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

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976 to provide the President and others within the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, health, foreign relations, the environment, and the technological recovery and use of resources, among other topics. OSTP leads interagency science and technology policy coordination efforts, assists the Office of Management and Budget with an annual review and analysis of Federal research and development budgets and serves as a source of scientific and technological analysis and judgment for the President concerning major policies, plans, and programs of the Federal Government. More information is available at https://www.whitehouse.gov/ostp.

About the National Science and Technology Council

The National Science and Technology Council (NSTC) is the principal means by which the Executive Branch coordinates science and technology policy across the diverse entities that make up the Federal research and development enterprise. A primary objective of the NSTC is to ensure science and technology policy decisions and programs are consistent with the President's stated goals. The NSTC prepares research and development strategies coordinated across Federal agencies to accomplish multiple national goals. The work of the NSTC is organized under committees that oversee subcommittees and working groups focused on different aspects of science and technology. More information is available at https://www.whitehouse.gov/ostp/nstc.

About the Subcommittee on Networking & Information Technology Research & Development

The Networking and Information Technology Research and Development (NITRD) Program has been the Nation's primary source of federally funded work on pioneering information technologies (IT) in computing, networking, and software since it was first established as the High-Performance Computing and Communications program following passage of the High-Performance Computing Act of 1991. The NITRD Subcommittee of the NSTC guides the multiagency NITRD Program in its work to provide the research and development foundations for ensuring continued U.S. technological leadership and for meeting the Nation's needs for advanced IT. The National Coordination Office (NCO) supports the NITRD Subcommittee and its Interagency Working Groups (IWGs) (https://www.nitrd.gov/about/).

About the Fast-Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics

The Fast-Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics is a multi-agency venue for building a whole-of-government approach, from research and development to pilot projects and adoption, to advance privacy-preserving data sharing and analytics technologies that enable collective data sharing and analysis while maintaining disassociability and confidentiality.

About This Document

This National Strategy to Advance Privacy-Preserving Data Sharing and Analytics is a cohesive national strategy to advance the research, development, and adoption of privacy-preserving data sharing and analytics technologies. A list of acronyms and abbreviations used in the document is included as a reference in Appendix A.

Copyright

Note: Any mention in the text of commercial, non-profit, academic partners, or their products, or references is for information only; it is not intended to imply endorsement or recommendation by any U.S. Government agency.

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

Chair

Arati Prabhakar, Director, Office of Science and Technology Policy (OSTP), Assistant to the President for Science and Technology

Acting Executive Director

Kei Koizumi, Principal Deputy Director for Policy, OSTP

SUBCOMMITTEE ON NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT (NITRD)

Co-Chair

Margaret Martonosi, Assistant Director for Computer and Information Science and Engineering, National Science Foundation (NSF)

Co-Chair

Kathleen (Kamie) Roberts, NITRD National Coordination Office (NCO)

Executive Secretary

Nekeia Butler, NITRD NCO

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

Co-Chairs

Tess DeBlanc-Knowles, Senior Policy Advisor for Artificial Intelligence, OSTP

James Joshi, Program Director, Computer and Information Science and Engineering, NSF

Naomi Lefkovitz, Senior Privacy Policy Advisor, National Institute of Standards and Technology (NIST)

Katelyn McCall-Kiley, Program Director, xD, U.S. Census Bureau

Technical Coordinator

Tomas Vagoun, NITRD NCO

Writing Team Members

Tess DeBlanc-Knowles, OSTP
Dylan Gilbert, NIST
James Joshi, NSF
Naomi Lefkovitz, NIST
Aaron Mannes, DHS
Katelyn McCall-Kiley, Census Bureau
Angela Robinson, NIST
Lisa Wolfisch, GSA

FTAC Members

Gil Alterovitz, VA
Lindsey Barrett, NTIA
Tim Bond, DARPA
Rafael Fricks, VA
Simson Garfinkel, DHS
Elena Ghanaim, NIH
Karyn Gorman, DOT
Michael Hawes, Census Bureau
Jeri Hessman, NITRD NCO
Kevin Herms, ED
Wu He, NSF
Brian Ince, ODNI
Luke Keller, Census Bureau
Brian Lee, CDC
Steven Lee, DOE
Nicholas Mancini, DOS
Kathryn Marchesini, HHS
Mark Przybocki, NIST
Andrew Regenscheid, NIST
Scott Sellars, DOS
Heidi Sofia, NIH

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.

- 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.

Introduction

Data drive scientific and technological breakthroughs, underpin policymaking, and power the global economy. Clinicians use data to identify the best treatments for their patients, farmers use data to predict and improve farm yields, researchers use data to generate new knowledge about natural and social phenomena, and public servants use data to create evidence-based policies. Artificial Intelligence (AI) and other emerging analytics techniques are amplifying the power of data, making it easier to discover new patterns and insights ranging from better prediction models to understand and mitigate the impacts of climate change to new methods for detecting financial crime.

Although data enable science, innovation, and insights, balancing the benefits of these data-derived insights with the imperatives of privacy, security, and other values is a longstanding challenge. For example, when developing new treatment options, medical researchers may benefit from broad access to electronic health records. However, those records may contain personal health information related to individual patients, compromising the privacy and safety of those patients as well as rights under health privacy laws and regulations on the protection of human subjects. Similarly, when researchers access authorized data without safeguards on how they access the data, privacy-sensitive information such as their location or the specific type of information they are accessing may be revealed. In many domains, collaborations that could improve AI model training and accelerate progress must be balanced with ethical and legal privacy concerns and intellectual property protection concerns.

Over several decades, privacy researchers, including statisticians and cryptographers, have been adapting various anonymization, statistical disclosure limitation, and private computation techniques to address privacy or confidentiality needs. As demands for granular data grow and as the amount of publicly available data on individuals has proliferated, so have the challenges to protecting against re-identification and record-linkage risks in protected datasets or membership inference attacks on analytical models built on sensitive data. New solutions to address these threats are increasingly needed.

Privacy-preserving data sharing and analytics (PPDSA) solutions include technical and sociotechnical approaches that employ certain types of privacy-enhancing technologies (PETs) to generate value from and enable analytics on data while protecting privacy and security. Some PPDSA approaches allow users (e.g., researchers and physicians) to gain insights from sensitive data without exposing the original data itself or allow them to access shared data without being tracked or identified. Other PPDSA approaches enable data sharing by obscuring personal data or making synthetic reflections of the original data that preserve the properties of interest in the data while protecting individual privacy.

What is meant by data?

For the purposes of this strategy, data are elements or values that convey information. Data can range from abstract concepts to physical measurements. They exist in a wide variety of types (e.g., text, image, audio, video) and contexts (e.g., location, time), including digital and non-digital forms. Data can vary in complexity from basic representations of information and graphs to multimodal, including those generated by complex human-computer interactions. Data that convey information about other data are known as metadata. Data may represent aggregate information or derived data. Numerous data operations, including those involving metadata, can have privacy implications for individuals and groups. Examples include data collection, access, analysis, sharing, transmission, and retention.

PPDSA technologies can unlock new forms of collaboration and new norms in the responsible use of personal data. By enabling the use of more comprehensive and diverse datasets, PPDSA technologies can help the global community tackle shared challenges and drive solutions in areas such as healthcare, climate change, financial crime, human trafficking, and pandemic response and achieve more equitable outcomes for underserved, marginalized, and vulnerable populations.

The Fast-Track Action Committee (FTAC) on Advancing Privacy-Preserving Data Sharing and Analytics was chartered under the National Science and Technology Council's (NSTC)[4] Networking and Information Technology Research and Development (NITRD)[5] Subcommittee in January 2022 to develop this National Strategy to Advance Privacy-Preserving Data Sharing and Analytics (Strategy) to advance the research, development, and adoption of PPDSA technologies.[6] This Strategy intends to chart a path forward for the Nation to steward responsible and accountable development and adoption of these transformative technologies.

What are PETs and PPDSA technologies?

For the purposes of this strategy, PETs refer to a broad set of technologies that protect privacy by removing personal information, by minimizing or reducing personal data, or by preventing undesirable processing of data, while maintaining the functionality of a system. PPDSA technologies refer to a subset of PETs that are essential for enabling data sharing and analytics in a privacy-preserving manner, such as secure multiparty computation, homomorphic encryption, differential privacy, zero knowledge proofs, synthetic data, federated learning, and trusted execution environments, which are discussed further in the document.

Applications of PPDSA Technologies

Deployments of PPDSA technologies have begun to show promising results. For example, the Boston Women's Workforce Council initiated a study of the Boston-area gender wage gap to assess the compliance of Boston-area employers with the Equal Pay Act.[7] A PPDSA technique, secure multiparty computation, was used to enable the joint analysis of sensitive salary data across Boston-area employers without disclosing those data to any party external to the employer. The 2020 U.S. Census used differential privacy, another PPDSA technique, to publish aggregate statistical data while protecting the privacy of individuals.[8] In another example, researchers at the University of Pennsylvania collaborated with nine other institutions using the PPDSA technique of federated learning to develop a machine learning (ML) model to analyze magnetic resonance imaging scans of brain tumor patients and distinguish healthy brain tissue from cancerous regions.[9]

Case Study: Boston Women's Workforce Council Wage Gap Study

Challenge: The Boston Women's Workforce Council (BWWC) sought to assess the compliance of Boston-area employers to the Equal Pay Act by computing the Boston-area gender and race wage gaps. The assessment required employers to share privacy-sensitive payroll data.

Approach: A traditional approach would have required Boston-area employers to release sensitive information about employees to a trusted entity for statistical analysis. Instead, the BWWC used secure multiparty computation. Boston-area employers were able to provide the salary information of their employees in a privacy-preserving way so that accurate wage statistics were computed even though underlying salary information was never disclosed.

Impact: Sixty-nine employers participated in the 2016 BWWC wage gap report, contributing wage information on 113,000 employees. By 2021, 134 employers contributed information on 156,000 employees' wages. This adoption of secure multiparty computation has allowed the BWWC to determine whether the gender wage gap is closing in Boston and track the progress each year.

In the public sector, PPDSA technologies are advancing the Federal Statistical System's ability to safely expand access to some of the nation's most sensitive data as intended under the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act).[10] They can also advance the mission articulated in the Federal Data Strategy[11] "to fully leverage the value of federal data for mission, service, and the public good"[12] by enabling data sharing and analysis across agencies to safely and responsibly share Federal data for research. They can support national priorities to advance open science by enabling federally funded researchers to maximize sharing of scientific data while respecting potential legal and privacy limitations.[13] The CHIPS and Science Act of 2022[14] includes a provision that establishes a National Secure Data Service (NSDS) demonstration which will, in concert with the statistical system's implementation of the Evidence Act, pilot a tiered access system, leveraging PPDSA technologies to facilitate access to Federal statistical data for evidence-based policymaking. There are also PPDSA applications in public health, real-time data for emergency response, traffic analysis, improved quality of human resources information, and urban planning, all of which can benefit from the analysis of sensitive data.

In the private sector, PPDSA approaches can, for example, enable companies to share data on cyber incidents without disclosing the identity of the company or competitive details of the company's operations. Banks can improve fraud detection by using PPDSA technologies for information sharing across organizational boundaries to identify suspicious patterns while better protecting customers' privacy. To enable more efficient and climate-friendly management of electricity or traffic through smart grids and smart cities, PPDSA technologies can facilitate understanding of consumption and travel patterns while not revealing privacy-sensitive information.

However, challenges related to inadequate understanding of how PPDSA technologies mitigate privacy risks and harms, limited access to technical expertise, trust, uncertainty about legal and policy compliance, financial cost, the usability and technical maturity of solutions, as well as how solutions map to different use cases have slowed adoption of PPDSA technologies.

PPDSA technologies have enormous potential to enable data collaboration, but strong policies and governance will be needed to ensure that they are developed and used in ways that protect multiple aspects of privacy, security, civil rights, and civil liberties. Many such safeguards are built into U.S. confidentiality and privacy laws and policies that apply to the Federal Government's use of data, and attention is needed to determine how to advance these protections and the use of PPDSA technologies together (see for example endnotes[15,16]). For example, PPDSA technologies may not fully resolve questions of whether data have been legitimately obtained or used and whether the nature or quality of the data could create harmful bias in the results.

Consideration of how individuals may control the collection and use of their data, including data deletion requests, should also factor into the design and use of PPDSA technologies. PPDSA technologies that are poorly designed or deployed could undermine the potential benefits. It is essential that solutions be applied in a manner that accomplishes the desired privacy goals. If employed without verifiable performance evaluations, transparency, and accountability, PPDSA approaches could create a false sense of privacy. In the private sector, deploying PPDSA technologies without appropriate safeguards could further empower data monopolies, undermining fair markets and increasing inequity in access. It is imperative that laws, regulations, and policy keep pace with innovations in PPDSA technologies to ensure that adequate protections are upheld as PPDSA technologies are implemented and used.

Privacy Risks and Harms in the Context of PPDSA

Establishing a precise definition of privacy is difficult because privacy is inherently multidisciplinary and encompasses diverse concepts, including confidentiality, consent, and control over multiple facets of data and identity.[17] Privacy needs shift with context, time, and individual or cultural differences. Understanding and defining privacy involves interdisciplinary research that brings together technologists, social scientists, economists, and privacy and legal professionals, among others. At its foundation, privacy safeguards basic human values such as autonomy, dignity, freedom of expression, freedom of association, and freedom to engage in intellectual exploration. Lack of privacy can undermine individual rights and freedoms and can worsen discrimination against marginalized groups. It is also important to note that ensuring security is critical to achieving privacy protection goals.

To improve the ability of organizations to assess privacy risks, the National Institute of Standards and Technology (NIST) has defined privacy risk as the likelihood that individuals will experience problems resulting from data processing, and the impact should they occur.[18] These problems can be expressed in various ways, but NIST describes them as ranging from dignity-type losses, which could include embarrassment or long-term reputational harm to more tangible harms such as discrimination; financial harm, which could result from identity theft; and losses in self-determination, which could include life-altering or threatening situations such as imprisonment or domestic violence.[19]

Privacy protection includes addressing various objectives, such as confidentiality, disassociability, predictability, and manageability.[20] Data confidentiality is well understood as a means of managing privacy risk and applies to the protection of data from unauthorized access or use, whether about an individual, group, or entity. Confidentiality protections are written into U.S. statistical laws and are accompanied by a suite of statistical disclosure limitations and tiered data access solutions. Many of the PETs that are used for PPDSA are cryptographic techniques that inherently meet the objective of confidentiality. The distinguishing characteristic of PPDSA approaches is their capability for meeting the privacy engineering objective of disassociability to prevent even authorized entities from making linkages between data and people's identities, which further enhances privacy with authorized use of the data.

For example, these include methods that encrypt data so that a third party can perform analytics on the encrypted information, those that add “noise” to real data or create synthetic datasets to allow for broad public release and analysis of the data, and those that use hardware-based solutions to protect data during analysis. Such technologies currently include, but are not limited to, secure multiparty computation, homomorphic encryption, zero-knowledge proofs, federated learning, secure enclaves, differential privacy, and synthetic data generation tools. When deploying PPDSA techniques, it is also critical to protect privacy in terms of what could be learned through privacy-eroding inferences from the output of the secure analysis.[21]

Privacy risks can arise when privacy objectives are not appropriately met when sharing or analyzing data. For example, loss of confidentiality may expose privacy-sensitive information. This may be the disclosure of an individual's identifying information or the privacy-sensitive attribute of a record/person. Various re-identification, data reconstruction, record linkage, or association attacks are possible even when some form of de-identification or anonymization technique is used, particularly by combining auxiliary information that is easily available.

Similarly, insufficient disassociability includes the potential for inferring privacy-sensitive information related to an individual. For example, from a published ML model, it may be possible to infer that a particular individual's data was included in the model's training dataset. An adversary may also observe

a data user's pattern of access to determine what type of information is being accessed and from where. In addition, despite the ability of the Onion Router (Tor) browsers to provide anonymous communication, website fingerprinting attacks can be used to predict a website visited by the user based on observed traffic patterns.[22] PPDSA solutions are typically designed by considering multiple threat models that characterize how an adversary may compromise a system. Addressing the possible risks often includes a combination of PPDSA techniques to safely carry out the analytics and to protect privacy in released results.

The Need for a National Strategy

As PPDSA technologies begin making the transition to practice and the understanding of their value grows, it is imperative that the Nation moves forward in a manner that harnesses the opportunities of these technologies and is grounded in a commitment to uphold democratic values. This Strategy sets out a vision for the future that responsible research, development, and adoption of PPDSA technologies can enable, and details a series of priorities and actions the public and private sectors can take to move toward that vision. Section 1 of the Strategy lays out the vision for a future data sharing and analytics ecosystem enabled by PPDSA technologies as well as the principles that should guide design, development, and implementation efforts. Section 2 provides context on the current state of PPDSA technology development and adoption, including the key challenges that must be overcome for the broader adoption of such technologies. Section 3 presents five strategic priorities and recommends actions within each that should be advanced to achieve the vision for the future of the PPDSA ecosystem defined in Section 1.

In developing the Strategy, the FTAC consulted a diverse group of participants and experts in the field from across government, academia, the private sector, non-profit, and civil society entities. Engagement activities included the following:

- A series of virtual roundtables, open to the public, to provide an opportunity for broad input. Three sessions were attended by nearly 300 total participants.
- A request for information[23] was issued on June 9, 2022, and garnered 77 responses[24] from the public.
- A data call to Federal agencies seeking information related to their use of PPDSA technologies.
- Informational presentations from a broad range of invited researchers, developers, and practitioners from government, academia, the private sector, and civil society.

The strategy integrates this broad set of perspectives and experiences to set out an inclusive path toward a future data ecosystem enabled by PPDSA technologies.

1: Vision and Guiding Principles

Vision

This Strategy strives to set the country on a path to a future where:

> *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 vision of the future applies broadly to individuals, groups, and society at large, including industry, civil society, academia, and government at all levels.

Guiding Principles

PPDSA technologies offer novel approaches to sharing and processing data to advance the physical and emotional well-being and prosperity of individuals and groups, as well as innovation-driven economic growth. However, on their own, PPDSA technologies cannot prevent the results of data analytics from being used in ways that may undermine democratic values, such as privacy, freedom, equity, accountability, and civil rights and liberties. These tenets must be reflected in all aspects of the development life cycle of PPDSA technologies if such technologies are to realize the vision of this Strategy.

The following principles are intended to guide the development and deployment of PPDSA technologies in support of progress toward achieving the 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.
- 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.

PPDSA technologies will be created and used in ways that protect privacy, civil rights, and civil liberties.

While PPDSA technologies can play an important role in protecting privacy, including confidentiality of data, they are not a complete solution. Without appropriate safeguards, the results from data sharing and analytics could be used in ways that degrade or infringe upon the privacy of individuals and groups or infringe upon core democratic values. Indeed, the confidentiality capabilities of PPDSA technologies may also make them more difficult to audit than conventional system designs because they restrict the information that parties can access or obtain. Furthermore, PPDSA technologies may not fully resolve questions of whether the data have been legitimately obtained or used (e.g., with meaningful individual consent) and whether the nature or quality of the data could create harmful bias in the results.

Public trust will hinge on the justified assurance that government, academic, and industry use of PPDSA solutions will respect privacy, civil liberties, and civil rights. The future PPDSA ecosystem must be transparent and inclusive and reflect privacy principles and preferences (e.g., individual participation and consent management). Achieving these outcomes will require developing systems informed by ethical, social, behavioral, and economic factors as well as human-centered design principles. Governance around the design, development, and deployment of PPDSA technologies will provide guardrails and accountability for responsible and ethical use. The Federal Chief Data Officers Council's adoption of the Data Ethics Framework with civil liberties and privacy inputs provides an encouraging way forward and helps Federal leaders and data users make privacy-conscious and ethical decisions as they acquire, manage, and use technology and data in support of their agencies' missions.[25] The Blueprint for an AI Bill of Rights[26] released by OSTP in October 2022 similarly guides the principles that should guide the design, use, and deployment of automated systems, including in terms of data privacy, data collection, and data use. Globally, the Federal Government can leverage partnerships to advocate for privacy-protecting technical standards and norms within and across international bodies.

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.

This Strategy envisions a future in which PPDSA technologies enable data-driven scientific research and innovations that will improve health, well-being, and prosperity. Individuals are not the only beneficiaries. Competing organizations in the market sector will be able to share and process data in ways not possible today without undermining individual privacy and intellectual property protections. However, it is imperative that everyone benefits equitably from the insights gleaned from the data. In keeping with the democratic values of equity and fairness, the entities creating and using these PPDSA technologies will prioritize equitable access to these benefits.

Ensuring that everyone benefits equitably requires the proper alignment of incentives to foster a data-enabled inclusive economy. When combined, the power of economies of scale, AI, and network effects could lead to market concentration or anticompetitive practices that undermine innovation or incentives for the responsible use of data. PPDSA technologies have the potential to enable smaller businesses or would-be market entrants to harness data to fuel responsible business models without being reliant on directly collecting, accessing, or licensing large amounts of data. Appropriate policy and technical guardrails and approaches to democratize PPDSA technologies and enhance equitable access to data can address these risks and foster innovation and robust market competition.

PPDSA technologies will be trustworthy and will be created and used in a manner that upholds accountability.

The use of PPDSA technologies will be built on a foundation of trust. PPDSA technologies need to be trustworthy, meaning that from a technical perspective, they need to be able to perform as intended. As PPDSA technologies evolve, it will be necessary to establish a framework for independent verification of these systems even as the technology continues to mature. To mitigate risk, organizations will need to have a means to rigorously test, evaluate, and continuously monitor the performance of PPDSA technologies. The development of risk and performance metrics and diverse testbeds will support such evaluation. Standards, certifications, frameworks, guidelines, and tools will further support the trustworthiness of PPDSA technologies, and in turn, provide accountability for their use.

However, accountability for the use of PPDSA technologies requires more than trustworthiness. As noted, the implementation of PPDSA technologies alone may not prevent the shared data and analytic

outputs from being used for non-democratic or otherwise harmful purposes. Entities engaging in data sharing and analytics must be accountable for their decisions and actions. Embedding the design, development, and deployment of PPDSA technologies in a larger framework that encompasses legal, regulatory, ethical, and policy mechanisms will help to create this level of accountability. This larger frame of accountability includes employing environmentally sustainable best practices in the design, development, and deployment of PPDSA technologies to minimize associated harms such as unsustainable energy use.

Technologies, laws, and policies that are developed in an open, inclusive, and transparent manner and leverage the expertise of a broad set of actively engaged collaborators, including the public sector, industry, academia, and civil society, will support accountability.[27] Design, development, and deployment of mechanisms to maximize openness and transparency within the technologies themselves (e.g., open-source software) should be prioritized. Organizations will also need the capabilities and procedures to respond to incidents in the event of sensitive data disclosures or reidentification risks.

Risk management can be used to support ethical decision-making to derive the benefits described in this Strategy while minimizing harmful impacts on individuals and society at large. Users of PPDSA technologies need the requisite knowledge, skills, and techniques to identify, prioritize, and respond to risks associated with data sharing and analytics activities. Therefore, effective education, awareness, and training for the responsible and ethical use of data and technologies, including applicable academic programs, workforce hiring pathways, and relevant government procurement activities will be needed. Conducting and publishing privacy impact assessments is another way to support transparency, demonstrate risk mitigation strategies, and maintain public trust. Existing risk management frameworks, such as the NIST Privacy Framework,[28] Cybersecurity Framework,[29] and AI Risk Management Framework[30] can be used to support these needs.

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.

As noted previously, the use of PPDSA technologies should not contribute disproportionately to unintended or harmful outcomes for underserved, marginalized, and vulnerable communities. Such harms may arise in diverse ways. In some cases, the implementation of the technology may create harm. For instance, adding noise in data to protect privacy while training an ML model typically reduces the accuracy of the trained model; such accuracy loss can be disproportionately worse for the underrepresented classes and subgroups in the training data.[31] In addition, when such noising techniques are used, the bias already present in the original data may be further amplified, which can lead to disproportionately increased harm to some groups represented in the data.

Ensuring equitable benefits from PPDSA technologies will require the inclusion of diverse viewpoints, expertise, and perspectives in PPDSA research and development (R&D). This includes domain experts that can best identify relevant statistics and insights to be preserved or generated from the data to inform the application of a PPDSA technology to maintain the necessary integrity of those measures. Taxonomies and methodologies for formal analysis of relevant risks and harms based on empirical studies will help provide foundations for advances that avoid negative outcomes. Auditing requirements and criteria, as well as rigorous benchmarks for success, will help facilitate meaningful and accountable implementation of this principle into the design, development, and deployment of PPDSA technologies.

Participants in the PPDSA Ecosystem

As noted, organizations that will benefit from an improved ability to share and analyze data include all levels of government, academia, and all sectors of industry. Relevant parties also include developers of PPDSA technologies, policymakers, and individuals who are the subject of or affected by data sharing and analytics. Many contributors within the PPDSA ecosystem can and should partner to design, develop, and deploy these technologies effectively and responsibly.

With the use of PPDSA technologies, individuals can feel more confident that their privacy will be maintained and may thus be more likely to participate in data sharing and analysis activities. PPDSA technology developers and deployers have a responsibility to communicate the technology performance capabilities and limitations in their systems, including their determinations about acceptable privacy-accuracy or functionality tradeoffs and any supporting measures to address or help mitigate these tradeoffs. Researchers, data scientists, or other third-party entities using the data and individuals or groups who could be harmed, especially already marginalized or vulnerable groups, must have a role in the design, development, and deployment process to communicate concerns or expose areas where more research is needed to improve the technologies or implementation solutions. This level of public participation can be facilitated through capacity building and the provision of tools and frameworks that facilitate communication and collaboration across technical and non-technical communities.

Civil society organizations and institutions can offer multidisciplinary subject matter expertise and diverse perspectives to support the activities necessary for achieving the vision put forward in this Strategy. Organizations dedicated to pertinent issues such as consumer rights, civil and human rights, privacy, data science, and emerging technologies can provide guidance to help ensure that the development and deployment of PPDSA solutions follow the guiding principles defined above and move the ecosystem toward the envisioned future. Civil society organizations also stand to benefit greatly from the insights that can be gleaned from PPDSA solutions to further public interest missions and goals, and it will be important to build their capacity to employ and engage with PPDSA technologies.

2: Current State

The theoretical foundations for many PPDSA tools and techniques (e.g., secure multiparty computation, homomorphic encryption) appeared in the late 1970s and 1980s alongside many of the advancements in computer science, networking, and the internet, often supported by Federal research funding. In parallel, statistical disclosure limitation techniques were advancing within the statistical profession, led by Federal statistical agencies.[32] As the global internet and advances in computing continued to mature throughout the 1990s and 2000s, techniques to prevent the identification or re-identification of data subjects became more relevant. Statistical disclosure limitation techniques are widely used today to manage the re-identification risk of publicly available data products in statistical organizations within and outside of government, along with a series of restricted access modalities such as physical and virtual data enclaves to provide secure access to data when public products are too risky.[33] In the pursuit of techniques that provide more provable guarantees for privacy, those same organizations are also investing in PPDSA-related technologies to expand their toolset.

Recent years have seen a dramatic shift in the role of PPDSA technologies more broadly, from generally beneficial to essential, although the relative maturity of each PPDSA approach varies considerably. As the value and demand for data continue to rapidly expand, so do privacy risks, creating a need for new solutions. Further, advances in AI systems have driven demand for data exponentially, with large language models and generative AI models reaching unprecedented performance through computationally intensive approaches based on large corpora of training data. Legislation and Federal initiatives have also elevated and prioritized the use of data for evidence-based policymaking and the sharing of scientific data from federally funded research. This landscape of technical capabilities, foundationally based on data, evolves daily and presents a unique moment for the convergence of techniques that fundamentally enhance privacy with those that are designed to derive insights from data. At the same time, the proliferation of data protection and privacy regulations and laws around the world and within the United States creates a complex regulatory environment, which may not account for new technologies.

The following examination of the current state captures some of the critical aspects of the legal environment, key challenges in adoption, and capabilities of specific tools and techniques.

Legal and Regulatory Environment

At the Federal level, the use and protection of personal data in the United States are governed by a set of risk-based, sector-specific laws and regulations that extend to different data types or uses. Some, such as the Privacy Act of 1974[34] and the Confidential Information Protection and Statistical Efficiency Act of 2002,[35] apply to how the Federal government handles certain information. Others, such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996,[36] the Children's Online Privacy Protection Act of 1998,[37] the Gramm-Leach-Bliley Act,[38] and the Family Educational Rights and Privacy Act of 1974[39] expand the scope of application to specified private-sector entities to protect specific types of personal data. Following the enactment of the California Consumer Privacy Act of 2018,[40] there are also an increasing number of state laws that provide more general protections.

Governing a key area for adoption of PPDSA approaches, the HIPAA Privacy Rule protects the privacy of individually identifiable health information held by covered entities, such as health care providers. The Privacy Rule requires authorization from individuals or a waiver of authorizations by an Institutional Review Board (IRB) or Privacy Board to disclose individually identifiable health information.[41]

More broadly, the Federal Policy for the Protection of Human Subjects,[42] referred to as the "Common Rule," sets forth requirements for the protection of human subjects in research.

Globally, data protection regulations have proliferated in the last few years following the introduction of the European Union's General Data Protection Regulation (GDPR),[43] including in countries such as Brazil,[44] China,[45] Canada,[46] and Singapore.[47] These create a complex regulatory landscape and cross-jurisdictional issues that add to the challenges of developing and deploying effective PPDSA technologies for cross-border research and applications.

In response to recent growth in AI and ML research and data-driven consumer products, several regulatory agencies, such as the Food and Drug Administration,[48] the Consumer Financial Protection Bureau,[49] and the Federal Trade Commission, have announced guidance and plans for advanced rulemaking[50] to address issues related to safety and privacy in the context of AI- and data-based systems and practices. However, similar regulatory efforts to clarify how PPDSA technologies do or do not meet various legal restrictions related to data sharing and protection have not yet been announced.

Key Challenges

While some PPDSA technologies have seen initial deployments in commercial products and public sector uses, others still have technical challenges to overcome to enable broader use (discussed in detail below). Nonetheless, a set of common challenges have hindered widespread adoption, including:

Limited awareness. Knowledge and understanding about PPDSA technologies in general are sparse among potential users in the public and private sectors. The capabilities afforded by PPDSA technologies are not widely known, understood, or appreciated outside of the research community. Furthermore, many potential users who have heard of the technologies are under the impression that there are no mature, useable PPDSA technologies.

Inconsistent definitions and taxonomy. An absence of standard definitions or taxonomies related to privacy and PPDSA technologies makes understanding privacy risks and considerations difficult and inconsistent among the parties. This has resulted in a lack of clarity about the appropriateness and use of available PPDSA technologies for various scenarios and how to manage privacy alongside functionality.

Inadequate understanding of privacy technology risks and benefits. Potential users of PPDSA technology need a way of understanding and evaluating the risks and benefits that the technology entails, including potential harms, and mapping technologies to relevant threat models. Few techniques offer absolute security, meaning that they cannot be compromised by an adversary who has unlimited computational resources and time. However, most practical techniques are designed to ensure that an adversary cannot compromise them in a reasonable amount of time. Mitigating risk, therefore, involves understanding the threat models and adversarial capabilities.

Lack of consensus standards. Few widely adopted standards specifying the mechanisms or use of PPDSA technologies exist at this time. Although an open initiative led by industry and academia has established a homomorphic encryption standard[51] and a standard for zero-knowledge proofs[52] is underway, more standards that specify foundational cryptographic primitives and other PPDSA technologies would facilitate adoption and trust in solutions. In addition, there are no widely adopted standards for data formats, application programming interfaces, or system architectures that are necessary to facilitate the interoperability and deployment of PPDSA technologies. Promotion of consensus standards, however, should not dissuade those considering PPDSA solutions from leveraging the tremendous flexibility that

many PPDSA techniques offer in terms of customizing implementation to best suit their data and use cases.

Varying stages of maturity. Existing PPDSA technologies are in varying levels of maturity with some having achieved initial success in terms of real-world adoption, particularly for simpler applications. Years of theoretical research in cryptography, for instance, have led to secure multiparty computation and homomorphic encryption techniques becoming more practical. However, these approaches still have scalability and efficiency challenges that need to be addressed in the context of a broader set of threat models to enable wider adoption. This is further complicated by a complex global regulatory environment with varying compliance requirements for personal data protection. In many cases, a combination of various PPDSA techniques needs to be employed to ensure end-to-end privacy that is measurable or formally provable, which remains a research challenge.

Insufficient usability of solutions. Current PPDSA approaches must be tailored to each specific deployment and require significant technical expertise to implement. There is not an appropriate set of usable tools and interfaces whose design is informed by human-centric design principles and social, economic, and behavioral factors that make the solutions easily deployable, configurable, accessible, and manageable by a diverse set of users with varying levels of capabilities. This hinders broader adoption among organizations that do not have specialized expertise and creates high deployment burdens, particularly in sectors that are already overtaxed or have low technical maturity. Even if significant specialized expertise is used to deploy such a technology, it can become ineffective if it is not used properly due to usability challenges, inconsistent assumptions about behavioral and economic issues, or misaligned incentives.

Inadequate implementation assessment capabilities and management of tradeoffs between privacy and other issues. Approaches that address privacy in a broader context of competing issues are underdeveloped. For instance, techniques such as differential privacy provide formally provable privacy guarantees but require more analysis to assess potential impacts on fairness and bias, as well as tradeoffs with accuracy in different implementations which can affect the degree of utility for data users. Effectively assessing the implementation of PPDSA solutions in terms of vulnerability in real-world deployment is another challenge, and the lack of standards makes it difficult for third-party auditors to evaluate the strength of privacy protections. There is also a lack of mature approaches for measurement and metrics concerning privacy risks and harms, which could help assess the efficacy of solutions and for devising risk-aware solutions.

Lack of clarity around regulatory compliance. Cross-functional partners (i.e., security, privacy, and general counsel) may struggle to determine when or if the use of PPDSA technology is consistent or in compliance with legal or regulatory requirements. Current laws and regulations do not adequately envision a role for PPDSA technologies and legal standards for privacy protection and interpretations of these standards do not directly address the question of whether the use of PPDSA technologies is sufficient to satisfy their requirements.

Impact on individuals, marginalized groups, and society. The use of PPDSA technology does not guarantee the safe use of data nor can it control for cases of illegitimate collection of data. Furthermore, certain approaches could produce analytical results that amplify or introduce bias, and such results could lead to decisions that are discriminatory or disproportionately harmful to certain subgroups represented in the data. PPDSA technologies in the early stages of development may be poorly understood and deployed or adopted in privacy-invasive and potentially unsafe ways that risk re-identification.

The above list is not meant to be exhaustive and must also be addressed within the context of today's fast-evolving emerging technological landscape.

Overview of PPDSA Capabilities

As a result of years of research, various PPDSA technologies have emerged that show promise for helping achieve the vision put forward by this Strategy. These include cryptographic and non-cryptographic techniques and span hardware- and software-based approaches. They address various privacy priorities, such as pseudonymization or anonymization of sensitive data and statistical disclosure control and promote confidentiality and disassociability. Table 1 reviews the key existing technical approaches that are essential for PPDSA and an in-depth discussion follows.

Table 1. Overview of Key Technical Approaches Essential for PPDSA.

Technique	Description	Value	Limitations
K-anonymity	Transforms a given set of *k* records in such a way that in the published version, each individual is indistinguishable from the others	Reduces the risk of re-identification	Vulnerable to reidentification attack if additional public information is available
Differential Privacy	Adds noise to the original data in such a way that an adversary cannot tell whether any individual's data was or was not included in the original dataset	Provides formal guarantee of privacy by reducing the likelihood of data reconstruction or linkage attacks	Limited to simpler data types; challenge in managing tradeoff between privacy, accuracy, or utility of data
Synthetic Data	Information that is artificially manufactured as an alternative to real-world data	Preserves the overall properties or characteristics of the original dataset	May still disclose privacy-sensitive information contained in the original dataset; difficult to mirror real-world data
Secure Multiparty Computation	Allows multiple parties to jointly perform an agreed computation over their private data, while allowing each party to learn only the final computational output	Increases the ability to compute over distributed datasets without revealing original data	Higher computational and communication costs/burdens, and difficult to scale
Homomorphic Encryption	Allows computing over encrypted data to produce results in an encrypted form	Only authorized users can see original and/or computed data	Higher computational cost and time
Zero-Knowledge Proof	Allows one party to prove to another party that a particular statement is true without revealing privacy-sensitive information	Increases ability to validate information without disclosing sensitive information	Cost and scalability
Trusted Execution Environment	Creates a secure, isolated execution environment parallel to the main operating system to process sensitive data	Allows faster secure analytics on data compared to encryption-based techniques	Introduces other ways sensitive data can leak
Federated Learning	Allows multiple entities to collaborate in building an ML model on distributed data without sharing original data	Minimizes data sharing while training a combined model	Various data reconstruction or inference attacks are still possible; require consistency across datasets held by multiple entities

Data anonymization and statistical disclosure limitation techniques. These techniques address privacy risks in publishing data by transforming the original data to limit the disclosure of sensitive information or prevent the re-identification of individuals or groups represented in the data. Challenges inherent in these approaches relate to how to balance privacy goals and the accuracy of the published data when used for various types of analysis. The capabilities of anonymization or disclosure techniques are currently limited to only simpler types of data, such as tabular data.

- *k-anonymity* is an anonymization technique that transforms a given set of k records in such a way that in the published version, each individual is indistinguishable from the others. Extensive research in k-anonymity and its extensions has shown that such techniques are vulnerable to reconstruction or linkage attacks[53] that can lead to the re-identification of a data subject if an adversary has relevant auxiliary information. k-anonymity has been recently used for the publication of COVID-19[54] datasets by the U.S. Centers for Disease Control and Prevention.[55]
- *Differential privacy*, a data perturbation approach, adds noise to the original data in such a way that an adversary cannot tell whether any individual's data was or was not included in the original dataset. Such approaches aim to ensure that statistical information computed using the published data remains statistically valid. Differential privacy has gained popularity because of its rigorous mathematical foundation. Initial deployments have not yet produced a generalizable method of how to best set the privacy parameter to control the strength of the privacy guarantee while optimizing for accurate analytic results. It has been used by the Census Bureau in the publication of 2020 Census data (see case study below).
- *Synthetic data* approach involves generating simulated data that preserves the overall statistical properties or characteristics of the original dataset without revealing its sensitive information. This approach is increasingly popular for training AI-based applications for image recognition[56] and healthcare,[57] as large volumes of synthetic data can be easily made available for training AI models. Research shows that synthetic data may still disclose privacy-sensitive information contained in the original dataset unless techniques such as differential privacy are used.[58]

Cryptographic techniques. Years of research have resulted in the development of various cryptographic techniques that support computation over private data. Scalability and efficiency issues remain key challenges for the practical deployment of many cryptographic techniques. Computation and communication costs become challenging for complex analytics (e.g., deep learning) or as the number of collaborating parties or data sizes increases.

- *Secure multiparty computation* allows multiple parties to jointly perform an agreed computation over their private data while allowing each party to only learn the final computational output. Computations can be agreed for a common (public) output, such as everyone learning an average salary, or can be tailored to distinct private outputs per party.[59] After over 30 years of theoretical research, secure multiparty computation is increasingly being seen as a viable PPDSA solution that can be used to support simple privacy-preserving computational tasks such as computing statistics or regression analysis. A special case of secure multiparty computation is private set intersection, which identifies overlapping elements from multiple datasets without revealing any other sensitive data.
- *Homomorphic encryption* allows computing over encrypted data to produce results in an encrypted form. Then, only users with appropriate keys can extract the result from its encrypted form. For example, additive homomorphic encryption allows computing a sum of two encrypted

values to produce the encryption of the sum of their original values. Fully homomorphic encryption supports arbitrary computation over encrypted data. Recent efforts led to the development of a homomorphic encryption standard in 2018.[60]

- *Zero-knowledge proof* is another well-known technique that allows one party, called a prover, to provide proof to convince another party, a verifier, that a particular statement is true without revealing privacy-sensitive information.[61] For example, some digital assets use zero-knowledge proofs to prove statements about transactions and balances without revealing additional metadata.[62] Zero-knowledge proofs can thus be used to support auditability or compliance checks. More recently, zero-knowledge proof schemes are being used for convolutional neural networks[63] to prove that the prediction task was carried out by the model itself, without disclosing any model information.
- *Functional encryption* allows computing a value of a mathematical function over some data using the encrypted form of that data. Special cases of functional encryption include identity-based encryption and attribute-based encryption, which allow encrypted data to be decrypted only by parties that satisfy specific criteria (identity or attributes).

Trusted execution environments. A trusted execution environment creates a secure, isolated execution environment parallel to the main operating system[64] to process sensitive data. It ensures verifiability of security regarding sensitive data processing or code execution, trusted input-output operations, and secure storage for data that can be accessed by only authorized entities at any time. Trusted execution environments may use hardware- or virtualization-based isolation techniques, and they can be made available on cloud platforms. Trusted execution environments provide a faster alternative to cryptographic approaches but bring additional privacy risks such as side-channel leakage of information.

Policy-based approaches. The development of machine-readable privacy policies allows a specification of rules that capture privacy protection requirements based on users' privacy preferences for sharing or analytic use of data. Such policies specify who can access, query, or use the privacy-sensitive information or analytics results for a specified purpose. Access control approaches, such as the purpose-aware access control approach,[65] and "sticky policies"[66] have been used to capture the purpose of data processing and operation to restrict authorized access to and use of shared data. Similarly, the World Wide Web Consortium has proposed the Platform for Privacy Preferences Project 1.0 (P3P),[67] which allows for a Web site to encode its data collection and data use practices in a machine-readable format. Attribute- or identity-based encryption, mentioned earlier, can be used to enable organizations to outsource data to third-party cloud service providers to share it with authorized entities in a privacy-preserving manner.[68] Ongoing research in this area focuses on effectively and efficiently specifying and enforcing dynamically evolving access and privacy policies. Enforcement of such policies may involve the application of other cryptographic or non-cryptographic techniques. Natural language processing techniques are being explored to analyze legal privacy policies to enable a system to understand privacy protection requirements[69] or generate machine-readable privacy policies that need to be enforced.

Other approaches. In addition to the above, other techniques exist for PPDSA that may use cryptographic or non-cryptographic approaches or some combination thereof. In many cases, data about individuals may be distributed across multiple organizations (e.g., different hospitals), each of which may be available in some anonymized form.

- *Privacy-preserving record linkage (PPRL)* allows secure and private linkage of data related to an individual from different datasets. For example, the NIH National COVID Cohort Collaborative[70] has created the largest national, publicly available patient-level limited dataset in U.S. history by

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.

Case Study: U.S. Census Bureau 2020 Census
Challenge: Data captured through the census contains valuable information for researchers nationwide. The Census Bureau is charged with publishing accurate statistics while protecting the confidentiality of personally identifiable information of any census participants. The Census Bureau sought ways to publish accurate statistics while protecting the privacy of individuals. In 2018, the Census Bureau conducted a simulated database reconstruction-abetted re-identification attack on the 2010 Census data and found that it can successfully re-identify a significant portion of the 2010 Census data. 2010 Census data was published using a technique that injected noise into the original "data by swapping records"[73] for de-identification. ***Approach***: To address the database reconstruction attack demonstrated on 2010 Census data, the Census Bureau developed a disclosure avoidance framework[74] based on differential privacy for publishing 2020 Census data. The differential privacy technique adds noise to the original data to obscure the presence or absence of any individual in a dataset while maintaining the statistical characteristics of the original data. Based on rigorous mathematical foundations, differential privacy allows tuning a privacy-loss budget to balance disclosure risk with accuracy loss. This technique allowed the Census Bureau to limit the disclosure risk for published data while maintaining accuracy. ***Impact***: The highest standards and emerging techniques in PPDSA technologies are being used to uphold privacy and disclosure avoidance for the American public. Compared to the use of differential privacy, the use of legacy statistical disclosure limitation techniques to comply with the statutory confidentiality obligations would require perturbing data with "noise" at levels that would make the published census data inadequate for many uses. Differential privacy also provides provable guarantees against a range of potential privacy attacks and allows the control of disclosure risk through the use of a privacy-loss budget. An example of the differentially private 2020 Census data is the formulation of redistricting plans for elected offices from the U.S. House of Representatives to local school boards in compliance with Federal voting rights laws.

3: Strategic Priorities and Recommended Actions

PPDSA technologies need to progress from the current state toward a future data ecosystem, where the envisioned technologies advance well-being and prosperity, promote innovation, and affirm democratic values. This requires concerted actions across the government, private sector, and civil society. To achieve this, five strategic priorities and their associated recommendations focus on accelerating research and maturing and advancing the adoption of PPDSA technologies.

Strategic Priority 1: Advance Governance and Responsible Adoption

PPDSA technologies offer tremendous potential for advancing science and innovation and protecting privacy. However, the adoption of PPDSA technologies on their own will not achieve the future state envisioned by this Strategy. Institutional efforts, national-level guidance, and proactive risk-mitigation measures are needed to ensure the advancement of PPDSA technologies in a manner that adheres to the Guiding Principles laid out in Section 1.

Recommendation 1.a. Establish a steering group to support PPDSA guiding principles and strategic priorities

A steering group that includes a broad range of representatives (e.g., public sector, industry, academia, civil society, and marginalized and vulnerable groups) could serve as a source of policy and technical expertise. The Federal Government could initially establish the steering group, set milestones, and measure progress, but over time transition the group to self-sustaining leadership. Multistakeholder efforts around digital identity that were initiated by the Federal government and eventually transitioned to community leadership in response to the National Strategy for Trusted Identities in Cyberspace[75] serve as a model for similar efforts to advance PPDSA. Ideally, the steering group will serve to convene broadly representative and inclusive partners to draw the greatest range of expertise in carrying out its mission and work to develop and maintain a healthy PPDSA ecosystem that will achieve the vision laid out by this Strategy. This would build a flourishing PPDSA ecosystem that respects democratic values.

The steering group should:

- Define the minimum rights and responsibilities of participants in the PPDSA ecosystem in alignment with the Guiding Principles set out in this Strategy, including:
    - addressing concerns about the use of analytic outputs in ways that could undermine democratic values
    - drawing on other frameworks to address broader privacy issues around collecting and controlling data
    - considering how to manage large private datasets to minimize the creation of data monopolies that could stifle innovation and amplify inequity in data-driven economies
- Administer processes for adopting standards and certification mechanisms that underpin the trustworthiness of PPDSA technologies.

Recommendation 1.b. Clarify the use of PPDSA technologies within the statutory and regulatory environments

PPDSA technologies are becoming increasingly feasible for real-world applications in many sectors. It is not clear, however, whether PPDSA technologies meet the requirements of existing privacy laws, policies, and regulations. These provide essential protections to the public but were primarily enacted

before PPDSA technologies were viable.[76] For example, it has not been determined to what extent PPDSA technologies meet specific de-identification standards or security safeguards that provide critical protections to individuals. Nor is it clear how the technologies will coalesce with important legal definitions or exceptions, such as "personally identifiable information" or "disclosure." In addition, PPDSA technologies will enable the processing of datasets protected by different privacy laws, complicating compliance with multiple legal regimes. To address this uncertainty, regulatory bodies should engage with developers, privacy professionals, and users of PPDSA technologies to ensure a mutual understanding of how the technologies work within the context of current regulations.

Adoption of PPDSA technologies also requires increased awareness of these technologies among policymakers and legal professionals. Agencies, professional associations, and legislatures should bolster their technical expertise to assist with writing laws, regulations, and guidance and assessment. Education and training should be prioritized to support the application of current laws and regulations, thus ensuring that any future privacy legislation is informed by technical developments.

Recommendation 1.c. Develop capabilities and procedures to mitigate privacy incidents

PPDSA technologies have great potential to protect privacy, but no technology can ensure that privacy incidents, such as the disclosure of sensitive data, will never occur. There are existing procedures to manage privacy incidents, but the technical complexity of PPDSA technologies may lead to situations where organizations do not realize that a privacy incident has occurred or understand how it happened. Organizations should undertake continuous monitoring and be prepared to respond to potential privacy risks or actual incidents. Risk management and active preparation to respond to information leakage are key components of responsible PPDSA technology implementation, which should be built into policies and procedures for both government and non-government deployments. As PPDSA technologies are developed, piloted, and implemented, work is needed to develop robust proactive-risk mitigation measures and incident response policies and procedures to maintain public confidence in PPDSA technologies.

Strategic Priority 2: Elevate and Promote Foundational and Use-inspired Research

In complement with Strategic Priority 1, Priority 2 highlights the importance of interdisciplinary research that combines technology, social science, law, policy, and domain expertise that will enable a sociotechnical approach to developing and implementing PPDSA solutions. This type of holistic approach will allow the legal and regulatory environment to evolve in tandem with PPDSA technologies, ensuring the two augment one another in protecting and preserving privacy rights.

Both foundational and use-inspired PPDSA R&D is necessary to keep pace with ever-growing privacy risks, the rapid growth of data collection, sharing, and analysis, and the development of emerging technologies, e.g., smart sensors, advanced wireless networks, AI/ML, and quantum computing. A successful PPDSA research strategy will enable security and privacy-by-design approaches ensuring PPDSA technologies are integrated at the earliest stages of system development. The Federal Government should promote increased engagement with Historically Black Colleges and Universities (HBCUs), Tribal Colleges and Universities (TCUs), and other Minority-Serving Institutions (MSIs), to strengthen, sustain, and broaden U.S. PPDSA research capacity and expertise. Through the expansion of PPDSA research capacity, students will gain knowledge and help create the technical capacity needed to provide solutions for a wide range of public and private sectors. A recent National Academies[77] report stated that "[m]ost of the current public scientific expertise in algorithm design, cryptanalysis, and other areas of applied cryptography is outside the United States." Thus, broadening research expertise is

critically important to secure U.S. leadership in PPDSA technologies. A broadly inclusive PPDSA research workforce can help advance goals of inclusion and equity in the privacy field.

Recommendation 2.a. Develop a holistic scientific understanding of privacy threats, attacks, and harms

It is important to develop scientifically well-founded taxonomies to systematically classify privacy threats, attacks, and harms in the context of various threat actors. This holistic approach should also provide a framework for assisting organizations in evaluating benefits alongside privacy risks. These taxonomies need to draw from the terminological and conceptual differences in key concepts around privacy between different contributors and disciplines (e.g., computer science, information science, law, policy, and social science). Interdisciplinary taxonomies are critical to building a systematic knowledge base and formal methodologies for reasoning about privacy guarantees provided by PPDSA solutions. Understanding these threats, attacks, and harms should be informed by contextual factors, i.e., socio-economic, and cultural issues that influence interactions with technology, have impacts on marginalized and vulnerable groups, and address regulations, including cross-border jurisdictional issues. R&D efforts should allow for risk- and harm-aware PPDSA solutions and support the development of appropriate metrics and measurements for privacy risks and harms.

Recommendation 2.b. Invest in foundational and use-inspired R&D for PPDSA technologies

Substantial and sustained investment in both the public and private sectors should support accelerated R&D that is focused on emerging PPDSA technologies and bold exploratory research targeted to create the next generation of PPDSA capabilities. Sustained research activities in areas of differential privacy, secure multiparty computation, homomorphic encryption, and other technologies have led to the development of PPDSA capabilities that are starting to transition into practice. More research from a broad set of stakeholders is needed so that the full potential of these PPDSA technologies is investigated and developed leading to scientifically well-grounded tools, models, methodologies, frameworks, and evaluation methods. Such advancement will enable data-driven collaborations and responsible data sharing, including maximizing public access to research data generated by federally funded research projects. Technical research into PPDSA technologies should be accompanied by research from a sociotechnical and human-centric design perspective that considers how the technology will interact with individuals, communities, institutions, and society as a whole.

Recommendation 2.b.1. Accelerate R&D for current and emerging PPDSA technologies. Current PPDSA capabilities, as presented earlier, include technologies at varying levels of maturity and limitations. Support for R&D should focus on addressing key challenges:

Complex and heterogeneous datasets. Several PPDSA techniques are limited by the types of data they can handle. Many of the current and emerging application areas for PPDSA technologies (e.g., social networks, personalized medicine, augmented and/or virtual reality, and smart cities) will generate massive amounts of diverse data types, such as multimedia (e.g., text, images, audio, and video), graphical (e.g., social networks), or real-time streaming data. Datasets to be shared or processed may be structured or unstructured and include multiple modalities. Significant R&D efforts are needed to advance PPDSA solutions that can deal with these complex datasets, such as the following:

- Research is needed to develop and effectively use techniques that will address tradeoffs between privacy guarantees, and utility or accuracy. It is also important to develop better cryptographic primitives and protocols to process multimodal datasets in various application contexts.

- Effective verification and validation approaches are needed to tackle accuracy and data quality issues when synthetic data are used. Research should address the effectiveness of such techniques when different types of datasets, analytics, or AI model training approaches are considered.
- The highly distributed nature of such heterogeneous data introduces significant research challenges for privacy protection when integrated (e.g., for data synthesis) or used for collaborative analytics or ML training.
- Research is needed to develop more effective approaches for granular specification; enforcement of access, use, or privacy policies; and the verification of correctness for policies across a variety of data types and applications. Correct enforcement of such policies may need to use combinations of cryptographic or non-cryptographic techniques.
- Techniques to extract rules from legal and organizational privacy policies or laws and convert them to machine-readable policies require additional development, especially considering the broad set of data types. As data objects are shared, effective ways to ensure that unauthorized sharing is prevented or detected are needed. An associated challenge is privacy-preserving access that will protect the privacy of a user accessing sensitive information.

Scalability and efficiency. Scalability and efficiency pose significant roadblocks to the practical deployment of many PPDSA technologies, particularly those that employ cryptographic techniques. Efficient and scalable approaches are needed to ensure their use in various technological contexts (e.g., IoT, edge and cloud platforms, or AI systems). The complexity of computation in emerging applications further adds to scalability and efficiency challenges, as does complex, real-time, heterogeneous, and distributed data. While more efficient cryptographic primitives need to be developed, hardware-based techniques, such as hardware accelerators or trusted execution environment-based approaches are viable faster alternatives. Sustained research is needed to further mature such hardware-based approaches.

Programmability and verifiability. An important research area is the development of high-level programming languages to improve the programmability of various PPDSA technologies (e.g., secure multiparty computation), which are typically hand-coded and optimized by experts. Such programming languages and their associated compiler toolchains would provide abstractions across multiple PPDSA solutions, a critical step to making them more practical. Research areas for innovative verification and validation are important to support the correctness of PPDSA techniques at design, implementation, and deployment stages. Many PPDSA techniques also need to be comprehensively investigated by considering a broader set of emerging threats across various sociotechnical and application contexts. For instance, in an honest-but-curious model, a server behaves honestly but may try to learn sensitive information. Existing secure multiparty computation solutions typically assume this type of model, which may be adequate for some applications. Often PPDSA solutions are developed in a piecemeal fashion or are standalone, but real-world deployment typically requires integrating different solutions to ensure comprehensive privacy and security. Integrating such solutions that are potentially designed from different threat models and then verifying the correctness of the integrated solution is an important research area. Formal and systems-integration techniques are needed to achieve end-to-end privacy and performance goals.

Metrics and measurements. Research needs to be accelerated to develop metrics for PPDSA technologies and effective measurement techniques—quantitative or qualitative—for privacy risks, accuracy, and associated unintended consequences or harms. Techniques (e.g., differential privacy) use

a privacy-loss parameter to capture the privacy disclosure that is acceptable, but there remains an inadequate level of understanding of how privacy parameter values should be set for different applications. Configuring privacy-related parameters becomes more challenging when different types of PPDSA techniques need to be integrated for comprehensive solutions, such as in privacy-preserving federated learning. Based on such metrics, effective assessment techniques are needed to define privacy risks and harms posed by a PPDSA solution for a specific application.

Fairness, transparency, and accountability. It is important to ensure that PPDSA solutions do not inadvertently amplify the bias that is already present in the original data. Such bias amplification issues are possible when certain techniques (e.g., k-anonymity or differential privacy) are used. Research should consider how PPDSA solutions ensure fairness. Research is also needed to address the lack of effective techniques to support and emphasize transparency and accountability when PPDSA solutions are used. For example, some regulations introduce additional data protection requirements, such as the "right to be forgotten." Solutions may further be augmented by a suite of innovative privacy auditing or compliance checking tools. For instance, very nascent research on machine unlearning techniques aims to address how to remove some training data elements and then retrain the associated AI/ML model efficiently, all in a privacy-preserving manner. These issues become even more challenging in federated learning scenarios. As PPDSA solutions are integrated into AI systems, explainability, and transparency should be emphasized.

Recommendation 2.b.2. Promote future-focused exploratory R&D with transformational goals. Current and emerging PPDSA technologies may need significant R&D efforts to enable their broader applicability and adoption. Additionally, concerted R&D should also support the more future-focused exploration of innovations in PPDSA solutions. The research should be geared towards fundamental technical approaches that focus on rapidly advancing or emerging technologies, e.g., next-generation AI, quantum computing, 6G and beyond networking, augmented/virtual reality, and digital assets. These new advances are expected to make computational and analytics capabilities, and the level of connectivity that facilitates information sharing to be orders of magnitude higher. While these environments present unprecedented opportunities for social good, bad actors will be equally empowered with immense computational, communication, and AI capabilities to infringe upon the privacy, security, and safety of individuals and communities. In such an anticipated future, it can be expected that the security and privacy threat landscape, or attack surface, will be substantially enlarged. Understanding emerging threats and designing appropriate defenses early is critical. Moreover, it is thus essential to pursue or support long-term, bold, and exploratory R&D efforts that are focused on developing the next generation PPDSA foundations (e.g., new cryptographic primitives), models, methodologies, and frameworks that can be resilient against a broad range of privacy threats. Research should emphasize privacy-by-design approaches that also consider how privacy may conflict with various other properties, such as security, resilience, and bias/fairness.

Recommendation 2.c. Expand and promote interdisciplinary R&D at the intersection of science, technology, policy, and law

While R&D highlighted in the preceding recommendations aims to provide key PPDSA technological foundations, it is important to emphasize that privacy is a highly multidisciplinary field, and significant R&D efforts should be fostered to develop and implement effective, usable, and socially responsible PPDSA solutions. PPDSA systems need to be understood as sociotechnical and human-centered systems grounded on ethical principles that will require foundational and applied research in social, behavioral, and economic sciences; human-centered design; and ethics, policy, law, and regulations. Sociotechnical approaches address not only how technology interacts with individuals, but also how it interacts with

broader social systems, organizations, and society at large. Such research is also needed to better understand what factors influence end-users and solution developers to adopt various privacy technologies.

Multidisciplinary research is critical to better understand user behavior, incentives, and privacy preferences at the individual level and needs at the collective level. Development of interdisciplinary theoretical foundations, such as the theory of contextual integrity,[78] is critical to ensure that PPDSA research is carried out holistically and incorporates multidisciplinary perspectives. Such sociotechnical and human factors research is also important to understand the social impact and potential unintended consequences of PPDSA solutions more comprehensively. Users need effective, usable, and inclusive tools that help them understand and evaluate the privacy implications of their technological interactions, ascertain the implications of using PPDSA technologies, and determine how to derive value from privately sharing data. Significant R&D is needed to design such usable and inclusive tools based on the insights from multidisciplinary research. Multidisciplinary research and participatory design approaches are specifically critical to understand how technological solutions can be made equitably available to various marginalized or under-represented groups, by ensuring that they are not disproportionately impacted through the use of PPDSA technologies. Sustained research in developing effective, inclusive, and equitable PPDSA solutions will empower and enable broader participation of such diverse groups in the PPDSA-enabled data economy and foster democratic values.

Similarly, research to understand how PPDSA solutions will interact or interface with existing organizational processes and culture is important for their effective design and eventual adoption. One critical area is collaboration between technical privacy researchers and regulatory entities, policymakers, and privacy policy experts. The absence of communication and collaboration among these groups has created a divergence in the way privacy is defined and treated by respective communities.

Existing state-of-the-art system design methodologies—for both hardware and software—typically treat laws and regulations as external requirements that systems must satisfy; however, such approaches fail to capture either the intricacies of underlying legal requirements or the dynamic nature of social norms and expectations. This makes designing systems that support PPDSA technologies challenging since these systems must handle evolving legal or regulatory landscape, provide provable compliance guarantees, and support transparency and accountability concerning the regulatory process, while also satisfying privacy expectations or requirements driven by users' preferences and social context.

Thus, significant research at the intersection of technology, policy, law, regulation, and social and economic sciences should be accelerated to foster a more consistent understanding of privacy threats or risks, harms, and legal implications, as well as other data protection compliance issues. A consistent understanding of privacy issues is critical to support better design, implementation, and compliant use of PPDSA solutions in a constantly evolving legal and regulatory landscape. R&D efforts are needed to ensure real progress toward achieving a future in which PPDSA technologies can be widely adopted for the benefit of individuals and society. Such research efforts should foster the use or development of techniques such as boundary objects methods. This allows entities or experts from different disciplines or technical, social, policy, or regulatory contexts to bring their local perspectives and collaboratively solve a global problem - to support reasoning about the development and deployment of PPDSA solutions.

Finally, an often-overlooked area of research that applies across the span of sociotechnical research is public awareness. For PPDSA technologies to be employed effectively, policymakers, organizations, and the general public will need an appropriate level of understanding of such technologies. Research is

needed to identify the most effective ways to communicate and foster public engagement on these concepts consistently to various audiences.

Different programs within the National Science Foundation[79] (NSF) provide potential models to foster such intensely multidisciplinary PPDSA research. The Secure and Trustworthy Cyberspace (SaTC) program is promoting broader social, behavioral, and economic research in the context of cybersecurity and privacy areas. The Designing Accountable Software Systems program also encourages research that addresses software system design approaches that are accountable within the context of regulatory and socio-cultural environments and requires multidisciplinary teams in research projects.

Strategic Priority 3: Accelerate Translation to Practice

While foundational and use-inspired R&D is critical for establishing rigorous scientific foundations to mature PPDSA solutions, it is equally important to cultivate an ecosystem that promotes a timely translation of theoretical results into real-world implementation and deployment. This requires significant applied research and engineering and systems development efforts in hardware and software as well as the promotion of pilot projects and tools. Actions that can establish pathways for translation and address deployment and adoption challenges are discussed below.

Recommendation 3.a. Promote applied and translational research and systems development

Many emerging PPDSA technologies have tremendous potential, but also have practical deployment challenges that need to be addressed to facilitate broader adoption. For example, innovative and better architectural solutions, or efficient and scalable protocols need to be developed for the deployment of secure multiparty computation, homomorphic encryption, zero-knowledge proofs, and trusted execution environments. In many cases, approaches to generate optimized solutions that are customized to specific application scenarios or constraints need to be developed (e.g., lightweight protocols for secure multiparty computation for IoT-Edge analytics). Federal funding opportunities and programs that support translational projects and early-stage startups should be expanded with a focus on PPDSA technologies to close the gap between theory and practice. NSF programs such as the Transition to Practice offering within the SaTC program, Innovation Corps, Partnerships for Innovation, Pathways to enable Open-Source Ecosystems (POSE),[80] as well as government-wide small business innovation research and small business technology transfer programs are important examples of programs that can be leveraged. Pilot projects, such as those being led by the United Nations PET lab (see case study in box); prize challenge competitions, e.g., the NIST Differential Privacy challenges[81,82] and US-UK PETs prize challenges (see Strategic Priority 5), and tech sprints, e.g., the Department of Veterans Affairs AI tech sprints,[83] also embody important approaches to accelerate the transition to practice and maturity of PPDSA technologies.

Recommendation 3.b. Pilot implementation activities within the Federal Government

Federal agencies should explore and identify opportunities to pilot PPDSA technologies and pursue new forms of data collaboration. For example, Federal statistical agencies could integrate secure multiparty computation into the implementation of the recently authorized NSDS demonstration, or science agencies can pilot techniques to safely use electronic health records to drive biomedical research. Agencies should assess their individual as well as cross-agency data sharing and analytics needs and opportunities that they are unable to pursue because of privacy concerns or security risks. Pilot projects allow agencies to demonstrate the value of PPDSA technologies for a specific use case, create and grow partnerships with technology providers and other participants, gain expertise, and identify gaps that need to be addressed by policy. However, a pilot project by itself is of limited value if the pilot ends and

the technology cannot be adopted by an agency. Pilot programs should include plans for how the PPDSA technologies will be transitioned for a successful deployment.

To complement the existing R&D of the Federal Statistical System to promote and advance PPDSA solutions, a Federal center of excellence could be established with specialized expertise to provide direct support to Federal agencies exploring PPDSA technology. The center could support agencies in pursuing pilot projects, as well as strategic planning, acquisition, and workforce development. There are several existing centers of excellence within the Federal Government that support the adoption of specific technologies. These include the General Services Administration's centers of excellence for AI,[84] Cloud Adoption,[85] and NIST's National Cybersecurity Center of Excellence.[86] These centers have a track record of demonstrated success, with NIST's National Cybersecurity Center of Excellence bringing together industry organizations, government agencies, and academic institutions to address the most pressing cybersecurity challenges using commercially available products. The General Services Administration's IT Modernization Centers of Excellence[87] reported millions of dollars in cost savings because of their work. Mirroring existing models, a PPDSA center of excellence would connect private sector innovation with government services, sharing best practices and expertise from both. To address the challenge of a lack of awareness of PPDSA technologies, the center could curate and publish a library of exemplary case studies and use cases and highlight PPDSA solutions' applicability and benefits to address agency needs. It would also be beneficial for the center to develop a playbook with a recommended roadmap for agencies exploring PPDSA adoption. Center of excellence support functions would be informed by the work of the Federal Statistical System's NSDS demonstration in advancing PPDSA solutions for Federal data.

Case Study: xD | U.S. Census Bureau and United Nations Privacy-Enhancing Technology Lab
Challenge: The United Nations Statistical Division realized the importance and opportunity of PETs for National Statistical Organizations across the world. Yet, it is often difficult to test the capabilities and realities of these possibilities. A small group of countries joined to form the UN PET Lab, a space to pilot the use of these technologies, experiment, share, and learn together. The Census Bureau represents the United States through its emerging technologies group, xD. ***Approach***: In 2022, the UN PET Lab ran a pilot study to test the ability to query data across a network made up of the countries and the UN without seeing the data itself. They did this by deploying an open-source software called PySyft and loading in open trade data to de-risk initial testing. Several PPDSA techniques were used including aspects of remote execution, secure multiparty computation, and differential privacy to enable each country to compare import and export statistics without fully accessing them. The Census Bureau also utilized aspects of Zero Trust as part of this deployment. ***Impact***: This pilot represents one of the first feasibility studies utilizing PPDSA techniques in this way and in an international context, illustrating the effectiveness and capabilities of PPDSA technologies today and encouraging their continued adoption and future potential. It also suggests that models like the UN PET Lab may be promising structures for organizing around PPDSA technologies and continuing to build out what further impact could look like.

Recommendation 3.c. Establish technical standards for PPDSA technologies

Voluntary, consensus-based standards will play a critical role in facilitating the adoption of PPDSA technologies. Standards for PPDSA technologies remain at a nascent stage of development, as industry, government, academia, and standards-developing organizations work to assess the maturity of the technologies and bridge the divide between theory and practice.

Existing sector-specific or use case-specific standards for security and privacy do not describe how organizations and implementers can or should use PPDSA technologies to meet overall system objectives. Without such standards, organizations seeking to use or implement PPDSA solutions lack

clear guidance that could ease deployment or provide confidence that such technologies would be accepted or meet sector-specific requirements. Further complicating the development of standards is the quick evolution of the field and the lack of standard taxonomy.

Standards-developing organizations and industry consortia, in collaboration with academia, should prioritize the creation of technical standards for PPDSA, and ensure that functional standards, protocols, or other specifications are drafted so that they will not become obsolete when PPDSA technologies are implemented. Establishing these standards can provide a common framework to guide PPDSA technologies' adoption, mitigate risks and biases, facilitate development of certifications, and promote implementation in commercial products and services while facilitating interoperability between vendors. Formalizing standards can also signal acceptance of PPDSA approaches at a national or international level and create benchmarks to help organizations evaluate solutions and measure successful deployments.

The Federal Government should strategically engage with academia, the private sector, and standards development organizations to study and develop standards for PPDSA technologies. In addition, participating groups should identify and prioritize use cases associated with PPDSA technical standards. A variety of application techniques exist across different categories of PPDSA technologies; however, many must be tailored for particular use cases to provide the necessary properties.

Recommendation 3.d. Accelerate efforts to develop standardized taxonomies, tool repositories, measurement methods, benchmarking, and testbeds

Having the right resources in place will facilitate and accelerate the responsible adoption of PPDSA technologies and enable participatory design approaches. Given the complexity and variety of both PPDSA technologies and privacy issues, a standardized taxonomy is needed to facilitate discussions at every point in the PPDSA lifecycle. Tools are needed to assist users in configuring and managing PPDSA solutions (e.g., differential privacy parameter tuning), understanding information leakage or auditing for privacy violations, visualizing data and analytics results in a privacy-preserving manner, and understanding privacy versus accuracy tradeoffs, etc.

Open-source software provides organizations the opportunity to build out PPDSA capabilities and processes to support business goals while gaining expertise in the use and application of PPDSA technologies. Open-source libraries for PPDSA technology implementations have been developed in recent years but have varying levels of maturity. Successful PPDSA technology deployments require high-quality software implementations that are correct, efficient, documented, and well-maintained so that when vulnerabilities are discovered, they can be rapidly addressed. If available with low-cost or free licenses via open source, one of the barriers to the adoption of PPDSA technologies could be lowered.

The Federal Government should explore opportunities to fund open-source efforts in support of the PPDSA ecosystem. For example, with its new POSE program, NSF is supporting organizations that will facilitate the creation and growth of a sustainable high impact open-source ecosystem around open-source research products. In addition, when Federal agencies develop PPDSA solutions, they should make those solutions publicly available whenever possible. For example, the Census Bureau published the complete code base for its differential privacy solution used for the 2020 Census.[88]

Accelerating PPDSA adoption will also require building testbeds and testing infrastructure to support experimentation with and evaluation of PPDSA solutions for different use cases and application scenarios. Such efforts should include the curation of a variety of datasets representing different use cases and scenarios that can be broadly applied for testing and exploration of sociotechnical approaches to testing, evaluation, verification, and validation of AI systems. These approaches connect the technology to societal values and provide recommended guidance for deployment.[89]

The NIST Privacy Framework can serve as an important baseline tool to help identify privacy priorities and risks, and whether PPDSA technologies can help meet these needs. In addition, the NIST Privacy Engineering Collaboration Space provides a repository for open-source tools and use cases.[90] However, additional tools are needed to help assess which PPDSA technologies could help meet privacy needs and mitigate identified risks. The Federal Government should support the development of tools such as decision aids and playbooks to guide how PPDSA technologies can be implemented in compliance with privacy requirements or governance models and facilitate participatory design approaches. For example, an authoritative PPDSA technology guide could be developed, which would act as a living reference for what PPDSA approaches are, the maturity levels of specific PPDSA technologies, and for which applications and problems technologies are well suited. Furthermore, the Federal Government should regularly collect, curate, and publish case studies of PPDSA solutions' deployments across the public and private sectors. A repository of successful deployments for case studies will enable interested parties to appreciate illustrative examples of how PPDSA technologies can be deployed, made compliant with privacy regulations and requirements, and managed with minimal privacy risks and harms.

Recommendation 3.e. Improve usability and inclusiveness of PPDSA solutions

Efficient design and implementation of PPDSA solutions, especially when considering making a product usable and inclusive, can require substantial customization and engineering efforts. Today's PPDSA technologies and tools are complex in how they are programmed, configured, and managed. They vary in the privacy-preserving functions and guarantees that are provided across different application scenarios. For example, addressing privacy concerns in an AI application may require integrating cryptographic (e.g., secure multiparty computation or fully homomorphic encryption) and non-cryptographic (e.g., differential privacy) techniques, each of which brings complexities in achieving a desired level of privacy. Such complexities can affect their usability, which in turn impacts decisions to adopt PPDSA solutions. Organizations need usable tools that facilitate the understanding of the performance of the PPDSA technology, the privacy implications of their use, and the value derived from their use.

Among solution developers and practitioners, usability improvements should focus on users and user groups with varying capabilities and limitations and should include efforts geared toward improving ease of use, clarity of results, and accessibility. Ideally, the development community would collaborate to determine accepted standards on the usability design of PPDSA technologies.

Significant and separate efforts are needed to foster the development of inclusive tools that are designed and implemented based on human factors and social, behavioral, and economic sciences. It is important to consider the diversity of participants in the PPDSA ecosystem, with an explicit focus on the underserved, marginalized, and vulnerable groups. To minimize or mitigate harm, inclusion should be evaluated throughout the development lifecycle of PPDSA. This might include activities that improve outreach to underserved communities about their needs for PPDSA solutions, mitigate bias in the data ingestion, aggregation, and refinement process, and determine what specific usability concerns and impacts underserved and marginalized groups have.

For both goals of usability and inclusiveness, human-centric approaches should be adopted and accelerated when developing PPDSA solutions. In such approaches, design choices consider both users and user groups with varying capabilities and limitations and specific underserved and marginalized groups. To achieve this, it is essential to co-design such technologies and tools with users by considering different user roles for data and their different needs and purposes, such as i) data owners or custodians who want to understand privacy and functionality parameters; ii) data participants who want to understand privacy and consent policies; and iii) data consumers who want to derive value from privacy-protected data.

Strategic Priority 4: Build Expertise and Promote Training and Education

Successful research, design, and implementation of PPDSA technologies require not only advanced technical knowledge, but also policy, data science, and domain expertise, as discussed above. Pulling together and cultivating cross-disciplinary expertise to effectively adopt PPDSA technologies remains a challenge for many organizations. From the decision makers who assess the risk and benefits of implementing PPDSA approaches to the program management and contracting personnel who manage the acquisition and evaluation of solutions to the technologists who implement the solutions—organizations need to build awareness and expertise across their workforce to effectively deploy and manage PPDSA technologies. A skilled, future-ready workforce needs to be strategically developed to facilitate progress toward the vision set out in this strategy.

Recommendation 4.a. Expand institutional expertise in PPDSA technologies

Federal and non-Federal entities have similar needs for recruiting and training personnel to meet the opportunities and challenges of adopting PPDSA technologies. Within the government, several agencies have demonstrated expertise in the application of PPDSA technologies to meet their missions. For example, the Census Bureau adopted differential privacy for the 2020 Census to modernize its approach to data protection and meet its requirement to ensure the privacy of respondents. The National Institutes of Health have been piloting privacy-preserving record linkage to bring clinical health data together without compromising patient privacy.[91]

However, a concerted effort is needed to expand expertise in PPDSA technologies across the Federal Government and within the private sector. NIST should develop a framework for the privacy workforce analogous to the National Initiative for Cybersecurity Education (NICE) framework[92] for the cybersecurity workforce. The NICE Framework provides a set of building blocks for describing the tasks, knowledge, and skills that are needed to carry out cybersecurity work. It enables organizations to develop their workforces to perform cybersecurity work, and it helps individuals engage in appropriate learning activities to develop their expertise. The development of a framework that defines the analogous set of tasks, knowledge, and skills needed for PPDSA solutions would assist Federal and non-Federal entities in cultivating expertise across interdisciplinary fields that contribute to PPDSA research and adoption. The framework could also be used to support direct-hire authorities for Federal positions with broader privacy expertise, including privacy engineering.

Building capacity across civil society and in communities affected by the use of PPDSA approaches will also be critical to providing civil society with tools to engage meaningfully in the PPDSA ecosystem and explore opportunities for employment of PPDSA solutions.

Organizations across sectors can leverage rotational opportunities to grow expertise among government, academia, and industry. Government programs provide the means to share expertise in the use of new technologies and approaches to solving governmental problems. The Intergovernmental

Personnel Act Mobility Program[93] and models for collaboration such as the Presidential Innovation Fellows,[94] the U.S. Digital Service,[95] and the Census Bureau's Emerging Technology Fellowship[96] should be leveraged and promoted to add external expertise to government agencies through permanent or term-limited appointments. Private organizations should similarly invest in building their internal privacy workforce by providing training opportunities for employees and establishing rotational programs with academia or other organizations.

In addition to these opportunities to grow expertise, organizations should consider using certification programs that have bodies of knowledge inclusive of PPDSA technology, privacy by design, and other key topics needed to meet workforce development goals for successful PPDSA adoption. These are available from different associations, institutions, and consortiums.

Recommendation 4.b. Educate and train participants on the appropriate use and deployment of PPDSA technologies

Privacy concepts are complex, as are PPDSA technologies. In many cases, even data curators are unsure about the privacy risks and what privacy solutions are best for various data sharing and analytics tasks. Thus, there is a need to adopt a holistic approach to better educate and train individuals of all ages and data professionals in all sectors to navigate the PPDSA landscape, building bridges between technical and non-technical communities.

It is important to develop effective ways to educate and train K-12 students, professionals in the workforce, members of civil society organizations, and marginalized and vulnerable populations. This should include the creation of specialized training and educational outreach activities and materials. Training should cover topics such as data sharing, PPDSA technologies, equity considerations, and procurement, and be tailored for different groups of participants and practitioners. These efforts should include and foster critical thinking and promote privacy and ethics literacy. In addition, investments should be made to support research in educational methods to develop innovative, effective techniques, or to transition such techniques from the broader field of computer science education to educate and train future generations on privacy and PPDSA.

Recommendation 4.c. Expand privacy curricula in academia

Privacy education in academia is still in its early stages, typically bundled as a small part of cybersecurity or computer or information science-related educational programs. Given the growing role of privacy in society and the pervasiveness of privacy concerns in the socio-technological ecosystem, it is imperative to significantly promote privacy education in academia, with a particular focus on PPDSA, not only for those pursuing technical studies but also in the context of ethics, law, social sciences, and other fields. Educational institutions should inform and integrate the growth of PPDSA-related topics into their curricula.

Federal agencies should establish incentives and funding mechanisms to foster privacy education. For instance, NSF has been funding projects focused on curriculum development as well as educational research in broad areas of cybersecurity for over two decades. Particular attention should be paid to support privacy education capacity building among HBCUs, TCUs, and MSIs.

In addition, the National Security Agency's Centers of Academic Excellence[97] program for designating educational institutions that meet cybersecurity educational standards and active research could be leveraged to build and promote privacy education, including possibly creating a CAE-Privacy designation program. CAEs have had a significant impact on the growth of formal cybersecurity education over the

past two decades, and similar approaches could help advance privacy education, including building ethical considerations into privacy approaches.

Toward fostering the next-generation workforce that can tackle challenges related to realizing the PPDSA vision, Federal agencies should explore leveraging existing scholarship for service programs and center-level partnerships with universities to expand and accelerate opportunities to build a PPDSA workforce. Special attention should be paid to developing PPDSA workforce opportunities within HBCUs, TCUs, and MSIs.

Strategic Priority 5: Foster International Collaboration on PPDSA

PPDSA technologies hold the potential to bring together governments, companies, and civil society to solve shared problems such as pandemics and climate change while complying with relevant privacy laws and regulations and upholding privacy rights. Creating a healthy global PPDSA ecosystem is central to this vision. To make this type of global collaboration a reality, the U.S. must work with other governments to develop, implement, and build domestic and international trust in PPDSA technologies. International cooperation will provide reciprocal benefits and help ensure that PPDSA R&D accommodates different legal and cultural understandings of privacy.

The Federal Government should cultivate international collaborations that support PPDSA technologies, open markets for U.S. industries working on PPDSA technologies, and the development and adoption of PPDSA technologies in a manner consistent with national interests and guiding principles detailed above. The following actions will promote strategic partnerships with developed and emerging economies aligned with U.S. interests and values to collaboratively address national and global challenges.

Recommendation 5.a. Foster bilateral and multilateral engagements related to a PPDSA ecosystem

The United States currently engages with allies and partners to promote 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. Accordingly, the State Department partners with many other Federal agencies to facilitate engagements with U.S. allies and partners to support the responsible development, deployment, use, and governance of inclusive and trustworthy PPDSA technologies.

U.S.-U.K. Privacy-Enhancing Technologies Prize Challenges[98]
To mature federated learning approaches and build trust in their adoption, the U.S. and U.K. Governments partnered on a set of prize challenges designed to strengthen privacy protections across the development and deployment of federated learning models. Innovators from academia, industry, and the broader public had the opportunity to participate in up to two separate tracks (improving the detection of financial crime and forecasting an individual's risk of infection during a pandemic) as well as the option to design one generalized solution that works for both scenarios for broader applicability. The challenges have been led by NIST and NSF, in cooperation with OSTP, in partnership with the U.K.'s Centre for Data Ethics and Innovation and Innovate UK. Participants competed for cash prizes from a total prize pool of approximately $1.6m/£1.3m and engaged through the challenge with regulators, government agencies, and global companies.

The Federal Government should increase touchpoints and mechanisms to collaborate with international partners and allies, from the leadership level down to individual scientists and program managers, to support PPDSA R&D and adoption activities. Specifically, the Federal Government should do the following:

Support international workshops and meetings of experts. The U.S. is uniquely positioned to convene international workshops and build partnerships with private and public sectors, as well as with civil society organizations, to drive innovation in the technological development and application of novel PPDSA techniques. These efforts can identify comparative strengths for these technologies through inclusive engagements and enable allies and partners to develop a common vision and expectations for the development and adoption of PPDSA technologies.

Pursue pilot projects and research collaborations. These collaborations can include pilot projects that assess the viability of implementing PPDSA technologies to work out technical and policy challenges in an iterative fashion, prize challenges and competitions that spur innovators to solve specific technical problems, and joint funding opportunities for U.S. teams to collaborate with counterparts in another country. Such initiatives could illuminate where and how PPDSA technologies can overcome existing barriers to international research collaboration posed by differing regulatory environments governing data sharing. The State Department engages in dialogues on a bilateral basis that strengthen partnerships, including through the development and implementation of Science and Technology Agreements. The State Department should add PPDSA R&D to the portfolios of these relationships and help Federal research funding agencies to collaborate on PPDSA research with their international counterparts.

Participate in bilateral and multilateral fora to advance the responsible adoption of PPDSA technologies. While setting policies to promote the development and adoption of PPDSA solutions domestically, it is important to recognize the diverse international views on privacy, technology, and data. PETs will require close international collaboration on technology standards, norms, and R&D. Both bilateral and multilateral discussions are needed to bring insights and positive examples on how to facilitate 21st-century cross-border data collaborations using PPDSA approaches. The development of PPDSA technology-focused bilateral discussions and agreements should be prioritized.

The power of data-sharing internationally will be greatest when many states can participate. To this end, the U.S. must encourage and lead multilateral efforts to develop PPDSA technologies that can operate under the privacy requirements of many nations. There are promising multilateral initiatives to test and research PPDSA technologies, and other multilateral groups have expressed interest in the potential of PPDSA technologies. Greater efforts are needed to create a vibrant international PPDSA ecosystem.

At a September 2022 Roundtable, the Data Protection and Privacy Authorities of the G7[99] recognized that PPDSA technologies can enable beneficial data sharing and stated their intention to advance their "responsible and innovative use." Their communique also called for the industry to develop technical standards and for governments to invest in PPDSA R&D.

Several multilateral fora are also incorporating PPDSA solutions into their agendas, such as the U.S.-EU. Trade and Technology Council and the Global Partnership on AI. The United States should continue to actively engage in these venues to facilitate progress on maturing PPDSA approaches, accelerating adoption, and putting in place guardrails for responsible use.

Recommendation 5.b. Explore the role of PPDSA technologies to enable cross-border collaboration

The growing internet connectivity and the digitization of the global economy have resulted in rapid increases in the collection, use, and transfer of data across borders. PPDSA technologies may be able to enable trusted cross-border data sharing to further enable U.S. practitioners to create a flexible, open information-sharing environment. This open model recognizes the value of data sharing, both within and across countries, and seeks to preserve cross-border data flows. PPDSA technologies can help the

world maintain this open model of data sharing while promoting data security and privacy. Certain PPDSA technologies can change the traditional conception of data flows by enabling the sharing of insights without sharing or exposing the actual data. The Federal Government should develop and enhance international collaborations across entities, sectors, and borders to help tackle shared challenges, such as health care, climate change, financial crime, human trafficking, and pandemic response—while mitigating privacy concerns.

A cohesive U.S. national approach for using PPDSA technologies to facilitate trusted cross-border data sharing will incubate relationships between academia and industry, both early-stage start-ups and leading technology companies. The Federal Government can build on the collective strengths of this robust group of innovators to enable the future of data collaboration and data flows that uphold U.S. core democratic values.

Effective implementation of PPDSA technologies requires not only foundational science and technology research but also the wider deployment of certification schemes that can both promote privacy and facilitate cross-border data flows and the development of these systems in a manner that can incorporate PPDSA technologies. The Cross-Border Privacy Rules (CBPR) certification model is an example of a voluntary certification process that allows companies to establish baseline compliance with different data protection laws across multiple jurisdictions through adherence to a core set of internationally-recognized common privacy protections, which can then be effectively enforced across all member jurisdictions that participate in the CBPR system and thus allow companies to transfer personal data across all participating jurisdictions. Bringing PPDSA technologies into such arrangements will build trust in the privacy protections of cross-border data flows, allowing their expansion and their benefits to be extended to larger and larger communities—while protecting fundamental rights.

Conclusion

PPDSA technologies have the potential to catalyze American innovation and creativity by facilitating data sharing and analytics while protecting sensitive information. This type of data sharing and analytics, leveraging advances in privacy-enhancing technologies, can advance research, unlock new insights, and enhance data-driven decision-making to address climate change, public health, social equity, and other challenges. Leveraging data at scale holds the power to drive transformative innovation. These important benefits must be promoted, while work is done to minimize potential problematic outcomes, i.e., violating individual privacy and undercutting fundamental rights. PPDSA technologies can play a critical role in protecting these ideals while enabling data analytics that will contribute to improvements in the quality of life of the American people.

While there are significant barriers to achieving this outcome and bringing PPDSA technologies to a mature state in which they can realistically be deployed at a large scale, this Strategy provides a path forward for impactful PPDSA solutions. R&D is needed, both to mature certain capabilities but also to deploy PPDSA technologies that are advanced but still present implementation challenges. A socio-technical approach toward PPDSA technologies that examine the interactions between technology and users, organizations, communities, and society as a whole will be critical to achieving the vision of this Strategy and upholding its guiding principles. Foundational multidisciplinary research is needed to better understand the nature of privacy and how PPDSA technologies can best augment privacy. This includes studying how people understand privacy and PPDSA approaches and how best to communicate these technical concepts in a manner that encourages their adoption.

There are immediate steps to be taken beyond research. Standards and taxonomies need to be developed to facilitate adoption. The workforce, particularly privacy professionals, needs to be educated about PPDSA solutions. Guidance is needed on how PPDSA approaches can meet statutory and regulatory privacy requirements. Institutions such as a steering group, large-scale privacy research centers, and centers for excellence will need to be established to ensure there are venues for the development and responsible adoption of PPDSA technologies. The development and deployment of PPDSA technology extend beyond U.S. borders making privacy-protecting cross-border data sharing invaluable as it can fuel global research, safety, security, and commerce. Finally, PPDSA technologies on their own will not prevent privacy incidents nor the insights derived from data sharing and analytics from being used in ways that could undermine democratic values. Strong governance and norms around use must support the growth of the PPDSA ecosystem.

This Strategy serves as a roadmap, providing background on the current state of PPDSA technologies and a series of recommendations to turn their potential for enabling privacy-protected data sharing and analysis into reality and fulfilling the vision statement that anchored this Strategy:

> *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.*

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

Appendix A: Abbreviations and Acronyms

AI	Artificial Intelligence
CBPR	Cross-Border Privacy Rules
CDC	Centers for Disease Control and Prevention
CHIPS	Creating Helpful Incentives to Produce Semiconductors for America
COVID-19	Coronavirus Disease 2019
DARPA	Defense Advanced Research Projects Agency
DHS	Department of Homeland Security
DOE	Department of Energy
ED	Department of Education
DOS	Department of State
DOT	Department of Transportation
FTAC	Fast-Track Action Committee
GDPR	General Data Protection Regulation
HBCU	Historically Black College or University
HHS	Department of Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act
IoT	Internet of Things
IRB	Institutional Review Board
ML	Machine Learning
MSI	Minority-Serving Institution
NICE	National Initiative for Cybersecurity Education
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NITRD	Networking and Information Technology Research and Development
NSDS	National Secure Data Service
NSF	National Science Foundation
NSTC	National Science and Technology Council
NTIA	National Telecommunications and Information Administration
ODNI	Office of the Director of National Intelligence
OSTP	Office of Science and Technology Policy
P3P	Platform for Privacy Preferences Project
PET	Privacy-enhancing Technology
PPDSA	Privacy-Preserving Data Sharing and Analytics
PPRL	Privacy-Preserving Record Linkage
R&D	Research and Development
SaTC	Secure and Trustworthy Cyberspace
TCU	Tribal Colleges and Universities
UK	United Kingdom
VA	Department of Veterans Affairs

Endnotes

[1] While protecting privacy is generally associated with individual persons, many of the data protection challenges and solutions discussed in this report apply equally to protecting the confidentiality of entities, such as businesses or other organizations.

[2] *NIH Workshop on the Policy and Ethics of Record Linkage: Workshop Summary | Data Science at NIH.* (2021, June). *NIH Policy and Ethics of Record Linkage Workshop.* https://datascience.nih.gov/nih-policy-and-ethics-of-record-linkage-workshop-summary

[3] Clifton, C. (n.d.). *A Roadmap for Greater Public Use of Privacy-Sensitive Government Data: Workshop Report*. arXiv.org. https://arxiv.org/pdf/2208.01636.pdf

[4] The White House. (2022, October 4). *National Science and Technology Council (NSTC) - OSTP - The White House*. https://www.whitehouse.gov/ostp/nstc/

[5] *The Networking and Information Technology Research and Development (NITRD) Program*. (n.d.). The Networking and Information Technology Research and Development (NITRD) Program. https://www.nitrd.gov/

[6] The FTAC also serves as a coordination mechanism for Federal initiatives related to privacy-preserving data sharing and analytics, as well as other domestic and international efforts that may arise in the near term.

[7] *Data — Boston Women's Workforce Council*. (n.d.). Boston Women's Workforce Council. https://thebwwc.org/wage-gap-studies

[8] U.S. Census Bureau. (2021, November 18). *Statistical Safeguards*. census.gov. https://www.census.gov/about/policies/privacy/statistical_safeguards.html

[9] Sheller, M. J., Edwards, et al. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-69250-1

[10] *Foundations for Evidence-Based Policymaking Act of 2018*. congress.gov. https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

[11] Federal Data Strategy. (n.d.). Federal Data Strategy. https://strategy.data.gov/overview/

[12] Office of Management and Budget (Executive Office of the President). (2019, June 4). *Federal Data Strategy - A Framework for Consistency*. whitehouse.gov. https://www.whitehouse.gov/wp-content/uploads/2019/06/M-19-18.pdf

[13] Office of Science and Technology Policy (OSTP), & Nelson, Dr. A. (2022, August 25). *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research*. whitehouse.gov. https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

[14] *CHIPS ACT OF 2022*. congress.gov. https://www.congress.gov/117/plaws/publ167/PLAW-117publ167.pdf

[15] The Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Office of Data Science and Sharing (ODSS). (n.d.). *Privacy Preserving Record Linkage (PPRL) For Pediatric Covid-19 Studies*. National Institute of Child Health and Human Development (NICHD). https://www.nichd.nih.gov/sites/default/files/inline-files/NICHD_ODSS_PPRL_for_Pediatric_COVID-19_Studies_Public_Final_Report_508.pdf

[16] Office of Science Policy. (2023, February 27). *NIH Workshop: Using Public Engagement to Inform the Use of Data in Biomedical Research*. https://osp.od.nih.gov/events/nih-workshop-using-public-engagement-to-inform-the-use-of-data-in-biomedical-research/

[17] As stated in the National Privacy Research Strategy: "Privacy is surprisingly hard to characterize. A full treatment of privacy requires a consideration of ethics and philosophy, sociology and psychology, law and government, economics, and technology. Embodying such broad considerations, the Federal Government's approach has been guided by the Fair Information Practice Principles (FIPPs), a framework for understanding stakeholder

considerations utilizing concepts of fairness, due process, and information security." Office of Science and Technology Policy (2016). *National Privacy Research Strategy.* https://www.nitrd.gov/PUBS/NationalPrivacyResearchStrategy.pdf

[18] National Institute of Standards and Technology. (2022, June 16). *The NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management*. nist.gov. https://www.nist.gov/privacy-framework/privacy-framework

[19] National Institute of Standards and Technology. (n.d.). *NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management*. NIST Privacy Framework. https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/resources

[20] Disassociability is focused on enabling the processing of data or events without association with individuals or devices beyond the operational requirements of the system. Manageability is focused on providing the capability for granular administration of data, including alteration, deletion, and selective disclosure. Predictability is focused on enabling reliable assumptions by individuals, owners, and operators about data and their processing by a system, product, or service. National Institute of Standards and Technology. NIST (2020, January). *The NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management.* https://www.nist.gov/privacy-framework/privacy-framework

[21] Big Data UN Global Working Group (n.d.). *UN Handbook on Privacy-Preserving Computation Techniques.* United Nations. https://unstats.un.org/bigdata/task-teams/privacy/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf

[22] Kwon, A., AlSabah, M., Lazar, D., Dacier, M., Devadas, S. et al. (2015). Circuit Fingerprinting Attacks: Passive Deanonymization of Tor Hidden Services. *Proceedings of the 24th USENIX Security Symposium*. USENIX. https://www.usenix.org/system/files/conference/usenixsecurity15/sec15-paper-kwon.pdf

[23] *Request for Information on Advancing Privacy-Enhancing Technologies*. (2022, June 9). Federal Register. https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies

[24] Networking and Information Technology Research and Development (NITRD) Program. (2022, September 6). *Public Input on Advancing Privacy-Enhancing Technologies*. The Networking and Information Technology Research and Development (NITRD) Program. https://www.nitrd.gov/87-fr-35250-responses/

[25] *Data Ethics Framework*. (n.d.). Data Ethics Framework. https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf

[26] Office of Science and Technology Policy. (2022). *The Blueprint for an AI Bill of Rights.* whitehouse.gov. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

[27] *Fifth U.S. Open Government National Action Plan*. (2022, December). U.S. Open Government Initiatives. https://open.usa.gov/assets/files/NAP5-fifth-open-government-national-action-plan.pdf

[28] National Institute of Standards and Technology. (2020, January). *The NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management*. nist.gov. https://www.nist.gov/privacy-framework/privacy-framework

[29] National Institute of Standards and Technology. (n.d.). *Cybersecurity Framework*. https://www.nist.gov/cyberframework

[30] National Institute of Standards and Technology. (2023, January 26). *AI Risk Management Framework*. nist.gov. https://www.nist.gov/itl/ai-risk-management-framework

[31] Bagdasaryan, E., Poursaeed, O., & Shmatikov, V. (2019, December). Differential privacy has disparate impact on model accuracy. ACM Digital Library. https://dl.acm.org/doi/abs/10.5555/3454287.3455674

[32] The Federal Committee on Statistical Methodology. (n.d.). *Subcommittee on Updating Statistical Methods for Safeguarding Protected Data*. FCMS.gov. https://www.fcsm.gov/resources/safe-guard-data/

[33] The Federal Committee on Statistical Methodology. (n.d.). *Subcommittee on Updating Statistical Methods for Safeguarding Protected Data*. FCSM.gov. https://www.fcsm.gov/resources/safe-guard-data/

[34] *Privacy Act of 1974*. congress.gov. https://www.congress.gov/93/statute/STATUTE-88/STATUTE-88-Pg1896.pdf

[35] *Confidential Information Protection and Statistical Efficiency Act of 2002.* congress.gov. https://www.congress.gov/107/plaws/publ347/PLAW-107publ347.pdf

[36] *Health Insurance Portability and Accountability Act of 1996.* congress.gov. https://www.congress.gov/104/plaws/publ191/PLAW-104publ191.pdf

[37] Federal Trade Commission. (n.d.). *Children's Online Privacy Protection Rule (COPPA)*. https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa

[38] *Gramm-Leach-Bliley Act.* congress.gov. https://www.congress.gov/106/plaws/publ102/PLAW-106publ102.pdf

[39] *Family Educational Rights and Privacy Act*. (1978, May 19). congress.gov. https://www.congress.gov/bill/95th-congress/house-bill/12795?q=%7B%22search%22%3A%22Family+ Educational+Rights+and+ Privacy+ Act+ of+1974%22%7D&s=8&r=2

[40] State of California Department of Justice. (n.d.). *California Consumer Privacy Act*. https://oag.ca.gov/privacy/ccpa

[41] HIPAA also defines a de-identification standard for protected information that requires there be no reasonable basis to believe the information can identify an individual. The regulation permits that health data de-identified according to this standard can be used without authorization by the individual. The standard can be met through either determination by a qualified expert (Expert Determination method) or the removal of 18 specific identifiers (Safe Harbor method).

[42] Protections, O. F. H. R. (2022, December 13). *Federal Policy for the Protection of Human Subjects ('Common Rule')*. HHS.gov. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html. The Common Rule expects informed consent be obtained from research participants to use their identifiable information, though there are exceptions, such as when an IRB grants a waiver of consent, the research falls under certain categories of exempt research, or the information is de-identified and not collected through intervention or interaction with the individual. The Common Rule defines information as de-identified when the identity of individual research participants cannot "readily be ascertained by the investigator." IRBs are often tasked with assessing whether research falls into the above exceptions to informed consent and generally whether privacy protections are adequate.

[43] Wolford, B. (2022, May 26). *What is GDPR, the EU's new data protection law?* GDPR.eu. https://gdpr.eu/what-is-gdpr/

[44] *LGPD Brazil - General Personal Data Protection Act. (n.d.).* https://lgpd-brazil.info/

[45] *Data Security Law of the People's Republic of China*. (n.d.). http://www.npc.gov.cn/englishnpc/c23934/202112/1abd8829788946ecab270e469b13c39c.shtml

[46] Office of the Privacy Commissioner of Canada. (2018, January 31). *Summary of Privacy Laws in Canada*. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/

[47] *Personal Data Protection Act 2012 - Singapore Statutes Online*. (2022, October 1). https://sso.agc.gov.sg/Act/PDPA2012

[48] U.S. Food and Drug Administration. (2022, September 28). *Guidance for Industry and Food and Drug Administration Staff*. U.S. Food and Drug Administration. https://www.fda.gov/media/109618/download

[49] *Consumer Financial Protection Circular 2022-04: Insufficient data protection or security for sensitive consumer information | Consumer Financial Protection Bureau*. (2022, August 11). Consumer Financial Protection Bureau. https://www.consumerfinance.gov/compliance/circulars/circular-2022-04-insufficient-data-protection-or-security-for-sensitive-consumer-information/

[50] *Trade Regulation Rule on Commercial Surveillance and Data Security*. (2022, August 22). Federal Register. https://www.federalregister.gov/documents/2022/08/22/2022-17752/trade-regulation-rule-on-commercial-surveillance-and-data-security

[51] Albrecht, M., Chase, M., Chen, H., & HomomorphicEncryption.org. (2017, November). *Homomorphic Encryption Security Standard*. homomorphicencryption.org. https://homomorphicencryption.org/standard

[52] ZKProof. (2022, November 29). *Homepage - ZKProof Standards*. ZKProof Standards. https://zkproof.org/

[53] A database reconstruction attack is a type of privacy attack on aggregate data that reconstructs a significant portion of a raw dataset. A linkage attack is an attempt to re-identify individuals in an anonymized dataset by combining that data with background information.

[54] Centers for Disease Control and Prevention (2022). *COVID-19*. CDC. https://data.cdc.gov/browse?tags=covid-19

[55] SAGE Publications. (n.d.). *Protecting Privacy and Transforming COVID-19 Case Surveillance Datasets for Public Use - Brian Lee, et. al, 2021*. SAGE Journals. https://journals.sagepub.com/doi/full/10.1177/00333549211026817

[56] Wood, E., Baltrusaitis, T., Hewitt, C., Dziadzio, S., Johnson, M. P., Estellers, V., Cashman, T. J., & Shotton, J. (2021). Fake it till you make it: face analysis in the wild using synthetic data alone. *International Conference on Computer Vision*. https://doi.org/10.1109/iccv48922.2021.00366

[57] Chen, R., et al. (2021). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, *5*(6), 493–497. https://doi.org/10.1038/s41551-021-00751-8

[58] The Advanced Computing Systems Association. Zhang, Z., et al. (2021, August 11). *PrivSyn: Differentially Private Data Synthesis*. www.usenix.org. https://www.usenix.org/system/files/sec21-zhang-zhikun.pdf

[59] Lindell, Y. (2020). Secure multiparty computation. *Communications of the ACM*, *64*(1), 86–96. https://doi.org/10.1145/3387108

[60] Albrecht, M., Chase, M., Chen, H., & HomomorphicEncryption.org. (2017, November). *Homomorphic Encryption Security Standard*. homomorphicencryption.org. https://homomorphicencryption.org/standard

[61] Goldwasser, S., Micali, S., & Rackoff, C. (1989). The Knowledge Complexity of Interactive Proof Systems. *SIAM Journal on Computing*, *18*(1), 186–208. https://doi.org/10.1137/0218012

[62] Ben-Sasson, E., & Chiesa, A., et al. (2014, May 18). *Zerocash: Decentralized Anonymous Payments from Bitcoin*. zerocash.org. http://zerocash-project.org/media/pdf/zerocash-extended-20140518.pdf

[63] Liu, T., Xie, X., & Zhang, Y. (2021). zkCNN: Zero Knowledge Proofs for Convolutional Neural Network Predictions and Accuracy. *Computer and Communications Security*. https://doi.org/10.1145/3460120.3485379

[64] Schneider, M., Masti, R. J., Shinde, S., Capkun, S., & Perez, R. (2022). SoK: Hardware-supported Trusted Execution Environments. arXiv.org. https://arxiv.org/pdf/2205.12742.pdf

[65] Xue, T., Wen, Y., Luo, B., Li, G., Li, Y., Zhang, B., Zheng, Y., Hu, Y., & Meng, D. (2022). SparkAC: Fine-Grained Access Control in Spark for Secure Data Sharing and Analytics. *IEEE Transactions on Dependable and Secure Computing*, 1. https://doi.org/10.1109/tdsc.2022.3149544

[66] Miorandi, D., Rizzardi, A., Sicari, S., & Coen-Porisini, A. (2020). Sticky Policies: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, *32*(12), 2481–2499. https://doi.org/10.1109/tkde.2019.2936353

[67] *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification*. (n.d.). https://www.w3.org/TR/P3P/

[68] Xu, R., Joshi, J., & Krishnamurthy, P. (2021). An Integrated Privacy Preserving Attribute-Based Access Control Framework Supporting Secure Deduplication. *IEEE Transactions on Dependable and Secure Computing*, *18*(2), 706–721. https://doi.org/10.1109/tdsc.2019.2946073

[69] Arora, S., et al. (2022, May 1). *A Tale of Two Regulatory Regimes: Creation and Analysis of a Bilingual Privacy Policy Corpus*. NSF Public Access. https://par.nsf.gov/biblio/10336685

[70] *National COVID Cohort Collaborative (N3C)*. (2022, September 26). National Center for Advancing Translational Sciences. https://ncats.nih.gov/n3c

[71] Aronson, J. (n.d.). *Landscape Analysis of Privacy Preserving Patient Record Linkage Software (P3RLS)*. National Cancer Institute Division of Cancer Control & Population Sciences. https://surveillance.cancer.gov/reports/TO-P1-PPRLS-Landscape-Analysis.pdf

[72] Kairouz, P. (2021). *Advances and Open Problems in Federated Learning*. arXiv.org. https://arxiv.org/pdf/1912.04977.pdf

[73] U.S. Census Bureau. (2021). *Disclosure Avoidance for the 2020 Census: An Introduction*. https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf

[74] U.S. Census Bureau. (2021). *Disclosure Avoidance for the 2020 Census: An Introduction*. https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf

[75] *National Strategy for Trusted Identities in Cyberspace*. (2011, April). Obama White House Archives. https://obamawhitehouse.archives.gov/sites/default/files/rss_viewer/NSTICstrategy_041511.pdf

[76] Walsh, J. M., Varia, M., Cohen, A., Sellars, A., & Bestavros, A. (2022). Multi-Regulation Computing. *Proceedings of the 2022 Symposium on Computer Science and Law*. https://doi.org/10.1145/3511265.3550445

[77] Cryptography and the Intelligence Community. (2022). *National Academies Press EBooks*. https://doi.org/10.17226/26168

[78] Nissenbaum, H. (n.d.). *Privacy as Contextual Integrity*. UW Law Digital Commons. https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10

[79] *NSF - National Science Foundation*. (n.d.). https://nsf.gov/

[80] *Pathways to Enable Open-Source Ecosystems (POSE)*. (2023, January 31). NSF - National Science Foundation. https://beta.nsf.gov/funding/opportunities/pathways-enable-open-source-ecosystems-pose

[81] *2018 Differential Privacy Synthetic Data Challenge | NIST*. (2022, July 19). NIST. https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2018-differential-privacy-synthetic

[82] *2020 Differential Privacy Temporal Map Challenge | NIST*. (2022, December 31). NIST. https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2020-differential-privacy-temporal

[83] *NAII AI Tech Sprints*. (2023, March 1). https://www.research.va.gov/naii/tech-sprints.cfm

[84] *Artificial Intelligence | GSA - IT Modernization Centers of Excellence*. (n.d.). https://coe.gsa.gov/coe/artificial-intelligence.html

[85] *Cloud Adoption | GSA - IT Modernization Centers of Excellence*. (n.d.). https://coe.gsa.gov/coe/cloud-adoption.html

[86] National Institute of Standards and Technology. (n.d.). *Working Together for Cybersecurity*. National Institute of Standards and Technology National Cybersecurity Center of Excellence. https://www.nccoe.nist.gov/

[87] *Centers of Excellence Home | GSA - IT Modernization Centers of Excellence*. (n.d.). https://coe.gsa.gov/

[88] U.S. Census Bureau. (n.d.). *GitHub - uscensusbureau/DAS_2020_Redistricting_Production_Code: Official release of source code for the Disclosure Avoidance System (DAS) used to protect against the disclosure of individual information based on published statistical summaries.* GitHub. https://github.com/uscensusbureau/DAS_2020_Redistricting_Production_Code

[89] Vassilev, A. (2022, November 9). *[Project Description] Mitigating AI/ML Bias in Context: Establishing Practices for Testing, Evaluation, Verification, and Validation of AI Systems*. CSRC. https://csrc.nist.gov/publications/detail/white-paper/2022/11/09/mitigating-ai-ml-bias-in-context/final

[90] *Collaboration Space | NIST*. (2020, June 1). NIST. https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space

[91] U.S. Department of Health and Human Services. (2022, September 2). *N3C Data Overview*. National Center for Advancing Translational Sciences. https://ncats.nih.gov/n3c/about/data-overview#privacy-preserving-record-linkage

[92] *NICE Framework Resource Center | NIST*. (2023, February 16). NIST. https://www.nist.gov/itl/applied-cybersecurity/nice/nice-framework-resource-center

[93] *Intergovernmental Personnel Act*. (n.d.). U.S. Office of Personnel Management. https://www.opm.gov/policy-data-oversight/hiring-information/intergovernment-personnel-act/

[94] *Presidential Innovation Fellows*. (n.d.). https://presidentialinnovationfellows.gov/

[95] *United States Digital Service*. (2023, February 28). United States Digital Service. https://usds.gov/

[96] U.S. Census Bureau. (2022, December 16). *Emerging Technology Fellowship Program*. census.gov. https://www.census.gov/about/census-careers/jobs/internships-fellowships/etf.html

[97] National Security Agency/Central Security Service. (n.d.). *National Centers of Academic Excellence*. https://www.nsa.gov/Academics/Centers-of-Academic-Excellence/

[98] *U.K.-U.S. prize challenges | Privacy-Enhancing Technologies*. (n.d.). https://petsprizechallenges.com/

[99] *Promoting Data Free Flow with Trust and Knowledge Sharing About the Prospects for International Data Spaces*. (2022, September 8). Roundtable of G7 Data Protection and Privacy Authorities. https://www.bfdi.bund.de/SharedDocs/Downloads/EN/G7/Communique-2022.pdf?__blob=publicationFile&v=1