urban planning
- Privacy tech to facilitate safety in edtech and children-directed technologies

4. White House OSTP RFI on specific regulations or authorities that could be used, modified, or introduced to advance PETs

Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking authorities under the Office of Management and Budget, the Federal Trade Commission, and financial regulatory bodies, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (*e.g.*, Other Transaction Authority) to procure PETs for development.

TROPT response to specific regulations or authorities that could be used, modified, or introduced to advance PETs:

- Federal privacy authorities like the FTC, CFPB, SEC, HHS, DoT, and DoJ
- Federal privacy laws, regulations, and standards like HIPAA, COPPA, GLBA, FERPA, FARs, FedRAMP, and SEC disclosure requirements
- NIST standards, perhaps a introducing a new framework for privacy technologies



5. White House OSTP RFI on specific laws that could be used, modified, or introduced to advance privacy tech

Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.

TROPT response to specific laws that could be used, modified, or introduced to advance privacy tech:

We recommend pursuing the following to promote privacy tech:

- The Promoting Digital Privacy Technologies Act (S.224 / H.R. 847)^{xxv} with recommended improvements^{xxvi}
- A comprehensive federal data protection law that includes the promotion of privacy tech *and* that closes loopholes on surveillance technologies used in public-private data sharing
- Revisions to COPPA that would promote privacy tech in protecting children, including requirements for educational settings
- Revisions to HIPAA to update the Security Rule and enhance the Privacy Rule to specifically address privacy tech



6. White House OSTP RFI on specific mechanisms, not covered above, that could be used, modified, or introduced to advance privacy tech

This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

TROPT response to specific mechanisms, not covered above, that could be used, modified, or introduced to advance privacy tech

At TROPT, we are open to exploring the following additional mechanisms to advance privacy tech:

- Public-private partnerships to fuel the emerging privacy tech industry
- Privacy tech startup innovation prize challenges, grants, and testbeds, potentially in partnerships with privacy tech startup accelerator and university incubator programs
- Creation of open-source protocols and technical guidance
- Privacy tech education and awareness initiatives, including TROPT whitepapers and TROPT events (*e.g.*, TROPT Data Privacy Day/Week, TROPT Privacy Tech Summit, and TROPT Webcasts) and USENIX's PEPR conferences



7. White House OSTP RFI on risks related to privacy tech adoption

Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

TROPT response to risks related to privacy tech adoption

As privacy tech industry leaders and privacy practitioners in the tech industry, we've seen the following privacy tech adoption risks:

- Conflicting technologies leading to lack of interoperability, adoption failure, and wasted resources
- Introduction of complexity, including process complexity, leading to adoption failure and wasted resources
- Introduction of new privacy and security vulnerabilities
- Lack of scalability leading to adoption failure and wasted resources
- Limitation or inhibition of future legitimate data uses

We at TROPT are willing to engage in conversations about the risks of privacy tech and PETs adoption, including demystifying false tradeoffs; measuring, quantifying, and remediating legitimate risks; and developing risk-benefit analysis guidance that the industry could use in their



adoption considerations. These could be especially helpful for emerging startups that lack the internal resources to develop their own privacy and security controls.

8. White House OSTP RFI on existing best practices that are helpful for privacy tech adoption

Information about U.S. policies that are currently helping facilitate adoption as well as best practices that facilitate responsible adoption. This includes existing policies that support adoption, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This also includes information about where and when PETs can be situated within tiered access frameworks for accessing restricted data, ranging from publicly accessible to fully restricted data.

TROPT response to existing best practices that are helpful for privacy tech adoption

We recommend the following resources for privacy tech adoption best practices:

- The foundational TROPT Defining the Privacy Tech Landscape Whitepaper 2021^{xxvii}
- The upcoming TROPT Privacy Tech Stack Whitepaper 2022
- Luk Arbuckle & Khaled El Imam's *Building an Anonymization Pipeline: Creating Safe Data*^{xxviii}

9. White House OSTP RFI on existing barriers, not covered above, to PETs adoption

Information about technical, sociotechnical, usability, and socioeconomic barriers that have inhibited wider adoption of PETs, such as a lack of public trust. This includes recommendations



on how such barriers could be overcome. Responses that focus on increasing equity for underserved or marginalized groups are especially welcome.

TROPT response to existing barriers, not covered above, to PETs adoption

As privacy tech industry leaders and privacy practitioners in the tech industry, we've seen the following privacy tech adoption barriers:

- Limited or competing budget and resources for privacy tech acquisition, development, and implementation
- Lack of (business & legal) incentives for privacy tech adoption
- Limited internal technical privacy expertise needed for privacy tech adoption
- Lack of awareness of privacy tech opportunities
- Largely academic versus practical privacy tech focus
- Privacy tech and PETs vendors' own inadequate privacy and security controls, with varying underlying causes (budget, buy-in, market focus, awareness, leadership and team without privacy and security expertise)
- Inconsistent definitions, as discussed in the previous section
- Existing silos between technologists, investors, academic researchers, practitioners, and buyer-users
- Lack of privacy (and greater ethical) development, leading to systemic biases resulting from inequity of available data (in other words, we need cross-functional experts from historically underserved minority populations to be instrumental in developing and addressing privacy tech – as engineers, creators, founders, advisors, and executives)



The solution for many of the above problems is appropriate incentives. Organizations are prone to engage in: 1) what is profitable (or satisfies their corporate mission); 2) what their customers want; and 3) what the law requires.

10. White House OSTP RFI on other information that is relevant to the adoption of PETs

Information that is relevant to the adoption of PETs that does not fit into any of the topics enumerated above.

TROPT response to other information that is relevant to the adoption of PETs

In addition to the above, we recommend exploring the following privacy tech initiatives:

- Small-to-medium business (SMB) insights and incentives in the form of grants, etc.
- Balanced/reasonable/supported barriers to privacy tech entry
- Right-sized privacy tech adoption resources toolkits for different organizations (*e.g.*, startups, SMBs, enterprise, Big Tech)

REFERENCES

ⁱ *Request for Information on Advancing Privacy-Enhancing Technologies*. (2022, June 9). Federal Register. <https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies>.

ⁱⁱ Office of the Privacy Commissioner of Canada, "Privacy Enhancing Technologies -- A Review of Tools and Techniques," November 2017, https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711/ (last accessed July 2022).

ⁱⁱⁱ *See i.*

^{iv} Turrecha, L., Ashley, E., et al. (2021). *The Rise of Privacy Tech Defining the Privacy Tech Landscape 2021*. <https://www.riseofprivacytech.com/wp-content/uploads/2021/11/TROPT-Defining-the-Privacy-Tech-Landscape-2021-v1.0-1.pdf>.

^v *See iv.*

^{vi} OECD Working Party on Information Security and Privacy, "Inventory of Privacy-Enhancing Technologies (PETs)," January 7, 2002,

<https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=dsti/iccp/reg%282001%291/final> (last accessed July 2022).

vii John J. Borking and Charles Raab. “Laws, PETs, and Other Technologies for Privacy Protection.” *Journal of Information, Law, and Technology*, 1:1–14, 2001.

viii S.224 Promoting Digital Privacy Technologies Act, Introduced February 4, 2021, <https://www.congress.gov/bill/117th-congress/senate-bill/224/text> (last accessed July 2022).

ix *See iv.*

x *See i.*

xi *See iv.*

xii Carnegie Mellon University CyLab *Directory*. Retrieved July 8, 2022, from <https://www.cylab.cmu.edu/directory/index.html>.

xiii *Harvard University Privacy Tools Project*. (n.d.). Retrieved July 8, 2022, from <https://privacytools.seas.harvard.edu/>.

xiv Apple Machine Learning Research. Retrieved July 8, 2022, from <https://machinelearning.apple.com/research>.

xv *Differential Privacy*. (2022). [C++]. Google. <https://github.com/google/differential-privacy> (Original work published 2019).

xvi Blumenthal, M., Brooks, S., Lefkowitz, N., & Vagoun, T. (n.d.). *National Privacy Research Strategy*. 34. <https://www.nitrd.gov/PUBS/NationalPrivacyResearchStrategy.pdf>

xvii Hartzog, W. (2018). *Privacy’s Blueprint*. Harvard University Press.

xviii Cavoukian, A. (2010). Privacy by Design The 7 Foundational Principles. *Information and Privacy Commissioner of Ontario*.

xix L. F. Cranor and N. Sadeh, "A Shortage of Privacy Engineers," in *IEEE Security & Privacy*, vol. 11, no. 2, pp. 77-79, March-April 2013, doi: 10.1109/MSP.2013.25.

xx Finneran-Dennedy, M., Fox, J., and Finneran, T.R. (2014). *The Privacy Engineer’s Manifesto Getting from Policy to Code to QA to Value*. Apress Berkeley, CA.

xxi Cronk, R.J. (2022). *Strategic Privacy By Design*. (2nd ed.). International Association of Privacy Professionals.

xxii Turrecha, L. (2021, October 7). The “Why?” Behind The Rise of Privacy Tech. *The Rise of Privacy Tech*. <https://medium.com/the-rise-of-privacy-tech/the-why-behind-the-rise-of-privacy-tech-2335f96a53be>.

xxiii *CyLab Security & Privacy Institute*. (n.d.). Retrieved July 8, 2022, from <https://www.cylab.cmu.edu/index.html>.

xxiv *PEPR ’22*. (2021, December 14). USENIX. <https://www.usenix.org/conference/pepr22>

xxv Stevens, H. M. (2022, May 12). *H.R.847 - 117th Congress (2021-2022): Promoting Digital Privacy Technologies Act (2021/2022)* [Legislation]. <http://www.congress.gov/>.

xxvi Turrecha, L. (2021, February 16). On Regulating Privacy Enhancing Technologies (PETs) Through the Promoting Digital Privacy Technologies Act (S.224)—What It Is, What’s Good About It, How To Improve It. *Privacy & Technology Publication*. <https://medium.com/privacy->



[technology/on-regulating-privacy-enhancing-technologies-pets-through-the-promoting-digital-privacy-ad2d63570551](#).

^{xxvii} *See iv.*

^{xxviii} Arbuckle, L. and El Emam, K. (2020). *Building an Anonymization Pipeline: Creating Safe Data*. Press by O'Reilly Media, Inc.



NOVEMBER 2021

THE RISE OF PRIVACY TECH DEFINING THE PRIVACY TECH LANDSCAPE 2021

A FOUNDATIONAL WHITEPAPER TO HELP
FUEL PRIVACY INNOVATION



AUTHORED BY

THE RISE OF PRIVACY TECH

Lourdes M. Turrecha and Emily Ashley

In collaboration with the TROPT Privacy Tech Landscape
Working Group members:

Nishant Bhajaria, Mary Berk, Michelle Finneran-Dennedy, Melanie Ensign, Debra Farber, Kiran Gopinath, Kelly Huang, Gilbert Hill, Fatima Khan, Cillian Kieran, Carolyn Krol, David Kruger, Caroline McCaffery, Kevin Riggle, and Alexandra Ross.

In consultation with leading privacy & technology domain experts and thinkers, some of whom are quoted in this whitepaper, and others who wish to remain anonymous in their valuable contributions.



THE RISE OF
PRIVACY TECH

The Rise of Privacy Tech's mission is to fuel privacy innovation. We do this by bringing together privacy tech founders, investors, domain experts, and advocates and bridging the tech-capital-expertise gaps in privacy.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	5
INTRODUCTION	6
DEFINING PRIVACY TECH	7
Defining the Operative Terms	7
What is privacy?	7
What are privacy principles?	9
What are TROPT's Privacy Tech Principles?	9
What is technology?	11
What is privacy tech, then?	12
Comparing Privacy Tech to Often Conflated Concepts	13
Comparing Privacy Tech to Adjacent Industries	15
Cybersecurity	15
Identity management	15
Data infrastructure	15
Martech & adtech	16
Legal tech	16
Data governance	16
Web 3.0	16
CATEGORIZING PRIVACY TECH: THE TROPT PRIVACY TECH STACK	17
The TROPT Privacy Tech Stack	17
B2B Privacy Tech Tools: Data Lifecycle Tools	19
Notice delivery and consent managers	19
Web & mobile tracker scanning	19
Data science, analytics, and sharing PETs	19
Data vaults	19
Data subject access rights (DSARs)	19
Identity management for privacy	20
Privacy program workflow automation & management	20
Vendor management for privacy	20
Security for privacy	20
Data governance for privacy	20
De-identification, pseudonymization, and anonymization PETs	21



TABLE OF CONTENTS

B2B Privacy Tech: Development Lifecycle Tools	21
Developer privacy training & awareness	21
Privacy requirements library	22
Technical privacy analysis	22
Developer PETs	22
DevPrivOps	22
Process and program management for privacy	22
Static code analysis	22
Dynamic code analysis	22
Final privacy review	22
B2C Privacy Tech	24
Data wallets	24
Data subject access rights (DSARs) & DSAR agents	24
Anti-trackers	24
Online pseudonyms	24
Privacy-first online services	24
B2B2C Privacy Tech	25
Centralized consent platforms	25
Data unions	25
BRIDGING GAPS BETWEEN PRIVACY TECH KEY PLAYERS	27
What Privacy Tech Buyers and Users Say	27
What Privacy Tech Startup Founders & Operators Say	32
What Privacy Domain Experts Say	34
What Investors Didn't Outright Say	36
FUELING THE FUTURE OF PRIVACY TECH	37
What Privacy Tech Founders Can Do	37
What Investors Can Do	39
What Privacy Domain Experts Can Do	40
What B2B Privacy Tech Buyers and Users Can Do	41
TROPT's Part	42
CONCLUSION	43
APPENDIX: SURVEY RESULTS	44
Privacy Tech Startups Survey	44
Privacy Domain Experts Survey	53
Privacy Tech User or Buyer Survey	63
ENDNOTES	72



EXECUTIVE SUMMARY

The emerging privacy tech industry is on the rise. For the first time in history, founders, technologists, investors, regulators, consumers, and enterprise customers are in agreement with privacy domain experts and consumer advocates that privacy is one of the most critical issues of our time. We need privacy innovation to help solve our privacy problems and the resulting mounting privacy technical debt. An increasing demand for privacy designed and engineered tools accompanies this fortuitous timing. In turn, an uptick in technologists and founders are rising to the occasion to build solutions to some of these privacy problems and seize the resulting market opportunities.

But at this nascent time, the privacy tech key players—founders, investors, domain experts, and user-buyers—aren't connected enough to move things forward at the pace the market needs. They're not even on the same chapter—let alone page—in their understanding of privacy or privacy tech. In this foundational whitepaper, we start to bridge the gaps between the key players in privacy tech.

We start off by defining privacy tech, which requires understanding privacy. We provide a summary of different privacy conceptualizations from leading privacy thinkers. We do this for several reasons. First, key players are unnecessarily talking past each other as they discuss the emerging privacy tech landscape. Second, this lack of a baseline understanding of privacy introduces friction, and, in certain contexts, prevents key players from moving forward with a partnership, purchase, investment, or other step towards fueling privacy tech and innovation.

As part of this exercise, we also summarize common privacy principles, introduce The Rise of Privacy Tech (TROPT)'s own set of privacy tech principles, and explore privacy tech's relationship with frequently conflated terms like cybersecurity, privacy engineering, privacy by design, and so on.

We then categorize privacy tech. To do so, we introduce the TROPT Privacy Tech Stack, pulling from foundational privacy and technology knowledge. We plan to update and use the TROPT Privacy Tech Stack to categorize the privacy tech landscape as it matures. We expect the privacy tech stack to evolve as the industry evolves, and we welcome feedback to help capture this evolution.

We also surveyed key players in privacy tech to hear about their different pain points, the privacy problems they're solving for, the solutions they're building or buying, and the challenges they face at this early stage in the privacy tech industry.

Lastly, we share insights from the forefront of the privacy tech industry, with recommendations on how we can collectively keep fueling privacy tech and innovation. We outline proposed responses to what we learned from this endeavor. Most importantly, we call for each key player to join us in fueling privacy tech, with specific recommendations on how to do so.



INTRODUCTION

There is increasing interest in the nascent privacy tech landscape.

Entrepreneurs are building solutions in response to privacy and data protection problems. As of the publication of this white paper, multiple indicators point to rapid growth in the privacy tech industry.

The Crunchbase database lists 945 companies that identify themselves as privacy companies.¹ The Rise of Privacy Tech (TROPT)'s own ongoing research identifies 745 privacy tech companies.²

Investor interest in funding privacy tech startups is equally strong and growing. Even in the midst of the pandemic when funding was reportedly stalled, funding towards privacy tech startups continued.³ Investors have reportedly poured in almost six billion US dollars (specifically, \$5.9B) in cumulative funding towards emerging privacy companies; four billion (\$4B) in the past three years alone.⁴

Recent years saw high growth and demand for privacy tech in the market. Last year, the privacy tech industry produced two unicorns,⁵ startups valued at more than a billion dollars. In the midst of the pandemic, the tech industry as a whole also embraced privacy tech in various forms, such as COVID-19 tools touting privacy features,⁶ including Apple and Google's joint contact tracing proposal.⁷ Apple's recent iOS 14⁸ and 15⁹ releases included a slew of privacy tech features.

In the consumer space, individuals' concern over their privacy led to their increased adoption of consumer privacy tech, including private messaging app, Signal, which reached peak status as the

top downloaded app¹⁰ during last year's wave of national protests in support of the Black Lives Matters movement. Signal downloads surged again during the last US national elections and in early 2021 after Elon Musk urged his Twitter followers to "Use Signal."¹¹

Consumers are increasingly becoming aware of other privacy tech tools available on the market and switching to privacy-protective alternatives like Brave, DuckDuckGo, and ProtonMail, which have reported similar increased adoption as privacy preserving alternatives to existing privacy-invasive browsers, search engines, and email. As of September 2021, the privacy-focused web browser Brave reported thirty-six million monthly active users. Brave's ad platform revenue has grown fourteen times in the last eighteen months.¹² The privacy-forward search engine, DuckDuckGo, reported over one-hundred million downloads since 2018, with over half of those downloads taking place in the past year.¹³ ProtonMail, which offers encrypted emails, reported five million users in 2018¹⁴ and have now surpassed fifty million users in 2021.¹⁵

Beyond consumer privacy tech, B2B privacy tech has taken off even faster. Most privacy tech tools are built in—and investments are poured into—the B2B side. This reality is reflected in the TROPT Privacy Tech Stack, which shows a more mature B2B side compared to the B2C and B2B2C sides.

These sample indicators clearly show that privacy tech is on the rise, but the understanding of what constitutes privacy tech remains low.

What exactly is privacy tech?



DEFINING PRIVACY TECH

As the privacy tech landscape continues to grow, it's critical to define and understand what qualifies as privacy tech for several reasons.

First, having a baseline understanding of privacy tech's scope helps facilitate healthy debate and avoid talking past each other. Some may argue that definitions are mere semantics, but being clear about what we're talking about has significant implications, for buyers purchasing privacy tech, investors investing money into privacy tech, and founders building privacy tech. A common starting point saves us valuable time and resources and enables us to instead focus on the greater goal of moving the needle on privacy and true privacy tech solutions.

Second, defining privacy tech forces transparency, which is a key privacy principle. Consumer and business customers deserve to know whether the privacy-branded tools they're using are actually privacy tech solutions.

Privacy tech boundaries also help customers better identify true privacy tech from tools that don't actually solve privacy problems. For example, while tools in the adjacent industries may sometimes overlap with privacy tech, customers deserve to know when they are and aren't privacy tech.

Having a clear understanding of what constitutes privacy tech helps fuel the privacy tech industry's success. By being clear about what privacy tech is, we weed out products that aren't privacy tech. This improves the privacy tech landscape's success as it matures and strengthens the privacy tech market's long-term viability. In turn, this means better privacy tech tools for both businesses and consumers.

DEFINING THE OPERATIVE TERMS

To define and understand privacy tech, we must first understand the operative terms: *privacy* and *technology*. Fortunately, and credit to leading privacy thinkers, both privacy and technology are well-covered and explored concepts.

What is privacy?

Different schools of thought have offered different privacy definitions and conceptualizations. We summarize some of the common ones in Figure 1 in the following page.



WHAT IS PRIVACY?

CONTRARY TO POPULAR MYTH, PRIVACY IS NOT JUST SECRECY OR ANONYMITY.



PRIVACY AS CONTROL

Perhaps the most commonly accepted definition of information privacy focuses on individual control over personal data, including how much of it is disclosed and to whom, as well as how it should be maintained and disseminated. Under this school of thought, privacy is individuals' claim to determine for themselves when, how, and to what extent personal information about them is communicated to others. *Alan Westin, Privacy and Freedom*



PRIVACY AS OBSCURITY

Obscurity is the idea that information is safe—at least to some degree—when it is hard to obtain or understand. *Evan Selinger and Woodrow Hartzog, Obscurity and Privacy*



PRIVACY AS TRUST

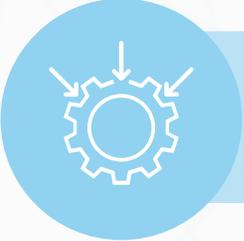
Privacy has also been conceptualized as trust, with four foundations: (1) the commitment to be honest about data practices, (2) the importance of discretion in data usage, (3) the need for protection of personal data against outsiders, and (4) the overriding principle of loyalty to the people whose data is being used, so that it is data and not humans that become exploited. Trust is important in forging information relationships in today's networked, data-driven society. *Neil Richards and Woodrow Hartzog, Privacy's Trust Gap*

Trust expectations make privacy expectations reasonable. Trust greases the wheels of effective sharing by encouraging interactions and information sharing. *Ari Ezra Waldman, Privacy as Trust: Sharing Personal Information in a Networked World*



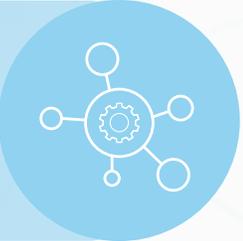
PRIVACY AS POWER

Privacy as power builds on privacy as choice and provides support for users to exercise their power and take back control over their personal data in a world where personal data is constantly being harvested and exploited through a surveillance economy. *Carissa Véliz, Privacy is Power*



PRIVACY AS CONTEXTUAL INTEGRITY

Privacy norms or expectations vary by context. Contextual integrity links privacy protection to norms for specific contexts, so that "information gathering and dissemination are appropriate to that context and obey the governing norms of distribution within it." *Helen Nissenbaum, Privacy as Contextual Integrity*



A TAXONOMY OF PRIVACY

As detailed above, privacy can mean different things to different people. Perhaps the most comprehensive conceptualization of privacy is Daniel Solove's A Taxonomy Privacy, a framework for understanding privacy in a pluralistic and contextual manner, grounded in the different kinds of activities that impinge upon privacy. This conceptualization shifts the focus away from the vague term "privacy" and toward the specific activities that pose privacy problems. *Daniel J. Solove, Understanding Privacy and A Taxonomy of Privacy*

Figure 1. What Is Privacy?

Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



TROPT embraces a multi-faceted understanding of privacy covering individual control over personal information,¹⁶ including the exercise of individual power,¹⁷ and facilitating trust in personal information relationships.¹⁸ We also embrace the comprehensive taxonomical approach to conceptualizing privacy based on the activities that impinge upon privacy: information collection, information processing, information dissemination, and invasions.¹⁹ We reject the limited conceptualization of privacy as secrecy, obscurity, or anonymity only.

What are privacy principles?

In addition to the different schools of thought defining privacy, privacy regulators and practitioners have historically talked about privacy in terms of privacy principles, often referred to as fair information practice principles or FIPPs. Different organizations have offered their own set of privacy principles. While these principles are not consistent across the board, they're largely similar and overlap, except for a handful.²⁰

Figure 2: Table of Different Privacy Principles²¹

Source: TROPT Defining the Privacy Tech Landscape 2021

	OECD	GDPR	DHS	IAPP	PIPEDA	FTC	TROPT
ACCESS / PARTICIPATION						✓	✓
ACCOUNTABILITY		✓	✓	✓	✓		✓
ACCURACY / DATA QUALITY	✓		✓	✓	✓		
CHALLENGING COMPLIANCE					✓		
CHOICE / CONSENT					✓	✓	
COLLECTION LIMITATION	✓			✓	✓		
DATA MINIMIZATION		✓	✓				✓
ENFORCEMENT / REDRESS						✓	
INDIVIDUAL PARTICIPATION	✓		✓	✓			
INDIVIDUAL CONTROL							✓
INDIVIDUAL ACCESS					✓		
INTEGRITY / CONFIDENTIALITY		✓					
INTEGRITY / SECURITY						✓	
LAWFULNESS, FAIRNESS, AND TRANSPARENCY		✓					
LIMITING USE, DISCLOSURE, AND RETENTION					✓		
OPENNESS / TRANSPARENCY / NOTICE / AWARENESS	✓	✓	✓	✓	✓	✓	
PURPOSE LIMITATION	✓						✓
PURPOSE SPECIFICATION			✓	✓	✓		
SECURITY	✓		✓	✓	✓		
STORAGE LIMITATION / RETENTION		✓					✓
TRANSPARENCY							✓
USE LIMITATION	✓		✓	✓			✓

What are TROPT’s Privacy Tech Principles?

We reviewed existing privacy principles and articulate seven TROPT Privacy Tech Principles. We chose six historical privacy principles that apply to privacy tech: transparency, purpose & use limitation, data minimization, individual control, security for privacy, and accountability. We add privacy by design given its importance in the design and development of products, systems, and processes that touch personal information and therefore impact privacy. We purposefully omitted lawfulness/fairness and individual redress from the TROPT Privacy Tech Principles given these two principles involve legal privacy issues, which are not solvable through technology.



TROPT'S PRIVACY TECH PRINCIPLES

TRANSPARENCY

The Transparency privacy principle requires organizations processing personal data to inform and obtain consent from individuals for the processing of their personal data. Transparency is typically required when information is collected and before the data processing scope is expanded beyond the original scope. That said, many believe that organizations should have transparent data processing practices throughout the entire data lifecycle.

Privacy tech examples: consent managers, just-in-time notices, and audio and video indicator lights that notify a user when an app is accessing audio or a green dot when an app is accessing video.

INDIVIDUAL CONTROL

The Individual Control privacy principle provides that individuals should have control over their personal data. Global privacy laws have defined individual privacy rights to include individuals' ability to access, delete, correct, and consent or object to the processing of the personal data that organizations have collected about them. Individual control is relevant throughout the entire data life cycle.

Privacy tech examples: data subject access requests (DSARs), consent management tools, data wallets, and the Global Privacy Control (GPC) protocol.

PURPOSE & USE LIMITATION

The Purpose & Use Limitation privacy principle means that organizations must limit the scope of their data processing solely to the purpose and use specified when data was collected from an individual. This principle comes into play after data is collected and remains a factor until that data is deleted.

Privacy tech examples: Data vaults, data wallets, and some privacy enhancing technologies (PETs) like zero knowledge proofs (ZKP).

DATA MINIMIZATION

Data Minimization requires organization to only process personal data that is relevant and necessary to meet their specified purposes. Data minimization applies throughout the data lifecycle, from collection and use to storage and retention.

Privacy tech examples: data disposition tools and PETs like differential privacy, secure multi-party computation (SMPC), ZKP, edge computing, and local processing address data minimization.

SECURITY FOR PRIVACY

The Security for Privacy principle refers to the need to specifically secure individual privacy, including individual personal data. In addition to building proper structures and using reasonable methods to secure data, security requires data accuracy, integrity, and quality.

Privacy tech examples: homomorphic encryption, access controls, and data disposition tools.

ACCOUNTABILITY

The Accountability privacy principle describes an organization's obligation to demonstrate the internal privacy program, processes, and controls it has in place for processing personal data. The need to be accountable for personal data spans across the entire data life cycle. Organizations need tools to help demonstrate their accountability, spurring many opportunities for privacy tech solutions in recent years, especially in light of global data protection laws like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

Privacy tech examples: data governance for privacy (discovery, inventory, mapping, and cataloging) tools, vendor management tools, privacy program workflow automation tools, and the broader data protection compliance tech solutions that automate privacy program workflows and manage data protection compliance.

Figure 3. TROPT Privacy Tech Principles

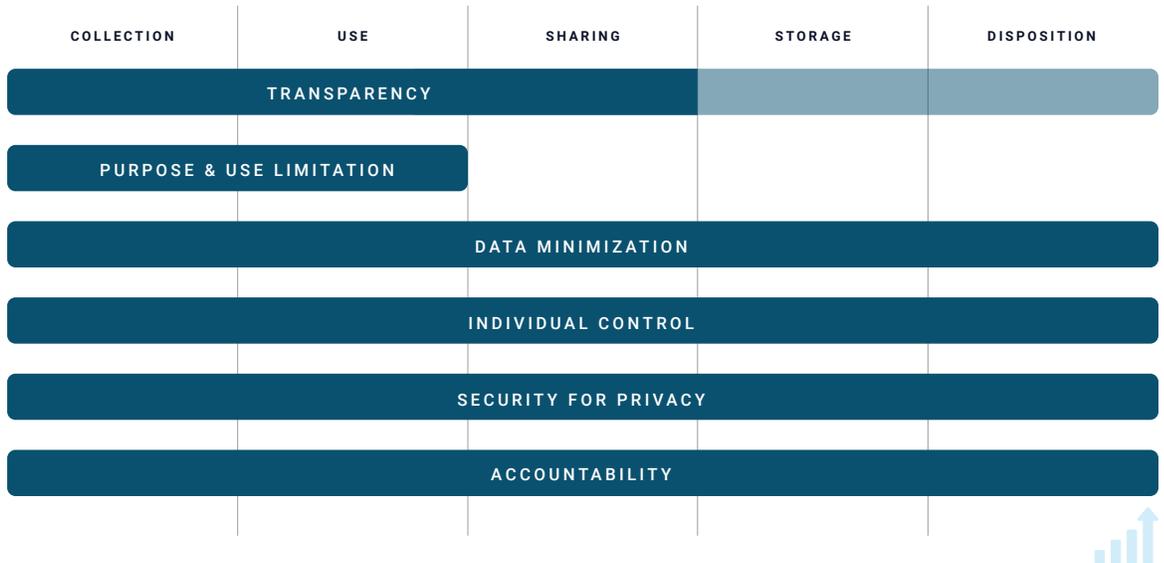
Source: TROPT Defining the Privacy Tech Landscape 2021



Each of the TROPT Privacy Tech Principles ties to at least one stage of the data life cycle. We illustrate this in Figure 4 below, which will come in handy in the next section, where we introduce the TROPT Privacy Tech Stack.

Figure 4. TROPT Privacy Tech Principles & the Data Lifecycle

Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What is technology?

Technology is “scientific knowledge used in practical ways in industry, for example in designing new machines.”²² Technology is also defined as “the use of science in industry, engineering, etc. to invent useful things or to solve problems.”²³



What is privacy tech, then?

Putting the two operative terms together, privacy technologies are simply solutions to privacy problems.²⁴

As explored above, these privacy problems could cover individual control over personal information,²⁵ including exercises of individual power,²⁶ or facilitating trust in personal relationships,²⁷ perhaps through accountability and compliance initiatives. These privacy problems could also cover any one of the seven TROPT Privacy Tech Principles: transparency, purpose & use limitation, data minimization, individual control, security for privacy, accountability, and privacy by design. Privacy technologies solve for any number of these privacy problems, which pop up in many different contexts.

privacy tech

pri·va·cy·tech | \ 'prī-və-sē'tek

1 (noun) technological solutions to privacy problems.

2 (noun) emerging industry of tech companies building solutions to privacy problems, in an effort to address the mounting privacy technical debt.



COMPARING PRIVACY TECH TO OFTEN CONFLATED CONCEPTS

Privacy and privacy tech are often conflated with several other concepts and terminologies. Below, we explore privacy tech's relationship with some of these concepts, including data protection, privacy enhancing technologies (PETs), privacy by design (PbD), privacy engineering, security, privacy-first, and anonymity.

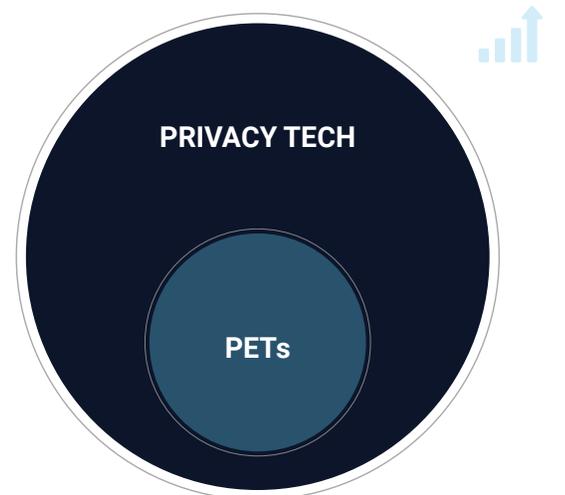
Data protection. Some privacy practitioners distinguish between the terms “privacy” and “data protection.” For example, the EU defines privacy as the “respect for private and family life” and data protection as the “protection of personal data.”²⁸ In practice, data protection focuses on improving organizations’ data protection practices, while privacy focuses on honoring individuals’ privacy preferences. While there are merits to making such a distinction, for the purpose of defining the privacy tech landscape, we hold the semantics and opt for simplicity given our intended broad audience of privacy tech founders and investors, in addition to the domain expert-advisors and user-buyers. In other words, we also consider as privacy tech those tools that solve for data protection problems.

PETs. PETs unfortunately do not have a universally accepted definition.²⁹ That said, the OECD defines PETs as “a wide range of technologies that help protect privacy,”³⁰ aiming to give the individual user (in a consumer context) or technology manager (in an organizational environment) the capability of controlling if, how much, or under what circumstances personal information is disclosed. Therefore, based on this commonly used OECD PETs definition, PETs are privacy tech. But privacy tech tools are not necessarily always PETs because privacy tech solves for other privacy problems beyond control over personal information. Another second widely cited definition of PETs is: a coherent

system of technology measures that protect privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system.³¹ Under this second common PETs definition, a technology that blocks online ad tracking would not be a PET because it interferes with system functionality, but it would certainly be privacy tech because it solves for a privacy problem: an individual’s choice not to be tracked. In comparison, a technology that allowed ad personalization while giving users control over their personal information would be a PET under this second definition and, therefore also qualify as privacy tech because it both solves for a privacy problem (individual control over their personal information) without losing ad functionality.³² Finally, S.224, the recently introduced Promoting Digital Privacy Technologies Act, defines PETs as “any software solution, technical processes, or other technological means of enhancing the privacy and confidentiality of an individual’s personal data in data or sets of data.”³³ S.224’s PETs definition also falls under privacy tech because any software solution, technical process, or other technological means that enhances the privacy and confidentiality of an individual’s personal information in data or sets of data qualify as technological solutions to privacy problems.

Figure 5. PETs & Privacy Tech

Source: TROPT Defining the Privacy Tech Landscape 2021



PbD. As covered in the previous section, lowercase privacy by design simply means privacy aforesight, whereas Dr. Anne Cavoukian's uppercase Privacy by Design articulates the following specific privacy design goals: 1) privacy should be proactive (not reactive) and preventative (not remedial); 2) privacy should be the default setting; 3) privacy should be embedded into design; 4) privacy should not impair full functionality (thus achieving a positive-sum, not zero-sum); 5) it should include end-to-end security to enable full lifecycle protection; 6) it should allow for visibility and transparency (keeping it open); and 7) design should be user-centric and pay respect to user privacy.³⁴ Bringing these concepts together, privacy tech is one of the tangible outputs of both theoretical lowercase privacy by design and uppercase Privacy by Design goals.³⁵

Privacy engineering. Privacy engineering refers to the discrete discipline or field of inquiry and innovation using engineering principles and processes to build controls and measures into processes, systems, components, and products that enable the authorized, fair, and legitimate processing of personal data.³⁶ Privacy engineering provides the discipline (or the how) for translating privacy by design principles (or the what), to an output, such as a process, a feature, or even a privacy tech tool.³⁷

Security. While security and privacy intersect, the two domains are distinct. Information privacy is concerned with the collection, use, dissemination, retention, and other processing of personal information, including the associated individual rights that empower individuals to take control over their personal information. Information security, on the other hand, is concerned with the confidentiality, integrity, and availability of information (not just personal information,

but also trade secrets, intellectual property, and other information that warrants securing) and the systems that process such information. Privacy has its set of principles, as covered above. Security also has its own set of principles, covering confidentiality, integrity, and accuracy (the CIA triad). Security is one of the main privacy principles—in other words, privacy requires security. Given the relationship between the two, there is a category of privacy tech tools that specifically address the security of personal information and therefore overlap with security solutions.

Privacy-first. In addition to PETs, privacy tech also includes “privacy-first” tools, a term used to describe products that provide their service in a privacy-forward way.³⁸ Their service's primary purpose is not to solve for a privacy problem. Instead, their primary purpose could be anything else, including search (in the case of DuckDuckGo), browsing (Brave), and messaging (Signal). That said, these products solve privacy problems despite their non-privacy primary purpose. Brave solves for privacy problems in web browsing; DuckDuckGo, in search; and Signal, in messaging. Because their primary purpose is not privacy, one could argue that the term “privacy-first” is an ill-fitting term for such tools. Other more appropriate terms could be “privacy-forward,” “privacy-friendly,” or simply, “privacy tech.” Setting aside the appropriate label, at the end of the day, these tools solve privacy problems (albeit as a secondary purpose), and therefore fall under the broader umbrella of privacy tech.

Anonymity. With today's government surveillance and surveillance capitalism double whammy, it's not surprising that a slew of anonymity tech tools have popped up. Anonymity tech are tools that help users remain anonymous online. These



include Tor,³⁹ certain Zero Knowledge Proofs (ZKPs) roll-ups to blockchain, and other cryptographic technologies that solve for anonymity. As covered above, privacy is by no means anonymity. Under the privacy as choice and privacy as power schools of thought, privacy covers other individual choices and exercises of power over personal information, beyond the choice to remain anonymous. Anonymity is merely one way to exercise privacy choice—it's an incomplete take on privacy but is still an important aspect of it. Anonymity tech solves for an individual's problem stemming from a desire to remain anonymous in certain contexts in this highly connected and surveilled world. Because privacy includes individual choice, the choice to remain anonymous in certain contexts often raises privacy problems when not honored. This means that anonymity tech is covered under the broader umbrella of privacy tech.

COMPARING PRIVACY TECH TO ADJACENT INDUSTRIES

There are several adjacent industries that overlap with privacy tech, but do not fall entirely within privacy tech. These industries include cybersecurity, data infrastructure, identity management, adtech and martech, legaltech, data governance, and Web 3.0. We explore and distinguish these industries from privacy tech below.

Cybersecurity

Privacy and cybersecurity are perhaps the closest overlapping industries, but are not one-and-the-same. Cybersecurity and privacy tech products protect personal information, but cybersecurity goes beyond protecting personal information

and extends more broadly to protecting systems, networks, devices, and infrastructure, in their entirety. Conversely, privacy tech products do more than just protect personal information; they also solve other types of privacy problems beyond insecurity, like transparency, data minimization, purpose and use limitation, etc. It bears noting that cybersecurity tools are not privacy-preserving by default; on the contrary, when not designed and engineered with privacy in mind, these tools often raise privacy issues such as surveillance and overcollection of personal information.

Identity management

Identity management technologies are tools for identifying, authenticating, and authorizing an individual or a group of individuals.⁴⁰ Identity management products include tools for governing and administering unique user digital representations.⁴¹ Identity management overlaps with privacy tech because both solve for problems relating to individual control over their identity, accuracy of personal information relating to identity, and secure access tied to identity-based personal information. That said, identity management tech goals extend beyond solving for privacy problems.

Data infrastructure

Data infrastructure tools promote data sharing and consumption.⁴² They cover data sources like SQL databases, cloud storage providers, data warehouses, and business intelligence tools. These tools process personal information, among other information types. Data infrastructure products were not historically built with privacy in mind. The data infrastructure space overlaps with the privacy tech space because some privacy tech tools address data infrastructure privacy problems.



Martech & adtech

Marketing technology, or “martech” is “a set of software solutions used by marketers to support mission-critical business objectives and drive innovation within their organizations. Martech solutions focus on content and customer experience, advertising, direct marketing, marketing management and marketing data and analytics.”⁴³ Martech products solve marketing problems and include digital asset management, marketing automation, chatbots, and social media monitoring tools. Advertising technology, or “adtech,” refers to analytics and technical tools used in the advertising space that help advertisers reach their target audiences.⁴⁴ Because adtech solutions generally rely on consumer data, privacy is a necessary consideration when building these products, especially in light of increasing global data protection laws. Similar to the data infrastructure space, adtech is adjacent to privacy tech because adtech raises privacy problems, some of which require privacy tech solutions. That said, the two spaces are distinct: adtech primarily solves for advertising problems, whereas privacy tech solves for privacy problems. Overlapping adtech and privacy tech startups provide solutions for privacy problems in advertising. Adtech and martech are getting disrupted by privacy tech, partly due to Google’s phasing out of third-party cookies, Apple’s rollout of platform privacy requirements and restrictions, and the passage of global data protection laws.

Legal tech

Legal tech generally covers products that legal teams use to facilitate their processes and improve their effectiveness, including maintaining simple, efficient legal operations. Legal tech provides operational and administrative solutions for document or contract management, legal research, case or practice management, and client intake automation. Legal tech is also on the

rise and has already seen over \$1 billion in funding in 2021.⁴⁵ Legal tech overlaps with privacy tech by offering data protection law compliance tools and privacy legal workflow automation tools.

Data governance

Data governance is “a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods.”⁴⁶ Data governance for privacy tools cover products that help Information Technology (IT), Governance, Risk, and Compliance (GRC), and privacy teams govern data, including data discovery, inventory, mapping, cataloging tools. Data governance overlaps with privacy tech where it uses technical solutions to manage privacy program management or data protection compliance, which often involve governing personal information.

Web 3.0

Web 3.0 is the next stage of the Internet. Because this stage is currently still being built, it has no universal definition yet. In practice, however, Web 3.0 refers to an Internet that is made possible by decentralized networks. The Web3 Foundation refers to Web 3.0 as “a decentralized and fair internet where users control their own data, identity, and destiny.”⁴⁷ Web 3.0 and privacy tech both have the shared goal of enabling user control over their personal information and identity. That said, privacy tech covers more than just Web 3.0 technologies, and vice versa.

Notwithstanding the intersections with the aforementioned adjacent industries, privacy tech is an emerging industry on its own, focusing on technical tools that specifically aim to solve privacy problems.

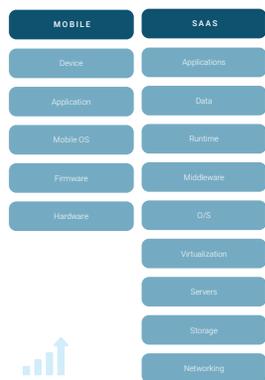


CATEGORIZING PRIVACY TECH: THE TROPT PRIVACY TECH STACK

But What is a Tech Stack?

A technology stack is simply a collection of different technologies that work together for some purpose, usually as a foundation for systems such as websites, mobile apps, and standalone desktop software.

Figure 6. Sample Existing Tech Stacks
Source: TROPT Defining the Privacy Tech Landscape 2021



In addition to defining privacy tech in the previous section, we now introduce the TROPT Privacy Tech Stack as a framework for categorizing privacy tech startups.

The challenge with creating a privacy tech stack is that personal information runs through almost every type of system: websites, mobile, desktop software, and SaaS. There are existing tech stacks for these different systems, into which different privacy tech tools can fit neatly.

Despite the challenge, we’ve taken a stab at visually articulating the emerging privacy tech stack, pulling from foundational privacy and technology knowledge.

We start with the B2B side of the TROPT Privacy Tech Stack, which organizes the privacy tech tools that organizations use to solve privacy problems throughout the personal data lifecycle. Data lifecycle privacy tech products are built for privacy lawyers, Information Security (InfoSec) Governance, Risk, and Compliance (GRC) professionals, data scientists, and other professionals whose work involves personal information. Then, we took a first stab at fleshing out the B2B development lifecycle side of the TROPT Privacy Tech Stack in line with the “shift left privacy” trend, which reflects the desire to solve

privacy problems earlier on, shifting left from the data lifecycle to the development lifecycle. This side includes privacy tech tools needed during the development cycle, when products and systems are built, before personal information is processed. Development lifecycle privacy tech products are built for professionals involved in the development lifecycle, which primarily include developers, program managers, and privacy engineers. We then cover the B2C and B2B2C privacy tech developments we’re seeing. B2C and B2B2C privacy tech are trailing behind the B2B side of privacy tech, which has seen most of the action in product development and investment dollars.

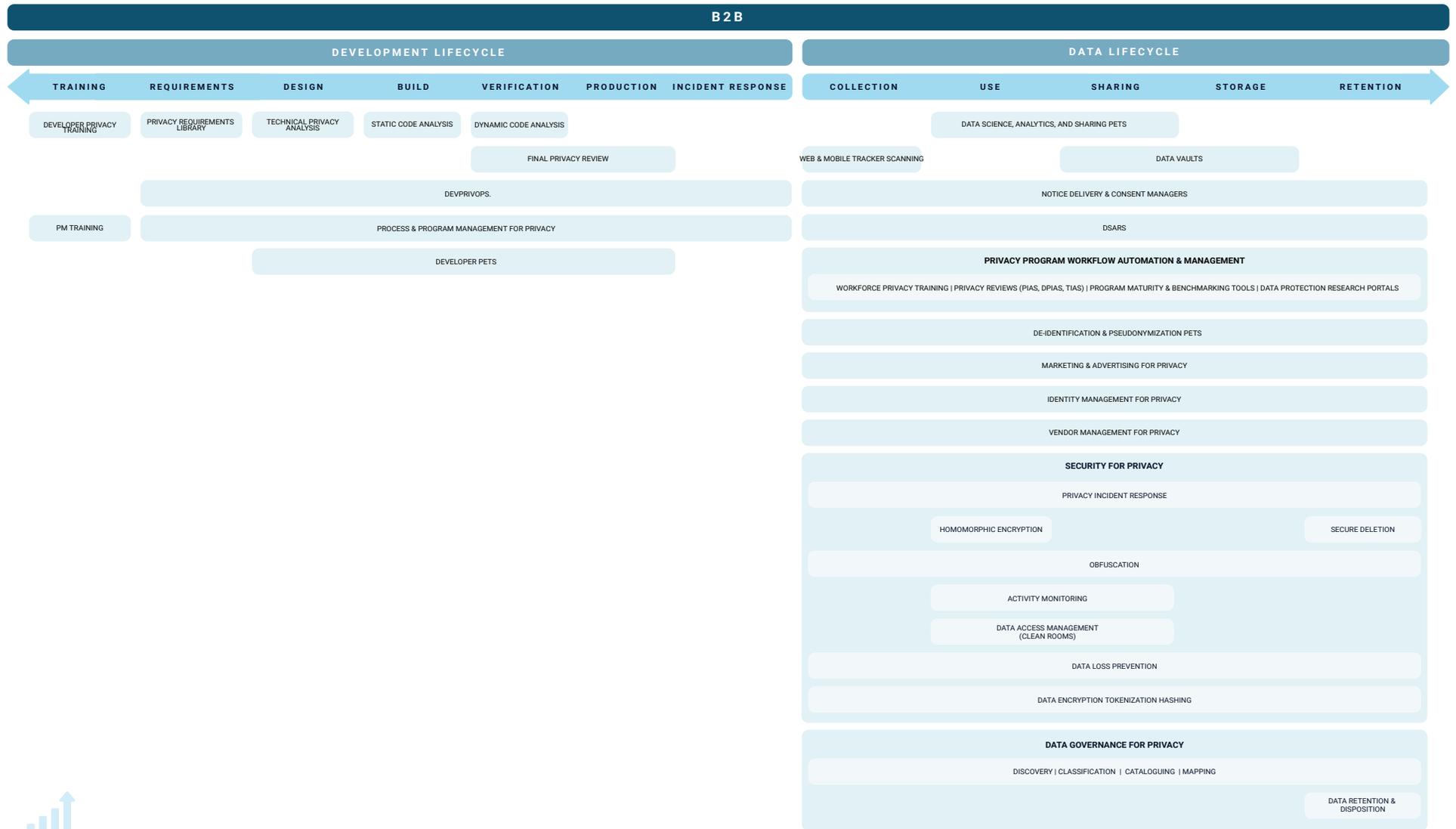
THE TROPT PRIVACY TECH STACK

The TROPT Privacy Tech Stack categorizes privacy tech based on where they fall under the data lifecycle (because data privacy is concerned with personal data) and the development lifecycle (because many privacy problems arise even before personal information is collected, when decisions are made at the development stage).



Figure 7. TROPT B2B Privacy Tech Stack

Source: TROPT Defining the Privacy Tech Landscape 2021



B2B PRIVACY TECH TOOLS: DATA LIFECYCLE TOOLS

Most of the mature privacy tech solutions fall under the B2B data lifecycle side of the TROPT Privacy Tech Stack. We break down the data lifecycle according to the following phases: collection, use, sharing, storage, and disposition. We cover some data lifecycle privacy tech categories in the following sections.

Notice delivery and consent managers

Notice delivery privacy tech tools inform individuals when their data is being collected or, if notice is required at a later stage of the data lifecycle, when their data is processed for a new purpose or shared with third parties beyond the scope of the original purpose. Examples include just-in-time notifications or the colored lights that indicate when an application is using a device's microphone or camera. Consent management tools are another example; they request, store, and enforce individual consent preferences. Cookie consent managers are a specific and common example of consent management tools. Notice delivery and consent managers apply throughout the data lifecycle.

Web & mobile tracker scanning

In 2020, at least eighty-seven percent (87%) of the most popular websites in the world engaged in some form of digital tracking, even when the website visitor did not sign into an account on the site.⁴⁸ Web and mobile tracking scanners are tools that give organizations visibility into the trackers on their websites, apps, and other assets. These tools scan websites and apps to reveal cookies, beacons, pixels, etc. crawling on the site or app.⁴⁹ These products typically sit at the data collection points.

Data science, analytics, and sharing PETs

Data science, data analytics, and data sharing privacy enhancing technologies (PETs) allow organizations to analyze and share personal information—and thus unleash its value—in a privacy-respectful way. These PETs provide data scientists, marketers, customer service teams, and any data-driven organization, insights about people, including their customers, while building in privacy into the analytics and sharing environment. Data science, analytics, and sharing PETs are used during the use and sharing phases of the data lifecycle.

These include products that incorporate differential privacy, Secure Multi-Party Computation (SMPC), Zero Knowledge Proofs (ZKPs), and other PETs.⁵⁰

Data vaults

Data vaults are secure data storage and data sharing systems that allow organizations to interoperate across different systems while minimizing data privacy and security risks.⁵¹ Data vaults often leverage PETs. Data vaults are used during the sharing and storage phases of the data lifecycle.

Data subject access rights (DSARs)

B2B Data Subject Access Rights (DSAR) tools facilitate organizations' responses to individual exercises of their data subject rights provided under global data protection laws, such as the General Data Protection Regulation (GDPR), the California Privacy Rights Act (CPRA), and the Virginia Consumer Data Protection Act (VCDPA).⁵² Data



subject rights empower people to access, correct, delete, port, or otherwise exercise control over the personal information that organizations have collected on them. B2B DSAR tools help organizations handle intake, routing, execution, and/or response to DSAR requests. These tools are used throughout the data lifecycle.

Identity management for privacy

As covered under the adjacent industries section, the scope of identity management extends beyond privacy, but privacy is a growing component of identity management programs, and incorporating privacy into these programs is becoming strategic for organizations who want to protect both their users and their companies.⁵³ Privacy goals can be achieved through identity management by verifying user identities with authentication solutions, and through awareness and education.⁵⁴ These tools are used throughout the data lifecycle.

Privacy program workflow automation & management

Privacy program workflow automation and management tools help privacy program teams manage compliance with data protection laws, regulations, and requirements, which are constantly evolving. Several privacy tech tools have been built in response to these compliance obligations and privacy program management pain points. These tools cover privacy reviews (privacy impact assessments, data protection impact assessments, transfer impact assessments, etc.), workflow automation tools, privacy program maturity and benchmarking tools, privacy and data protection research portals, etc. These tools are used throughout the data lifecycle.

Vendor management for privacy

Vendor management for privacy tools help assess third-party vendors' privacy and security policies, access controls, and information storage practices. They include tools that automate vendor privacy questionnaires, score or benchmark vendor privacy controls, and audit vendor privacy controls. These tools are used throughout the data lifecycle.

Security for privacy

Security for privacy tools is a category intersecting both cybersecurity and privacy tech. It covers tools that specifically secure personal information. It includes incident response, data loss prevention (DLP), obfuscation, and access management tools. Data obfuscation includes techniques meant to "camouflage [personal data] in ways that render it useless if it were accidentally exposed or stolen."⁵⁵ Data obfuscation includes encryption, tokenization, and hashing technologies. Access management tools govern access to personal data (and other types of data). DLP tools ensure that personal data or other sensitive data is not lost or misused. These tools are used throughout the data life cycle.

Data governance for privacy

Data governance tools help with process or policy management, data glossaries, metadata management, or document classification.⁵⁶ There is a subset of data governance tools that solve for data protection problems. They include data discovery, classification, and cataloging tools, which are the building blocks for managing data processing and data



protection compliance. Data discovery tools help identify personal data. Discovery should be continuous, as definitions and practices evolve. Data classification tools help categorize personal data and match discovered personal data to the specific categories. Once classified, data cataloging helps establish safeguards to properly protect personal data.

Data disposition tools help organizations dispose of personal data, taking into consideration data protection and other regulatory requirements,⁵⁷ retention policies, consumer requests, legal holds and legitimate business needs.⁵⁸ Automated records retention tools, in particular, can help organizations track retention policies and schedules, to determine whether to retain or dispose of personal data. These tools are used throughout the data lifecycle.

De-identification, pseudonymization, and anonymization PETs

De-identification and pseudonymization PETs help render individuals less identifiable, while anonymization PETs render them anonymous. From a data level, de-identification removes identifying data associated with a particular data subject from a data set.⁵⁹ Whereas, pseudonymization masks direct identifiers from personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information.⁶⁰

Meanwhile, anonymization goes further to render personal data anonymous in such a way that the data subject is not or no longer identifiable. De-identification, pseudonymization, and anonymization PETs include encryption, tokenization, hashing, differential privacy, etc.⁶¹ These tools are used throughout the data lifecycle.

Marketing & advertising for privacy

Marketing & advertising rely heavily on personal data, raising privacy

problems. These adjacent industries are experiencing an upheaval, resulting to privacy tech solutions popping up at the intersection of these industries to address marketing and advertising privacy problems.

B2B PRIVACY TECH: DEVELOPMENT LIFECYCLE TOOLS

Perhaps one of the more exciting developments we're seeing in privacy tech—at least in the B2B space—is the emergence of developer privacy tech tools. We break down the development lifecycle according to the following phases: training, requirements, design, build, production, and incident response. Development lifecycle privacy tech tools help solve for privacy problems before they materialize, or at least before data is collected, shifting left from the data lifecycle to the development lifecycle. This B2B development lifecycle privacy tech trend is in line with the adjacent shift left security trend.⁶² These products equip developers, privacy engineers, and other technical privacy teams with the tools they need to design and engineer privacy into the development lifecycle, build privacy-respecting products, and avoid common privacy pitfalls in developing products, software, systems, and technologies, in general.

Developer privacy training & awareness

Developer privacy training & awareness platforms provide training and awareness tools for developers, engineers, and other technical teams on common privacy problems that may arise during the development process. These tools are used during the training phase of the development lifecycle.



Privacy requirements library

Privacy requirements libraries provide developers with a library of privacy rules and requirements to translate into code in developing products and systems that touch personal information. These tools are used during the requirements phase of the development lifecycle.

Technical privacy analysis

Technical privacy analysis tools help engineers identify potential privacy problems. The tools help create repeatable rules to address privacy gaps, which developers in turn can implement during the build stage. Technical privacy analysis tools help scale privacy engineers and other technical privacy domain experts by making guidance available across the platform to engineers. These tools are used during the design phase of the development lifecycle.

Developer PETs

Developer PETs are tools developers use to enhance privacy during the development lifecycle. For example, developers could use “dummy data,” or data sets with randomly generated, made-up data points. Developers could also choose to use “synthetic datasets,” which are data created artificially instead of being generated from the world. Or they could choose to deploy other PETs like differential privacy, SMPC, homomorphic encryption, etc. in the development of new systems and products. These tools are used during the design, build, QA, and production phases of the development lifecycle.

DevPrivOps

DevPrivOps tools are a suite of tools for developers to help build in privacy design and engineering into products. They cover tools from the requirements stage to the production stage of the development lifecycle.

Process and program management for privacy

Process and program management for privacy tools are tools for technical privacy program managers and privacy engineers. They cover the entire development lifecycle.

Static code analysis

Static code analysis privacy tech tools scan static (non-running) source code for privacy bugs or privacy problems in the code. These tools analyze code against a set (or multiple sets) of coding rules for privacy. They're typically used during the build stage of the development lifecycle.

Dynamic code analysis

Dynamic code analysis tools scan code while the software is running to improve the identification and correction of privacy bugs. They're typically used during the verification stage of the development lifecycle.

Final privacy review

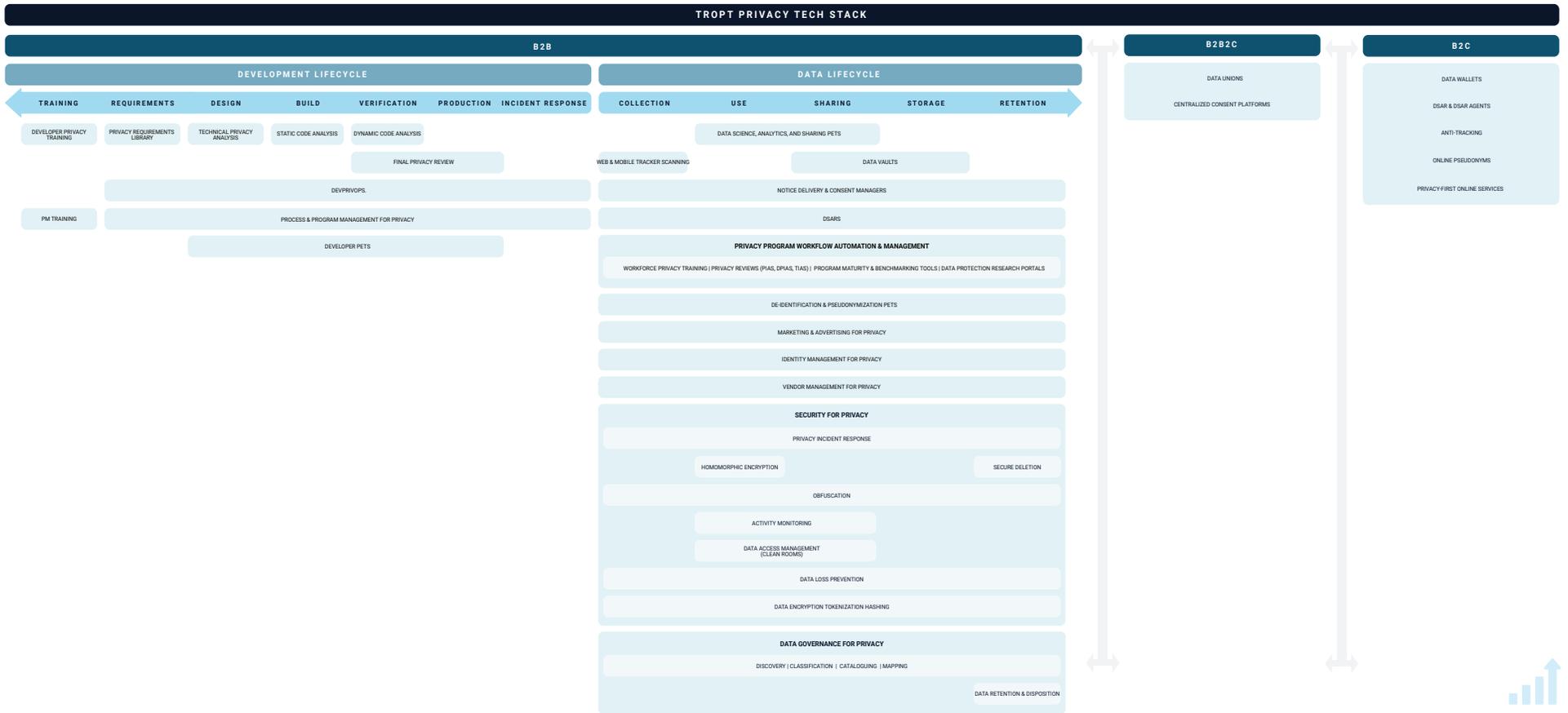
Final technical privacy review tools facilitate the examination of all privacy requirements, privacy problems raised, privacy bugs identified in the code, and mitigations put in place during the early stages of the development lifecycle. Final technical privacy review tools are used just before or at the beginning of production.

Despite privacy's central focus on individuals, privacy tech tools for organizations (i.e. B2B) as outlined above have gained more momentum than their consumer privacy (B2C) counterparts, which we turn to in the next section. This is due to several factors, including global data protection laws regulating organizations' privacy practices, budget, competing business models in the consumer space, etc.



Figure 8. TROPT Privacy Tech Stack

Source: TROPT Defining the Privacy Tech Landscape 2021



B2C PRIVACY TECH

Despite trailing behind their B2B counterparts, B2C privacy tech products are slowly emerging and gaining increasing adoption. Common B2C privacy tech products include data wallets, data subject access rights (DSARs) and DSAR agents, anti-trackers, online pseudonyms, and privacy-first online services.

Data wallets

Data wallets allow users to securely store and encrypt all of their personal data in one unique electronic ecosystem, like a wallet, and then choose which third parties can access which data points through the wallet. Their goal is to give individuals complete control over who can access their personal data.⁶³

DSARs & DSAR agents

In addition to the B2B DSAR tools previously covered, there are also available DSAR tools for individuals that empower them to exercise their rights to access, correct, delete, port, and exercise general control over their personal information. DSAR agents are services that execute DSARs on consumers' behalf. B2C DSAR tools currently have not gained as much traction as their B2B counterparts. This is likely due to a couple of factors: 1) the general sentiment that consumers should not be paying for privacy; and 2) perhaps relatedly, the historical investment focus on the B2B side of privacy tech.

Anti-trackers

Anti-tracking tools help consumers block tracking of their personal information, activities, and behavior in different

contexts, including at the browser, app, and device levels. Anti-trackers include private browsers, browser plug-ins, operating system features blocking cross-app tracking, and VPNs. Anti-trackers address consumer problems relating to control, including the preference to remain anonymous in certain contexts.

Online pseudonyms

Online pseudonyms are tools that allow individuals to interact pseudonymously online. Avatars are a common example of these tools. Other examples include products that automatically generate random email addresses, phone numbers, or other personal information, in lieu of sharing real information.

Privacy-first online services

We're also seeing privacy-first online services emerge as alternatives to the surveillance-heavy status quo. In the private messaging space, as covered above, Signal gained widespread adoption among consumers. Among email service providers, ProtonMail provides an end-to-end encrypted email platform. Brave addresses private Internet browsing and video conferencing. DuckDuckGo provides a search engine that doesn't collect personal information, track online search, or deliver ads.



B2B2C PRIVACY TECH

B2B2C privacy tech tools bridge the B2B and B2C sides of the privacy tech landscape. These tools service both business and consumer-side users. Today, the most common B2B2C privacy tech platforms we're seeing are centralized consent platforms and data unions.

Centralized consent platforms

Centralized consent platforms aggregate consumer consent preferences. Organizations that wish to process personal information for various reasons can engage with centralized consent platforms to source data that has been collected legitimately, oftentimes with explicit consent and specific parameters for data processing purposes.

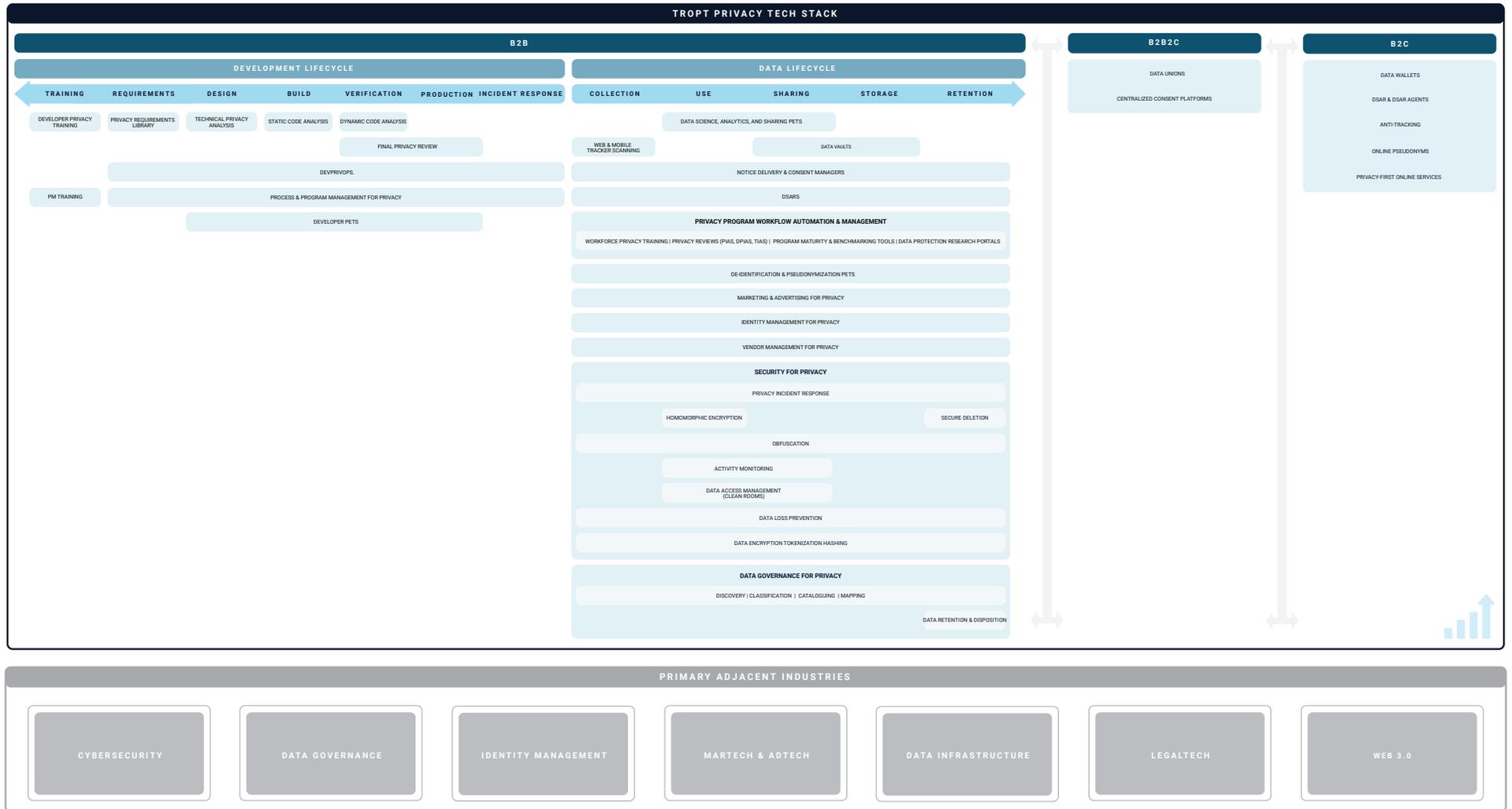
Data unions

Beyond centralized consent platforms, there are data union platforms that allow for the bundling of personal information with others. They distribute a share of the revenue when someone pays to access it. Some studies have shown that individual personal information on its own fetches a nominal price. When combined in a data union, aggregated data increases in value. Data unions typically allow individuals to choose which types of personal information they want to share and the amount of value they wish to receive in exchange for sharing their data, if any. Oftentimes, the data union dividend is in the form of cryptocurrencies. Data unions allow individuals to have more control over the monetization of their personal information.



Figure 9. TROPT Privacy Tech Stack + Adjacent Industries

Source: TROPT Defining the Privacy Tech Landscape 2021



BRIDGING GAPS BETWEEN PRIVACY TECH KEY PLAYERS

In addition to defining and categorizing privacy tech in the previous sections, we also surveyed privacy tech key players to bridge the tech-capital-expertise gaps in privacy tech. We highlight some of the findings in this section and share the full survey responses in the Appendix towards the end of this whitepaper. As this is the first year that we’re conducting this research, we anticipated receiving a small sample size of responses to our inaugural survey. We addressed this by conducting more in-depth interviews with select privacy tech leaders, some of whom we quote in the following sections.

WHAT PRIVACY TECH BUYERS AND USERS SAY

B2B privacy tech users and buyers—from engineers to lawyers—shared their insights on the privacy tech space. They shared the following privacy problems for which they’re trying to find technology solutions: data minimization, retention, vendor management, de-identification, accountability tools, and “shift left privacy” tools for developers. We highlight some of these pain points in the following sections.

Figure 10. Privacy Tech Buyers and Users Survey: Privacy Problems They Need Solved Through Tech



Source: TROPT *Defining the Privacy Tech Landscape* 2021

Data Minimization. With today’s data-driven business models, strategies, and products, cross-functional teams (typically privacy, security, IT, and data governance professionals) are oftentimes faced with the uphill battle of pushing their organizations to adopt data minimization, the privacy principle and legal requirement to only collect, use, and store personal information that is relevant and necessary to meet specific legitimate purposes. They are looking for tools that address data minimization in an automated fashion, at earlier stages in both the data and development lifecycles. Such tools not only help minimize privacy risks; they also force organizations to only process relevant data that truly holds value for them, making it easier for them to unleash such data’s value, instead of incurring data liabilities and drowning in proverbial data lakes.

Data Retention and Disposition. It’s unsurprising that respondents listed data retention as a privacy problem for which they need privacy tech solutions. End-of-life data retention and disposition is one of the more difficult problems privacy, data governance, and GRC professionals are facing. There are a few reasons for this. First, these practitioners are going up against the default resistance to getting rid of data, partly due to the ingrained Big Data mindset of collecting all the data, all the time. Second, data retention and disposition sit at the very end of the data lifecycle, away from the prying eyes of customers and the public. Privacy practitioners tend to prioritize more public-facing privacy problems like transparency (notice and consent)



and DSARs. Third, most data protection laws contain retention and deletion obligations, which add to their criticality. Lastly, the proliferation of shadow IT systems makes retention and deletion technically difficult to solve.

Vendor and Third-Party Management.

Privacy tech users and buyers raised vendor and third-party management as another problem area that requires privacy tech solutions. This is unsurprising given the increased regulatory⁶⁴ and customer scrutiny over how organizations govern personal information sharing with their vendors and other third parties.

DevPrivOps (or Shift Left Privacy Tools for Developers). Perhaps one of the most exciting trends we're seeing in the privacy tech space is the desire to solve privacy

problems earlier on in the data lifecycle. We see the "shift left privacy" trend extend even before the ingestion point or the beginning of the data lifecycle. We're seeing the trend move further left, before data is collected, shifting as early as the development lifecycle, when products are developed and critical decisions affecting privacy are made. Privacy tech users and buyers want visibility into how code and algorithms affect privacy. They want code governance tools for developers to use. In other words, they want to stop the privacy problems before they're created.

"Vendor and third-party management have been one of the most challenging things. I can't build a team of developers that is so big that we're able to build everything in house, to ensure that everything is built with privacy by design. We have to use other services and, unfortunately, there are not that many options yet in all those services to find something that is also building with privacy by design ... Privacy and security reviews are particularly painful, not just painful for vendors, but also for the companies asking for such reviews. They're causing internal strife. They're causing tension between stakeholders because the privacy team, or information security team, or whoever's requiring the privacy and security reviews, enter into conflict with their own co-workers. They're slowing down their co-workers from getting the tools that they need."

Caroline McCaffery, CEO & Founder, ClearOPS

"A lot of existing tech startups tend to optimize for privacy when it comes to data at rest. They will look at the warehouse and they'll say we'll do deletion, we'll do categorization, inventory, and discovery. The problem is at that point, think of a horizontal funnel going from left to right, the narrow end on the left-hand side, if you approach it on the right-hand side, the size of data has grown significantly. I'm asking privacy tech startups, 'Can you intercept data at the point of ingest, before any engineer gets their hands on it?' so you can tag the data early on, you can delete some data before it gets used, made copies of, or before— heaven forbid it—data gets shared with a third party. I'm looking at startups that are trying to catch data right at the point of ingestion ... 'shift left' as people say." That's something I optimize for.

Nishant Bhajaria, Head of Privacy Engineering and Architecture, Uber



Privacy tech buyers and users reported the following challenges in navigating the privacy tech space in general: ineffective solutions, visibility into privacy tech products, inability to integrate with existing systems, nonexistent solutions, and evolving data protection landscape.

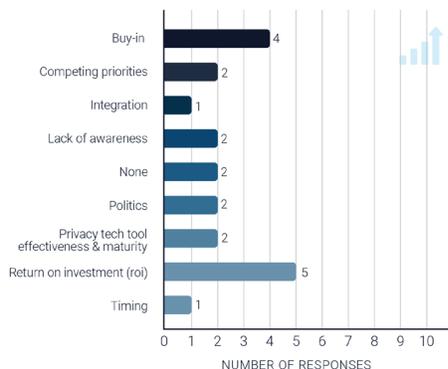
Figure 11. Privacy Tech Buyers and Users Survey: Challenges in Navigating the Privacy Tech Space



Source: TROPT Defining the Privacy Tech Landscape 2021

They also disclosed facing the following specific challenges in securing budget for privacy tech tools: return on investment (ROI), buy-in, competing priorities, politics, product maturity, integration, and timing. Their reported team and company budgets varied widely, as illustrated in Figures 13 and 14.

Figure 12. Privacy Tech Buyers and Users Survey: Challenges in Securing Privacy Tech Budget



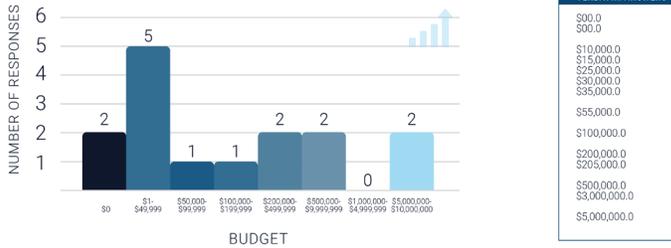
Source: TROPT Defining the Privacy Tech Landscape 2021

“If you buy a tool, who is going to wield the political clout within the company to make sure that there is onboarding? It’s very critical before you onboard a third-party [privacy tech] tool, that you have a clear understanding of how onboarding is going to work. What is the tech stack looking like? Is there any platform compatibility? It’s important to make sure that the engineers are in the room and the decision is being made. What typically happens is the engineers have to do a proof of concept after somebody else has made the decision to acquire the tool. This creates inefficiencies and disconnect. So, my recommendation would be, bring the engineers to the discussion sooner, because at the end of the day, if you don’t, the engineers will leave. Or they’ll do a bad job of onboarding it.”

Nishant Bhajaria, Head of Privacy Engineering and Architecture, Uber



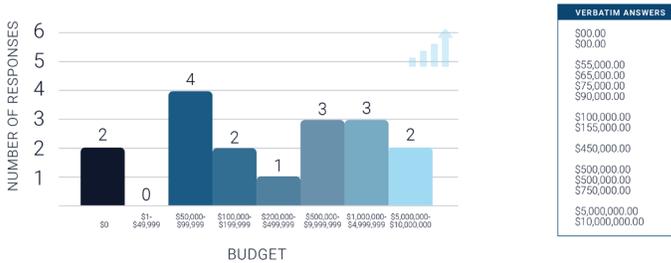
Figure 13. Privacy Tech Buyers and Users Survey: Privacy Tech Team Budget



Average team privacy tech tool budget: \$655,357.14
 Median team privacy tech tool budget: \$45,000.00

Source: TROPT Defining the Privacy Tech Landscape 2021

Figure 14. Privacy Tech Buyers and Users Survey: Privacy Tech Company Budget



Average company total privacy tech tool budget: \$1,267,142.86
 Median company total privacy tech tool budget: \$127,500.00

Source: TROPT Defining the Privacy Tech Landscape 2021

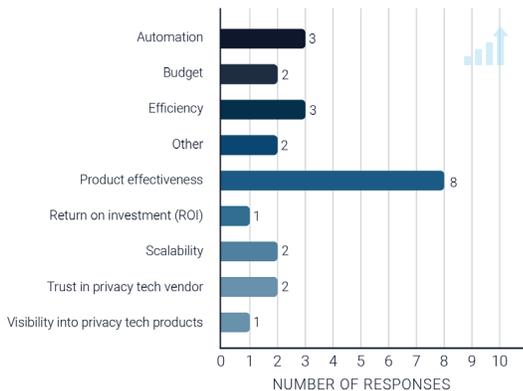
Buyers and users shared the following factors that affect their interest in buying privacy tech tools: product effectiveness, automation, efficiency, budget, scalability, vendor trust, and ROI.

“The privacy tech space is ripe for growth and opportunity - it’s very important that new solutions focus on creating scalable processes that cleanly interoperate with an organization’s existing tech stack.”

Fatima Khan, Director, Corporate Counsel, Privacy & Product, Okta



Figure 15. Privacy Tech Buyers and Users Survey: Factors Affecting Interest in Buying Privacy Tech



Source: TROPT Defining the Privacy Tech Landscape 2021

"I think there's a lot of pitches and there's a lot of promises that the vendors generally make and that's their business. I don't fault them for doing that. But I think when you're out laying some money and you're working on implementation of a new product especially in the privacy sphere where there's just not a lot of history there, you're trusting that what the marketing speak is around a particular feature or functionality of privacy tech is actually going to work in practice and it's going to make somebody's life better or easier. Someone who's working in the privacy program, a privacy engineer or data analyst, whoever you have on your operations team, it's going to make their life better, easier, get them the information in an easier way ... I'm not expecting perfection, I'm not expecting the silver bullet. I'm not expecting that a particular product is going to make a company compliant. I actually don't appreciate when the privacy tech guarantees compliance because I think there's always going to be some choices that you're going to have to make in terms of how you implement a tool."

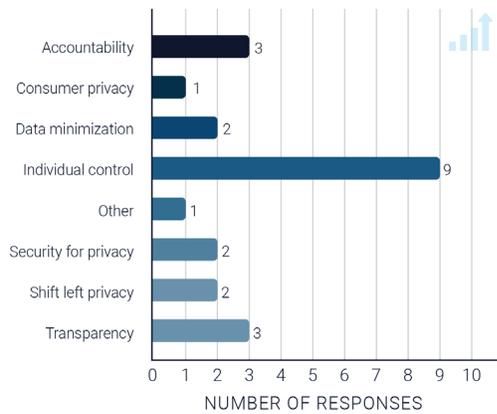
Alexandra Ross, Senior Director, Senior Data Protection, Use, & Ethics Counsel, Autodesk



WHAT PRIVACY TECH STARTUP FOUNDERS & OPERATORS SAY

Privacy tech startup founders & operators also shared their own takes on the privacy tech landscape. Respondents shared that they are building tech solutions to address the following privacy problems: individual control, demonstrating accountability, transparency, security for privacy, shifting privacy left, and data minimization, amongst other things.

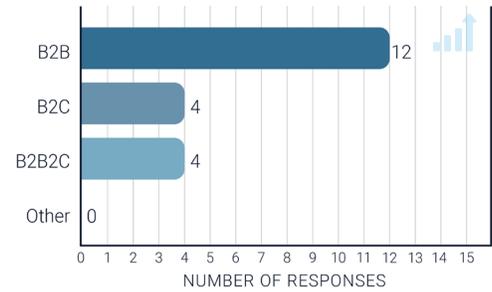
Figure 16. Privacy Tech Startup Survey: Privacy Problems They're Solving



Source: TROPT Defining the Privacy Tech Landscape 2021

Most surveyed privacy tech founders reported building for the B2B side of the privacy tech market. This aligns with the maturity of the TROPT B2B Privacy Tech Stack, compared to the B2C and B2B2C sides. Founders on the B2B side indicated that they were building primarily for Legal & Compliance and IT & Data Governance teams, with Marketing, Engineering, InfoSec, and other teams trailing behind. B2C privacy tech startups are also lagging behind their B2B counterparts.

Figure 17. Privacy Tech Startup Survey: Business Model

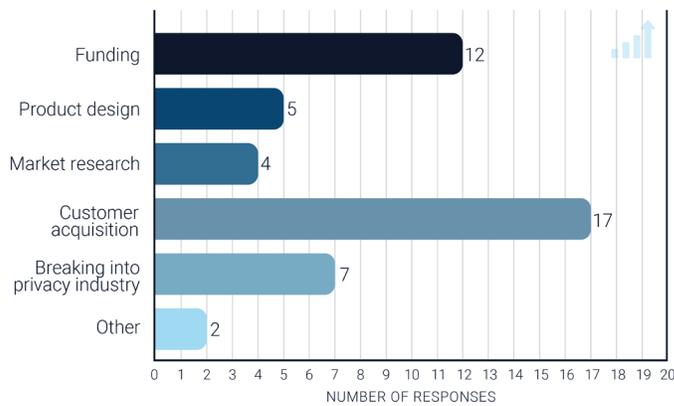


Source: TROPT Defining the Privacy Tech Landscape 2021



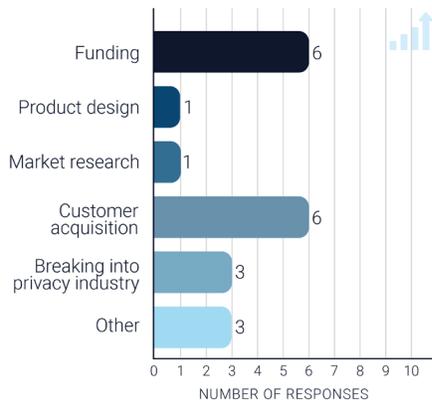
Responding privacy tech startup founders & operators named customer acquisition, funding, breaking into the privacy field, product design, and market research as their pain points, with funding and customer acquisition tied as their single biggest pain points.

Figure 18. Privacy Tech Startup Survey: Startup Pain Points



Source: TROPT Defining the Privacy Tech Landscape 2021

Figure 19: Privacy Tech Startup Survey: Single Biggest Startup Pain Points



Source: TROPT Defining the Privacy Tech Landscape 2021

“Two years ago, no one understood the pain-point we were solving. About 6 months ago, if you were trying to raise funding as a privacy tech company, you were immediately put in the bucket of a cybersecurity company. However, when we got to cybersecurity investors, they didn’t think we qualified as a true cybersecurity [startup] investment, either, because we’re not a product that plugs neatly into the security tech stack, so that was a big challenge. Right now, in talking to investors, I’m finding that there is a huge variation in the levels of research that they have done into the privacy tech landscape. Their levels of understanding vary and there’s no cohesion yet. I have rarely met an investor who has dug into privacy tech as a standalone sector.”

Caroline McCaffery, CEO & Founder, ClearOPS



WHAT PRIVACY DOMAIN EXPERTS SAY

We also surveyed and interviewed privacy domain experts who are involved in the privacy tech space. They cited the following reasons for their interest in the emerging privacy tech landscape: advisory roles, consulting roles, in-house privacy roles, purchase opportunities, and Angel investment opportunities. Most offer direct privacy tech experience, while others cite their experience in building privacy programs, working in relevant industries (B2B, B2C, adtech, etc.) as additional expertise they offer. Privacy domain experts named development of privacy thought leadership and privacy product design as the top challenges with which they can help privacy tech startups.

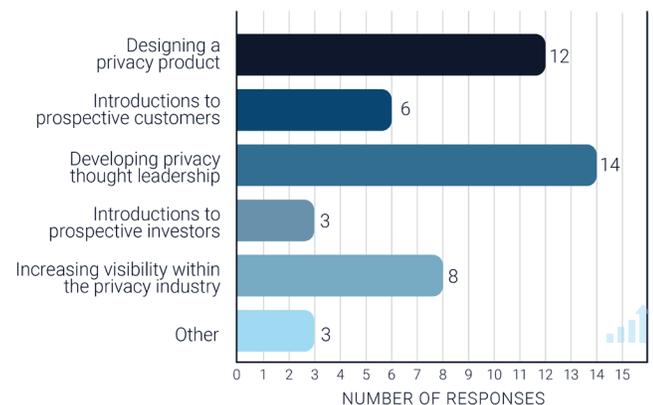
“I’ve learned that you really want to make sure that you’ve got a great working relationship with the founders. So, I now ask myself, “Are these founders people I want to work with long term? Are they coachable? Do they want my advice or does the team simply want to use my name and reputation for fundraising and marketing purposes?”

Debra Farber, CEO, Principled LLC

“When assessing whether to work with a privacy tech startup, I think a lot of the personality of some of the founders and the people that you’re going to be working with at that startup ... So much of it comes down to the actual execution, the founders, and whether I click with them. Do I get a sense that they’re really going to pursue some of their goals? That’s what sort of attracted me and led me to partner with some of the companies that I have advised over the years ... It’s really that sort of human connection with the people involved that are working on implementing that technology.”

Alexandra Ross, Senior Director, Senior Data Protection, Use, & Ethics Counsel, Autodesk

Figure 20. Privacy Domain Expert Survey: Challenges Experts Help Privacy Tech Startups With

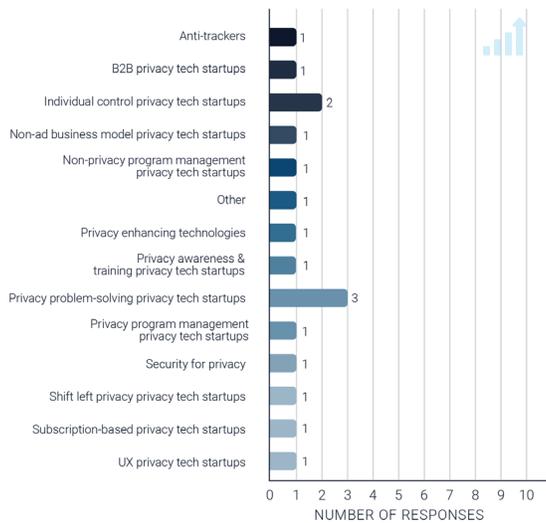


Source: TROPT Defining the Privacy Tech Landscape 2021



Most experts viewed privacy tech startups that focus on solving for privacy problems in general and, specifically, individual control problems, as the most promising ones. This is noteworthy given our definition of privacy tech as *technological solutions to privacy problems*. Other experts picked startups working on anti-trackers, PETs, security for privacy, and, in line with the trends we’re seeing, “shift left privacy” as most promising.

Figure 21. Privacy Domain Expert Survey: Privacy Tech Startups Experts Deem Promising



Source: TROPT Defining the Privacy Tech Landscape 2021

“I will not promote privacy tech vaporware or unsubstantiated privacy claims, so the first step is to conduct a technical analysis so we’re all on the same page about what the technology does and doesn’t do. That gives us a single source of truth in terms of what is real from a technical perspective, which allows our clients to maintain credibility in their communications.”

Melanie Ensign, CEO, Discernible Inc.

When asked about their biggest challenges in working in the privacy tech space, they named ineffective solutions, low privacy tech understanding, mistrust in privacy tech vendors, rapid pace of innovation, immaturity of solutions, and unpaid advisory roles as some of their prominent pain points.



WHAT INVESTORS DIDN'T OUTRIGHT SAY

We surveyed investors, but we didn't receive enough responses to report any significant insights from them. Amongst the key players in privacy tech—founders, investors, expert-advisors, and buyers—investors are seemingly the least interested in digging into and understanding privacy. Most of the investors we've engaged with through TROPT view privacy tech as an insignificant niche falling under cybersecurity, which, as we've discussed earlier in this whitepaper, is wholly inaccurate. We predict that as more and more privacy tech unicorns emerge and as the current ones prepare for IPO, investors will necessarily follow. We salute the early visionaries who were first to funding privacy tech, at its nascency. We encourage emerging privacy tech investors to jump in and help fuel privacy tech by funding the next round of privacy tech unicorns.



FUELING THE FUTURE OF PRIVACY TECH

As we move from defining and categorizing the privacy tech landscape, we take the next steps in bridging the gaps between privacy tech key players and fueling privacy innovation. Each privacy tech key player has a part to play in this endeavor. We lay out our recommendations for how each key player can help fuel the future of privacy innovation.

WHAT PRIVACY TECH FOUNDERS CAN DO

We need more privacy tech founders to build solutions to today's privacy problems. We outline some specific recommendations for privacy tech founders as follows:

More privacy tech founders should build privacy tech products that address privacy problems earlier in the data and development lifecycles.

As the data lifecycle side of the privacy tech stack gets crowded and as existing solutions mature, plenty of remaining privacy problems—or privacy tech opportunities—remain in the development lifecycle side. In line with the “shift left privacy” trend, development lifecycle privacy tech needs to catch up with data lifecycle privacy tech.

More privacy tech founders should solve for consumer privacy problems.

Consumer (B2C) privacy tech sorely needs to catch up with business (B2B) privacy tech. Privacy is inherently about people, after all. Moreover, there is a clear marketplace demand for privacy, not just in the B2B space, but also in the B2C space.⁶⁵

Privacy tech founders should start with understanding privacy problems before building technical solutions for them.

Founders should begin with understanding the privacy problems that privacy tech buyers and users have, and then build privacy tech solutions that address those problems. Further, they should understand that privacy problems go beyond compliance, security, or other limited takes on privacy. By understanding privacy, the breadth of privacy problems that exist, and the specific privacy problems that privacy tech users-buyers have, founders can build better privacy tech products.



Privacy tech founders should walk their privacy talk in building their startups.

It's critical that privacy tech founders demonstrate their commitment to privacy, not just in their marketing speak, but also in terms of outcomes and metrics. Founders need to be thoughtful about how they build their products, choose their business model, and conduct their business operations. As previously covered, trust is central to privacy. Privacy tech founders who fail to walk their privacy talk will break customer and public trust in their startup, products, and brand.

"It's very challenging to see the difference between a company that is taking advantage of privacy, whether it's a regulation or a branding type of exercise, to gain customers versus a startup that actually believes in pushing privacy forward as a human right. I do think of those as two very different types of companies... I would actually hope that they would be doing both, but the ones that are sort of exploiting are the ones I think are challenging specifically for investors because they may be doing very well from a revenue perspective or show promise on that side, and yet they're not actually pushing the rights that someone like myself so strongly believes in... I think it's very difficult, on the investor side, for them to assess and do due diligence on a company's own privacy."

Caroline McCaffery, CEO & Founder, ClearOPS

Privacy tech founders should learn to effectively communicate privacy tech's value proposition.

Doing so means being able to help customers and prospects make the case for privacy tech and secure the budget and buy-in they need.

"I see two common challenges among privacy tech startups when learning to communicate their product's value proposition. First, they don't always understand what's valuable from the perspective of their user or customer. Often, their messaging is optimized for investors with a heavy emphasis on compliance requirements, but that's usually not what differentiates your product from a sales perspective. Things like interoperability, auditability, quick and painless implementation, and support for cross-functional stakeholders don't typically make it into the pitch deck for investors, but they're critical factors for customers considering a vendor solution. The second challenge I see often is not being able to effectively engage with the technical privacy community directly. This community right now is so small, in terms of where the influence comes from and whose opinion matters in driving adoption of new technologies, that if you are not actively involved in the community, you miss the opportunity to build relationships with folks who can not only lend your product technical credibility, but also advise your product teams on what the market is eager to buy."

Melanie Ensign, CEO, Discernible, Inc.



WHAT INVESTORS CAN DO

As much as we need more privacy tech founders to build solutions to some of today's privacy problems, we equally need investors to recognize the rise of privacy tech and invest in the emerging privacy tech industry.

Investors should understand that there is a clear marketplace demand for privacy, both in the B2B and B2C sides.⁶⁶ Failing to do so means missing out on investing in upcoming privacy tech unicorns. It also means failing to step up and address the mounting privacy technical debt created by the tech industry, particularly by technologies that were built without much regard to privacy. This mounting privacy technical debt partly fuels the marketplace need for privacy tech.

Investors should recognize that privacy tech opens up environmental, social, and governance (ESG), social responsibility investments (SRI), and impact investing opportunities. This is game-changing for venture capital. Today's limited partners (LPs) demand that their money make a positive impact on society and the world at large.

By now, investors should understand that privacy tech is its own emerging industry. Privacy tech is not a cybersecurity or other industry subcategory. We hope that this whitepaper helps illustrate the big picture of this nascent but emerging space.

Put simply: investors should invest in privacy tech. There's a clear marketplace demand for privacy tech. Investing in startups that solve privacy problems is the right thing to do. Other forward-thinking investors have already poured almost six billion US dollars (specifically, \$5.9B) in cumulative funding towards emerging privacy companies, with four billion (\$4B) in the past three years alone.⁶⁸

"Today's investors no longer care solely about returns. More and more limited partner (LP) investors — investors who invest in venture funds — want their money to make a positive impact on society and the world at large, including when it comes to privacy.

It turns out that SRI accounted for more than \$1 out of every \$4 under professional management in the United States. Investors also used ESG criteria to evaluate companies for investment, and these accounted for \$16.6 trillion in investment assets in 2020. It seems LPs care enough about their investment dollars to the tune of a \$715 billion impact investing market, and some LPs are even looking to link a fund's carried interest to quantifiable impact metrics."⁶⁷

**Lourdes M. Turrecha, TROPT
Founder and privacy tech Angel
investor**



WHAT PRIVACY DOMAIN EXPERTS CAN DO

Privacy domain experts should help build privacy tech solutions. We need more privacy domain experts to jump in as privacy tech founders, advisors, operators, and even investors. Why? Because they are intimately familiar with privacy and understand the biggest privacy problems. With privacy domain experts helping build privacy tech, we avoid with ineffective privacy tech solutions or, worse, privacy tech vaporware.

Privacy domain experts should help better articulate the biggest privacy pain points they're seeing that need privacy tech solutions. Their intimate understanding of privacy and privacy problems will immensely help guide the privacy industry.

Privacy domain experts should weigh in on both the promising and the ineffective privacy tech products they're seeing. To help highlight the best privacy tech products in each category and weed out ineffective products, privacy domain experts should weigh in on the privacy tech products that are effectively solving privacy problems. Conversely, they should voice when they see products that market themselves as privacy tech but don't actually do a good job at solving privacy problems. This saves privacy tech user-buyers from wasting their valuable privacy tech budget, internal political capital, and other resources required to implement such products.



WHAT B2B PRIVACY TECH BUYERS AND USERS CAN DO

Privacy tech buyers and users should better articulate their biggest privacy pain points. This will not only help guide privacy tech founders in building tools that solve privacy problems, but also directly help buyer-users get the tools they need to solve their privacy problems.

Privacy tech buyers and users should provide public feedback on what works and doesn't work in the privacy tech products they're testing, using, and buying. This will help ensure that privacy tech startups have the information they need to improve their products. Even if they don't act on this information, this gives competitors the insight required to build better, competing products. In the end, this direct feedback helps fuel the privacy tech landscape.

Privacy tech buyers should learn how to better articulate the value of solving some of their privacy problems through privacy tech tools. This will help get cross-functional teams on board with privacy tech goals and help secure the privacy tech budget they need.

Privacy tech buyers and users should work cross-functionally in selecting and implementing privacy tech products. This will help ensure that their investment in their privacy tech purchase is not wasted and avoid internal strife.

"Even within companies, you see privacy engineers being disconnected from other engineers. You see legal and privacy engineering being disconnected and then public policy is further separate. And the comms team is sitting again, further out. You have all of these different folks trying to do the same thing, the right thing, which is find that balance between customers, between trust and compliance, but they don't talk to each other. So sometimes when it comes time to make the decision, "Do I or do I not onboard this tool?" Do I or do I not collect this data?" Sometimes those decisions are left to an engineer who may not have the right context."

Nishant Bhajaria, Head of Privacy Engineering and Architecture, Uber



TROPT'S PART

At TROPT, we will endeavor to continue our mission of fueling privacy tech and innovation. We specifically make the following commitments.

We commit to continuing to define and categorize the privacy tech landscape as it matures. Next year, we will begin categorizing startups into the TROPT Privacy Tech Stack.

We commit to continuing to bring together the privacy tech key players—founders, investors, experts, and user-buyers—to bridge the tech-capital-expertise gaps in privacy innovation. As covered in the previous section, startups disclosed funding, customer acquisition, and breaking into the privacy space as some of their biggest challenges. To address these challenges, we will continue connecting privacy tech startups with privacy domain experts and investors through our startup-advisor matching program and our startup-investor matching program.

We commit to continuing to spread awareness on privacy tech and innovation. We already do this through our events and programs. Current TROPT events include our conferences, TROPT Data Privacy Day (in January) and TROPT Virtual Summit (in June). Current programs include TROPT Innovators, our membership program for privacy tech key players who want year-round involvement with the TROPT privacy tech community in between events. Beyond our existing initiatives, we have several projects in the works, and we are open to feedback from the privacy tech community on how to best help privacy tech key players fuel privacy innovation.

We commit to helping increase privacy tech investments. We will do so by increasing investor awareness in and understanding of this nascent privacy tech industry. We will also continue to make available the TROPT privacy tech deal flow tool⁶⁹ for investors who want to invest in privacy tech startups and support the TROPT privacy tech community.

“The TROPT Defining the Privacy Tech Landscape Whitepaper is a necessary foundational step towards understanding, defining, and categorizing the nascent privacy tech industry. Now that we’ve set a baseline understanding of what constitutes privacy tech, we are better equipped to identify privacy tech products and act on the resulting marketplace opportunities. And because we believe in privacy’s value, we see the numerous opportunities for privacy tech key players: privacy tech product ideas for founders, partnerships for privacy tech startups, investments for investors, privacy tech solutions for buyer-users, and advisory roles for domain experts, just to name a few. TROPT was founded to create a place for the industry key players to come together, realize privacy tech’s value, and move the needle on privacy. With law typically lagging behind technology, we don’t have to keep waiting for policymakers and regulators to give us the privacy solutions we’ve been asking for. As privacy innovators, we can build technological solutions to some of our privacy problems, and help fuel the privacy tech industry.”

Lourdes M. Turrecha, TROPT Founder and privacy tech Angel investor



CONCLUSION

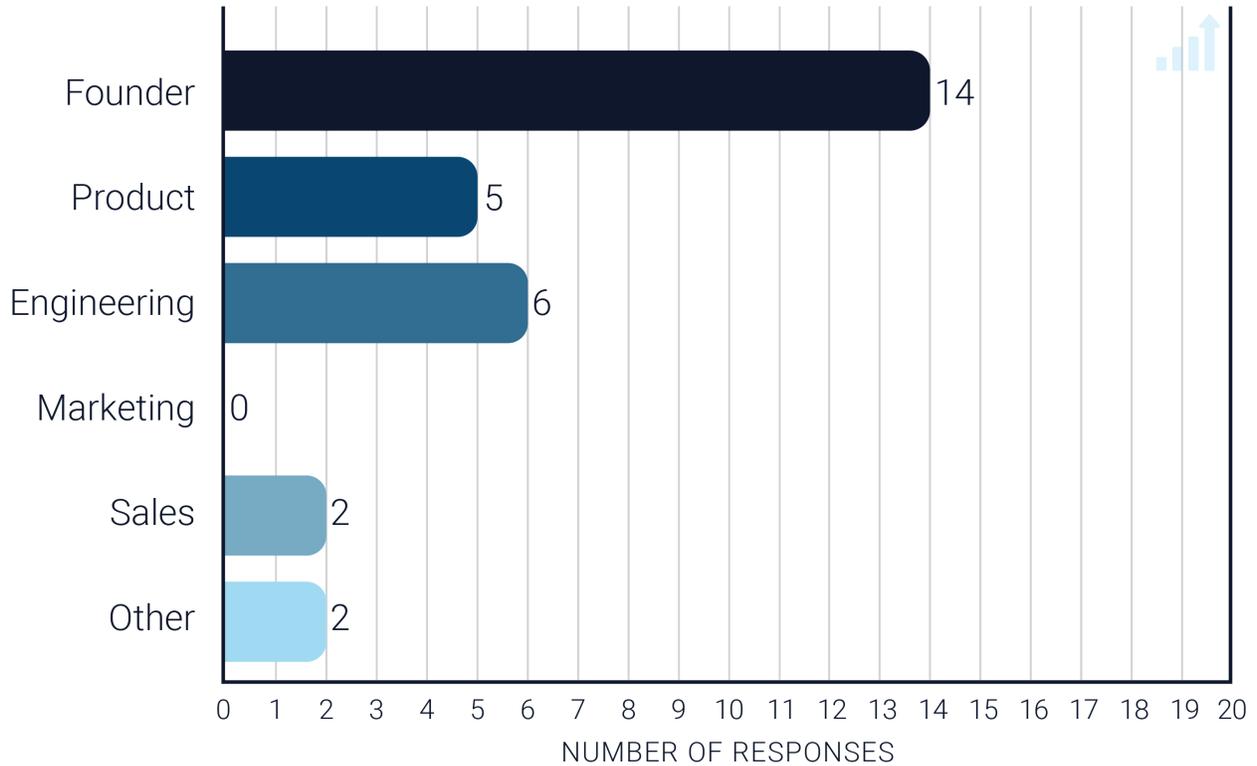
Our primary aim for this whitepaper is to define and categorize privacy tech. In the process of doing so, we ended up creating the TROPT Privacy Tech Stack to visually articulate the breadth of privacy tech, beyond defining its scope. We also sought to demonstrate privacy tech's value by seeking qualitative feedback from the different privacy tech key players. Future versions of this whitepaper will explore how the different key players' pain points, challenges, and focus change over time. TROPT will continue to work with TROPT Innovators members and the greater TROPT community of privacy tech key players to define, categorize, and fuel the privacy tech landscape. For additional information about TROPT's privacy tech industry research, please contact community@riseofprivacytech.com.



APPENDIX: SURVEY RESULTS

What is your role at your privacy tech startup?

Privacy Tech Startup Survey
Answered: 20



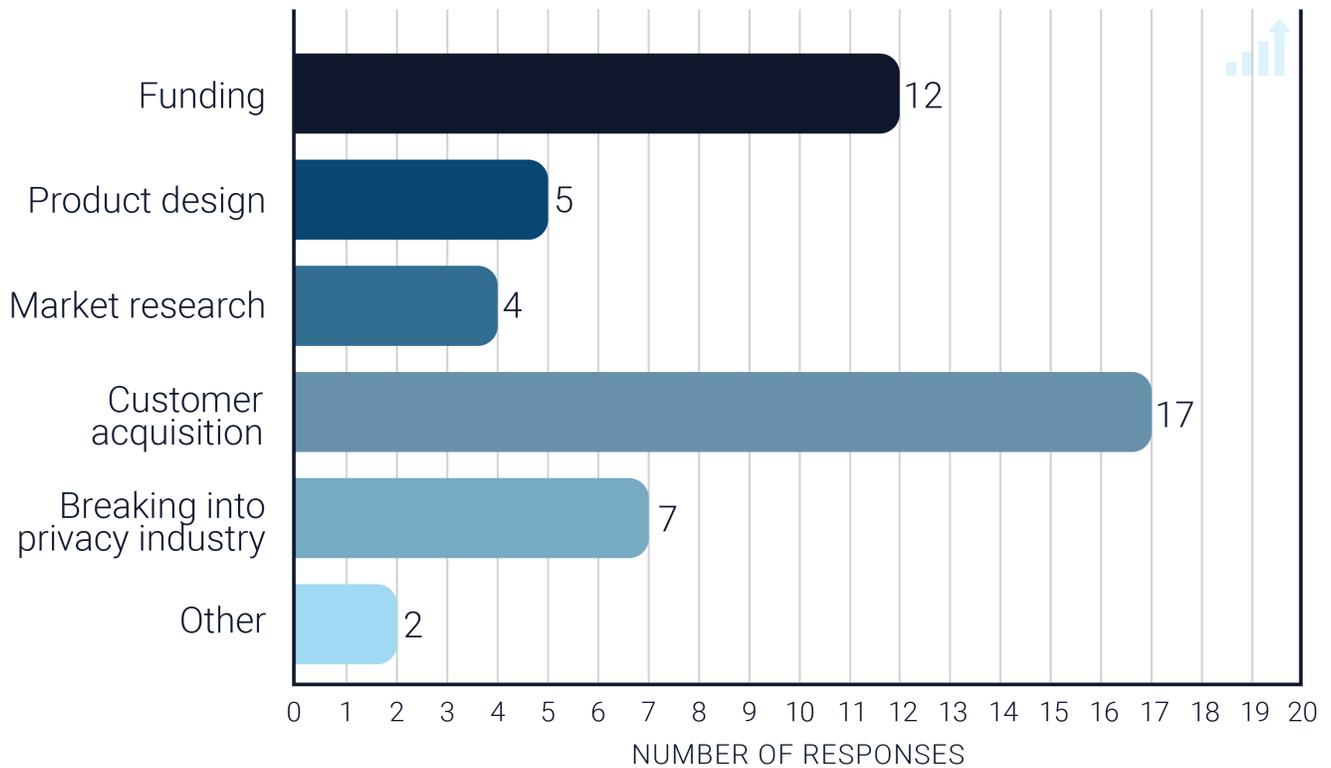
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



Which of the following are pain points for you? (You can pick more than one)

Privacy Tech Startup Survey

Answered: 20



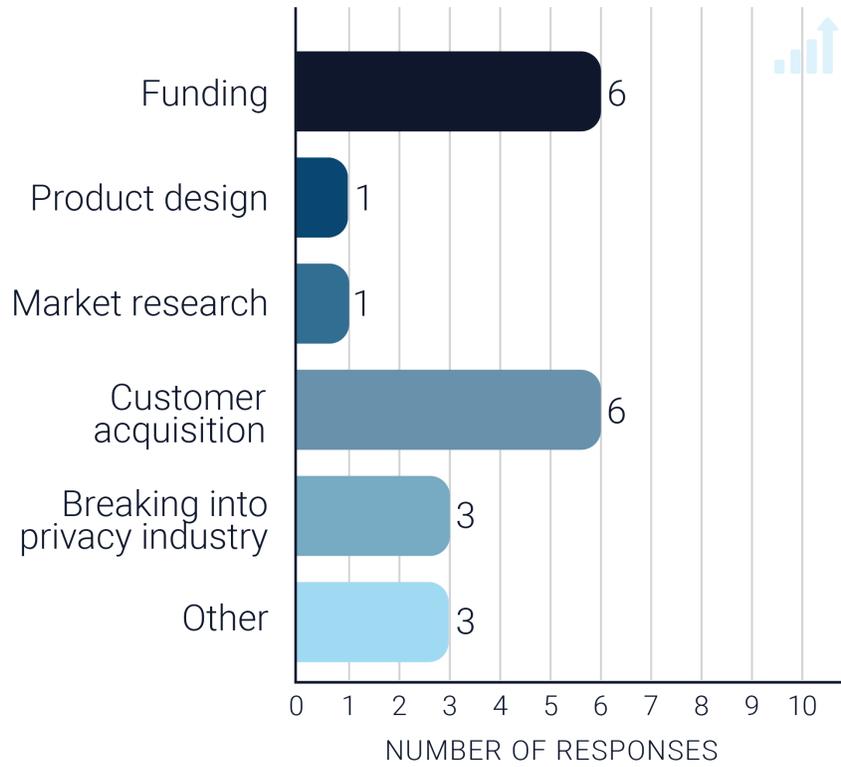
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What is your single biggest privacy tech pain point? (Pick one)

Privacy Tech Startup Survey

Answered: 20



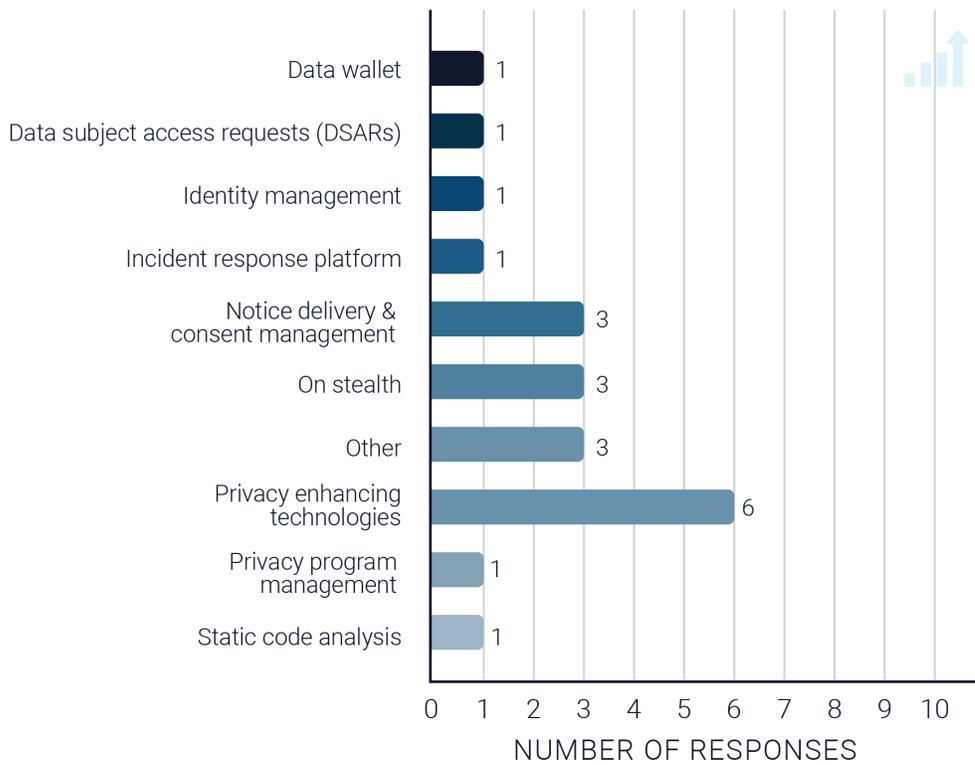
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What privacy tech product are you currently working on?

Privacy Tech Startup Survey

Answered: 20



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Privacy Policy Evaluation Item	Notice Delivery and Consent Management
Legal compliance tech	Privacy Program Management
Zero-Party Data collection, unification, and activation to power Privacy-First Personalization.	PETs
Private data layer for decentralized applications	PETs
Confidential	On stealth
A Platform for Private Cloud Apps	Other
Consumer subscription service to help people take back control of their personal data from companies that shouldn't have it	DSARs
On stealth	On stealth
Data wallet	Data wallet
Privacy Incident Response Platform	Incident Response Platform
	Other
The one	Other
Synthetic data	PETs
Computation over encrypted data - we have two products: [redacted] and [redacted]	PETs
Homomorphic encryption	PETs
Static Code Analysis for Privacy	Static Code Analysis
On-device contextual targeting for advertisers that does not rely on identifiers	PETs
Decentralized Identity	Identity management
T&Cs for SMBs, and Consent Management for SMBs	Notice Delivery and Consent Management
Consent management, privacy policy automation	Notice Delivery and Consent Management

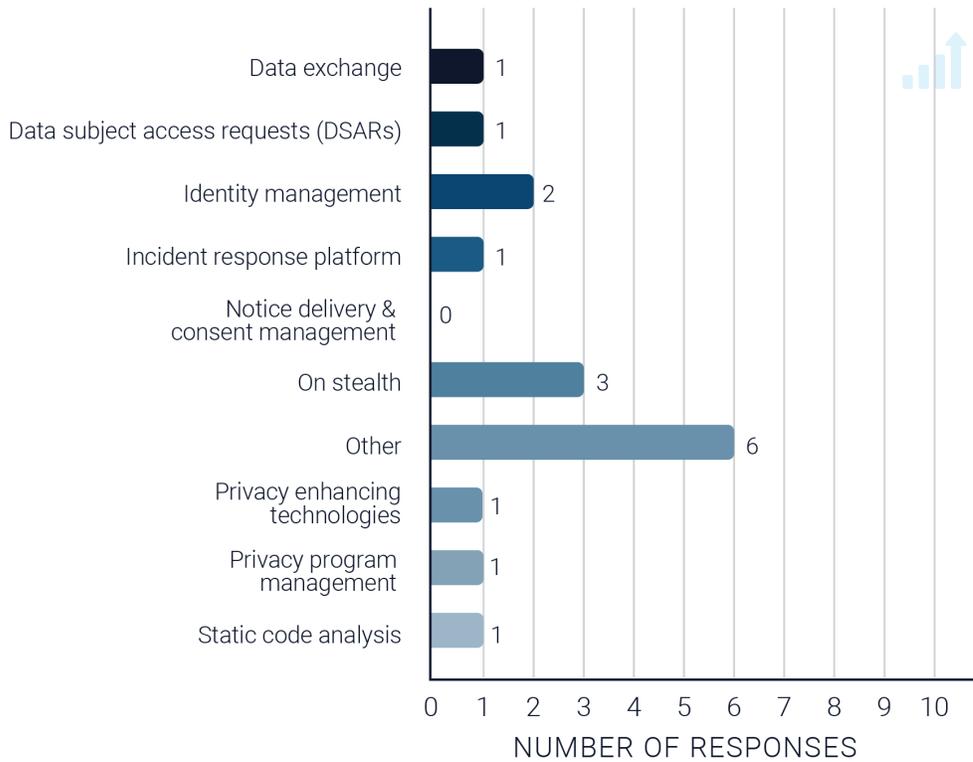
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What privacy tech product are you planning to build in the future, if any?

Privacy Tech Startup Survey

Answered: 20



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Automatic Privacy Policy Evaluation & Consent Management	Notice Delivery and Consent Management
Full privacy management software	Privacy Program Management
Product is in market now, & we are actively scaling go-to-market.	Other
Self sovereign identity	Identity Management
Confidential	On stealth
A Platform for Private Cloud Apps	Other
Consumer subscription service to help people take back control of their personal data from companies that shouldn't have it	DSARs
On stealth	On stealth
Decentralized data marketplace	Data Exchange
Continually improving our Privacy Incident Response Platform	Incident Response Platform
N/A	Other
A social net	Other
N/A	Other
We will develop out the suite around our two launched products.	Static Code Analysis
N/A	Other
Extending the use cases around static code analysis for privacy.	Static Code Analysis
Can't disclose at the moment	On stealth
Decentralized Identity	Identity management
T&Cs for SMBs, and Consent Management for SMBs	Notice Delivery and Consent Management
Consent management, privacy policy automation	Notice Delivery and Consent Management

Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What privacy problem are you currently solving for?

Privacy Tech Startup Survey

Answered: 20



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Obscure Privacy Policy	Transparency
How to plan a full management scheme	Accountability
Enabling brands to deliver customer experiences that are always welcome, relevant and compliant. Empowering consumers with transparency, choice and control over the data they share with a given brand in order to improve their experience with that brand.	Transparency, Individual Control, Accountability
Consumer control of data (automate the 8 rights of individuals) & GDPR compliance automation for organizations	Individual Control
Personal Privacy	Individual Control
Consumers need options to own & control their data, not just software for-rent behind an EULA	Individual Control
Consumer lack of control over their personal data	Individual Control
Consumer privacy	Consumer Privacy, Individual Control
Individual Personal data monetization	Individual Control
Incident Response	Security for Privacy
Ugandans who can't have access to cheap quick loans	Other
Dev	Shift Left Privacy
The need for safe and useful data for testing and development	Shift Left Privacy
Browser based encryption, privacy and security issues around cookie deprecation	Security for Privacy
Analytics privacy	Data minimization
Privacy by Design. We help companies ship code fast without breaking user privacy. We do this by scanning the source code continuously and alerting privacy teams when a code change has privacy issues. We detect personal data, third party and hidden data flows.	Shift Left Privacy
Eliminating the reliance of digital identifiers to deliver personalized ads (and content)	Data minimization
Allowing patients to control their identity attributes and by extension provide granular consent of their health data	Individual Control
T&Cs for SMBs, and Consent Management for SMBs	Accountability
Automating consent	Transparency, Individual Control

Source: The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021



What privacy problem are you interested in solving with these future products?

Privacy Tech Startup Survey

Answered: 20



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Explicit User Consent	Transparency
Easy step-by-step guidance in legal compliance	Accountability
Enabling brands to deliver customer experiences that are always welcome, relevant and compliant. Empowering consumers with transparency, choice and control over the data they share with a given brand in order to improve their experience with that brand.	Transparency, Individual Control, Accountability
Minimizing data exposure; zero-knowledge identity verification	Data Minimization
User Experience	Individual Control
Most privacy problems are amenable to giving users private ownership & control of their data, and making it easy for non-experts to run decentralized software.	Individual Control
Consumer lack of control over their personal data	Individual Control
Consumer privacy	Consumer Privacy
Lack of equity for consumers	Individual Control
Incident Response	Security for Privacy
Marketing	Transparency, Individual Control, Data Minimization
Social net	Other
N/A	Other
Better data privacy in the advertising industry.	Other
Big data privacy	Data minimization
Automatically update RoPA and PIA/DPIAs with code scans.	Shift Left Privacy, Accountability
Privacy first mobile advertising	Transparency, Individual Control, Data Minimization
Lack of privacy and consent	Transparency, Individual Control
T&Cs for SMBs, and Consent Management for SMBs	Accountability

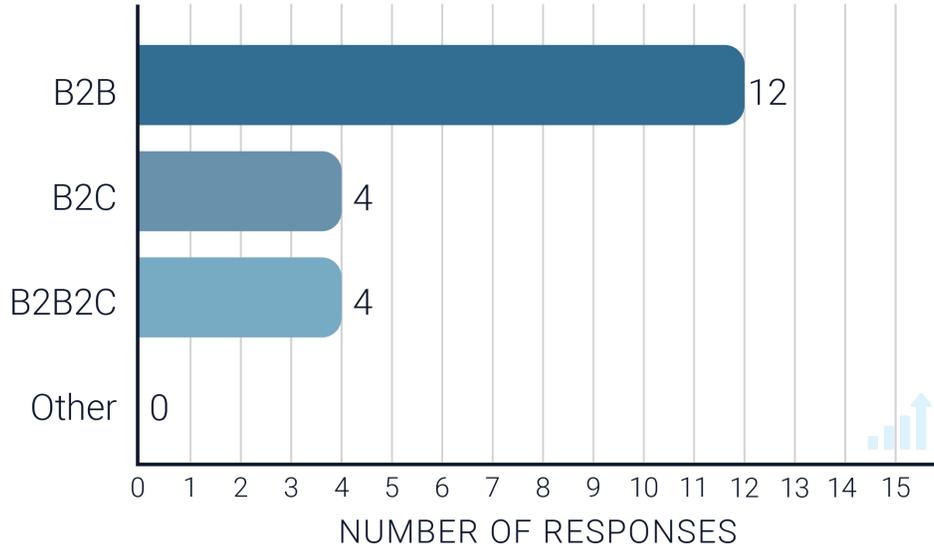
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What is your business model?

Privacy Tech Startup Survey

Answered: 20



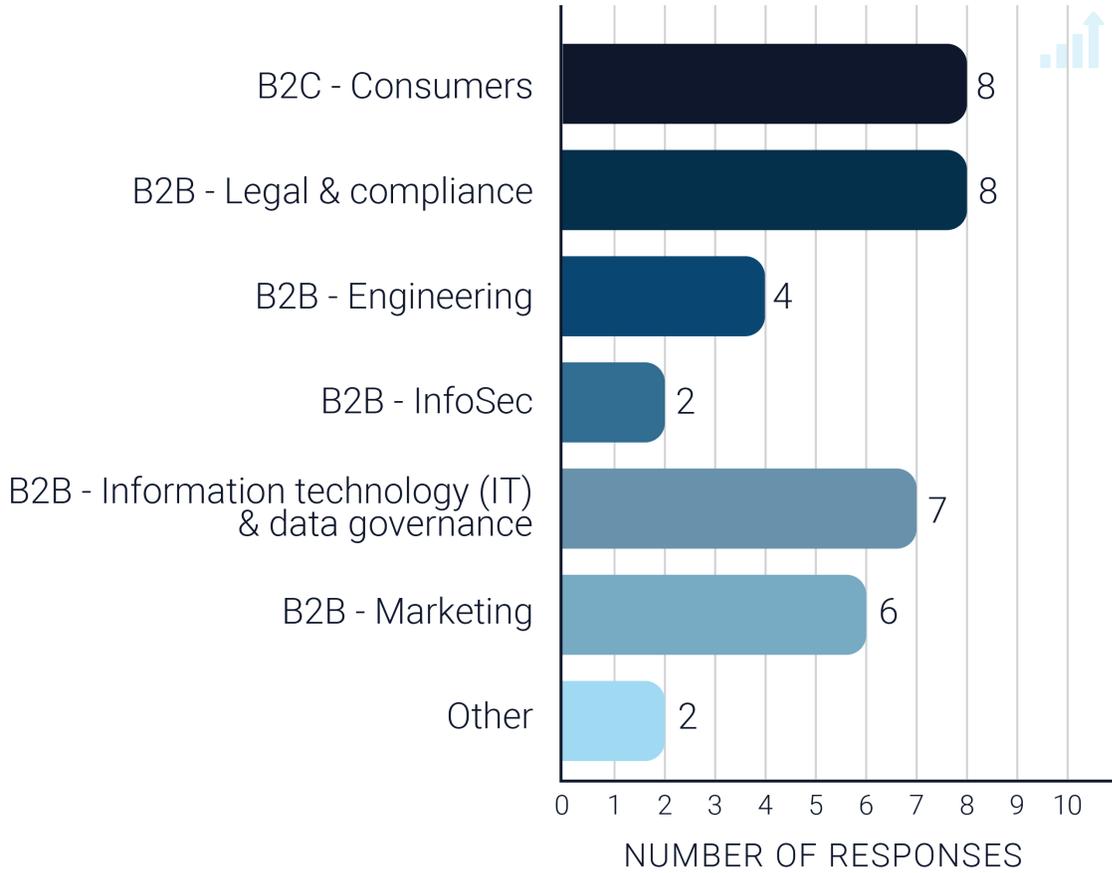
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



Who is your privacy tech startup building for?

Privacy Tech Startup Survey

Answered: 20



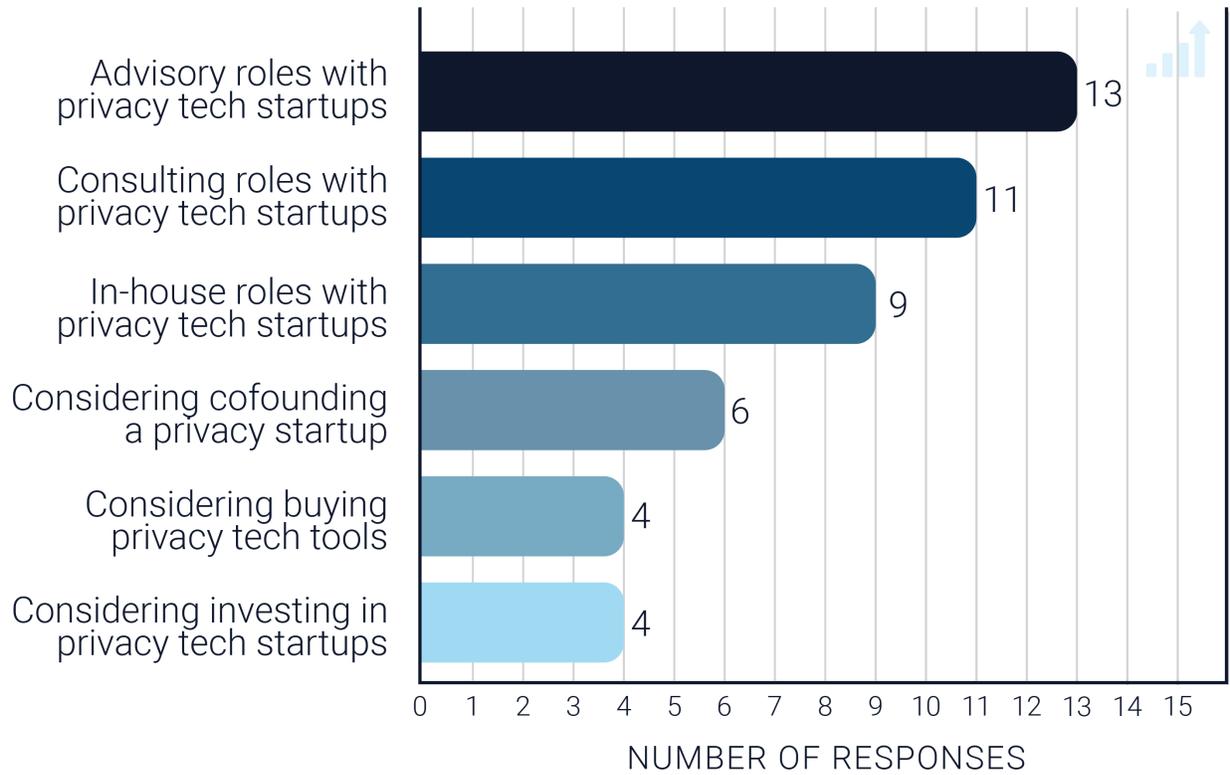
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What interests you about the emerging privacy tech landscape?

Domain Experts Survey

Answered: 15



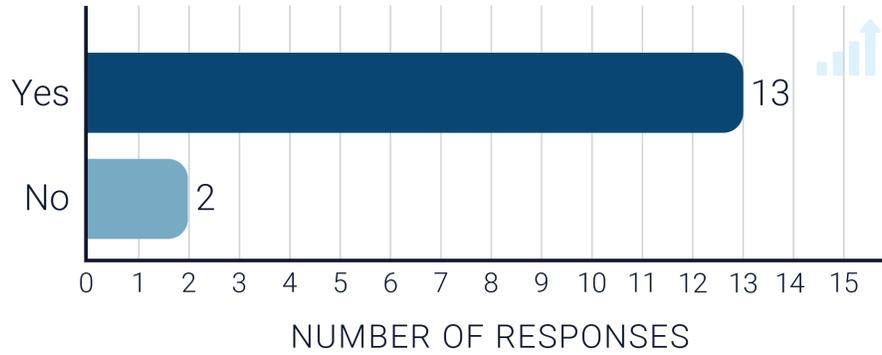
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



Are you currently working in the privacy tech landscape?

Domain Experts Survey

Answered: 15



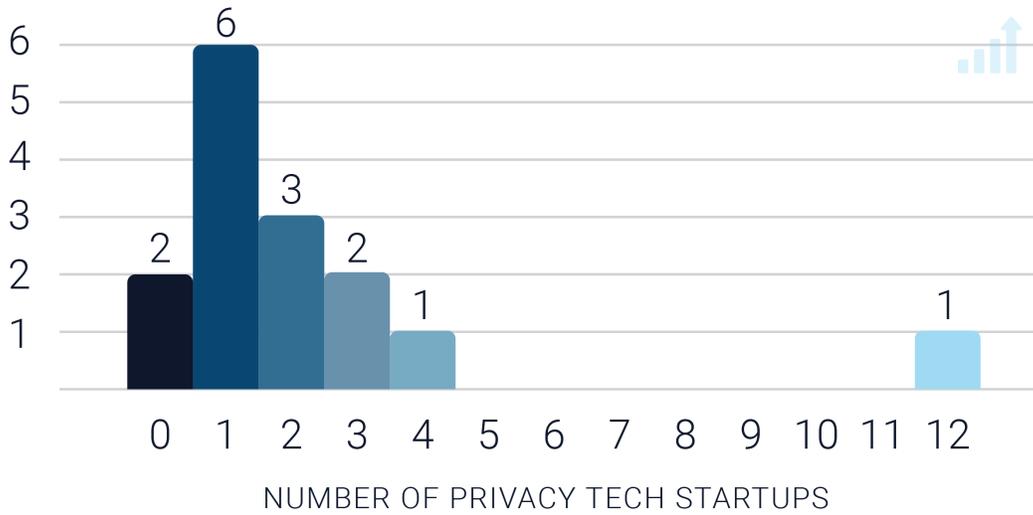
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



How many privacy tech startups are you working with?

Domain Experts Survey

Answered: 15



Average number of privacy tech startups: 2.27

Median number of privacy tech startups: 1.00

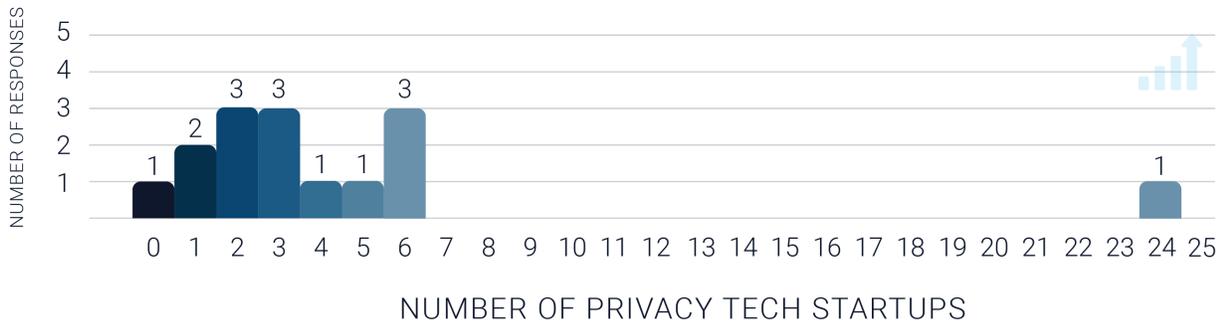
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



How many privacy tech startups do you anticipate working with in the next 12 months?

Domain Experts Survey

Answered: 15



Average number of anticipated privacy tech startups: 4.53

Median number of anticipated privacy tech startups: 3.00

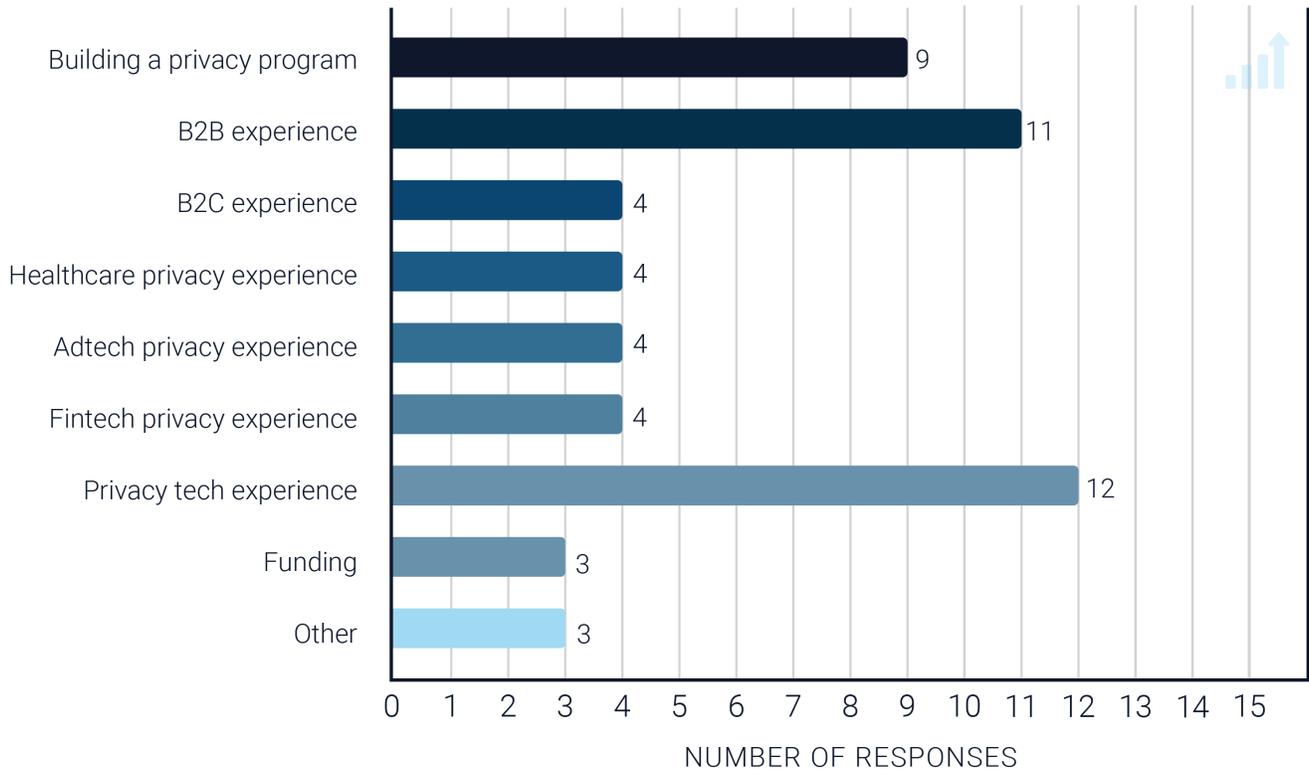
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



Which areas of expertise do you offer privacy tech startups?

Domain Experts Survey

Answered: 15



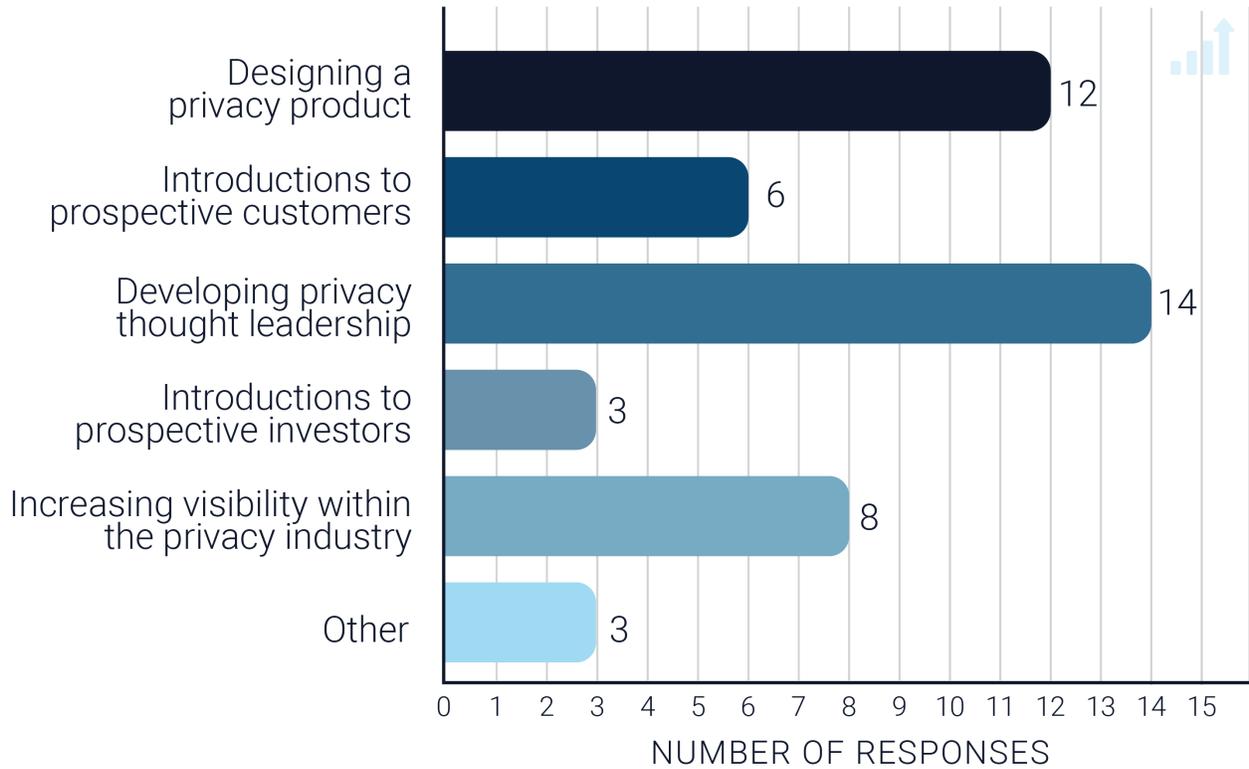
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



Which challenges can you help privacy tech startups with?

Domain Experts Survey

Answered: 15



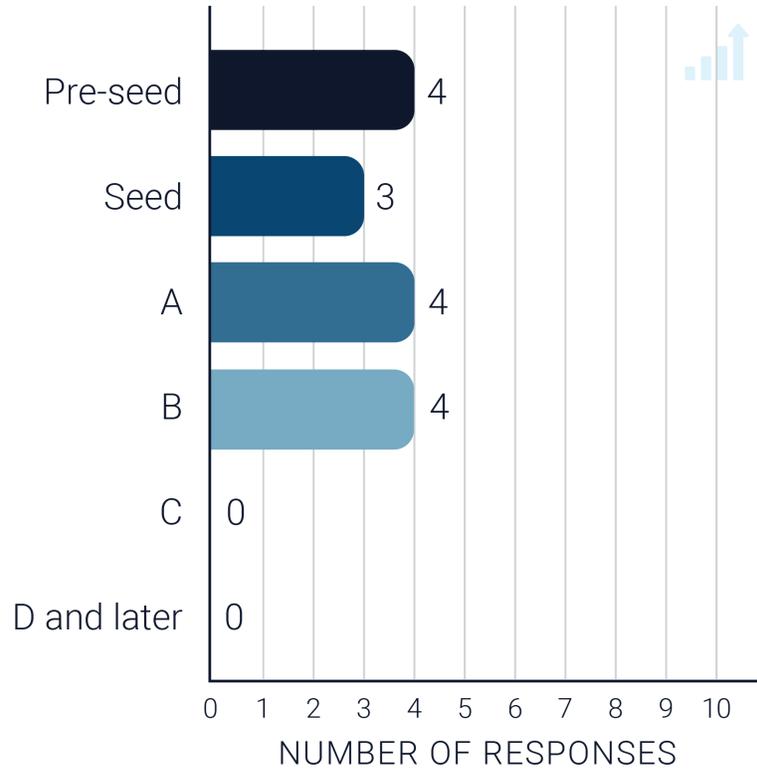
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What startup stage are you interested in?

Domain Experts Survey

Answered: 15



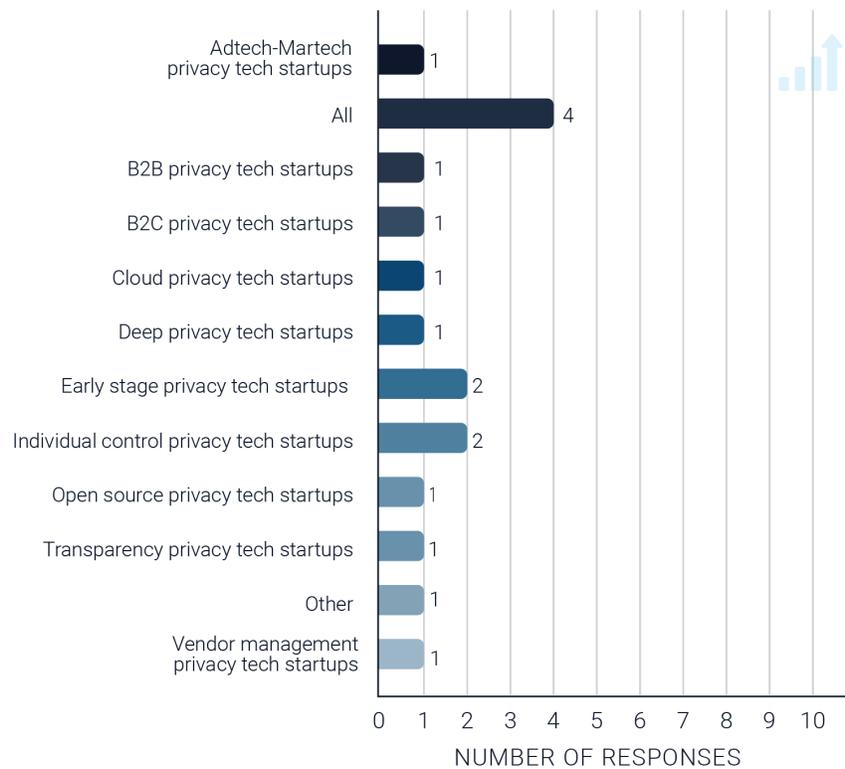
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What types of privacy tech startups are you looking to work with?

Domain Experts Survey

Answered: 15



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Pre-Series A, B2C	B2C Privacy TechStartups, Early Stage Privacy TechStartups
All	All
Data ownership, Synthetic data, Digital identity, Blockchain & privacy, other emerging privacy tech	Individual Control Privacy Tech Startups
Ground-breaking innovative martech/adtech	Adtech-Martech Privacy Tech Startups
Those that need help understanding the privacy implications of their products	All
B2B privacy tech startups that are solving current business problems vs. theoretical ones.	B2B Privacy TechStartups
Those addressing consent and data control challenges	Transparency Privacy Tech Startups, Individual Control Privacy Tech Startups
Those operating in the cloud with a SAAS service	Cloud Privacy Tech Startups
Engagement and education	Other
Deep privacy tech	Deep Privacy Tech Startups
Hardware, open-source	Open Source Privacy Tech Startups
Open source project, NGO, probably role in round A startup	Early Stage Privacy Tech Startups
Really all kinds, especially given the vast number that will come under privacy law purview	All
Vendor management tools.	Vendor Management Privacy Tech Startups
Any!	All

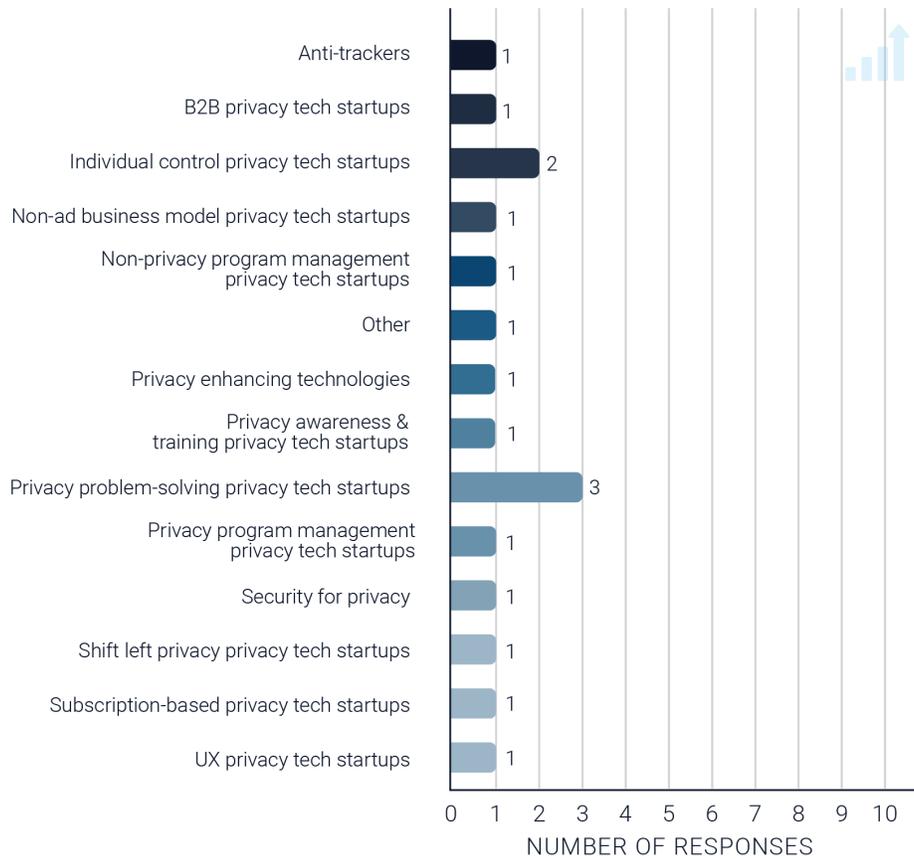
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



Which types of privacy tech startups do you think are most promising?

Domain Experts Survey

Answered: 15



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Anti-tracking and encryption	Security for Privacy, Anti-Trackers
Those solving immediate needs of customers	Privacy Problem-Solving Privacy Tech Startups
Data ownership	Individual Control Privacy Tech Startups
Rethinking the role of adtech/martech and giving back the control to the user	Individual Control Privacy Tech Startups
Anyone doing something other than compliance	Non-Privacy Program Management Privacy Tech Startups
Those that are solving real pain for businesses	Privacy Problem-Solving Privacy Tech Startups
Those with strong UX chops, even if not consumer facing	UX Privacy Tech Startups
Those that compete with OneTrust and help organize, coordinate, and automate compliance needs.	Privacy Program Management Privacy Tech Startups
Engagement and education	Privacy Awareness & Training Privacy Tech Startups
DevPrivOps	Shift Left Privacy Privacy Tech Startups
Subscription-based models not relying upon ad revenue	Subscription-based Privacy TechStartups, Non-Ad Business Model Privacy Tech Startups
SaaS with MPC, HE, FL and synthetic data solutions	Other
Ones that combine data monetization with purpose-driven use they will endure tough challenges	Privacy Problem-Solving Privacy Tech Startups
Privacy Techstartups that solve privacy problems of users, at scale	Vendor Management Privacy Tech Startups
B2B	B2B Privacy Tech Startups

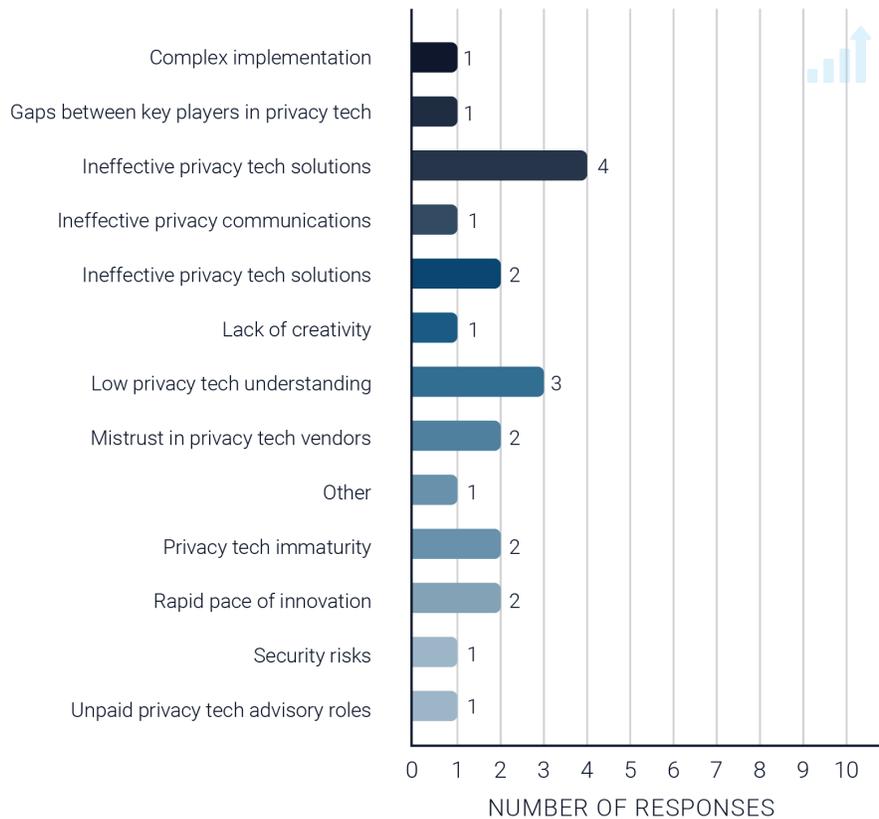
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What are your biggest challenges in working in the privacy tech space?

Domain Experts Survey

Answered: 15



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Rapidly changing tech landscape and lack of understanding by the global community (low tech individuals/consumers)	Rapid Pace of Innovation, Low Privacy Tech Understanding
Thinking out of the box	Lack of Creativity
The biggest challenge in privacy tech is keeping up with the pace of innovation -while- finding the best implementations of those ideas and identifying the real breakout ideas in that mess of creativity.	Rapid Pace of Innovation, Privacy Tech Immaturity
Confusion! It's the intersection of security, engineering, marketing, legal... Lots of misleading/incomplete info, very complex ecosystem, etc.	Low Privacy Tech Understanding, Ineffective Privacy Tech solutions
Misunderstanding by new entrants	Low Privacy Tech Understanding
██████████ sucks, we need better technology options	Ineffective Privacy Tech solutions
A few advocates in the community with a lot of passion forget that it's not a substitute for effective messaging	Ineffective Privacy Communications
Doing my current advisory roles pro bono	Unpaid Privacy Tech Advisory Roles
Application	Unknown
Bridging gaps	Gaps Between Key Players in Privacy Tech
Avoiding snake oil	Ineffective Privacy Tech solutions
Explainability: the complexity of the services and products and that it's hard to get a overview from the user/business perspective (legal, data scientists and business roles). That it's hard to match a specific business need to a specific pet, and to assess the maturity of the product or service. To create trust in the Privacy Enhancing Technologies products and services of the end user side (creating the right pr spin). Match security and privacy in the pet space (during development lifecycle and knowledge of Privacy Enhancing Technologies among infosec people)	Ineffective Privacy Tech solutions, Mistrust in Privacy Tech vendors
Too many players who are just faking it	Ineffective Privacy Tech solutions, Mistrust in Privacy Tech vendors
Implementation is a huge issue. It shouldn't take 6-8 months. Internal resources and political capital to deploy. Ease/risk of implementation (will it break or negatively impact other systems). Risks include security, availability of systems.	Ineffective Privacy Tech solutions, Complex Implementation, Security Risks
A one-stop vendor that has many viable features to help in the b2b space	Privacy Tech Immaturity

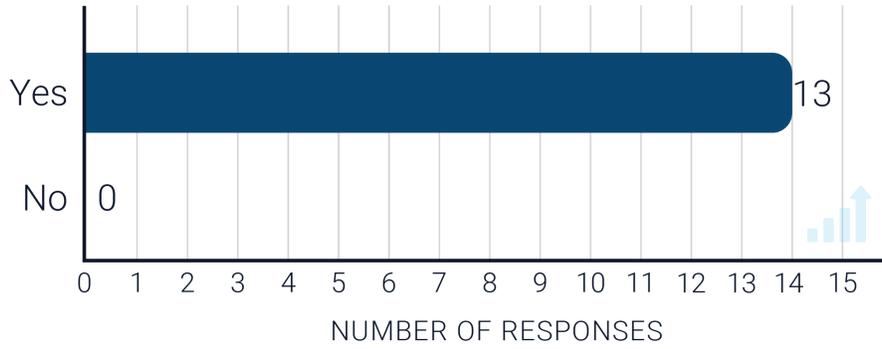
Source: The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021



Are you interested in buying privacy tech?

User or Buyer Survey

Answered: 14

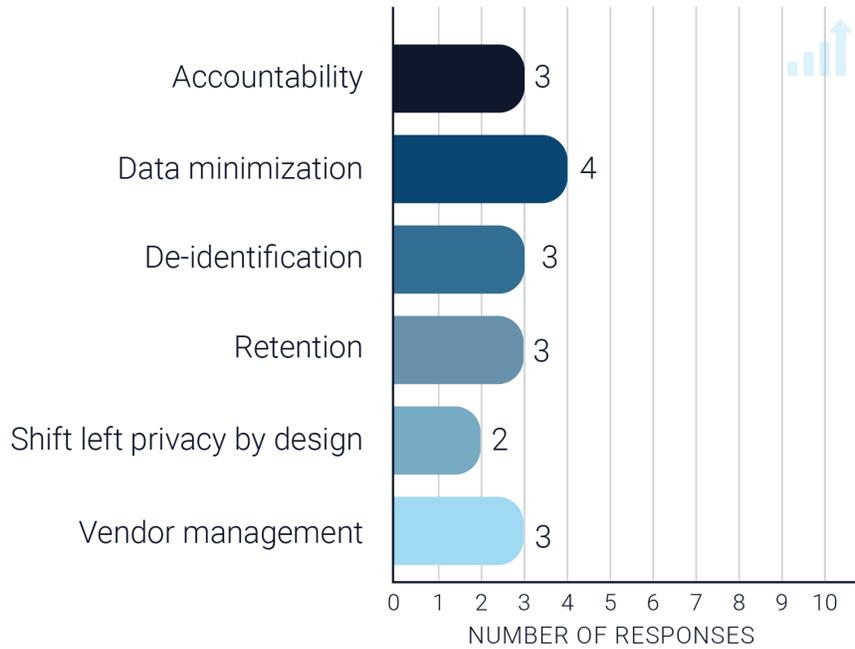


Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What types of privacy problems do you have that you wish there was a tool for?

User or Buyer Survey
 Answered: 14



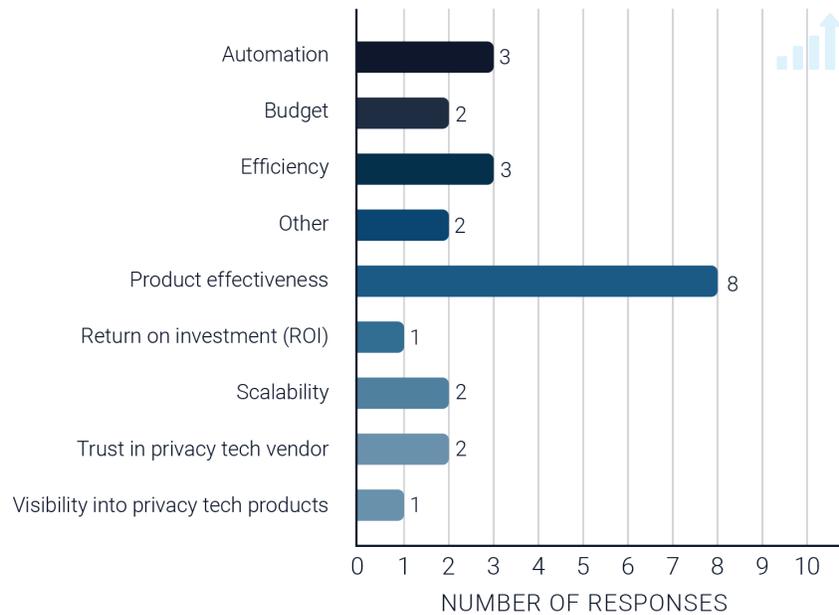
VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Data retention	Retention
Data identification, vendor management	Vendor Management, De-Identification
Vendor assessment	Vendor Management
Algorithm/code governance tool	Shift Left Privacy by Design
Differential privacy tools	Data minimization, De-identification
Data deletion and retention policy enforcement	Retention
Third party privacy questionnaire/reviews	Vendor Management
De-identification, eyes-off model training	De-identification
Data minimization	Data minimization
Internal privacy rules repository	Accountability
Data deletion & retention	Retention
Code privacy governance	Shift Left Privacy by Design
1. Realtime data maps 2. Reliable data inventory 3. Privacy preserving data analytics	Accountability, Data Minimization
1) Cradle-to-the-grave data lifecycle data management tool 2) Data processing analytics tool, covering data analytics 3) Privacy audit tools. 4) Age verification. 5) Privacy trust marks.	Accountability, Data Minimization

Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



Why are (or aren't) you interested in purchasing privacy tech right now?

User or Buyer Survey
 Answered: 14



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Effective features	Product Effectiveness
Automation, scalability, documentation	Automation, Scalability,
Interest: manage vendors & other program pain points	Product Effectiveness
Maturity, ROI	Product Effectiveness, ROI
Availability and effectiveness of products	Product Effectiveness
Too many shady tools	Product Effectiveness
Not enough information about what's available	Visibility into Privacy Tech Products
Many organizations struggle with getting started with data science due to privacy concerns. Even before applying DP, FL it anything else, they would need to start looking at data but are blocked on PII/PHI limitations)	Other
Interest: availability of resources, budget. Disinterest: trust in vendor, efficacy of solutions,	Budget, Product Effectiveness, Trust in Privacy Tech Vendor
Interested: to free up headcount, manage privacy program, and become more efficient	Automation, Efficiency
Very interested in purchasing technical solutions to complex privacy tech problems	Product Effectiveness
Interested in deep privacy tech	Other
Building a privacy program from the ground up at a startup, looking for tech to jumpstart program development and increase program maturity year over year. Initial focus on compliance solutions. Budget and strategy yet to be defined.	Automation, Efficiency, Budget
Interest: efficiency and scalability, plug-and-play; disinterest: having to build products for the vendor, promising PbD without following through.	Efficiency, Scalability, Product Effectiveness, Trust in Privacy Tech Vendor

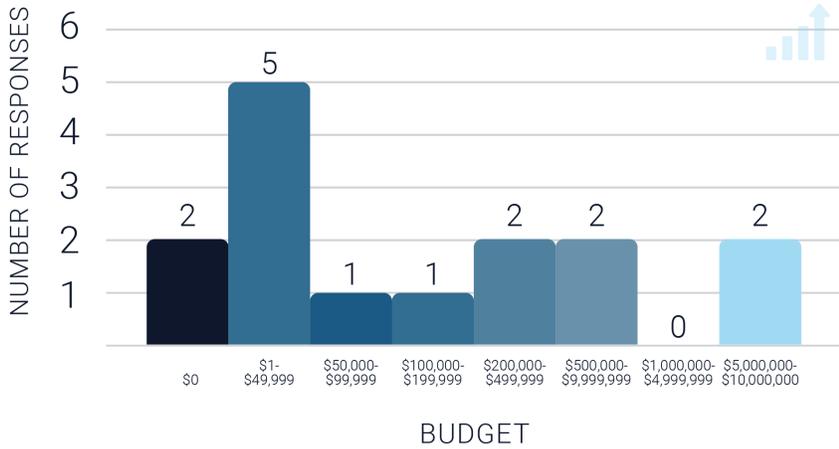
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What's your team's privacy tech tool budget?

User or Buyer Survey

Answered: 14



VERBATIM ANSWERS
\$0.0
\$00.0
\$10,000.0
\$15,000.0
\$25,000.0
\$30,000.0
\$35,000.0
\$55,000.0
\$100,000.0
\$200,000.0
\$205,000.0
\$500,000.0
\$3,000,000.0
\$5,000,000.0

Average team privacy tech tool budget: \$655,357.14

Median team privacy tech tool budget: \$45,000.00

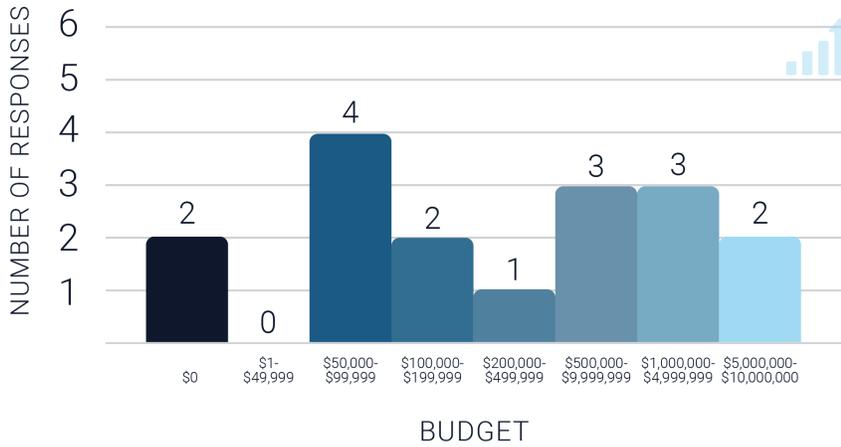
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What's your company's total privacy tech tool budget?

User or Buyer Survey

Answered: 14



VERBATIM ANSWERS
\$0.00
\$00.00
\$55,000.00
\$65,000.00
\$75,000.00
\$90,000.00
\$100,000.00
\$155,000.00
\$450,000.00
\$500,000.00
\$500,000.00
\$750,000.00
\$5,000,000.00
\$10,000,000.00

Average company total privacy tech tool budget: \$1,267,142.86

Median company total privacy tech tool budget: \$127,500.00

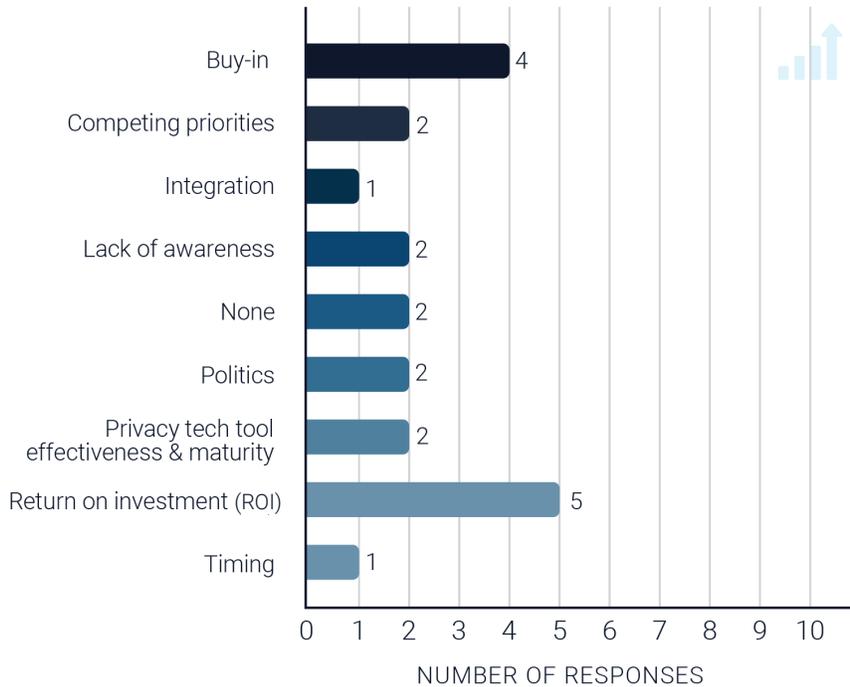
Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What are some of your challenges in securing a privacy tool budget?

User or Buyer Survey

Answered: 14



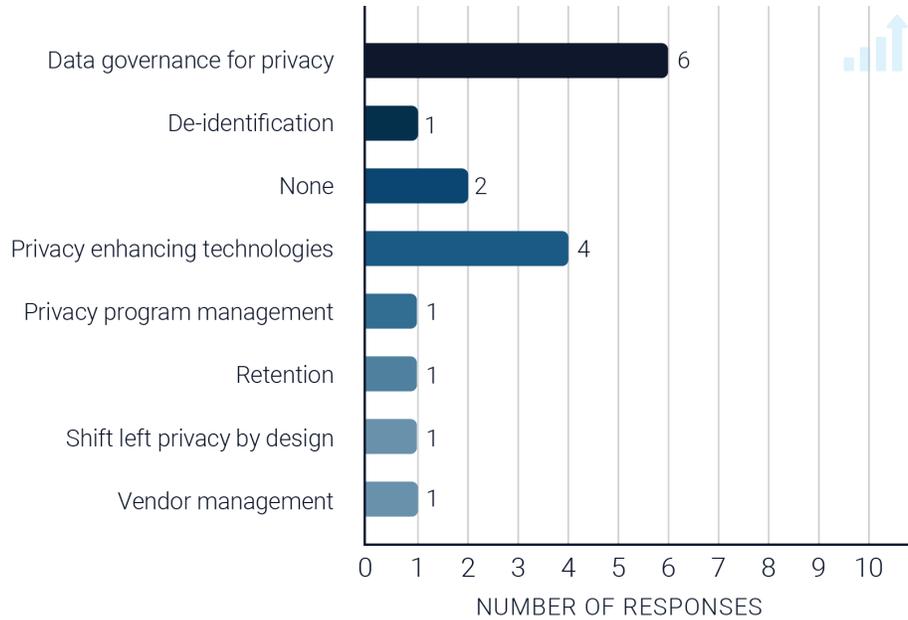
VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Timing, team and leadership support, demonstrating value	Timing, Buy-In, ROI
Awareness of risk, other projects take priority	Lack of Awareness, Competing Priorities
Leadership buy in	Buy-In
Demonstrating ROI.	ROI
None	None
Recency bias - FB got away with it	ROI
Integration	Integration
Approval from various stakeholders that the tool covers their compliance needs	Buy-In
Immature products. No products are solving hard technical problems at scale.	Privacy TechTool Effectiveness & Maturity
No challenge. We're able to get budget when we demonstrate a product's value, such as saving on additional headcount.	None
Finding tools that work. Justifying a tool's ability to solve an important business pain point.	Privacy TechTool Effectiveness & Maturity
Internal politics	Politics
1. Initial focus is on compliance, getting budget for privacy tech that falls outside of compliance(GDPR, CCPA) scope is difficult to advocate for. 2. Shared ownership, privacy tools are used by various teams like data infra, analytics, adtech teams and aligning on 1 cost center is difficult. 3. Privacy is not a core business focus area	ROI, Buy-In, Politics, Competing Priorities
Lack of understanding by stakeholders/budget approvers, demonstrating value.	Lack of Awareness, ROI

Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What privacy tech tools would you buy if you had the budget?

User or Buyer Survey
 Answered: 14



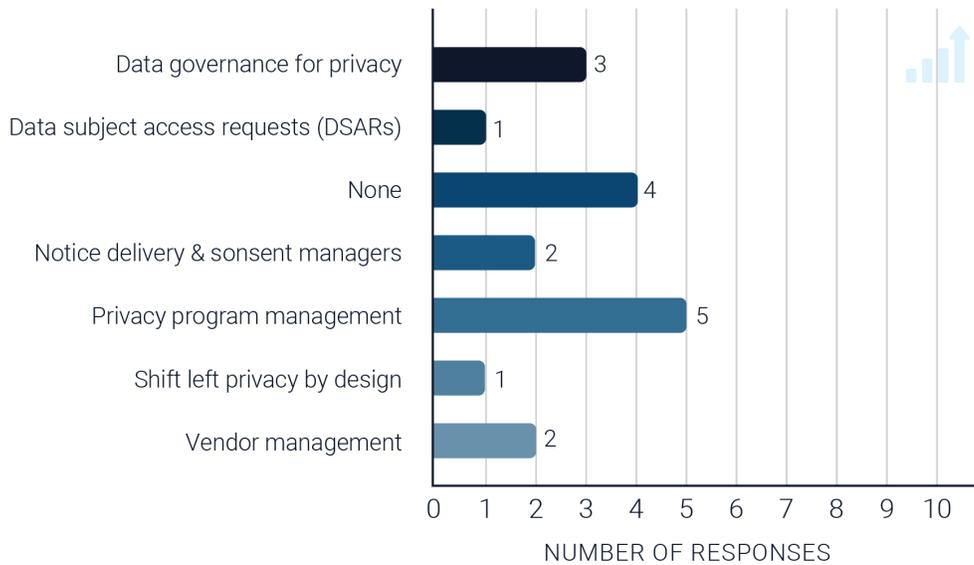
VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Data sharing governance	Data Governance for Privacy
Data identification, vendor management	De-Identification, Vendor Management
Data catalogue tool	Data Governance for Privacy
Data discovery	Data Governance for Privacy
Privacy Enhancing Technologies for data analytics	Privacy Enhancing Technologies
Hmm, not sure yet	None
Data inventory	Data Governance for Privacy
N/A	None
Differential privacy tools.	Privacy Enhancing Technologies
Article 30 data inventory tool	Data Governance for Privacy
Data deletion and retention tools	Retention
All the mature Privacy Enhancing Technologies available	Privacy Enhancing Technologies
Privacy by design tools, differential privacy tools	Privacy Enhancing Technologies, Shift Left Privacy by Design
Digital asset management privacy tech tool, to demonstrate good data stewardship and transparency, like a privacy trust mark	Data Governance for Privacy, Privacy Program Management

Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What privacy tech tool have you budgeted for?

User or Buyer Survey
 Answered: 14



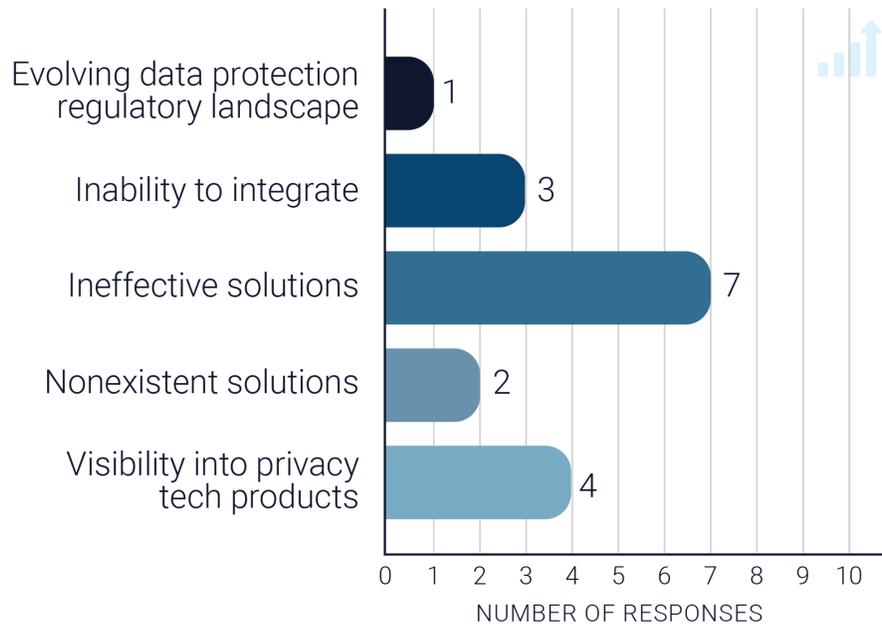
VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
DPIA	Privacy Program Management
None	None
Vendor review tool	Vendor Management
Consent management	Notice Delivery & Consent Managers
N/A	None
N/A	None
PIA	Privacy Program Management
N/A	None
Vendor management tools.	Vendor Management
Data inventory, privacy policy rules repository, PIA tools	Data Governance for Privacy, Privacy Program Management
Privacy engineering training tool	Shift Left Privacy by Design
Privacy training	Privacy Program Management
Data discovery tools (BigID), Cookie Management tools, Consent Management tools	Data Governance for Privacy, Notice Delivery & Consent Managers
PIAs, DSARs, data discovery.	Privacy Program Management, DSARs, Data Governance for Privacy

Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



What are your biggest challenges in the privacy tech space?

User or Buyer Survey
 Answered: 14



VERBATIM ANSWERS	TRANSLATION FOR GRAPHIC
Not enough available privacy tech tools with good features	Nonexistent Solutions, Ineffective Solutions
Velocity of new laws and regulations and lack of interoperability of standards	Inability to Integrate, Evolving Data Protection Regulatory Landscape
Knowing which tools do what	Visibility into Privacy Tech Products
Product and feature maturity	Ineffective Solutions
Finding promising privacy tech tools	Visibility into Privacy Tech Products
Too many simplistic solutions	Ineffective Solutions
Integrating with existing tools and systems	Inability to Integrate
It's too theoretical and doesn't necessary align with blockers and issues in industry	Ineffective Solutions
Getting market research directly from users/buyers.	Ineffective Solutions
Cutting through the marketing noise and figuring out best in breed	Visibility into Privacy Tech Products
Maturity of products	Ineffective Solutions
Not enough available privacy tech products for technical pain points	Nonexistent Solutions, Ineffective Solutions
Lack of awareness about the tools that exist outside of the big players like [redacted]	Visibility into Privacy Tech Products
Silo'd solutions. Ch integrations.	Inability to Integrate

Source: *The Rise of Privacy Tech (TROPT) Defining the Privacy Tech Landscape 2021*



ENDNOTES

- ¹CrunchBase, Privacy Companies: Summary, <https://www.crunchbase.com/hub/privacy-companies> (Last Accessed October 28, 2021).
- ²TROPT Privacy Tech Database, <https://airtable.com/shrNEnZFaT4q12ezs>.
- ³Chandy Gopalakrishnan, "Who's Funding Privacy Tech?" April 20, 2021 for Tech Crunch, <https://techcrunch.com/2021/04/20/whos-funding-privacy-tech/> (Last Accessed October 28, 2021).
- ⁴CrunchBase, Privacy Companies: Summary, <https://www.crunchbase.com/hub/privacy-companies> (Last Accessed October 28, 2021).
- ⁵Lourdes Turrecha, "Defining Privacy Tech," May 23, 2021, <https://medium.com/privacy-technology/defining-privacy-tech-ae7b022888ec> (Last Accessed October 28, 2021).
- ⁶[See, for example NIST, "NIST Develops Privacy-Preserving Encounter Metrics That Could Help Slow Down Future Pandemics," March 29, 2021, <https://www.nist.gov/news-events/news/2021/03/nist-develops-privacy-preserving-encounter-metrics-could-help-slow-down> (Last Accessed October 28, 2021), or UCSF, "UCSF Launches Pilot of Privacy-Preserving Smartphone Tool to Help Slow Spread of COVID-19," September 30, 2020, <https://www.ucsf.edu/news/2020/09/418646/ucsf-launches-pilot-privacy-preserving-smartphone-tool-help-slow-spread-covid> (Last Accessed October 28, 2021).
- ⁷Apple and Google, "Privacy-Preserving Contact Tracing," <https://covid19.apple.com/contacttracing> (Last Accessed October 28, 2021).
- ⁸Emily Ashley & Lourdes Turrecha, "Apple's iOS 14 privacy labels are rolling out today – and other notable iOS 14 privacy features," December 8, 2020, <https://medium.com/privacy-technology/ios14s-privacy-labels-are-being-rolled-out-today-plus-its-other-notable-privacy-features-3e528db63901> (Last Accessed October 28, 2021).
- ⁹Apple Newsroom, "Apple advances its privacy leadership with iOS 15, iPadOS 15, macOS Monterey, and watchOS 8," June 7, 2021, <https://www.apple.com/newsroom/2021/06/apple-advances-its-privacy-leadership-with-ios-15-ipados-15-macos-monterey-and-watchos-8/> (Last Accessed October 28, 2021).
- ¹⁰Jason Aten, "How Signal Became the Most Popular App in the World Overnight," January 16, 2021 for Inc.com, <https://www.inc.com/jason-aten/how-signal-became-most-popular-app-in-world-overnight-why-it-matters.html> (Last Accessed October 28, 2021).
- ¹¹Rae Hodge, "What is Signal? Everything you need to know about Elon Musk's app recommendation," January 28, 2021 for Cnet Tech, <https://www.cnet.com/tech/services-and-software/what-is-signal-everything-you-need-to-know-about-elon-musks-app-recommendation/> (Last Accessed October 28, 2021).
- ¹²Brave, "Brave Passes 36 Million Monthly Active Users," September 2, 2021, <https://brave.com/36m-mau/> (Last Accessed October 28, 2021).
- ¹³Gilad Edelman, "DuckDuckGo's Quest to Prove Online Privacy is Possible," June 16, 2021 for Wired, <https://www.wired.com/story/duckduckgo-quest-prove-online-privacy-possible/> (Last Accessed October 28, 2021).
- ¹⁴Nick Lucchesi, "Protonmail hits 5 Million Accounts and Wants Users to Ditch Google by 2021," September 17, 2018 for Inverse, <https://www.inverse.com/article/49041-protonmail-ceo-andy-yen-interview> (Last Accessed October 28, 2021).
- ¹⁵Natasha Lomas, "Proton, the Privacy Startup Behind E2E Encrypted ProtonMail, Confirms Passing 50M Users," May 19, 2021 for TechCrunch, https://techcrunch.com/2021/05/19/_trashed-13/ (Last Accessed October 28, 2021).
- ¹⁶See Alan F. Westin, "Privacy and Freedom," March 1, 1968 <https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=3659&context=wlufr&ref=hackernoon.com> (Last Accessed October 28, 2021).
- ¹⁷See Carissa Veliz, "Privacy is Power: Why and how you should take back control of your data," <https://www.penguinrandomhouse.com/books/673341/privacy-is-power-by-carissa-veliz/>.
- ¹⁸See Neil M. Richards and Woodrow Hartzog, "Privacy's Trust Gap," January 17, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2899760 (Last Accessed: October 29, 2021) and Ari Ezra Waldman, "Privacy as Trust: Sharing Personal Information in a Networked World," August 14, 2013, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2309632 (Last Accessed: October 29, 2021).
- ¹⁹See Daniel J. Solove, "Understanding Privacy," May 5, 2008, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1127888 (Last Accessed October 28, 2021).
- ²⁰For example, purpose and use limitation are hotly debated as out-of-date in today's data-driven world.
- ²¹OECD: Privacy Principles, <http://oecdprivacy.org>; Information Commissioner's Office, Guide to the GDPR: Principles, <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles>



ples/; Department of Homeland Security: The Fair Information Practice Principles, <https://www.dhs.gov/publication/privacy-policy-guidance-memorandum-2008-01-fair-information-practice-principles>; IAPP: Fair Information Practice Principles, <https://iapp.org/resources/article/fair-information-practices/>; PIPEDA Fair Information Principles, https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/p_principle/; FTC: Privacy Online (A Report to Congress), <https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf> (All links last Accessed October 28, 2021).

²²Oxford Learner's Dictionaries, "Technology," <https://www.oxfordlearnersdictionaries.com/us/definition/english/technology> (Last Accessed October 28, 2021).

²³Merriam-Webster, "Technology," <https://www.merriam-webster.com/dictionary/technology> (Last Accessed October 28, 2021).

²⁴Lourdes Turrecha, "Defining Privacy Tech," May 23, 2021, <https://medium.com/privacy-technology/defining-privacy-tech-ae7b022888ec> (Last Accessed October 28, 2021).

²⁵See Alan F. Westin, "Privacy and Freedom," March 1, 1968 <https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=3659&context=wlu&ref=hackernoon.com> (Last Accessed October 28, 2021).

²⁶See Carissa Veliz, "Privacy is Power: Why and how you should take back control of your data," <https://www.penguinrandomhouse.com/books/673341/privacy-is-power-by-carissa-veliz/>.

²⁷See Neil M. Richards and Woodrow Hartzog, "Privacy's Trust Gap," January 17, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2899760 (Last Accessed: October 29, 2021) and Ari Ezra Waldman, "Privacy as Trust: Sharing Personal Information in a Networked World," August 14, 2013, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2309632 (Last Accessed: October 29, 2021).

²⁸European Commission's Charter of Fundamental Rights, https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_en (Last Accessed October 28, 2021).

²⁹Office of the Privacy Commissioner of Canada, "Privacy Enhancing Technologies – A Review of Tools and Techniques," November 2017, https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711/#fn9 (Last Accessed October 28, 2021).

³⁰OECD Working Party on Information Security and Privacy, "Inventory of Privacy-Enhancing Technologies (PETs)," January 7, 2002, <https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=dsti/iccp/reg%282001%291/final> (Last Accessed October 28, 2021).

³¹John J. Borking and Charles Raab. Laws, PETs and Other Technologies for Privacy Protection. *Journal of Information, Law and Technology*, 1:1–14, 2001.

³²Lourdes Turrecha, "Defining Privacy Tech," May 23, 2021, <https://medium.com/privacy-technology/defining-privacy-tech-ae7b022888ec> (Last Accessed October 28, 2021).

³³S.224 Promoting Digital Privacy Technologies Act, Introduced February 4, 2021, <https://www.congress.gov/bill/117th-congress/senate-bill/224/text> (Last Accessed October 28, 2021).

³⁴Ann Cavoukian, "Privacy by Design: The 7 Foundational Principles," January 2011 for Information & Privacy Commissioner of Ontario, Canada, <https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf> (Last Accessed October 28, 2021).

³⁵Lourdes Turrecha, "Defining Privacy Tech," May 23, 2021, <https://medium.com/privacy-technology/defining-privacy-tech-ae7b022888ec> (Last Accessed October 28, 2021).

³⁶Michelle Finneran Denny, Jonathan Fox, and Thomas R. Finneran, "The Privacy Engineer's Manifesto," <https://link.springer.com/book/10.1007/978-1-4302-6356-2> (Last Accessed October 28, 2021).

³⁷Lourdes M. Turrecha, "Defining Privacy Tech," May 23, 2021, <https://medium.com/privacy-technology/defining-privacy-tech-ae7b022888ec> (Last Accessed October 28, 2021).

³⁸Id.

³⁹Jake Frankenfield, "Tor," August 4, 2021 for Investopedia, <https://www.investopedia.com/terms/t/tor.asp> (Last Accessed October 28, 2021).

⁴⁰Techopedia Dictionary, "Identity Management (ID Management)," <https://www.techopedia.com/definition/13636/identity-management-id-management> (Last Accessed October 28, 2021).

⁴¹Gartner Glossary, "Identity Management," <https://www.gartner.com/en/information-technology/glossary/identity-management> (Last Accessed October 28, 2021).

⁴²Techopedia Dictionary, "Data Infrastructure," <https://www.techopedia.com/definition/31682/data-infrastructure> (Last Accessed October 28, 2021).

⁴³Gartner Glossary, "Marketing Technology." <https://www.gartner.com/en/marketing/glossary/marketing-technology> (Last Accessed October 28, 2021).

⁴⁴Gartner Glossary, "Ad tech," <https://www.gartner.com/en/information-technology/glossary/ad-tech> (Last Accessed October 28, 2021).

⁴⁵Chris Metinko, "Legal Tech Makes Its Case With Venture Capitalists, Tops \$1B In Funding This Year," September 23, 2021



for Crunchbase News, <https://news.crunchbase.com/news/legal-tech-venture-investment/> (Last Accessed October 28, 2021).

⁴⁶The Data Governance Institute, "Defining Data Governance." <https://datagovernance.com/defining-data-governance/> (Last Accessed October 28, 2021).

⁴⁷Web3 Foundation, "About" page <https://web3.foundation/about/> (Last Accessed October 28, 2021).

⁴⁸Geoffrey A. Fowler, "87 percent of websites are tracking you. This new tool will let you run a creepiness check," September 25, 2020 for The Washington Post, <https://www.washingtonpost.com/technology/2020/09/25/privacy-check-blacklight/> (Last Accessed October 28, 2021).

⁴⁹The Markup: Blacklight, <https://themarkup.org/series/blacklight> (Last Accessed October 28, 2021).

⁵⁰At a high level, differential privacy provides a mathematically rigorous definition of privacy that quantifies risk. It introduces carefully calibrated "noise," or random and meaningless data, to mask personal data. See Harvard University Privacy Tools Project: Differential Privacy, <https://privacytools.seas.harvard.edu/differential-privacy>

Secure multi-party computation (SMPC) allows multiple parties to "jointly compute an arbitrary functionality without revealing their own private inputs and outputs." See Zhao et al., "Secure Multi-Party Computation: Theory, Practice, and Applications," <https://www.sciencedirect.com/science/article/abs/pii/S0020025518308338>.

Zero Knowledge Proofs (ZKPs) are a cryptographic method that allows one party to prove (the prover) to another (the verifier) that they know a statement to be true without sharing additional information. See Massachusetts Institute of Technology Theory of Computation: Zero-knowledge Proofs, <https://toc.csail.mit.edu/node/218>.

⁵¹Transmute, "Encrypted Data Vaults for Trusted Data Access," January 14, 2020, <https://medium.com/transmute-techtalk/encrypted-data-vaults-c794055b170e> (Last Accessed October 28, 2021).

⁵²Data Protection Commission, "The Right of Access," <https://www.dataprotection.ie/en/individuals/know-your-rights/right-access-information> (Last Accessed October 28, 2021); Stinson LLP, "Preparing for the CRPA," January 22, 2021 for JD Supra, <https://www.jdsupra.com/legalnews/privacy-2021-preparing-for-the-cpra-7038388/> (Last Accessed October 28, 2021); The Colorado Privacy Act (Bill Text), https://leg.colorado.gov/sites/default/files/2021a_190_signed.pdf (Last Accessed October 28, 2021); and Virginia Consumer Data Protection Act (Bill Text), <https://lis.virginia.gov/cgi-bin/legp604.exe?212+sum+SB1392> (Last Accessed October 28, 2021).

⁵³Shawn Keve, "Privacy Must Be Central to Identity and Access Management," May 7, 2020 for Forbes, <https://www.forbes.com/sites/forbestechcouncil/2020/05/07/privacy-must-be-central-to-identity-and-access-management/?sh=82e-d578696ea> (Last Accessed October 28, 2021).

⁵⁴CPO Magazine, "Working Together for Privacy Protection: Identity Management Best Practices for Companies and Consumers," June 29, 2021, <https://www.cpomagazine.com/cyber-security/working-together-for-privacy-protection-identity-management-best-practices-for-companies-and-consumers/> (Last Accessed October 28, 2021).

⁵⁵Mark Settle, "Privacy Debt: The Achilles Heel of Every IT Organization," September 10, 2020 for Forbes, <https://www.forbes.com/sites/marksettle/2020/09/10/privacy-debt-every-it-organization/?sh=3eb5ead87a3d> (Last Accessed October 28, 2021).

⁵⁶Thor Olavsrud, "Data Governance: A best practices framework for managing data assets," March 18, 2021 for CIO Magazine, <https://www.cio.com/article/3521011/what-is-data-governance-a-best-practices-framework-for-managing-data-assets.html> (Last Accessed October 28, 2021).

⁵⁷PwC, "Five steps to meeting the CPRA's new data retention requirements," <https://www.pwc.com/us/en/services/consulting/cybersecurity-privacy-forensics/library/cpra-data-retention-preparation.html> (Last Accessed October 28, 2021).

⁵⁸EY, "How to Minimize Risk through Data Disposition," February 19, 2021, https://www.ey.com/en_us/consulting/how-to-minimize-risk-through-data-disposition (Last Accessed October 28, 2021).

⁵⁹NIST Computer Security Resource Center Glossary, "De-identification" https://csrc.nist.gov/glossary/term/de_identification (Last Accessed October 28, 2021).

⁶⁰Bryan Cave Leighton Paisner, "What is 'pseudonymized' data?," January 17, 2020, <https://www.bclplaw.com/en-US/insights/what-is-pseudonymized-data.html> (Last Accessed October 28, 2021).

⁶¹NIST Privacy Engineering Program, "De-identification Tools," <https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/focus-areas/de-id/tools> (Last Accessed October 28, 2021).

⁶²Richard Seiersen, "A Modern Shift-Left Security Approach," January 4, 2021 for Forbes, <https://www.forbes.com/sites/forbestechcouncil/2021/01/04/a-modern-shift-left-security-approach/?sh=19fab35a7729> (Last Accessed October 28, 2021).

⁶³Steve O'Hear, "Digi.me and Personal merge to put you in control of the nascent 'personal data ecosystem,'" August 17, 2017 for TechCrunch, <https://techcrunch.com/2017/08/17/digi-me-and-personal-merge/> (Last Accessed October 28, 2021).

⁶⁴GDPR, CCPA, HIPAA, amongst many other data protection laws require organization to pass down privacy and security



requirements to their vendors and other third parties processing personal information.

⁶⁵See, for example: TROPT Data Privacy Day 2021, The Marketplace Demand for Privacy https://www.youtube.com/watch?v=rR7npMhyp_w (Last Accessed October 25, 2021); Andrew Perrin, "Half of Americans have decided not to use a product or service because of privacy concerns," April 14, 2020 for Pew Research Center, <https://www.pewresearch.org/fact-tank/2020/04/14/half-of-americans-have-decided-not-to-use-a-product-or-service-because-of-privacy-concerns/> (Last Accessed: October 29, 2021); and Consumer Reports, "Consumer Reports Study Finds Marketplace Demand for Privacy and Security," October 21, 2020, <https://www.consumerreports.org/media-room/press-releases/2020/10/consumer-reports-study-finds-marketplace-demand-for-privacy-and-security/> (Last Accessed: October 29, 2021)

⁶⁶TROPT Data Privacy Day 2021, "The Marketplace Demand for Privacy," https://www.youtube.com/watch?v=rR7npMhyp_w (Last Accessed: October 25, 2021).

⁶⁷Figures pulled from: The Forum for Sustainable and Responsible Investment (US SIF), "Sustainable and Impact Investing Overview," <https://www.ussif.org/files/2018%20Infographic%20overview%20%281%29%281%29.pdf> (Last Accessed: October 29, 2021); US SIF, "2020 Report on US Sustainable and Impact Investing Trends," <https://www.ussif.org/files/US%20SIF%20Trends%20Report%202020%20Executive%20Summary.pdf> (Last Accessed: October 29, 2021); and Anne Field, "With \$715B in AUM, Impact Investors Stay the Course, Despite the Pandemic, Says the GIIN," June 16, 2020 for Forbes, <https://www.forbes.com/sites/annefield/2020/06/16/with-715b-in-aum-impact-investors-stay-the-course-despite-the-pandemic-says-the-giin/?sh=377bde166d3c> (Last Accessed: October 29, 2021).

⁶⁸CrunchBase, Privacy Companies: Summary, <https://www.crunchbase.com/hub/privacy-companies> (Last Accessed October 28, 2021).

⁶⁹TROPT Privacy Tech Deal Flow Tool. For startups: <https://www.riseofprivacytech.com/founders/>. For investors: <https://www.riseofprivacytech.com/investors/>.



ABOUT THE RISE OF PRIVACY TECH

The Rise of Privacy Tech's (TROPT) mission is to fuel privacy innovation. We do this by bringing together privacy tech founders, investors, experts, and user-buyers to bridge the tech-capital-expertise gaps in privacy tech.

TROPT was founded by Lourdes M. Turrecha. TROPT's inaugural advisors include Michelle Dennedy, Melanie Ensign, Debra Farber, Gilbert Hill, and Fatima Khan.



Contact information for The Rise of Privacy Tech (TROPT) can be found at www.riseofprivacytech.com.

Published November 2021.

© 2021 The Rise of Privacy Tech. All rights reserved.

The Rise of Privacy Tech and The Rise of Privacy Tech logos are trademarks of The Rise of Privacy Tech. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner or member does not imply a partnership relationship between The Rise of Privacy Tech and any other company.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Turner, Sean and Zucker-Scharff, Aram

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Statement from Aram Zucker-Scharff and Sean Turner to OSTP on PATCG

FILING DEADLINE: Jul 8, 2022

This is a response to OSTP's Request for Information on Advancing Privacy-Enhancing Technologies [Doc. 2022-12432] on behalf of Aram Zucker-Scharff and Sean Turner - members of the public.

BACKGROUND: Private Advertising Technology in the W3C

Historically, web-based advertising has involved the collection and processing of significant amounts of data about web users by a variety of parties within the digital advertising ecosystem. In recent years, a number of these parties (including Mozilla, Apple, Meta, Google, and others) have begun to redesign the mechanisms used to carry out a variety of advertising-related functions in more privacy-preserving ways.

In October of 2021, participants in the World Wide Web Consortium (W3C) chartered the Private Advertising Technology Community Group (PATCG) [1][2][3][4] to incubate web features and application programming interfaces (APIs) that support advertising while acting in the interests of users, in particular providing strong privacy assurances. The CG welcomes participation from browser vendors, OS vendors, mobile application vendors, advertisers, publishers, ad buyers, advertising platforms and intermediaries, privacy advocates, web application developers, and other interested parties.

The CG is developing use cases and requirements to better scope the private advertising problem space. A threat model is being developed that will enumerate the actors in the digital advertising ecosystem, their capabilities, potential attacks, and mitigation strategies. Mitigation strategies the CG is already investigating include technologies that rely, in part or wholly, upon multi-party computation (MPC) and/or trusted execution environments (TEEs). The CG is also investigating multiple APIs including:

- Interoperable Private Attribution (IPA), which proposes a mechanism for advertisers to understand the performance of their ad campaigns that is based on MPC and upholds differential privacy guarantees;
- Private Click Measurement (PCM), which aims to allow for ad clicks to be recorded without tracking users across websites; and
- Topics, which is intended to allow advertisers to target ads to users based on users' interests, without personally identifying the users.

TECHNICAL STANDARDS AND PRIVACY TECHNOLOGIES

Technical standards are crucial in the deployment of protocols for sharing data and analytics, because the mechanisms used for preserving privacy fundamentally involve multiple parties that need to interoperate. This need is even more pronounced in the context of protocols that rely on a non-collusion assumption. Such protocols require at least two entities to communicate for each measurement or collection of data, while strictly limiting the extent of their coordination. In the

absence of relevant technical standards, these parties would have to collaborate deeply to establish communication, threatening the non-collusion assumption.

RECOMMENDATIONS

To advance private advertising technology in the OSTP's work, we offer the following recommendations:

- Support the development of new private advertising techniques.
- Encourage private advertising developers and researchers to make their work compatible with open standards.
- Encourage deployment of open standards to support private advertising technology as opposed to proprietary solutions.
- Encourage privacy-respecting practices that leverage private advertising technologies among marketing and advertising companies.
- Ensure that federal procurement of private advertising technologies is based on open technical standards when possible.
- Encourage federal privacy regulators to monitor and support the development of private advertising technologies based on open standards.
- Support research into privacy-enhancing technologies that can improve the state of the art and its public understanding.

[1] <https://www.w3.org/community/patcg/>

[2] <https://patcg.github.io/>

[3] <https://www.w3.org/community/patcg/participants>

[4] Community Groups are participant-driven community fora for discussion and incubation prior to (potential) formal standardization activity at W3C.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

University of Southern California Information Sciences Institute

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

RFI Response: Request for Information on Advancing Privacy-Enhancing Technologies

July 8, 2022

These comments are submitted on behalf of the University of Southern California's Information Sciences Institute.

1. Response to Topic # 1: Specific research opportunities to advance PETs

1.1. Understanding and Regulating How Companies Collect, Store and Use Private Data

Startup landscape today is mostly unregulated when it comes to private data collection. What data is collected and how, how it is stored and how it is used is left fully to companies to decide. At best companies must inform users of data collection and get them to accept the terms of services. Companies misuse this by asking too many privileges from users, and justifying this by saying that users have freedom to not use a service. It is well known that users are not rational when it comes to deciding which service to use and how to reason about various privacy/utility trade-offs. It is also very likely that companies collect more data than they need, and store it in an ad-hoc manner. When that data is leaked, customers' privacy is put at risk.

We need thorough analysis of business practices around private data collection, storage and use. We need to understand how much of the data is really needed for each business model, and how that data can best be protected. We then need common APIs for data collection and common, secure services for storing and querying this data. Finally, we need to ensure that customers can view and delete their data at will and that they can revoke access to certain data while still being able to access services (potentially at a lower utility level). For example, a customer may revoke access of an online service to their address, which should lead to deletion of this address from the company's systems. The user should still be able to browse and order items, but would then be prompted to input their address each time they check out.

Further, we need studies on how users reason about privacy choices, and how well they can meet their privacy goals with current user interfaces. We also need to shift to a private-by-default model for user data. Instead of asking users to give access to their private data to the company offering a free service, the company should ask for data access only when

absolutely necessary. No data should be assumed public and no data should be kept longer than absolutely necessary.

2. Response to Topic # 2: Specific technical aspects or limitations of PETs

2.1. Privacy-infringement of PETs

Privacy preserving technologies enable sharing of sensitive information while preserving privacy. However, there are open questions about the privacy implications of some of the used technologies. In this response, we specifically discuss two of these technologies and the questions around them.

The first technology is using AI generated face images to replace real face images in the shared data. The problem with this technology is that there is no guarantee that the generated face images do not correspond to real people. Therefore, if AI generated images are used to replace real images, the privacy of an individual might be compromised in order to preserve the privacy of another individual. In a pilot study we conducted at USC ISI, we estimated that around 90% of face images in the CelebA-HQ face dataset can be faithfully generated by a face image generation model (StyleGAN2). That indicates the high recall of such a model with respect to the distribution of real face images. Estimating the precision of these models is a more interesting and challenging task that requires further research.

The second technology is using adversarial perturbation to obfuscate the real identity of a face image. In this process, the face image is made to match an identity that is different from the original one by adding an imperceptible perturbation. However, the privacy of the new identity can be infringed by the practice. Therefore, methods for adversarial face image obfuscation should be designed such that there is some level of guarantee that the target identity of the obfuscation process does not match a real identity.

2.2. Privacy-preserving Entity Resolution

Entity resolution is the task of identifying records in different datasets or databases that refer to the same entity in the real world. For example, consider patient databases in two hospitals. The entity resolution task in this case is to identify the records of the same patients in the two databases. To accomplish this task, an algorithm will compare the information present in the records to determine if the records refer to the same person. The algorithm would compare the names, home or work address, phone numbers, etc. to make the determination. The principal challenge is that the two hospitals may record the information in different formats. For example,

one hospital may use separate fields for first name and last name, and the other hospital may store the full name in a single field, including a middle name. The privacy preserving entity resolution task is to solve the entity resolution task without revealing sensitive information. Specifically, in the privacy preserving setting, each party will know which of its records are present in the databases of the other parties, but will learn nothing about any other records that the other parties may have in their databases. For example, if one hospital has records about Mary, Joe and Bob, and the other hospital has records about Joe, Bob and Susan (and assuming that Joe and Bob are the same person), then the first hospital will know that the second hospital has records about Joe and Bob, but will not learn any information about the two people unless explicitly disclosed by the second hospital. Furthermore, the first hospital will not know that the second hospital has a record about a person named “Susan”. In fact, the first hospital will not learn anything about the distribution of names, locations or any other information present in the database of the second hospital. Likewise, the second hospital will not learn any information about the first hospital’s database beyond knowing that it has records for Joe and Bob.

Privacy preservation creates a significant challenge for entity resolution as the data must be encrypted to preserve privacy, but once encrypted, similarity comparisons become difficult. This is especially the case for fuzzy or approximate matching, which is required to account for differences in formats and data structures. While entity resolution is a well-studied area of computer science (a query for “entity resolution ” in Google Scholar returns over 2 million results), privacy preserving entity resolution has received little attention in the scientific community.

The University of Southern California’s Information Sciences Institute (ISI) carries out basic and applied research in artificial intelligence, networks and cybersecurity, high- performance computing, microelectronics, and quantum information systems. Its \$100M annual external funding comes from the NSF, DoD, IC, NIH, DoE, industry, foundations, and other sponsors. ISI is home to the first quantum computer in academia. Part of the USC Viterbi School of Engineering, ISI has more than 400 personnel that includes 28 faculty that advise 65 PhD students.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Visa Inc

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



U.S. Office of Science and Technology Policy (OSTP) Request for Information on Advancing Privacy-Enhancing Technologies

Notice Published on the [Federal Register](#) on June 9, 2022

Visa Inc.'s Response to OSTP's RFI on Advancing Privacy-Enhancing Technologies (PETs)

Submitted by Visa's Global Privacy Office and Global Strategic Initiatives Teams

July 8, 2022

Visa welcomes the opportunity to provide input on privacy-enhancing technologies ("PETs") and their emerging role in data-sharing and collaborative analytics. We look forward to continuing this important dialogue as an industry stakeholder.

If you would like to discuss any of our responses in greater detail, please contact the Visa Chief Privacy Officer, Leigh Feldman, or Visa Global Privacy Counsel, Sunny Seon Kang.

Please find Visa's comments below.

1. Specific research opportunities to advance PETs:

Hackathons, tech sprints, and regulatory sandboxes provide opportunities for cross-functional stakeholders to collaborate on a problem statement and navigate solutions through the participants' diverse expertise and perspectives. The value of these workshops is in welcoming multidisciplinary approaches to a common challenge, such as applying a technological solution to a traditionally legal and regulatory problem.¹ The role of hackathons in validating modern PETs and associated cryptographic methods has been demonstrated as early as 2006 with the Netflix Prize Dataset de-identification challenge, which tested the robustness of anonymization techniques against adversarial attacks.²

PETs are currently at a juncture of evolving from academic research to practical and scalable application. Research by the Financial Action Task Force identified one of the main barriers to PETs adoption as the lack of certainty on compliance standards.³ Therefore, the development of PETs would benefit from

¹ For example, the U.S. Consumer Financial Protection Bureau (CFPB) published a request for information in 2019 on utilizing "Tech Sprints as a means to encourage regulatory innovation and collaborate with stakeholders in developing viable solutions to regulatory compliance challenges" and cited the UK Financial Conduct Authority's hackathon they hosted on PETs solutions for anti-money laundering and financial crime challenges. See at, <https://www.federalregister.gov/documents/2019/09/18/2019-20201/request-for-information-regarding-tech-sprints>.

² Narayanan and Shmatikov, *How To Break Anonymity of the Netflix Prize Dataset*, Cornell University Database (November 2007): <https://arxiv.org/abs/cs/0610105>

³ Financial Action Task Force (FATF), *STOCKTAKE ON DATA POOLING, COLLABORATIVE ANALYTICS AND DATA PROTECTION* (July 2021): <https://www.fatf-gafi.org/media/fatf/documents/Stocktake-Datapooling-Collaborative-Analytics.pdf>

examining their proof of concept in hackathons which focus on a real-life use cases, often through regulatory supervision and partnership.⁴

As the White House Office of Science and Technology Policy (“OSTP”), U.S. National Science Foundation, and U.S. National Institute of Standards and Technology (“NIST”) jointly work with the UK government to develop PETs challenges in the coming year,⁵ we believe it would be beneficial to encourage industry collaboration with academic researchers and universities in the challenges, and to work towards an outcome that furthers technical guidance and best practices for privacy-preserving analytics.

2. Specific technical aspects or limitations of PETs:

PETs collectively refer to a variety of computing techniques, including but not limited to: differential privacy, use of synthetic data, federated learning, homomorphic encryption, and secure multi-party computation. Although many of these techniques have a long history of academic research and development, their deployment in the enterprise environment is still relatively nascent and exploratory in certain sectors.

It is important to note that each privacy-preserving technique entails constraints and tradeoffs. Certain methods of de-identification can diminish the usability and accuracy of data. For example, differential privacy injects noise to protect data subjects from identifiability, but doing so decreases the accuracy of analytics done with that data.⁶ Additionally, computational overheads associated with cryptographic methods often cause delays in processing, limiting their application to more complex machine learning models such as decision trees or neural nets.⁷

Some of the constraints listed above have been mitigated by combining two or more types of PETs, such as federated learning with differential privacy, secure multi-party computation and homomorphic encryption, and so forth.⁸

3. Existing barriers related to PETs adoption; how regulatory frameworks may address them:

A technical understanding of PETs should underpin the policy framework. This would ensure that regulatory guidance on PETs is neither quickly outpaced by innovation nor overly prescriptive such that it would inhibit new technical approaches.

Organizations interested in exploring PETs as an additional safeguard would benefit from regulatory guidance which clarifies compliance standards (i.e. in defining what constitutes anonymization versus pseudonymization) while also providing flexibility on the technical approach to achieve that standard.

⁵ <https://www.whitehouse.gov/ostp/news-updates/2021/12/08/us-and-uk-to-partner-on-a-prize-challenges-to-advance-privacy-enhancing-technologies/>

⁶ Bagdasaryan et al, *Differential Privacy Has Disparate Impact on Model Accuracy*, 33rd Conference on Neural Information Processing Systems (NeurIPS 2019): <https://proceedings.neurips.cc/paper/2019/file/fc0de4e0396fff257ea362983c2dda5a-Paper.pdf>.

⁷ Damgard et al, *Perfectly Secure Multiparty Computation and the Computational Overhead of Cryptography*, International Association for Cryptologic Research (2010): https://link.springer.com/content/pdf/10.1007/978-3-642-13190-5_23.pdf

⁸ Civil society groups such as Future of Privacy Forum and the International Association of Privacy Professionals have published on market and regulatory frameworks propelling PETs. See, *Future of Privacy Forum, Privacy Tech’s Third Generation A Review of the Emerging Privacy Tech Sector* (June 2021): https://fpf.org/wp-content/uploads/2021/06/FPF-PTA-Report_Digital.pdf

Regulatory clarity on utilizing PETs would accelerate PETs adoption as organizations would better understand their obligations when approaching a new way of computing or sharing data.

4. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:

The advancement of PETs is particularly important to privacy-preserving machine learning and network analysis in investigating financial crimes. The emerging role of PETs in the financial sector has been examined by the World Economic Forum⁹ and the Royal United Services Institute.¹⁰ The United Nations has published a white paper on privacy-preserving computation techniques which examines their applicability to statistics and model development.¹¹

5. Best practices that are helpful for PETs adoption:

Data privacy and security through PETs can achieve greater data accountability and informational autonomy for data subjects. Privacy-by-design, especially through PETs, should be emphasized as a critical step to proactively mitigating privacy risks.

Technological advances in privacy-preserving tools are creating an opportunity for organizations to work towards preventing, rather than reacting, to data-related harms. Thus, the focus of best practices should be reframed from avoiding consumer harms to championing consumer empowerment through de-identification and PETs.

⁹ World Economic Forum, *The Next Generation of Data-Sharing in Financial Services: Using Privacy Enhancing Techniques to Unlock New Value* (September 2019): https://www3.weforum.org/docs/WEF_Next_Gen_Data_Sharing_Financial_Services.pdf

¹⁰ RUSI, *Future of Financial Intelligence Sharing (FFIS) Innovation and discussion paper: Case studies of the use of privacy preserving analysis to tackle financial crime* (January 2021): https://www.future-fis.com/uploads/3/7/9/4/3794525/ffis_innovation_and_discussion_paper_-_case_studies_of_the_use_of_privacy_preserving_analysis_-_v.1.3.pdf

¹¹ United Nations, *UN Handbook on Privacy-Preserving Computation Techniques*: <https://unstats.un.org/bigdata/task-teams/privacy/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf>

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

VMware

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.



July 8, 2022

Dr. Alondra Nelson
Director, Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, DC 20504

Dear Dr. Nelson:

VMware appreciates the opportunity to submit these comments in response to the Office of Science and Technology Policy's Request for Information (RFI) on Privacy-Enhancing Technologies (PET). Below we provide a high-level overview of VMware's research and development approach and vision relating to the data economy as well as answers to several of the questions posed in the RFI.

VMware's Vision: Confidential Computing Technologies as PET Foundation for the Data Economy

The data economy The data economy is defined by Wikipedia as "a global digital ecosystem in which data is gathered, organized, and exchanged by a network of vendors for the purpose of deriving value from the accumulated information". The data economy is estimated to comprise 1% of US GDP¹. Importantly, there is a growing understanding that shared data is more valuable than unshared data; for example, Gartner reports that "Data and analytics leaders who share data externally generate three times more measurable economic benefit than those who do not"². Data sharing increases the importance of both data security (the ability to keep secrets) and data privacy (the ability to control who sees those secrets and under what conditions).

Inhibitors of a data economy Consumers and enterprises are understandably reluctant to share data without privacy assurances, arguably holding back the full potential of the "data economy". There have been numerous instances of unauthorized data sharing³ that have reduced trust by data owners. We hypothesize that technical privacy guarantees will unleash a more extensive and effective data economy.

Desired future state We believe that the future state should be one in which data owners retain full privacy controls **implemented with trustworthy technology foundations, not merely operating on trust**. Data producers/owners should have technically grounded control over who sees their data, when it is available for use, how it is combined with other data, when it is deleted, etc. Enterprises that participate in the data economy, and who implement appropriate technological controls, will become

¹ https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/FTQ_1_Jan_2019.pdf

² <https://www.gartner.com/smarterwithgartner/data-sharing-is-a-business-necessity-to-accelerate-digital-business>

³ <https://iapp.org/resources/article/u-s-state-data-breach-lists/>



not just “trusted” but “trustworthy”. Ideally, a virtuous cycle of deployment will be induced: the first enterprises to deploy trustworthy technology may become preferred service providers, and those who do not will be viewed with a skeptical eye by an increasingly large number of customers/partners.

Beneficiaries Data producers and owners will benefit, both intangibly (e.g., from superior services resulting from the controlled sharing of data) and tangibly (e.g., from the possibility of directly monetizing the controlled sharing of data in a data economy). Enterprises that offer products and services based on shared, combined data will benefit economically through competitive advantage and more customized services. Companies that provide infrastructure for the data economy will benefit. A workforce that specializes in implementations of the data economy will benefit. However, there could also be dislocations. Corporations whose products and services rely on the availability of uncontrolled data might find that it is less available and/or that they would need to compensate data owners for the use of that data. In addition, there could be costs associated with implementing technical privacy controls; it is our hypothesis that the cost of such controls is less than the economic and societal benefit of a robust data economy business opportunity. Finally, the widespread deployment of confidential computing has security advantages beyond use in the data economy; thus national security will eventually be strengthened.

Technology

Confidential computing as a foundation for security and privacy. It is our hypothesis that “confidential computing” (CC) technologies are necessary foundations for enabling comprehensive data security and privacy. Confidential computing is a principled, hardware-based security mechanism for distributed computing. CC protects the integrity and confidentiality of processing and data, wherever programs run, from malicious programs or careless insiders. CC provides confidentiality for data in use and enables, for the first time, a general-purpose end-to-end encryption of data (at rest, in transit, during processing) that was not feasible prior to the recent introduction of “trusted execution environments” (TEEs) such as Intel’s Software Guard Extensions (SGX), AMD’s Secure Encrypted Virtualization, Intel’s Trust Domain Extensions (TDX), and Arm “Realms”. Through isolation and encryption, CC provides a general-purpose means of grounding data security in the data economy.

Measurement and attestation: under-appreciated aspect of confidential computing. Two under-appreciated properties of confidential computing are “measurement” and “attestation”, which determine both the programs (applications) and the hardware that can be “trusted”. Through these mechanisms, CC grounds data privacy capabilities that could unleash a vibrant data economy. Without measurement and attestation, unauthorized programs could be arbitrarily used on private data by malicious insiders, careless operators, and/or predatory participants in a data economy. With measurement and attestation, it becomes possible to control who can see or modify the data and under what circumstances.

Hardcoded privacy and chains of trust. The security concept of “confidentiality” is a building block for privacy in the sense that privacy controls must be correctly implemented through encryption, certificate management, and proper operation of the software that directly manipulates the decrypted data. Many details need to be successfully addressed in order to implement privacy controls, i.e., to specify who is



given access to data and under what circumstances. In general, formal policy languages are not yet used to specify data privacy controls which means that manual techniques must be used to verify that a given system obeys an implicit privacy specification. Not only are these manual mechanisms error prone, but they are also impossible to verify automatically as use evolves. The full range of details that needs to be considered is enormous, even without the risks associated with coding errors and malicious exploits. Confidential computing does not inherently address these complexities, but it does offer a foothold for addressing these problems that cannot be easily subverted.

One important contemporary cloud security problem that has privacy implications involves the simple use case of protecting a cloud application (and its data) from a cloud provider hosting that application. This is a case where in principle a careless or malicious operator has the potential to introduce a privacy violation because both the cloud provider and the application developer are a “chain of trust”. One general class of privacy violation that needs to be addressed is preventing unauthorized access, sharing and manipulation by a cloud provider; in other words, to remove that cloud provider from the chain of trust.

To leverage the foothold provided by confidential computing, VMware has researched and implemented a “certifier” framework that simplifies the range of issues that need to be addressed in confidential system design, such that trust policy is separated from implementation and (with few additional lines of code) the developer’s role in ensuring security is reduced to providing a correctly written program and specifying an access policy that represents their intentions. It is VMware’s intention to contribute this certifier framework to the open source community, to enable both “hard-coded” privacy such as trustworthy cloud computing as well “data economy” applications that we hope will eventually support sophisticated end-user “Data Use Controls”.

Data use controls (DUCs). The measurement and attestation of programs opens the door to a privacy-preserving data economy infrastructure. A given program can be inspected to ensure that it does not programmatically divulge, reveal, or misuse data. Measurement and attestation of this program can then ensure that only acceptable programs operate on data in question. However, initially, data providers will not have machine-readable specifications concerning rights and limitations on use of their data. Therefore, that specification will necessarily be encoded in the logic of the attested program. A better option exists, in which each data provider can explicitly specify the uses and restrictions that apply to their data, and those directives are enforced automatically and monitored. The bridge to this next step involves policy languages that have been researched under the term “data use controls” (DUCs).⁴ By bundling data with their DUCs, a data provider’s policy concerning the permitted uses of their data could potentially be unconditionally enforced by the attested program. Such a program could be verified to correctly implement data use controls in a trustworthy and verifiable manner. Over time, there is potential for automated verification techniques over increasingly sophisticated DUCs policy languages, which are already being shown⁵ as capable of supporting the full generality of data privacy requirements such as the Health Insurance Portability and Accountability Act (HIPAA) and the EU’s

⁴ <https://www.usenix.org/conference/hotosxiii/do-you-know-where-your-data-are-secure-data-capsules-deployable-data-protection>

⁵ cf <https://arxiv.org/abs/1909.00077>



General Data Protection Regulation (GDPR). As described below, DUCs represent an important direction for use-inspired basic research and in some cases may be ready for research translation.

The role of cryptographic approaches to secure multiparty computation. Cryptographers are sometimes skeptical of hardware-based security mechanisms such as TEEs in part because they may contain opaque implementation details; in principle, the manufacturer is in the chain of trust, and can introduce security flaws deliberately or mistakenly. There is a significant body of research in cryptographic approaches to privacy that do not require trust in the hardware manufacturer (for example, secure multi-party computation, homomorphic and semi-homomorphic encryption, and zero-knowledge proofs). However, there are often performance and complexity challenges with these approaches. In addition, present research has not addressed the incorporation of DUCs into these approaches. Nevertheless, these cryptographic approaches can potentially enhance specific privacy challenges and complement CC as a form of defense in depth even if they do not become a general-purpose foundation for the data economy. More research is needed.

Blockchain platforms. Blockchain-based infrastructures are often proposed as a foundation for digital commerce and may play a role in the “data economy”. Blockchain platforms involve a combination of technical components including consensus mechanisms (e.g., proof-of-work, proof-of-stake, proof-of-authority), immutable data structures (e.g., Merkle trees), and transaction languages that execute commercial exchange. These foundations have been used to create marketplaces for digital objects (e.g., non-fungible tokens or NFTs). Blockchains can also be useful as a foundation for other important aspects of a data economy, such as auditing of configuration changes, security compliance, and data access transparency. Although blockchains have been identified as one foundation for privacy-preserving operations such as financial “know your customer” inquiries, the full machinery of modern blockchains may not be needed to support the data economy. For example, it is possible for two parties to execute commercial agreements without involving immutable data structures, complex consensus mechanisms, or formal transaction languages. Therefore, the door should be left open to new discoveries about the role of blockchain machinery in the data economy, but blockchains do not need to be an assumed foundation.

Technology priorities and challenges. To accelerate the data economy, the first priority is to encourage the widespread deployment of TEEs that underpin confidential computing. Concurrently, incentives and programs should encourage the use of CC technologies for privacy protection in the context of enterprises that participate in the data economy. Third, research should be accelerated in the area of data use controls. Initially, enterprises will likely need to specify their customers’ intent on behalf of the customers, but eventually the control of that specification should devolve to the originator/owner of the data. NIST has played an important leadership role in promulgating the idea and approach to “zero trust” security⁶; the agency could arguably play a similar leadership role in developing the architecture and practical approach to a data economy. Ultimately, it will be desirable to standardize DUCs policy languages to increase interoperability.

⁶ <https://csrc.nist.gov/News/2022/planning-for-a-zero-trust-architecture-white-paper>



With the above context, we now turn to commentary on selected RFI Questions:

1. *Specific research opportunities to advance PETs*

A modified version of Donald Stokes’ 4-quadrant research framework⁷ can help characterize the initiatives needed to enable a data economy. To complement Stokes’ original research framework, NSF’s new Technology and Innovation Partnerships (TIP) directorate has proposed a new dimension for *research translation* which we represent as “TIP’s dimension”; this dimension focuses on the work of innovation, i.e., translating inventions into practice. A comprehensive governmental initiative would invest in a range of research programs touching on each quadrant and dimension.

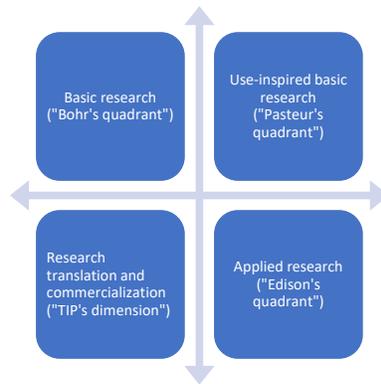


Figure 1: A modified version of Donald Stokes' framework

Confidential computing is sufficiently advanced that it can be immediately put into practice, while new research can subsequently enhance and deepen the latent and emerging opportunities. Therefore, VMware suggests pursuing each quadrant in parallel.

Research translation and commercialization: vendor-neutral, standardized infrastructure for CC and DUCs. It would be valuable to encourage development of common infrastructure for confidential computing and data use controls across producers and exchangers of data. Today, there are a plethora of TEEs under production, and, every cloud provider has a different approach to confidential computing (in particular to the level of support that is provided for key privacy-related features of confidential computing - measurement and attestation).

Thus, a different approach is needed to remove each specific cloud provider from the chain of trust, which creates an onerous burden for most cloud users. As mentioned above, VMware intends to bring certain vendor-neutral technologies to the open source community in order to help with this problem. Much work will remain, however, ranging from new data exchange interfaces for data lakes and warehouses, to common policy languages for expressing privacy desires.

⁷ https://en.wikipedia.org/wiki/Pasteur%27s_quadrant



Patterned on its work in “zero trust” security⁸, NIST could be of significant help in developing best practices and incremental approaches to vendor-neutral, standardized data exchange infrastructure that could be used by enterprise that wish to become privacy leaders on behalf of their employees and customers.

Research translation and commercialization: seeding the data economy with exemplar products and services Although many processor vendors have introduced hardware abstractions for confidential computing, industry is only beginning to exploit these capabilities for security and privacy. A key goal of research translation should be to proliferate creative new privacy-related use cases and implementations of confidential computing for purposes of privacy.

One mechanism to advance this goal would be to introduce new Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) initiatives that encourage the development of confidential computing products and services that leverage secure multi-party data sharing. In addition, the National Science Foundation (NSF)’s new Technology, Innovation and Partnerships (TIP) directorate could contemplate launching innovative public-private partnerships to stimulate the data economy and accelerate research translation in CC and DUCs, leveraging not only the Convergence Accelerator model but also looking at other models including joint solicitations.⁹

Use inspired basic research: data use controls and other advances As mentioned elsewhere, foundational PET research has already been funded and is being advanced by NSF as well as digital infrastructure providers such as VMware. Examples of this research include portable TEE abstractions, separation of attestation policy and implementation, certification services for end-to-end trust establishment in distributed confidential computing systems, and research into data use controls policy languages. Some of this work may lead to standards and implementation in open source under the guidance of forums such as the Confidential Computing Forum. However, there is additional use-inspired basic research that would be highly desirable in the long run. Important topics include DUCs frameworks; parameterizing automated CC “attestation” with data use controls; mapping prose data use policies into machine-readable DUCs specifications; identifying synergies between CC and other PETs; and extending CC to emerging hardware such as machine learning (ML) accelerators, remote memory, tagged architectures, etc.

Applied research There is a fair bit of applied research that is needed to deploy confidential computing, including, for example, performance and correctness work related to specific implementations. Although industry will likely carry out much of this applied research without government funding,

⁸ <https://csrc.nist.gov/News/2022/planning-for-a-zero-trust-architecture-white-paper>

⁹ VMware regularly engages in such public-private research partnerships with the NSF and has co-funded joint solicitations in [Security](https://www.nsf.gov/pubs/2016/nsf16582/nsf16582.htm) (<https://www.nsf.gov/pubs/2016/nsf16582/nsf16582.htm>), [Edge](https://www.nsf.gov/pubs/2018/nsf18540/nsf18540.htm) (<https://www.nsf.gov/pubs/2018/nsf18540/nsf18540.htm>), [Sustainability](https://nsf.gov/pubs/2020/nsf20594/nsf20594.htm) (<https://nsf.gov/pubs/2020/nsf20594/nsf20594.htm>), and [Next G Telecommunications](https://www.nsf.gov/pubs/2021/nsf21581/nsf21581.htm) (<https://www.nsf.gov/pubs/2021/nsf21581/nsf21581.htm>).



government can support the applied research necessary in other ways, such as by serving as a convening body for industry, academia, and government to discuss applied research developments.

Basic research Finally, there is a need for additional basic research in the area of privacy. Of necessity, the term is a subjective, and is viewed and experienced differently by different populations. For example, Prof. Seny Kamara (Brown University) has called attention to the importance of considering marginalized communities in the context of cryptography.¹⁰ The NSF Social, Behavioral, and Economics (SBE) directorate is positioned to engage in relevant social science and ethnography research, to identify societal impacts of privacy or its absence, and to understand economic and incentive structures leading to a vibrant data economy. In addition, NSF CISE directorate is well-positioned to engage in more basic research on novel PETs.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

Sectors: Although there is a segment of the economy that does (or could be seeded) to specifically support a “data economy” (e.g., data aggregators), the “data economy” could also be an aspect of many existing industries. Sectors that could participate in a CC-based data economy (and benefit from PETs more broadly) include:

- **Companies that aggregate and analyze data** as their core business
- **Web-enabled enterprises** that wish to remove themselves from the chain of trust in respecting privacy controls
- Enterprises in industries like **finance** and **healthcare** that require data aggregation but have regulatory requirements around privacy-preserving data sharing
- **Government agencies (defense, intelligence, etc)** that have multilevel security challenges

Workloads: Privacy-preserving workloads that could be implemented with today’s confidential computing capabilities range from the scientific (e.g., privacy-preserving genomics analysis) to commerce (e.g., privacy-preserving auctions) to new businesses (e.g., consumer monetization of privacy data). Moreover, confidential computing offers a general-purpose and performant way to implement many of the workloads being discussed today in the context of cryptographic analytics (e.g., secure multi-party computation, homomorphic and semi-homomorphic encryption, private set intersection, zero knowledge proofs) and certain forms of federated machine learning. Programs aimed at exploring the use cases should help motivate others by showing feasibility and inspiring new commercial opportunities for trustworthy data analytics.

7. Risks related to PETs adoption

¹⁰ <https://www.youtube.com/watch?v=Ygq9ci0GFhA>



We see a number of “incubation risk” unknowns involved in promoting the data economy. Many of these can potentially be mitigated, but as with all initiatives it will be important to monitor and adjust the nation’s investments and plans accordingly as more is learned about these risks.

Risk: The concept of “privacy” is extremely broad, and there may be aspects that cannot be supported with technology.

- Mitigation: Nevertheless, the end-to-end encryption enabled by confidential computing, combined with technical data user controls over data dissemination, will be a useful advance for many classes of data. In addition, basic research over the next few years may reveal unexpected techniques and insights to close selected gaps.

Risk: IOT and “unpermissioned” data collection can circumvent technical user controls when third parties are collecting the data.

- Mitigation: Nevertheless: (1) even if user controls can be circumvented, that limitation doesn’t mean other user data shouldn’t be protected and user-controlled; (2) with the right ecosystem hacks “trustworthy” companies may offer those controls voluntarily; (3) this is an area where regulatory interventions may usefully complement technology; and finally (4) basic research may reveal useful new techniques and approaches (cf. the Stanford Secure Internet of Things center’s ideas around auditing IoT devices¹¹).

Risk: Derived data is particularly challenging to control from a user control perspective and from a differential privacy perspective.

- Mitigation: Attestation and data use controls could potentially be extended to expose data only to programs whose derivations are “acceptable”. In addition, provenance information could be included in the DUCs to retain controls over the use of derived data.

Risk: The data economy sounds like an idea that presumes Blockchain, not just confidential computing – is there a risk-increasing dependency?

- Mitigation: Our view is that Blockchain could be one of several mechanisms for exchanging value, managing identity, etc. but need not be an assumed underpinning of any solutions. Not every data economy application needs strict consistency or immutable ledgers, for example. Even so, perhaps the machinery of digital commerce (NFTs) could be part of a data economy solution.

Risk: Unlike pure algorithmic cryptographic approaches (secure multiparty computation, etc.) confidential computing puts the hardware vendor in the chain of trust. Some may argue that the community should focus on algorithmic approaches to data exchange instead.

- Mitigation: Trust in the processor company is a form of fate sharing. If one can’t trust the processor, all is lost anyway. Moreover, even if cryptographic approaches are used, the input data ought to be controlled with confidential computing and DUCs. Our view

¹¹ <https://dl.acm.org/doi/10.1145/3081333.3081342>



is that there are many techniques that can be added to confidential computing as a form of defense in depth, although more research is needed into the specifics.

Risk: The community has discovered vulnerabilities in certain TEEs and it will take time to deploy successor technologies. Moreover, new vulnerabilities are likely to emerge.

- Mitigation: Our view is that there is enough momentum around TEEs that the vendors are likely to improve over time with widespread adoption. Meanwhile even vulnerable TEEs are hard to exploit, so there is significant CC value-added even in the presence of exploits.

Risk: Industry lacks sufficient support for a common data economy architecture (common infrastructure APIs, interfaces, etc).

- Mitigation: Addressing this risk would be the point of a comprehensive public-private initiative. Our view is that we need to find an “ecosystem hack” to bootstrap this industry. We also believe it would be valuable for NIST to do for the data economy what it has done for the concept of “Zero Trust”.

Risk: Today’s cryptography may be broken by Quantum techniques and/or flawed implementations

- Mitigation: In addition to the use of post-Quantum techniques where feasible, we advocate crypto agility at the enterprise scale to allow for comprehensive system updates as flaws are found and to assist with mapping appropriate ciphers to data based on data classification – VMware is doing research in this space¹² and recently demonstrated quantum safe crypto agility at VMworld 2021¹³.

9. Existing barriers, not covered above, to PETs adoption

As mentioned above, Prof. Seny Kamara’s thoughtful keynote “Crypto for the People”¹⁴ articulated the need for research (and approaches to research) that would better address the needs of marginalized communities. It would be desirable if data use controls and confidential computing could be leveraged by marginalized communities to provide new opportunities for user control, provenance, access transparency, and forensics related data misuse. However, research is needed to understand these opportunities.

10. Other information that is relevant to the adoption of PETs.

To guide the development of a data economy, it may be helpful to develop a maturity model for security and privacy in the data economy. A potential draft scaffolding is as follows. Many industry players are currently working to achieve “level 1”, and there are selected instances of “level 2”.

¹² <https://research.vmware.com/projects/cryptographic-agility>

¹³ <https://octo.vmware.com/its-time-for-crypto-agility/>

¹⁴ <https://www.youtube.com/watch?v=Ygq9ci0GFhA>

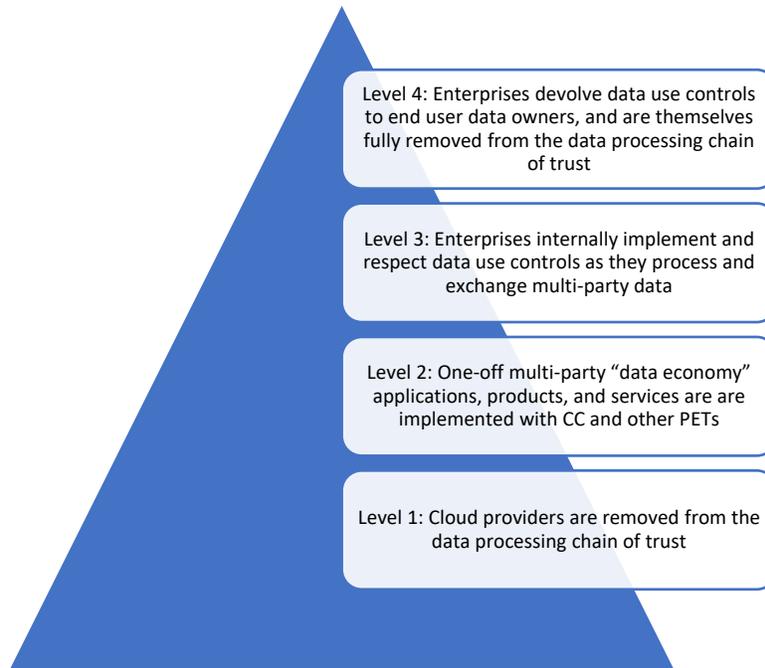


Figure 2 A maturity model of privacy controls in the data economy

Conclusion

Thank you again for this opportunity to provide comments on PETs as the Office of Science and Technology Policy embarks on the development of a national strategy on privacy-preserving data sharing and analytics. VMware looks forward to continuing the conversation and contributing to this important effort.

Sincerely,

Chris Ramming
Senior Director, Research and Innovation

VMware Inc.
3401 Hillview Avenue
Palo Alto, CA 94304

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Wehbe, Joseph

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

July 8, 2022

Office of Science and Technology Policy (OSTP)

on behalf of the Fast Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics of the Subcommittee on Networking and Information Technology Research and Development (NITRD) of the National Science and Technology Council, the National Artificial Intelligence Initiative Office, and the NITRD National Coordination Office

Re: Request for Information on Advancing Privacy-Enhancing Technologies
(Document Number: 2022-12432)

Submitted by:

Joseph Wehbe



“World Economic Forum

Recognized Artificial Intelligence
& Entrepreneurship Expert”

AI Ecosystem Builder

Joseph Wehbe is an American artificial intelligence ecosystem builder. Led the #1 winning team of a Massachusetts Institute of Technology (MIT) Challenge (knowledge-economy) in 2020. He received an AI master’s degree recognized by the leading research institute in Canada in which Dr. Geoffrey Hinton (the Godfather of AI) is the Chief Scientific Advisor. Joseph is also an ambassador for Stanford Women in Data Science.

Dear Sir/Madam,

Effectively developing a national strategy on privacy-preserving data sharing and analytics, and associated policy initiatives requires:

1. Operationalizing the strategy at a Federal, State, & Local Government using a software / infrastructure layer.
2. Building a pipeline of talent & projects as a Government to Grassroots AI value network to execute on this strategy.
3. Redesigning the entry margin for verified stakeholders, marginalized and underrepresented groups along with the non-consumers to have a way to participate.

Context for my response:

According to the AI Index Report 2022, the US ranks 1st globally in Talent, Research, Development, and Commercial. (see graphic below)

While we rank 35th in operating environment, and 17th in government strategy. Our low ranking in these 2 areas are the basis of all my feedback. This translates into the USA having the best talent, entrepreneurs, innovators, and companies in the world, but the strategies do not effectively impact the local government level. Why? Because the innovation and artificial intelligence ecosystems in the US are built around the greatest learning institutions in the world, who carry the responsibility. They don't revolve around the people. In AI development, while technology and algorithms have been commoditized, skilled workers that are able to create solutions to AI problems are the most important factor. There is a demand for a whole generation of workers with capacities in artificial intelligence. This will be the generation of talent that will support national interest in aerospace, defense, education, housing, transportation, public safety, supply chain, manufacturing, and many other industries critical to the American homeland. We can improve the 35th and 17th running using a software and infrastructure layer.

USA Global Index Ranking

*source:
<https://www.tortoisemedia.com/intelligence/global-ai/>

Tortoise

Intelligence | Global AI Index



United States Of America

1	Talent
4	Infrastructure
35	Operating Environment
1	Research
1	Development
17	Government Strategy
1	Commercial
1	Total Rank

My Responses to Question 9 & 10 combined below:

9

→ **[Response by JW]** Existing barriers, not covered above, to PETs adoption: Information about technical, sociotechnical, usability, and socioeconomic barriers that have inhibited wider adoption of PETs, such as a lack of public trust. This includes recommendations on how such barriers could be overcome. Responses that focus on increasing equity for underserved or marginalized groups are especially welcome.

→ **[Response by JW]** *Other information that is relevant to the adoption of PETs: Information that is relevant to the adoption of PETs that does not fit into any of the topics enumerated above.*

Based on my proprietary frameworks, in building artificial intelligence ecosystems there are 8 stakeholders that effectively connect as a value network. While government policy and advocates will call to increase diversity of talent by lowering the barriers of participation for all regardless of organizational affiliation, this is a recipe for disaster.

If this happens, it means we are removing educational merit if we want security, and accountability. My frameworks redesign the entry margin / the entry barrier to Artificial intelligence (and other technologies) not lower the barrier. The existing barrier to PETs adoption can be solved by redesigning the entry barrier for each stakeholder.

In my experience as a World Economic Forum artificial intelligence and entrepreneurship expert, I believe there's a demand for a whole generation of workers skilled in technology. These technologists must emerge from innovation ecosystems. There are systematic, structural and institutional barriers that many times and almost always limit opportunities that are also applicable to PETs. Once we identify the attributes of each innovation ecosystem, we can use artificial intelligence to identify the specific barriers affecting PETs.

We can't generalize or compartmentalize the barriers. Successful innovation ecosystems that we know of in Boston, Silicon Valley, and Seattle all have educational merit. Why? Because they are anchored by the most successful entrepreneurs and ventures in the world.

Innovators from underrepresented backgrounds and underserved communities do not have a pathway to achieve their innovation goals. Capital is not the barrier in these ecosystems. It's the lack of intellectual infrastructure in the region that's the main barrier. Using advanced technology and artificial intelligence, we can identify the attributes of each community. Generalizing the results is doing a dis-service to the community. And relying on human knowledge alone does not do the work / results justice.

Using a software layer, the US Federal government must identify the stakeholders who are eligible to participate in the innovation ecosystem. Think of those "eligible" as the "total addressable market." From this pool of eligible stakeholders, we identify those who have shown a willingness to become entrepreneurs, innovators, or technologists using various artificial intelligence methods that we can identify. To provide context, a Harvard Business School professor's definition of entrepreneurship is "the pursuit of opportunity beyond resources controlled." In this case, my definition of underrepresented, underserved, or marginalized is an individual or group who has knowledge and education of the said technology or innovation but lacks the intellectual infrastructure to realize their goals. The education component is critical because everything we do, must have educational merit. If our targets are truly innovators/technologists, these stakeholders must have educational merit. We are not expecting them to have all the knowledge, but should have access to the intellectual infrastructure. Using artificial intelligence, we can identify those most likely to succeed innovators/technologists and unlock their potential to succeed. Unfortunately, using human knowledge alone to realize this goal is difficult.

Entrepreneurs/technologists from marginalized groups and underserved communities all exist / live in communities with community colleges, technical schools, vocation schools, colleges, universities, and high schools. Using technology and a software layer, we can leverage these institutions to identify the specific barriers. But we must “redesign the entry barrier” so that these stakeholders (innovators or specific user type) understand “how” to achieve their goals, otherwise, they won’t have the willingness to share and engage. Again, we can’t “lower” the barrier to engage, because that won’t have educational merit, we must redesign the barrier.

I will reinforce again that capital is NOT the barrier for marginalized or underserved innovators to achieve their goals. If we throw capital at the problem, but the intellectual infrastructure does not exist, then we don’t achieve our goal. Let us use technology and artificial intelligence to solve this problem, redesign the entry barrier for stakeholders, and build a pipeline of innovators so they are no more underrepresented, underserved, or marginalized. Our goal is to have upward mobility for each of these stakeholders so that they are no more identified as such.

Highlights of AI expertise I offer:

- The benefits of innovation and artificial intelligence ecosystems are distributed unevenly across the US & don’t exist in the heartland. I can fix that.
- I can provide expertise to dismantle the structural, institutional, and systematic barriers that limit opportunities for stakeholders and bring educational merit to the technology workforce.
- Redesign entry margin for stakeholders so the US can build a pipeline of innovation / artificial intelligence talent so we are all known as American innovators.

Yours faithfully,

Joseph Wehbe
“Artificial Intelligence Ecosystem Builder”

**“The recipe is straightforward,
let us invest in AI Education, and
AI Research & Development.”-JW**



World Economic Forum
Recognized Artificial Intelligence
& Entrepreneurship Expert

AI Ecosystem Builder

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Western Governors' Association

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

June 23, 2022

Dr. Alondra Nelson
Director
Office Of Science and Technology Policy
The White House
1600 Pennsylvania Avenue NW
Washington, DC 20500

RE: Request for Information on Advancing Privacy-Enhancing Technologies (87 FR 35250)

Dear Director Nelson:

In response to the Office of Science and Technology Policy's (OSTP) Request for Information (RFI) on Advancing Privacy-Enhancing Technologies (PETs) ([87 FR 35250](#)), attached please find the following Western Governors' Association (WGA) policy resolutions:

- 2021-02, Utilizing State Data in Federal Decision Making;
- 2021-01, Strengthening the State-Federal Relationship; and
- 2020-07, Rural Development.

In these policy resolutions, Western Governors address the following Topics included in the RFI:

- 3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs;
- 5. Specific laws that could be used, modified, or introduced to advance PETs;
- 6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs;
- 7. Risks related to the adoption of PETs; and
- 9. Existing barriers, not covered above, to the adoption of PETs.

Specifically, in WGA Policy Resolution [2021-02, Utilizing State Data in Federal Decision Making](#), Western Governors urge the federal government to consult with states and state agencies as partners – rather than ordinary stakeholders – in the collection, stewardship, analysis, and use of data to inform federal decision-making processes. Further, federal agencies should consult with states to develop policies that more effectively and consistently incorporate state data into federal decision making. The resolution also recognizes: the importance of state privacy and data stewardship laws; the need for federal consultation with states on the development of data and data sharing policies; and the advisability of uniformity in data standards and stewardship policies across all levels of government.

I appreciate your consideration of these comments as they articulate Western Governors' bipartisan policy positions and recommendations on these important issues.

Western Governors submit these remarks through the public notice and comment process for administrative recordkeeping purposes. The Governors, however, maintain that this process is an insufficient channel for state-federal communication on federal actions that may affect state

Dr. Alondra Nelson

June 23, 2022

Page 2

authority or administrative activity. Western Governors strongly urge you to engage in meaningful, substantive, and ongoing consultation with states in advance of prospective decisions or related public processes. Such consultation will result in more effective, efficient, and resilient federal policy benefiting the Western Governors', and the OSTP's, constituents.

Please contact me if you have any questions or require further information. In the meantime, with warm regards and best wishes, I am

Respectfully,

James D. Ogsbury
Executive Director

Attachments



Policy Resolution 2021-02

Utilizing State Data in Federal Decision Making

A. BACKGROUND

1. State and federal agencies increasingly rely on quantitative and qualitative data to inform evidence-based policymaking, improve service delivery, more effectively manage resources, improve regulatory enforcement, and more accurately measure program performance and effectiveness.
2. States serve a critical function as a primary sources and stewards of economic, social, geospatial, scientific, technical, and other datasets that support a wide array of federal agencies and programs. State agencies often have the best available science, expertise, and other institutional data resources for purposes of federal decision-making processes.
3. States are both sovereignties and the primary administrators of numerous federal administrative and regulatory programs under a system of cooperative federalism, which distinguishes them from other non-federal sources of data.
4. State agencies also rely on timely access to current and accurate federal datasets to inform their own decision-making processes, develop more effective policy, improve service delivery and public communication, and to administer federally-delegated administrative programs.
5. In addition to federal laws and regulations, state agencies operate under their own privacy and data stewardship laws, regulations, and policies that protect personal and confidential information from public disclosure or other inappropriate use or disclosure. These protections help establish public trust that ultimately improves government effectiveness.
6. Public access to datasets that serve as the basis for federal agency actions promotes transparency and accountability in the decision-making process. Nevertheless, blanket requirements to make publicly available all data considered by federal agencies – particularly if this data consists of raw data provided by states – may infringe upon states' statutory imperatives to protect personally identifiable and otherwise sensitive information. It may also infringe upon fundamental privacy and data stewardship principles like purpose specification and data minimization. Even where there is no state legal barrier to disclosure of raw data, state agencies may maintain significant reservations about the public release of raw data.
7. The 2019 Federal Data Strategy directs federal agencies to “[e]ffectively, routinely, transparently, and appropriately use data in policy, planning, and operations to guide decision-making [and] share the data and analyses behind those decisions.” Additionally, agencies are directed to “[f]acilitate data sharing between state, local, and tribal governments and the Federal Government, where relevant and appropriate and with

proper protections, particularly for programs that are federally funded and locally administered, to enable richer analyses for more informed decision-making.”

8. Improvements in intergovernmental data sharing, stewardship, integration, protection, and utilization will require robust federal investments in a modern data infrastructure, technology, and training.

B. GOVERNORS' POLICY STATEMENT

1. State data serves a critical role in the successful implementation of a variety of federal programs and in federal agencies' fulfillment of their statutory missions and directives.
2. Subject to state laws and other requirements for data protection and transparency, federal agencies should be required to incorporate state and local data and expertise into their analysis and decision-making processes. This data should include geospatial, scientific, technical, economic, social, and other information relevant to issues the agency is trying to address.
3. Congress and the Executive Branch should look to states and state agencies as partners – rather than ordinary stakeholders – in the collection, stewardship, analysis, and use of data to inform federal decision-making processes. Federal agencies should recognize the existence and limitations of state privacy and data stewardship laws, regulations, and policies and work with states to develop strategies that encourage effective state-federal data sharing while appropriately protecting data according to state law.
4. State data – particularly non-aggregated raw data – is subject to differing levels of protection under various state laws, regulations, and policies. Western Governors encourage Congress and federal agencies to recognize the limitations on complete transparency of state data in federal decision making and to work with states to identify ways in which protected data can inform federal decision-making processes without conflicting with applicable state laws, regulations, or policies.
5. Federal agencies should consult with states – on a government-to-government basis – in the development and implementation of policies, programs, and strategies to more effectively and consistently incorporate state data into federal decision making, including implementation of applicable federal statutes and programs, as well as the Federal Data Strategy and development of annual Federal Data Strategy Action Plans.
6. Federal agencies should also consult with states to ensure that state and local partners have access to timely and reliable federal datasets for purposes of informing state and local decision-making processes.
7. Congress and the Executive Branch should support, and work with state toward, the modernization of our nation's data infrastructure and intergovernmental data-sharing and analysis capabilities. Data infrastructure should be based on best practices for data stewardship and must properly protect personal and confidential information in accordance with state and federal law. Federal agencies should consult with states to develop guidelines for intergovernmental data-sharing agreements and other protocols that include commitments to fundamental privacy and data stewardship principles like purpose specification and data minimization.

8. Federal agencies should work with state and local partners to develop uniform data standards, where appropriate, to maximize data quality and facilitate intergovernmental data use, access, sharing, and interoperability.
9. Western Governors support congressional efforts to broaden statutory exemptions under the Freedom of Information Act to protect personally identifiable and sensitive state-shared data from disclosure.
10. Western Governors urge the Executive Branch to develop uniform privacy and data stewardship policies based on best practices and uniform interpretations of federal privacy and data stewardship laws, regulations, policies, and other directives applicable to data received from states, as well as other non-federal sources.

C. GOVERNORS' MANAGEMENT DIRECTIVE

1. The Governors direct WGA staff to work with congressional committees of jurisdiction, the Executive Branch, and other entities, where appropriate, to achieve the objectives of this resolution.
2. Furthermore, the Governors direct WGA staff to consult with the Staff Advisory Council regarding its efforts to realize the objectives of this resolution and to keep the Governors apprised of its progress in this regard.

This resolution will expire in December 2023. Western Governors enact new policy resolutions and amend existing resolutions on a semiannual basis. Please consult <http://www.westgov.org/resolutions> for the most current copy of a resolution and a list of all current WGA policy resolutions.



Policy Resolution 2021-01

Strengthening the State-Federal Relationship

A. **BACKGROUND**

1. Western Governors are proud of their unique role in governing and serving the citizens of this great nation. As the chief elected officials of sovereign states, they bear enormous responsibility and have tremendous opportunity. Moreover, the faithful discharge of their obligations is central to the success of the Great American Experiment.
2. It was the states that confederated to form a more perfect union by creating a national government with specific responsibilities for common interests. In this union, the states retained their sovereignty and much of their authority.¹
3. Under the American version of federalism, the powers of the federal government are narrow, enumerated and defined. The powers of the states, on the other hand, are vast and indefinite and encompass all powers of governance not specifically bestowed to the federal government by the U.S. Constitution. This principle is memorialized in the Tenth Amendment, which states: "The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people."
4. This reservation of power to the states respects the differences between regions and peoples, recognizes a right to self-determination at a local level, and provides for flexible, tailored solutions to policy challenges. It also requires the federal government to engage with states – our nation's dynamic laboratories of democracy – on a government-to-government basis befitting their co-sovereign status.
5. In addition to states' reserved sovereign authorities, Congress has recognized state authority in federal statute by: (1) directing the federal government to defer to state authority, including such authority over land and water use, education, domestic relations, criminal law, property law, local government, taxation, and fish and wildlife; and (2) delegating federal authority to states, including the regulation of water quality, air quality, and solid and hazardous waste.
6. [Executive Order 13132](#), Federalism, reinforces these constitutional, statutory, and judicial principles and directs federal agencies to have an accountable process to ensure meaningful and timely input from state officials in developing policies with federalism implications.
7. The relationship between state and federal authority is complex and multi-dimensional. There are various contexts in which these authorities manifest and intersect:

¹ The U.S. Supreme Court has confirmed that, "[d]ual sovereignty is a defining feature of our Nation's constitutional blueprint" and "States entered the Union with their sovereignty intact." *See, e.g., Sossamon v. Texas*, 563 U.S. 277, 283 (2011).

- a) **State Primacy** – All powers not specifically delegated to the federal government in the Constitution. In the absence of Constitutional delegation of authority to the federal government, state authority should be presumed sovereign. *Examples: groundwater, wildlife management (outside of the Endangered Species Act), natural resources management, electric transmission siting.*
 - b) **Shared State-Federal Authority** – Fact patterns in which federal authority and state primacy intersect. *Examples: wild horses and burros on federal lands, interstate water compacts.*
 - c) **Federal Authority Delegated to States** – Federal authority that Congress has delegated to states by statute. Many such statutes require federal agencies to set federal standards (and ensure those standards are met) but authorize states to implement those standards. *Examples: water and air quality, solid and hazardous waste.*
 - d) **Federal Statutory or Other Obligations to States** – Where the federal government has a statutory, historical, or moral obligation to states. *Examples: Payments in Lieu of Taxes; Secure Rural Schools Act; shared mineral royalties; agreements to clean up radioactive waste that was generated by federal nuclear weapons production.*
 - e) **Exclusive Federal Authority** – Powers enumerated in the Constitution as exclusive powers of the federal government. In areas of exclusive federal authority, state law can be preempted if Congress clearly and unambiguously articulates an intent to occupy a given field or to the extent it conflicts with state law. *Examples: national defense, production of money.*
8. In contravention of the Founders’ design, the balance of power has shifted toward the federal government and away from the states. Increasingly prescriptive regulations tie the hands of states and local governments, dampen innovation, and impair on-the-ground problem-solving. Failures of the federal government to consult with states reflect insufficient appreciation for local knowledge, preferences, and competencies. In many cases, these federal actions encroach on state legal prerogatives, neglect state expertise, and/or infringe on state authority.
 9. The federal government often requires states to execute policy initiatives without providing the funding necessary for their implementation. State governments cannot function as full partners if the federal government requires them to devote their limited resources to compliance with unfunded federal mandates.
 10. State authority and autonomy is also eroded when prescribed federal policies become effectively mandatory through the contingency of federal funding streams that states depend on to deliver critical services.
 11. Too often, federal agencies: solicit input from states after a decision is already made or a public process is started; ask states to provide feedback on a proposed action without providing details or documents regarding what the agency is proposing; or do not respond to state input or incorporate feedback from states into their decisions. This does not afford states with the respect and communication required by law, and states currently have no

recourse for an agency's failure to consult except for litigation on the merits of a federal decision.

12. Congress and Executive Order 13132 currently require federal agencies to document the effects of their actions on states in certain circumstances. In practice, federal agencies rarely prepare these prescribed federalism assessments or statements. Even when federal agencies prepare such documents, they are not ordinarily informed by input from affected states. In addition, these documentation requirements only apply at the end of the rulemaking process and cannot substitute for early and meaningful consultation with states.
13. Federal agencies have suggested to states that there are legal or other barriers to state consultation, such as: federal agency policies restricting ex parte communications; concerns about the applicability of Federal Advisory Committee Act (FACA) procedures to meetings between state and federal officials; and issues with sharing information that would otherwise be exempt from disclosure under the Freedom of Information Act (FOIA).
14. Federal agencies do not adequately incorporate state data and expertise into their decisions. This can result in duplication, inefficiency, and federal decisions that do not reflect on-the-ground conditions. Consideration and incorporation of state, tribal, and local data and analysis will result in federal actions that are better-informed, more effectively coordinated among all levels of government, and tailored to the communities they affect.
15. Many of these issues stem from a profound misunderstanding throughout the federal government regarding the role and legal status of states. Over the past several years, Western Governors have worked to improve the federal government's understanding of state sovereignty, authority, and state-federal consultation; meaningful structural change, however, has yet to occur.

B. GOVERNORS' POLICY STATEMENT

1. A good faith partnership between states and the federal government will result in more efficient, economic, effective, and durable policy, benefiting the Governors' and the federal government's shared constituents and resulting in a nation that is stronger, more resilient, and more united.
2. Improving state-federal communication and coordination is a goal that transcends party lines, and it is among the Governors' highest priorities. The Governors urge Congress and the Executive Branch to make fundamental changes to realign and improve the state-federal paradigm.

State Sovereignty and Authority

3. States are co-sovereigns with the federal government pursuant to the Tenth Amendment of the U.S. Constitution and other federal law. Congress and federal agencies must recognize state sovereignty and must not conflate states with other entities or units of government. States should not be treated as stakeholders or members of the public.
4. State authority is presumed sovereign in the absence of Constitutional delegation of authority to the federal government.

- a) Federal legislative and regulatory actions should be limited to issues of national significance or scope, pursuant to federal constitutional authority. Preemption of state laws should be limited to instances of necessity.
 - b) Where Congress preempts state law (acting pursuant to federal constitutional authority), federal law should accommodate state laws, regulations, and policies before its enactment and permit states that have developed alternate standards to continue to enforce and adhere to them.
 - c) Federal agencies should construe federal law to preempt state law only when a statute contains an express preemption provision or there is some other compelling evidence that Congress intended to preempt state law.
5. Congress and federal agencies should respect the authority of states to determine the allocation of state administrative and financial responsibilities in accordance with state constitutions and statutes. It should further:
- a) Ensure that federal government monitoring is outcome-oriented;
 - b) Minimize federal reporting requirements; and
 - c) Refrain from dictating state or local government organization.
6. When a state is meeting the requirements of a delegated program, the role of a federal agency should be limited to the provision of funding, technical assistance and research support. States should have the maximum discretion to develop implementation and enforcement approaches within their jurisdiction without federal intervention. Federal agencies should recognize and credit states' proactive actions.
7. Congress and federal agencies should avoid imposing unfunded federal mandates on states. In addition:
- a) Federal assistance funds, including funds that will be passed through to local governments, should flow through states according to state laws and procedures;
 - b) States should have the flexibility to transfer a limited amount of funds from one grant program to another and to coordinate the administration of related grants;
 - c) Federal funds should provide maximum state flexibility without specific set-asides; and
 - d) Governors should have the authority to require coordination among state executive branch agencies, or between levels or units of government, as a condition of the allocation or pass-through of funds.
8. Congress and the Executive Branch should create or re-establish entities to discuss and act on federalism issues, such as the Speaker's Task Force on Intergovernmental Affairs, the U.S. Advisory Committee on Intergovernmental Relations, the Subcommittee on Intergovernmental Affairs, or a federalism office within the White House. These entities should have the ability and resources to make recommendations to improve the state-

federal relationship and include states in their membership or actively involve states in their discussions.

State-Federal Consultation

9. Federal agencies must engage in consultation with states on a government-to-government basis in accordance with states' legal status. Congress should clarify and promote the need for state-federal consultation.
10. Improving state-federal consultation will result in more effective, efficient, and long-lasting federal policy for the following reasons:
 - a) Governors have specialized knowledge of their states' environments, resources, laws, cultures, and economies that is essential to informed federal decision-making;
 - b) Federal agencies can reduce duplication through the use and incorporation of state expertise, data and documentation;
 - c) Authentic communication and information exchange will help federal agencies determine whether an issue is best addressed at the federal level; and
 - d) Through meaningful dialogues with affected states, federal agencies can also avoid unintended consequences and address or resolve state concerns.
11. Each Executive department and agency should have a clear and accountable process to provide each state – through its Governor or their designees – with early, meaningful, substantive, and ongoing consultation in the development of federal policies that affect states. The extent of the consultation process should be determined by engaging with affected states. At a minimum, this process must involve:
 - a) Conducting consultation through federal representatives who can speak or act on behalf of an agency;
 - b) Inviting states to provide input outside of a public process and before proposals are finalized;
 - c) Enabling states to engage with federal agencies on an ongoing basis to seek refinements to proposed federal actions prior to finalization;
 - d) Providing robust information and documents (including non-final, non-public, draft, and supporting documents) about potential federal actions, including proposed rules, to Governors or their designees;
 - e) Addressing or resolving, where possible, state issues, concerns, or other input unless precluded by law;
 - f) Documenting how state concerns were resolved or why they were unable to be resolved in final decisions; and

- g) Making reasonable efforts to achieve consistency and avoid conflicts between federal and state objectives, plans, policies, and programs.
12. Governors affirm their reciprocal role in advancing a clear, predictable, timely, and accountable consultation process. Governors or their designees must continue to provide clear expectations for the appropriate scope and scale of consultation and must work with federal agencies to make consultation processes as efficient as practicable. As chief executives, Governors must also ensure the views of the state are clearly and consistently conveyed throughout the consultation process by prioritizing significant issues and resolving competing viewpoints across state government.
 13. In many cases, federal agencies are required – whether by statute, executive order, regulation, policy, or other mandate – to consult, cooperate, and coordinate with states before taking action. However, due to states’ unique legal status, the need for federal-state engagement is not limited to express directives and should extend to any federal actions that may have direct effects on states, on the relationship between the federal government and states, or on the distribution of power or responsibilities among the various levels of government. Federal agencies should consult with states regarding what types of agency actions typically affect states and the extent of consultation required for these types of actions.
 - a) These actions include the implementation of federal statutes and the development, prioritization, and implementation of agency policies, rules, programs, reviews (e.g., Governor’s Consistency Reviews), plans (e.g., resource management plans), budget proposals and processes, strategic planning efforts (e.g., reorganization), and federal litigation or adjudication that affects states.
 - b) When a federal agency proposes to enter into any agreement or settlement that affects states, the agency should provide all affected Governors or their designees with notice of the proposal and consult with, and seek the concurrence of, Governors or their designees who respond to the notice.
 14. Congress and the Executive Branch should require federal agencies to promulgate regulations in consultation with Governors, setting forth their procedures to ensure meaningful, substantive consultation with states on federal actions that affect states. This direction should also clarify that, for rulemakings affecting states:
 - a) An agency’s satisfaction of rulemaking requirements under the Administrative Procedure Act (including the solicitation of public comments) does not satisfy an agency’s obligation to consult with states; and
 - b) Consultation should occur before publication of a notice of proposed rulemaking or before an advanced notice of proposed rulemaking is submitted to the Office of Management and Budget (OMB).
 15. Congress and the Executive Branch should consider the following additional accountability measures:

- a) Requiring the designation of a federalism official with the responsibility for implementing state-federal consultation and publish this official's name, title, and contact information on the agency's website;
 - b) Requiring OMB to regularly submit a report to Congress and Governors on state-federal consultation and implementation of agency consultation rules;
 - c) Requiring federal agencies to provide a summary of their efforts to consult with states, including a discussion of state input and how that input was considered or addressed, in any proposed and final rules;
 - d) Creating a process where Governors can notify OMB of an agency's failure to consult or comply with their consultation procedures; and
 - e) Providing an opportunity for Governors or their designees to seek judicial review of an agency's failure to consult.
16. Congress and the Executive Branch could make federalism reviews more effective by:
- a) Working with Governors to develop specific criteria and consultation processes for initiating and performing these reviews.
 - b) Providing Governors with an opportunity to comment on federalism assessments before any covered federal action is submitted to OMB for approval.
17. Congress and federal agencies should take the following actions to clarify that *ex parte* policies, FACA, and FOIA are not barriers to consultation:
- a) Federal agencies should (and Congress should require them to) clearly identify and provide rationale for any perceived barriers to consultation;
 - b) Federal agencies should clarify that consultation with state officials does not qualify as *ex parte* communications and that *ex parte* communications are not prohibited at any point during an informal rulemaking process;
 - c) Congress should clarify that meetings held exclusively between federal personnel and state elected officials or their designees acting in their official capacities or in areas of shared responsibilities or administration (and not for the purpose of obtaining collective advice) do not qualify as requiring compliance with FACA procedures; and
 - d) Congress should clarify that FOIA's exemptions apply to federal records shared or exchanged with states (as if those records were shared, exchanged, or created solely within the federal government) and create a statutory exemption to FOIA disclosure for state records in instances where publication of state records provided to federal agencies would violate existing state law.

State Data and Expertise

18. Federal agencies should utilize state data, expertise, and science in the development of federal actions that affect states.
19. Congress and the Executive Branch should, subject to existing state requirements for data protection and transparency, require agencies to incorporate state and local data and expertise into their decisions. This data should include scientific, technical, economic, social, and other information on the issue the agency is trying to address.
20. States merit greater representation on all relevant committees and panels advising federal agencies on scientific, technological, social, and economic issues that inform federal regulatory processes.

C. GOVERNORS' MANAGEMENT DIRECTIVE

1. The Governors direct WGA staff to work with congressional committees of jurisdiction, the Executive Branch, and other entities, where appropriate, to achieve the objectives of this resolution.
2. Furthermore, the Governors direct WGA staff to consult with the Staff Advisory Council regarding its efforts to realize the objectives of this resolution and to keep the Governors apprised of its progress in this regard.

This resolution will expire in December 2023. Western Governors enact new policy resolutions and amend existing resolutions on a semiannual basis. Please consult westgov.org/resolutions for the most current copy of a resolution and a list of all current WGA policy resolutions.



Policy Resolution 2020-07 Rural Development

A. BACKGROUND

1. Vibrant and prosperous rural communities are essential components of western states and the nation. Rural communities in the West grow and supply food, steward natural resources, contribute disproportionately to the armed services, and are critical to state economies.
2. Rural communities in the West are richly diverse and face varying threats and opportunities. They do, however, share common characteristics, such as low population density and distance from urban centers, that create challenges for economic development.
3. Nationally, the rural population is increasing slowly after a period of decline from 2010-2017, although there is great variation in demographic trends at the county level. Many rural western counties are experiencing population growth due to net migration of retirees as well as people seeking quality of life and amenities.
4. Most rural communities have a higher proportion of older residents than urban and suburban communities. Rural communities are preparing for an anticipated wave of retirements, leading to a reduction in skilled workforce and potential closure of local businesses.
5. Western states have many of the highest per-capita veteran populations in the nation. It is estimated that approximately 25 percent of all veterans live in rural areas. These veterans have lower rates of employment compared to veterans living in urban areas and to their non-veteran rural colleagues.
6. Most rural communities have experienced slower economic and job growth than their urban counterparts and have not yet fully recovered from the 2008 recession. Many rural economies rely heavily on a few industries. This makes these communities more sensitive to trends affecting those industries and can make it more difficult to recover from disruptions. Rural communities have also been significantly affected by the COVID-19 public health threat and will suffer distress from any associated economic downturn.
7. Federal programs for rural development are spread across multiple agencies. Some agencies are responsible for rural infrastructure investments while others focus on economic development. The multiplicity of administering entities and the lack of consistency across agencies: renders it difficult for rural stakeholders to navigate federal programs; compounds time and expense needed to apply for funding opportunities; and creates inefficiencies in the distribution of resources.
8. The vast majority of federal loan and grant programs for rural communities are targeted to physical infrastructure. Little funding exists to support development of local capacity, including strengthening community organizations, nonprofit entities, and other groups serving rural communities.

9. By offering opportunity, connectivity, and quality of life, rural communities can thrive and prosper while improving economic stability and protecting the rural character and natural resources that draw people to these areas.
10. Small businesses, including farming and ranching, are the foundation of rural economies, generating tax revenue, creating jobs, providing essential goods and services, and contributing to the culture and character of small towns. Building a successful business in a rural community is particularly challenging due to limited access to capital, available workforce and smaller customer bases. Many successful rural businesses operate on thin margins and if they close, are extremely difficult to replace.
11. High-speed internet, commonly referred to as “broadband,”¹ is the critical infrastructure of the 21st century and a modern-day necessity for individuals, businesses, schools and government. Many rural western communities lack the business case for private broadband investment due to the high cost of infrastructure and the low number of customers in potential service areas. This has left many rural businesses and citizens at a competitive disadvantage compared to urban and suburban areas with robust broadband access.
12. Transportation connectivity is critical to strengthening economies and improving quality of life. Air service is particularly important to connect remote western communities to urban hubs. Pilot shortages, infrastructure constraints, and airline consolidation have negatively impacted rural air connectivity. The airline industry has been significantly impacted by COVID-19 and rural service, which is typically the least profitable, has been reduced.

B. GOVERNORS' POLICY STATEMENT

1. Western Governors believe that strengthening social infrastructure in rural communities is the best strategy to ensure rural quality of life and prosperity. Congress and federal agencies should increase the proportion of rural economic development and infrastructure funding that goes toward capacity-building. Accordingly, Western Governors call for ample and consistent federal funding for institutions, training, and technical assistance. Robust social infrastructure is fundamental to economic and community development and maximizes the impact of state and federal resources.
2. Social infrastructure is especially critical during disasters or crises. The COVID-19 pandemic and associated economic crisis have illustrated the challenge of rapidly deploying resources to the most urgent needs. Western Governors are committed to strengthening the resilience of rural communities by helping to foster local leadership and strengthen networks and connections within and among rural communities across the West.
3. Western Governors believe that many federal programs for rural development and distressed communities include unintended barriers for rural individuals and entities that need assistance most. Western Governors urge federal agencies to work with states to: thoroughly evaluate program requirements; identify barriers for rural applicants; and revise onerous requirements in a manner that recognizes the limited resources and capacity of rural applicants. In particular, Western Governors are concerned by:

¹ The Federal Communications Commission defines fixed “broadband” as service offering minimum speeds of 25 Megabits per second (Mbps) down and 3 Mbps up.

- a. Scoring criteria that relate to numerical size and impact, such as the number of jobs created or the number of people served, which disadvantage small and isolated communities;
 - b. Requirements that applicants partner with other institutions like community colleges or foundations, which may not operate in the rural community seeking assistance;
 - c. Financial match or cash-on-hand requirements that rural organizations cannot meet; and
 - d. Overly complicated or technical applications that deter rural customers from applying.
4. Western Governors also urge federal agencies to use state data for eligibility determinations when requested by states. States often have more up-to-date and granular data for rural communities than federal sources.
5. Western Governors recognize and support efforts at the federal and state level to coordinate the deployment of resources, leverage funding, and create one-stop application processes for rural customers. Western Governors are interested in exploring strategies to expand those models to include more funders and further enhance coordination between agencies and between states and the federal government.
6. Western Governors believe that changes in our economy, labor force, and technological innovations require fundamental changes in economic development strategies. Western Governors promote rural development policies that focus on quality of life and the support of small businesses and entrepreneurs. This will develop rural communities that are attractive places to live and work while protecting their rural character, natural resource-based industries, and natural areas.
7. In the wake of disasters in rural communities, including the COVID-19 pandemic, providing small businesses with the tools and resources to survive disruption and prosper again is critical for economic recovery and social stability. The impacts of COVID-19 mitigation measures have illustrated the urgent need for working capital to stabilize rural businesses and ensure their continued viability during periods of prolonged disruption related to disasters and emergencies. The federal response to declared disasters must include sufficient and accessible business stabilization funds, including grants and long-term, forgivable loans. Congress should also examine how such funds are distributed by the Small Business Administration (SBA), U.S. Department of Agriculture Rural Development (USDA RD), and the Economic Development Administration (EDA) to ensure that adequate systems are in place to handle demand during widespread emergencies.
8. During COVID-19 response, many small lenders could not access federal funds to support their communities because they did not meet program thresholds or were unable to compete against large banks. Small lenders, including community development financial institutions (CDFIs), community banks, and credit unions, are often the only lender serving rural and tribal communities. In the ongoing response to the economic impacts of COVID-19, Western Governors encourage Congress to set aside emergency relief funds for small

lenders and consider program requirements that allow those entities to participate through a streamlined process to rapidly distribute resources while maintaining fiscal accountability.

9. Western Governors are eager to work with public universities, community colleges, and the business community to expand opportunities for young people to stay in their rural communities. There is a high demand for skilled workers in rural communities and states should work together on regional solutions that provide the appropriate training and skills for the jobs that are available in rural communities where possible. Western Governors are also committed to increasing employment among veterans and the disabled community in the rural West.
10. Western Governors encourage increased flexibility in the use of federal economic development resources (particularly EDA funds) to facilitate investments in quality of life and amenities in rural communities. Governors believe that metrics based solely on the absolute number of jobs created do not reflect the important economic benefits of investments in community assets that make rural communities attractive places to live. Nor do they account for the relative impact of job creation in less populated rural communities or areas with high unemployment or poverty rates.
11. Western Governors strongly support improving and increasing broadband connectivity in the rural West through significant federal investments in mapping, deployment, and adoption. The Governors have highlighted substantive policy recommendations in the Western Governors' Association policy resolution addressing broadband connectivity.
12. Western Governors have developed robust policies addressing the challenge of providing services and maintaining infrastructure essential to communities across the vast expanse of the rural West. These policies address broadband access, healthcare, surface transportation and water quality. Western Governors are committed to working with Congress and federal agencies to improve the efficacy of federal and state programs to support critical infrastructure in the rural West.
13. Western Governors recommend further changes in approach to supporting water and wastewater infrastructure in the West:
 - a. Funding for communities not served by water systems is critically needed. Western Governors recommend set-asides to develop innovative solutions for communities and tribes that cannot be served by traditional systems.
 - b. Western Governors are concerned about shortages of certified water system operators and request a coordinated effort to increase training opportunities through the U.S. Department of Agriculture, U.S. Environmental Protection Agency, U.S. Department of Labor and college and university programs that develop these skilled workers to ensure that existing water access in rural communities can be maintained.
14. Western Governors emphasize that air service is a necessity in many rural communities. As the airline industry recovers from the economic impacts of COVID-19, Western Governors urge Congress to consider measures to address challenges in rural air connectivity, including funding for infrastructure and service subsidies.

15. Western Governors support the use of cooperative business models to preserve rural businesses and fill community needs for childcare, home care, main street businesses, housing, and more. Western Governors recognize the need for substantial technical assistance and education in developing new cooperative businesses and support federal funding of such efforts.
16. Western Governors are concerned by food security challenges in rural communities. Rural grocery store closures jeopardize livability and community health. Western Governors are interested in exploring strategies to ensure rural food security by strengthening local agricultural economies and developing regional approaches to rural food supply chains.
17. Housing remains a critical challenge in rural communities across the West. Western Governors are committed to working together to share best practices and effective solutions for housing preservation and development in the rural West.
18. The Cooperative Extension System, which serves every county in western states, is an important asset for rural development. Western Governors believe that Cooperative Extension can play a more meaningful role in economic development efforts in distressed communities and support continued investment in the system as it responds to the changing needs of rural communities. Western Governors are committed to maximizing the efficacy of Cooperative Extension in their states.

C. GOVERNORS' MANAGEMENT DIRECTIVE

1. The Governors direct WGA staff to work with Congressional committees of jurisdiction, the Executive Branch, and other entities, where appropriate, to achieve the objectives of this resolution.
2. Furthermore, the Governors direct WGA staff to consult with the Staff Advisory Council regarding its efforts to realize the objectives of this resolution and to keep the Governors apprised of its progress in this regard.

Western Governors enact new policy resolutions and amend existing resolutions on a bi-annual basis. Please consult www.westgov.org/resolutions for the most current copy of a resolution and a list of all current WGA policy resolutions.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Work, Paul R.

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to the Office of Science and Technology Policy (OSTP) Request for Information (RFI) on Advancing Privacy-Enhancing Technologies (PETs)

From: Paul R. Work

Organization: Member of the Public

Date: 7 July 2022

Topics Addressed: 3. *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs;* 8. *Existing best practices that are helpful for PETs adoption;* 9. *Existing barriers, not covered above;* 10. *Other information that is relevant to the adoption of PETs*

Background: I am a retired person who spent almost 50 years working in the field of computer systems applied to everything from life and healthcare insurance, to toll road technologies, to logistics, to the defense sector, and other national security spaces. I spend the last 15+ years working closely with academia and small businesses on research and development efforts for organizations like the Defense Advanced Research Projects Agency (DARPA), Intelligence Advanced Research Projects Activity (IARPA), Office of Naval Research (ONR), Air Forces Research Laboratory (AFRL), and the Semiconductor Research Corporation (SRC), just to name a few. That said, while I have been a systems engineer and certified systems architect for the latter half of those 50 years, espousing holistic, multi-discipline, systems thinking and development, focused on “Chips to Data Centers, and all the Algorithms There On”, and while my credentials in PETs are minimal, I do have a general understanding of some of and insight into the techniques and technologies of PETs, such as Multi-Party Computation (MPC) and Differential Privacy (DP).

I want to thank the leaders and members of the Fast Track Action Committee (FTAC) on Advancing Privacy-Preserving Data Sharing and Analytics (APPDSA) for their Subgroup Roundtables held virtually back in early June. There were a good number of contributors on the Zoom calls and a great discussion with researchers and organizations involved in the development and fielding of PETs. This was very helpful in catching up with a sense of the state of the art and the state of the research in PETs and the thoughts and concerns that they shared with the implementation and fielding of PETs.

Response to Topic 3: *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs.*

As it may come out in the other responses below, my current interest in the development of PETs is focused on applications to folks like me, senior citizens on fixed incomes, and of an age where the majority of my peers, while users today of technologies like iPhones, iPads, Ring doorbells, Roku streaming devices, and Amazon Prime, were adults during the development of these tools. We have seen the maturation of these technologies, while our younger generations literally grew up with them. While some of us may have been involved in the “computer business”, and a group of us are skilled practitioners in the arts and science of cybersecurity and data privacy, I would venture that the majority are simply users of the technology without a deep understanding of the nuances of it. Back in 2019, the Consumer Technology Association (CTA) conducted a study on “Active Aging: Consumer Perceptions and

Response to the Office of Science and Technology Policy (OSTP) Request for Information (RFI) on Advancing Privacy-Enhancing Technologies (PETs)

Attitudes”¹ where they found that while “virtually all seniors want to live independently and stay home as they age,” seniors view “cost and concern over learning new technologies could be barriers to adoption.” The study also found that “Over two-thirds of both seniors and caretakers have concerns about privacy with health monitoring devices for seniors. With the plethora of news regarding cybersecurity breaches and how they affect nearly everyone, it’s easy to understand this fear.”

We depend on the good will and best efforts of researchers and practitioners who develop these sorts of technologies and tools. An example of this is the encouragement from the cybersecurity community to leverage the use of SSL (Secure Sockets Layer) encryption, and its more modern and secure replacement, TLS (Transport Layer Security) encryption to protect data sent over the Internet. For us that means simply use links which have the prefix: “https://” and depend upon our browsers to check for this and signal to use when we are not using it, to ensure our safety and security on the various sites we may go to.

As for the privacy of our data, the current approach seems to be the use of “pop-ups” which ask us (for every site!) what sort of “tracking” we want to allow (if they are nice enough to detail any of it.) Why can’t we have browser and app technologies that are (1) required to honor already defined “answers” to these tracking questions and (2) if we do want to “share” our data with them, do so using PETs to ensure that our data is still protected beyond simply sharing in the “raw”, even if it is encrypted during transmission from our devices? It appears that most websites and apps make use of upward to dozens of ad tracking tools, understandably to ensure they have a way to pay for their “free” services. If Amazon or CVS are already getting our monies for products or services we procure via their sites and apps, why do we have to allow third parties to have access to our data (including our IP address) without our permission.

One of the most egregious offenders are the companies who are willing (by asking us) to put us on their emailing list, then send us an email with tantalizing headlines along with a sentence or two to entice us to follow a link and then when we do follow the link, we are “given to opportunity” to pay to read the rest, behind a “pay wall.” All of this because we are using their email service for free. So, we have freely given us our email address, which they will sell to others, who will then gather more of our data and preferences when we hit their websites. We need to have basic infrastructure in the technology stack to ensure that only the details we truly wish to share are done so, and without having to make the decision again and again for every site we “touch.” Can we have operating system hosted tools to configure our level of data sharing and does not simply block us from getting to a site or service or product (like Chrome will do if we don’t use a https:// address and if it does the site actually has a valid certificate)?

My *recommendation* is the development of an Internet style (e.g., Internet Engineering Task Force (IETF)) standard that is implemented in every operating system (covering all devices from servers to mobile and Internet of Things (IoT) devices) and is contained in both human and machine-readable format, along with contextual semantics to ensure that there are sufficient opportunities to share our data, as well as, to secure it from sharing. This should also be extended to other tech stacks (e.g.,

¹ CTA Study: “Active Aging: Consumer Perceptions and Attitudes”, <https://www.cta.tech/Resources/i3-Magazine/i3-Issues/2019/May-June/Active-Aging-Perceptions-and-Attitudes>, 5 June 2019, last accessed: 2 July 2022.

Response to the Office of Science and Technology Policy (OSTP) Request for Information (RFI) on Advancing Privacy-Enhancing Technologies (PETs)

containerization, and Application Programming Interface (API) management) to ensure completeness of implementation and scalability.)

This sort of capability, to configure my “default” privacy settings and have all sites and apps honor them, through the use of standard interfaces, also applies, and maybe even more so, to artificial intelligence (AI) / machine learning (ML), where our private data may be consumed and used without our knowledge. There needs to a basic requirement for all AI/ML applications to have an “explainability” (think: XAI²) feature such that the privacy settings we configure can be understood by us (humans) and in a way that we can configure our data use to the benefit of the AI/ML tools. This will be essential for healthcare and financial applications, to ensure that traceability of our data is controlled by us. Existing laws and rules, such as the Health Insurance Portability and Accountability Act (HIPAA) still allow for a great deal of our personal data to be seen and shared so that we can receive medical services and products, yet currently with many more people than is really necessary able to see it. If we had mandatory PET use for exchange of this data, we would have higher confidence to share it, e.g., for medical or pharmacology purposes, to achieve the gains promised by the users of our data. Say if we have the privacy standard configuration and context data stored on mobile devices, we could share it with our medical providers, researchers, and application developers, with a high confidence of correct and proper use of our data, to the benefit of all. And to work with companies like Apple, Firefox, and DuckDuckGo to pilot these PET enabled capabilities, since they already have very well-known privacy advocacy activities within their organizations.

Another *recommendation* is to have end users to be able to leverage PETs to be able to conduct business when talking with human customer service representatives (reps) (and which are being replaced with automated “intelligent” chat avatars). We often have to respond with either publicly available data which the customer reps have access to, or with “security” questions which again the human reps have access, so not actually very secure or private. The use of Authentication tools (such as Authy) provides a two-factor sort of service, yet this is only used in more automated interfaces (e.g., login) and isn’t used in the verbal exchanges with customer reps. The intent here being to go beyond “authentication” in the exchange processes, so that information exchanges requiring Personally Identification Information (PII) (to include credit card information, driver’s license number and IP addresses), so that any sort of “phishing” like approach or “man-in-the-middle” action can be reduced. Developing a solution along these lines should also aide in government -to -government exchanges (whether they be Federal, State or local) to minimize the amount of private data (even if it available in public databases) from being shared in the “raw” and to support the exchanges since not only is the PII protected, the actual data needed (e.g., law enforcement data) can support an inquiry being made without sharing it in all of its details.

Response to Topic 8: Existing best practices that are helpful for PETs adoption.

Recommendation: Work on ways to support PET implementation by leveraging/mirroring Data Labeling efforts done in the Machine Learning (ML) community for supervised and reinforcement learning, in order to collect and use these for context capture and development necessary for PETs to work well.

² See Samek, Wojciech, et al., eds. Explainable AI: interpreting, explaining and visualizing deep learning. Vol. 11700. Springer Nature, 2019.

Response to the Office of Science and Technology Policy (OSTP) Request for Information (RFI)
on Advancing Privacy-Enhancing Technologies (PETs)

Recommendation: Leverage organizations like the American Association of Retired Persons (AARP) and their safeguarding efforts like AARP BankSafe and the AARP Fraud Watch Network Helpline³ programs. One such activity could be to approach AARP to propose to do a joint pilot program with them to roll out a customer service interactions approach with PETs in play to reduce the exposure of private data during interactions over the phone or in person. This pilot could also address the lack of basic understanding of PET capabilities, so training and other learning materials for all ages, similar to those developed for the cybersecurity “hygiene” materials, could be developed and disseminated, as well as work with one or two technology developers to prototype products that can then be tested.

Response to Topic 9: Existing barriers, not covered above.

Recommendation: Incorporate PET capabilities into Computer Science curriculum, at all age levels, so that we go from awareness to practice as a “normal” vs “exception” approach to incorporation of data privacy technologies. Understanding the challenge of curriculum changes from elementary to post-secondary, this is no small task, yet data privacy is just as important as cybersecurity and will need the development of “champions” to achieve this result.

Recommendation: Address the lack of tool suites for review of code and coding practices that enforce a “data privacy first” approach to developing both software, firmware, and hardware. This can be done by first identifying what commercial tools are being developed (and maybe already being used) as well as to continue the funding of both academic research as well as public-private partnerships to develop the tools and work with NIST on the development of the standard processes, similar to what is currently done in the cybersecurity space. AND then pilot these both in government and in private sector efforts to “wring them out.” Leveraging an “open source” approach may be a good way to get greater adoption faster. Also incorporating this effort with the recommendation above, will certainly achieve multipliers (no pun intended) in results.

Recommendation: There is need for hardware accelerators, not just for cryptological purposes, to support the computational resource needs of PETs at both the “edge” (e.g., IOT or mobile devices) and the datacenters (cloud servers) and possibly even within the communications devices (e.g., routers & switches). These can be in general purpose computing devices, like CPUs as well as purpose build hardware accelerators, similar to Graphics Processing Units (GPUs) (from Nvidia), Tensor Processing Units (TPUs) (from Google), Adaptive Compute Acceleration Platform (ACAPs) (from Xilinx), or simply purpose built chipllets (from Intel and AMD). Again, a pilot program with one or more of the hardware vendors (e.g., at both the chip and the application level) would not just address development of prototypes, it would also cover the development of business models necessary to make these viable, as well as provide something with which to feed to the technology social media (e.g., Wired, TechCrunch, Ars Technica, and others) to raise both interest and awareness.

Response to Topic 10: Other information that is relevant to the adoption of PETs.

In the field of systems and software (also applicable to firmware and hardware) is the use of architectural design languages (e.g., SAE International standard AS5506D Architecture Analysis & Design

³ AARP .org Bulletin, June 2022, p33 “AARP Expands Safeguards” and p34 “Hear these words? HANG UP!”, <https://press.aarp.org/2022-6-3-AARP-Bulletin-Exclusive-AARP-National-Geographic-Second-Half-of-Life-Research>, last accessed: 2 July 2022.

Response to the Office of Science and Technology Policy (OSTP) Request for Information (RFI)
on Advancing Privacy-Enhancing Technologies (PETs)

Language (AADL)⁴ are used to address the holistic treatment of systems development.

Recommendation: Perhaps development of a Privacy Protection Analysis and Design Language (PPADL, pronounced “paddle” :) would be useful for both researchers and practitioners, for a holistic approach to privacy preservation. This would address APIs at both the hardware as well as the software.

Recommendation: Application of PETs to Digital Twins (DTs) and other Intellectual Property exchanges to allow for both additional levels of design as well as manufacturing while ensuring privacy of data in the models or DTs. This would allow the models to be shared with multiple suppliers in such a fashion as to ensure the information needed by the individual organizations was available to complete their tasks (e.g., complete additional design work or manufacture a particular part) while keeping other information protected. This might leverage the work done in the area of “Cognitive Digital Twins”⁵ in order to capture the contextual model information to be incorporated into the PET exchanges.

Lastly, *Recommendation:* Develop of a standard lexicon of terms to include and help establish the concept of “Data Ownership” which can be monetized by both government / private industry business purposes as well as “consumers” of their business services and products. This last bit being difficult given the current state of development and implementation of Web 2.0 & 3.0 efforts, yet still in need for both communities to deal with fairness to both the individual and the corporations. Have these terms defined and agreed upon, then allow us to develop the technologies suggested in the prior responses.

⁴ <https://www.sae.org/standards/content/as5506d> and https://en.wikipedia.org/wiki/Architecture_Analysis_%26_Design_Language; both last accessed 2 July 2022

⁵ Ali, Patel, Breslin, Harik and Sheth, “Cognitive Digital Twins for Smart Manufacturing”, IEEE Intelligent Systems, Volume: 36, Issue: 2, March-April 1 2021, DOI: 10.1109/MIS.2021.3062437.

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

World Data Exchange

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

POWERED BY



digi.me



WORLD DATA
exchange

LEADING THE WORLD IN PERSONAL DATA CONTROL AND PRIVACY

RFI Response

Advancing Privacy-Enhancing Technology

Submission for the US Office of Science and Technology Policy (OSTP)

World Data Exchange Holdings Pty Ltd

8TH JULY 2022

Disclaimer – the views expressed in this submission have been provided in good faith and are based on our own beliefs and industry knowledge which is subject to change without notice due to the rapid evolution of the Data Economy, legislation, and world events. E.&.O. E

Office of Science Technology Policy (OSTP)

Submission emailed to | to PETS-RFI@nitrd.gov

July 8th, 2022

Advancing Privacy-Enhancing Technologies

World Data Exchange (WDX) is an innovative private Company offering consumer-centric personal data sharing technology infrastructure with embedded Privacy, Security and Consent functionality that places the individual at the centre of their data controls.

As an industry stakeholder in Human centred data sharing innovation we welcome the opportunity to comment on advancing Privacy-Enhancing Technologies (PETs) for the consideration of the White House Office of Science Technology Policy on behalf of the Fast Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics of the Subcommittee on Networking and Information Technology Research and Development (NITRD) of the National Science and Technology Council, the National Artificial Intelligence Initiative Office, and the NITRD National Coordination Office.

Whereas we believe industry, academic advocacy, and public commentary are vital to assist Government on informed development of a national strategy on privacy-preserving data sharing and analytics, along with associated policy initiatives.

We support National strategy which seeks to put forth increased vision about the responsibly of harnessing privacy-preserving data sharing and analytics to benefit individuals and society. We recognise the aligned purpose to propose actions from research investments to training and education initiatives, to the development of cohesive standards, policy, and regulations needed to achieve that vision, in a manner that advances current Privacy Regulations and promotes data protections for all, as part of a comprehensive regulatory approach towards digital democracy to which Privacy Enhancing Technology will play a key role.

About us

World Data Exchange is a Sydney based Australian company established to introduce a range of privacy enhancing technologies (PETs) and digital rights management solutions empowering consumers to protect and mobilise their data for their benefit. Our technologies provide consumers with the means to control and manage their data using methods such as tokenised consent certificates and decentralised (self-controlled/ encrypted) data wallets as well as assisting connected businesses to transact permissioned data sharing in a compliant “privacy by design” manner.

WDX is committed to the adoption of ethical consent management and related security frameworks that provide an interoperable and consumer-centric approach to personal data

sharing in a digital environment. As an emerging industry leader our team is happy to discuss any of the issues covered in our submission in further detail upon request by OSTP.

PETs by way of data facilitation platforms

The digi.me personal data exchange platform from WDX offers infrastructure specifically designed to enable individuals to securely regain and self-aggregate, view and exchange their personal information from multiple sources. Then with their consent, to decide how it may be shared with others.

The digi.me PET technology is classified as an intermediated dual sided infrastructure Platform/App marketplace designed to;

- provide individuals with the means by which they can connect, access and keep their personal information up to date in real time from multiple different data sources
- link to multiple data sources via Apps, web-based services, and other structured and unstructured databases through Application Programming Interfaces (APIs)
- enable people to securely download their personal information and to store it in a device or other cloud storage location of their choosing, with the entire process safeguarded using strong encryption
- enable individuals to understand better the nature, extent, and value of their personal information through a dashboard that provides both a visual representation of it as well as curating it so that people can ‘make sense’ of their own data holdings
- provide individuals with functionality that easily empowers them to share their personal information safely if they choose to do so on the terms that they determine

digi.me does not see, touch or hold user data



Data encryption and normalization happens inside the app without **digi.me** ever being able to see or access user data.



Only the user has the credentials to access their **digi.me** library and must provide credentials directly to data sources.



digi.me stores no user data. The user chooses their own location where encrypted data is stored.

Our privacy by design and distributed architecture reduce costs & liability

In short, it is a platform rich in privacy preserving technology (PPT) that gives people awareness of, and control over, their personal information.

The digi.me platform uses an innovative decentralised architecture approach to implement multi-sectoral data facilitation routing. This means that at no time does digi.me touch, hold

or see its users' data. The digi.me platform does not sell or trade user data. It does not tell its users how or where to share their data. The platform provides the data pipes between individuals and the holder of their data and those to whom they wish to share their data in a way that is private and secure.

As a private and secure data exchange platform, digi.me enables **data portability**; complies with the requirements of the European Union's (EU's) *General Data Protection Regulation* (GDPR) and has been subject to an EU data protection impact assessment (DPIA). To satisfy EU and other international privacy requirements, it adopts a 'Privacy by Design' approach to protecting personal information. digi.me also meets or exceeds international security standards.

Digi.me functionalities are flexible and scalable. They are not limited to any categories or to industry or social sectors. Currently, they include health, finance, social, wearables and entertainment.

Input on the current privacy regulation environment

Alongside members of organisations such as the [MyData Global](#) forum, and a multiplicity of other organisations providing PPTs or privacy enhancing technologies (PETs), World Data Exchange recognise that individuals' views on privacy may vary widely, depending upon the presence or absence of trust, their degree of knowledge about privacy issues and risks, their ability to protect their privacy (technical 'know how'), and whether or not they obtain a benefit or advantage from disclosing their information.

In short, there is no one-size-fits-all approach to privacy, with individuals changing their views over time, between organisations, and in accordance with their preferences or personal beliefs. This spectrum of views produces demand for a range of technical solutions, including new or enhanced ways of managing consent.

Currently, it is difficult for individuals to have a clear view of the way in which their information is collected, used, disclosed, stored, and otherwise handled despite the fact that privacy legislation is pervasive and technical solutions exist for a range of privacy issues and risks. Currently, there is a limit to any individual's ability to manage all relevant privacy issues and risks on their own. We believe that it is important not to underestimate the number of people who would like to adopt more privacy-aware data practices but do not have the knowledge or technical ability to do so. Privacy regulators, backed by appropriate legislative frameworks, can help individuals take steps to protect themselves and develop greater trust in the online environment.

Increasing concerns around transparency and fairness, as well as a renewed interest in ethics, are a direct response to increasing complexity, monetisation, and opacity in the personal information ecosystem. We are seeing regulators around the world questioning whether or not the approach to personal data taken by large digital platforms is lawful and fair. We are seeing governments around the world taking another look at their privacy legislation, asking whether or not it remains fit-for-purpose and making changes in response. In addition to the EU, this has included the UK, Brazil, New Zealand, Singapore,

California, and Canada. Previously ‘dormant’ legal policy issues – such as online tracking, surveillance and dark patterns – have become dominant.

Almost continuous data breaches, as well as the impact of the GDPR and Californian privacy legislation, have also contributed to a commercial shift away from the non-transparent collection and handling of personal information, towards more consumer-centric approaches. Multiple market responses are emerging to address these concerns. Whether this involves established players like Google changing their practices or Apple mandating pro-privacy requirements – or newer, ‘privacy first’ companies like World Data Exchange developing innovative platforms, products, and services – significant change is underway.

This activity demonstrates that PPTs and PETs are continuing to develop and evolve in response to privacy-intrusive products and services. In the mid-1990s, [Privacy by Design](#), along with its 7 Foundational Principles, emerged in response to Pets’ inability to gain traction. Almost 25 years later, in 2022, PPTs and PETs are no longer untested or failing to gain traction – they provide a component of an overarching privacy solution, alongside legislation, regulation, organisational practices and Privacy by Design or privacy engineering approaches. Any updates to the Privacy Act should be informed by these developments while maintaining the current principles-based and technology neutral approach.

Key input on forming Privacy Policy

- We support the evolution of Privacy Policy in review by several Governments and welcome any regulatory led changes or enhancements to that enable it to remain up-to-date and meaningful, capable of tackling the specific privacy issues and risks facing us today
- We note the importance that Policies within such Regulation remaining technology neutral. In particular privacy legislation:
 - Policy should not impede the development of multiple technical solutions, reflecting the multiplicity of individual views on privacy; and
 - Policy should not prevent businesses from offering PPTs/PETs as part of, or the primary component of, their commercial offerings
- The Privacy policies should maximise consistency in its operation, including through the removal of unnecessary exemptions that result in increased uncertainty for individuals and organisations
- Recognising that consistency remains key to streamlining international privacy compliance requirements, Governments just like in our home Country of Australia should continue working towards a global standard of privacy protection, thereby enabling a small technology business to operate through compatible platforms and standards designed for global reach and cross border frameworks. Likewise, it is important that Governments and Business provides clients/citizens with equivalent protection when travelling with cohesive jurisdictional regulations – this point,

rather than ‘adequacy’, should provide the policy basis for pursuing consistency on both a National and International levels

- There is an increasing need to separate out issues relating to ‘consent’ within policy principles, ‘consent’ within an entity’s Terms and Conditions, and what may prove to be the ‘death throes’ of consent-based cookies and related tracking technologies. Each of these involves different legal, regulatory and policy issues. They should not be conflated. A failure to separate out the various definitional, policy and legal/regulatory issues relating to consent is likely to result in further confusion and, potentially, lead to reforms that prove to be undesirable in the medium-to-long term
- Claims around ‘consent fatigue’ in media we believe are vastly overstated and avoid tackling the real issues faced by individuals. This aspect within Privacy Policy requires further, detailed contextual analysis. We note that numerous technical, regulatory, and operational options are available now and can be used by entities and individual consumers to streamline consent processes. The key focus here should be protecting the individual’s autonomy
- There are two core elements from the GDPR that we suggest should be considered as part of the US Privacy policy reforms processes which seek to include PETs
 1. A straightforward right to data portability that is consumer-centric and enables all entities to participate subject to meeting a minimum set of technical requirements (ensuring data portability is viewed first and foremost as an individual right)
 2. Explicit and informed consent (supplemented by the regulator as/if required) to ensure that individuals are protected from online tracking, surveillance, and dark patterns (amongst other current issues)

Industry focus | PETs and Patient Centric Healthcare

eHealth as an “in focus” industry sector presents the opportunity to propose significant step change by utilising PETs through emerging Patient Centric health models staging the entrée towards the future of digital health-care services.

Patient Centricity becoming “the new normal” fundamentally transforms the health industry’s ability to provide and lead cohesive digital transformation through enabling platforms which unite and underpin the eHealth sector across our wider eco-system via the critical element of secure, private, and interoperable data (as a raw material) delivered in a patient centric manner through robust industrial scale systems.

The outcome is headlined by the provision of human centred connected technologies that shall shift the current model from reactive to proactive health-care management by enabling countless interwoven solutions seamlessly. Such a transformational shift changes behaviours towards self-empowered health and ignites a new breed of patient centric

products and services that will save lives and reduce the ever increasing burden on traditional health-care systems that are continually challenged with an aging population. Emerging data centric ehealth approaches harness holistic data management for ehealth to extend across technologies such as A.I., Machine Learning, Wearables, Telemedicine, Genomic and IoT interfaces - to name a few - means that in decades from now Government and Industry can be equipped to better address and even hope to eradicate or increasingly prevent diseases such as diabetes, heart disease and cancer plus provide more personalised direct care and diagnoses through more efficient and effective forms – including in home care – by positioning the individual to sit and act at the centre of their health requirements through data insights, innovative products and hyper personalised services.

During times of crisis as seen with the COVID-19 Pandemic ehealth data facilitation “utility” can service our Nation faster during a cycle whereas “unknowns” change overnight requiring the urgent need to deploy agile solutions at scale with sometimes only days to respond to the evolving requirements via point solutions that navigate technology interconnects plus systems supports which demand interoperability and go far beyond initial Contact Tracing apps or help make it truly effective.

During the Pandemic Countries have needed to become more agile in providing a range of COVID-19 solutions to assist front line workers, Allied Health, Hospitals, Researchers, Clinicians, GP’s, First Responders, and care givers whilst not forgetting our citizens, particularly those who are more vulnerable in our society such as seniors, regionally displaced, indigenous and individuals with disabilities or those who reside in full time care facilities.

Data facilitation platform technologies with embedded PET assist in placing Government Health Agencies at the epicentre with an enhanced ability to rapidly respond to ever increasing demands. New Patient Centric step change models are flexible for fast and agile implementation with the ability to interoperate across data hungry systems and cascade through end to end frameworks with API’s that bridge existing systems in a compliant manner. With Governance and seamless delivery “direct to patient” services can be routed with interconnecting B2B distribution through a multi-dimensional eco-system operating on agnostic data rails and internet connectivity. Not to mention federated A.I. and Machine Learning aspects.

Utilising PET to onboard a game changing moment in eHealth.

Today’s mobile population navigates a complex health landscape, crossing between specialties, organisations, sectors, regions or even countries. However, while the individual traverses the health ecosystem their data tends not to follow them. Our health also extends well beyond traditional healthcare settings and includes our lifestyle and behaviours whether at home, work and at play. Much of our habits and behaviour are locked up in data which surrounds our lives. World events such as COVID-19 create an unprecedented challenge for governments when it comes to surveillance and tracking while creating an

imperative to help individuals manage their life. Patient centricity is the philosophy of equipping individuals with data and technology enabling them to better manage their lives while respecting their privacy and rights.

WDX consumer centric solutions which feature PETs represent new approaches to enable the rapid adoption of digital health services to support the COVID-19 effort with two key aims:

- Equip individuals with a data wallet to gather up data across their life.
- Provide a COVID-19 related set of apps and services.

This approach enables industry to collaborate globally to bring together a unique set of competencies and capabilities to help address the crisis and evolve Health-care ecosystems beyond COVID-19.

When trust is restored via PETs that offer citizens a personal data platform which provides individuals with the ability to own and control their data which promotes trusted handling of personal or sensitive data. To engender trust, we engage with award-winning market leading compliance framework with platforms for jurisdictional accreditation and distribution of health apps, enabling the safe and sustainable adoption of digital health apps and services at scale.

Rapidly deploying these technologies will provide a foundation of capabilities which can not only support the immediate COVID-19 response but additionally offer:

- Monitoring and self-assessment of individuals.
- Supporting self-management for at home care.
- Connecting people with health, wellbeing, and condition specific digital services.
- Medical research now and in the future.

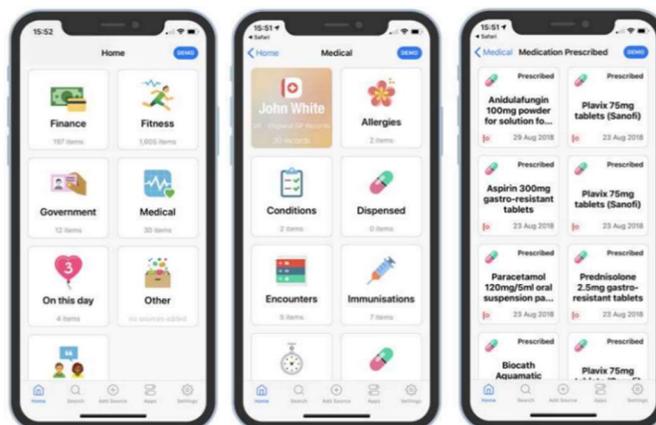


Figure 1 - digi.me personal data library platform

The core aim for PET based capabilities is to underpin digital eHealth futures in terms of:

1. Enabling the accelerated the adoption of digital health through better access to data and a trusted complaint framework of services. Equipping individuals with the data and tools they need to better support themselves reducing the burden on the healthcare and providing a basis for wider national distribution.
2. Putting in place core set of capabilities and competencies that would allow a much more rapid and immediate technological and data driven response to future pandemics or emergency situations.

With existing privacy enhancing technology, such core capabilities can be deployed today to provide the foundation of an accelerated digital health adoption.

PETs can also be immediately applied into several other industry sectors such as Social, Finance, AdTech and other markets by supporting innovation leaders with the capability to address priority needs. In eHealth this might be applications for anxiety amid mental health, Asthma, Chronic obstructive pulmonary disease (COPD) Diabetes and other priority areas escalating within the public health system.

Overall, we remain convinced Privacy Enhancing Technologies and well-considered Privacy reforms will drive digital transformation and enhance the data economy by leveraging the Regtech sector to ensure consumer centric technologies are the cornerstone of citizen data rights.

As Privacy and Data Sharing platform leaders, we hope this submission is of assistance to OSPT, please feel free to contact us for any further clarification or continued input in relation to this submission.

We look forward to continued participation.

Kind regards,

Joanne Cooper

CEO

World Data Exchange Holdings Pty. Ltd.

Level 5, 11 York Street
Sydney NSW 2000
Australia