Request for Information (RFI) on

Advancing Privacy Enhancing Technologies

Datavant

Request for Information on Advancing Privacy-Enhancing Technologies
Office of Science and Technology Policy (OSTP)
submitted via email PETS-RFI@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 SoftwareAmerican 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.