National Strategy to Advance Privacy-Preserving Data Sharing and Analytics

A Report by the

FAST-TRACK ACTION COMMITTEE ON ADVANCING PRIVACY-PRESERVING DATA SHARING AND ANALYTICS

NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT SUBCOMMITTEE

of the

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

March 2023

Table of Contents

Executive Summary

Data are vital resources for solving society's biggest problems. Today, significant amounts of data are accumulated every day—fueled by widespread data generation methods, new data collection technologies, faster means of communication, and more accessible cloud storage. Advances in computing have significantly reduced the cost of data analytics and artificial intelligence, making it even easier to use this data to derive valuable insights and enable new possibilities. However, this potential is often limited by legal, policy, technical, socioeconomic, and ethical challenges involved in sharing and analyzing sensitive information. These opportunities can only be fully realized if strong safeguards that protect privacy[1]—a fundamental right in democratic societies—underpin data sharing and analytics.

Privacy-preserving data sharing and analytics (PPDSA) methods and technologies can unlock the beneficial power of data analysis while protecting privacy. PPDSA solutions include methodological, technical, and sociotechnical approaches that employ privacy-enhancing technologies to derive value from, and enable an analysis of, data to drive innovation while also providing privacy and security. However, adoption of PPDSA technologies has been slow because of challenges related to inadequate understanding of privacy risks and harms, limited access to technical expertise, trust, transparency among participants with regard to data collection and use,[2] uncertainty about legal compliance, financial cost, and the usability and technical maturity of solutions.[3]

PPDSA technologies have enormous potential, but their benefit is tied to how they are developed and used. Existing confidentiality and privacy laws and policies provide important protections to individuals and communities, and attention is needed to determine how to uphold these protections through the use of PPDSA technologies and maintain commitments to equity, transparency, and accountability. Consideration of how individuals may control the collection, linking, and use of their data should also factor into the design and use of PPDSA technologies.

Recognizing the untapped potential of PPDSA technologies, the White House Office of Science and Technology Policy (OSTP) initiated a national effort to advance PPDSA technologies.

This National Strategy to Advance Privacy-Preserving Data Sharing and Analytics (Strategy) lays out a path to advance PPDSA technologies to maximize their benefits in an equitable manner, promote trust, and mitigate risks. This Strategy takes great care to incorporate socioeconomic and technological contexts that are vital to responsible use of PPDSA technologies, including their impact on equity, fairness, and bias—and how they might introduce privacy harms, especially to disadvantaged groups.

This Strategy first sets out a vision for a future data ecosystem that incorporates PPDSA approaches:

> *Privacy-preserving data sharing and analytics technologies help advance the well-being and prosperity of individuals and society, and promote science and innovation in a manner that affirms democratic values.*

This Strategy then lays out the following foundational guiding principles to achieve this vision:

- PPDSA technologies will be created and used in ways that protect privacy, civil rights, and civil liberties.
- PPDSA technologies will be created and used in a manner that stimulates responsible scientific research and innovation, and enables individuals and society to benefit equitably from the value derived from data sharing and analytics.

harmonizing electronic health record data from health systems across the US and using PPRL to connect different COVID-19-related patient data from multiple organizations. Various techniques including secure multiparty computation or data perturbation approaches can be used to support PPRL.[71]

- *Private information retrieval* is another useful technique that allows a client to retrieve data from a database server without the server knowing what was retrieved or queried. This protects a user's access privacy by making sure the data owner cannot track what content or types of information the user is accessing. Similar access privacy issues relate to a server learning about what a user is accessing based on observable access patterns. Oblivious random access memory is a technique that has been used to address such access privacy issues.
- *Federated learning* allows multiple entities to collaborate in building a machine learning model on distributed data. It provides inherent privacy protection as participants do not have to share their raw data. Instead, each participant trains a local model on their data which is then integrated into the collaborative model. Recent research[72] has identified persistent privacy risks in federated learning, which are also found more generally in ML, such as model inversion attacks that can reconstruct the private training data or membership inference attacks that can identify if a data sample is part of the training dataset. Research is ongoing in combining some of the above-referenced cryptographic techniques to close these vulnerabilities and create privacy-preserving federated learning.

Summary and challenges. The *t*echniques mentioned above are some of the key existing technical approaches relevant for privacy-preserving data publishing or analytics, but not an exhaustive list. The technologies described are at different levels of maturity with some such as differential privacy or secure multiparty computation seeing initial, limited success in deployment, and others still in earlier stages of development. Cross-cutting technical challenges such as those related to understanding and quantifying disclosure risks, scalability and efficiency, and verification and validation approaches to ensure the correctness of design, implementation, and deployment present barriers to broader adoption. Furthermore, in many application scenarios, the integration of various techniques will be needed to support end-to-end privacy. Additional work is also needed to determine how issues of fairness, transparency, and accountability can be assured while achieving privacy guarantees.

- PPDSA technologies will be trustworthy, and will be created and used in a manner that upholds accountability.
- PPDSA technologies will be created and used to minimize the risk of harm to individuals and society arising from data sharing and analytics, with explicit consideration of impacts on underserved, marginalized, and vulnerable communities.

Based on the guiding principles, this Strategy identifies the following strategic priorities for the public and private sectors to progress toward the vision of a future data ecosystem that effectively incorporates PPDSA technologies:

- **Advance governance and responsible adoption** through the establishment of a multi-partner steering group to help develop and maintain a healthy PPDSA ecosystem, greater clarity on the use of PPDSA technologies within the statutory and regulatory environments, and proactive risk mitigation measures.
- **Elevate and promote foundational and use-inspired research** through investments in multidisciplinary research that will advance practical deployment of PPDSA approach and exploratory research to develop the next generation of PPDSA technologies.
- **Accelerate translation to practice** through pilot implementations, development of consensus technical standards, and creation of user-focused tools, decision aids, and testbeds.
- **Build expertise and promote training and education** through concerted efforts to expand PPDSA expertise across the public and private sector and foster privacy education opportunities from K-12 through higher education, with particular attention to capacity building in underserved communities.
- **Foster international collaboration on PPDSA** through promotion of partnerships and an international policy environment that furthers the development and adoption of PPDSA technologies and supports common values while protecting national and economic security.

PPDSA technologies have the potential to catalyze American innovation and creativity by facilitating data sharing and analytics while protecting sensitive information and individuals' privacy. Leveraging data at scale holds the power to drive transformative innovation to address climate change, financial crime, public health, human trafficking, social equity, and other challenges, yet it also holds the potential to violate privacy and undercut the fundamental rights of individuals and communities. PPDSA technologies, coupled with strong governance, can play a critical role in protecting democratic values and mitigating privacy risks and harms while enabling data sharing and analytics that will contribute to improvements in the quality of life of the American people. This Strategy serves as a roadmap for both the public and private sectors to responsibly harness the potential of PPDSA technologies and move together toward the vision that anchors this Strategy.

OSTP, in partnership with the National Economic Council, will focus and coordinate Federal activities to advance the priorities put forward in this Strategy.