Federal Register Notice 87 FR 35250, https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies, June 9, 2022

Request for Information (RFI) on

Advancing Privacy Enhancing Technologies

PUBLIC Group International Ltd

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/

PUBLIC

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 (not exhaustive)	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	• Permits outsourced computation where **data re-identification is not a concern**, but **trust is low** • Able to handle complex big data processing	• 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	• Permits data collaboration with **low or intermediate trust**, where accuracy of analysis does not need to be perfect	• 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	• Where **trust is low or nonexistent,** empowers collaboration and computations without ever decrypting	• 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**	• Multi-party analytics where **trust is low** or **data transfer risk is high** • Commonly **used with DP** • Training models with small datasets	• **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.