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

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.

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