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

Asemio

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/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. 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/.