SPECIAL COMPETITIVE STUDIES PROJECT

SOCIETY

Interim Panel Report

December 2022

Contributors

SCSP LEADERSHIP

Dr. Eric Schmidt, Chair
Ylli Bajraktari, President & CEO

BOARD OF ADVISORS

Michèle Flournoy
Dr. Nadia Schadlow
William "Mac" Thornberry III
Robert O. Work

SOCIETY PANEL

Rama Elluru, Senior Director
Chuck Howell, Senior Director
Ben Bain, Director
Jenilee Keefe Singer, Director
Annabelle Darby, Research Assistant
Evan Miller, Research Assistant
Rebecca Stachel, Research Assistant
Daniel Trusilo, Research Assistant
Michael Garris, Special Contributor

SOCIETY ADVISORS

David Danks
Mignon Clyburn
Eric Horvitz
Gina Neff
Lynne Parker

The Society Panel Interim Panel Report (IPR) is the last of six interim reports from the overall work that the Special Competitive Studies Project (SCSP) has conducted over the past year and that was summarized in our Mid-Decade Challenges to National Competitiveness report published on 12 September 2022. This report benefited greatly from insights and expertise by a number of individuals to whom we are deeply grateful. It aims to reflect many, though not all, of those insights. It was prepared by SCSP staff and, as such, it is not a consensus document of all the experts who assisted.

SOCIETY PANEL INTERIM PANEL REPORT

National Data Action Plan

Introduction

The United States must approach its data assets with the same logic and urgency driving America's new industrial policy, reshaping the Pentagon's defense strategy, and reorienting U.S. foreign policy. In a strategic competition with China that is at once a geopolitical contest, a juxtaposition of democratic vice authoritarian systems of government, and a race for technological leadership, the ability to access, analyze, and act on data-driven insights is a competitive advantage. However, the United States remains far from leveraging data in any of these contests despite housing more data centers than any other country, serving as the home of the world's largest technology companies, dominating the big data and business analytics market, and being the world's largest data producer.[1]

In the sixteen years since a British mathematician coined the phrase "data is the new oil,"[2] the United States has still not developed the data vision equivalent of a sustainable energy strategy. The landscape is a series of ad hoc and under-resourced federal efforts, a patchwork of state legislation, and a mosaic of United States Government (USG), private sector, and academic efforts to realize data-derived benefits. We have no comprehensive data privacy guidance at the national level despite widespread agreement of its desirability, no scalable model for aggregating private and public sector data and analytics to solve overarching societal problems despite universal acknowledgement of the untapped potential, and no adequate infrastructure to lower the barriers to data sharing despite knowing them. Federal data may be open, but it is not easily accessible, the open data ecosystem is difficult to navigate, and datasets are often outdated. Some of the most important data for solving scientific and societal problems remains siloed in corporate data centers or is too expensive for all but the largest companies and universities to derive benefit. Even when isolation serves no commercial benefit, companies are reluctant to share data given regulatory uncertainties and lack of incentives. Meanwhile the ability of companies and data brokers to obtain, transfer, and exploit individuals' data remains a wild west, vulnerable to privacy violations and national security threats.

The result is a data environment too permissive to garner public trust, too restrictive to unlock private data for the public good, and too inefficient to maximize publicly-held data for

[1] Bhaskar Chakravorti, et al., Which Countries Are Leading the Data Economy?, Harvard Business Review (2019).
[2] Charles Arthur, Tech Giants May Be Huge, But Nothing Matches Big Data, The Guardian (2013).

commercial or public use. The absence of an executable data strategy has left the United States at a competitive disadvantage. We have now reached the point where the chasm between required and actual data practices poses risks to national competitiveness and national security.

China, meanwhile, believes that learning from the world's data, while walling off its own, is the path to competitive advantage. Mirroring such an approach would be disastrous for the United State and its allies, and incompatible with its values. The free flow of data – with appropriate safeguards for privacy and security – is the logical extension of the larger American project of encouraging the free flow of goods, ideas, and people. More data available to more individuals, universities, companies, and government at all levels will foster innovation, improve the delivery of services, lower costs for services, and incentivize the kinds of public-private partnerships that are the foundation of the U.S. competitiveness model.

The Opportunity: Leveraging Data for National Advantage

In our increasingly networked world, vast amounts of data[3] are created every day.[4] The proliferation of data can be used for good or harm. It can beneficially enable improved decision making and innovation, as well as expand opportunities for better government performance, economic growth, and public good.[5] The United States must provide a data governance model that shapes both beneficial outcomes and upholds democratic values. The use of data, inside and outside of government, to support national interests like economic and social prosperity, while upholding values such as privacy, is critical to demonstrating the advantages of liberal democratic responses to data opportunities.

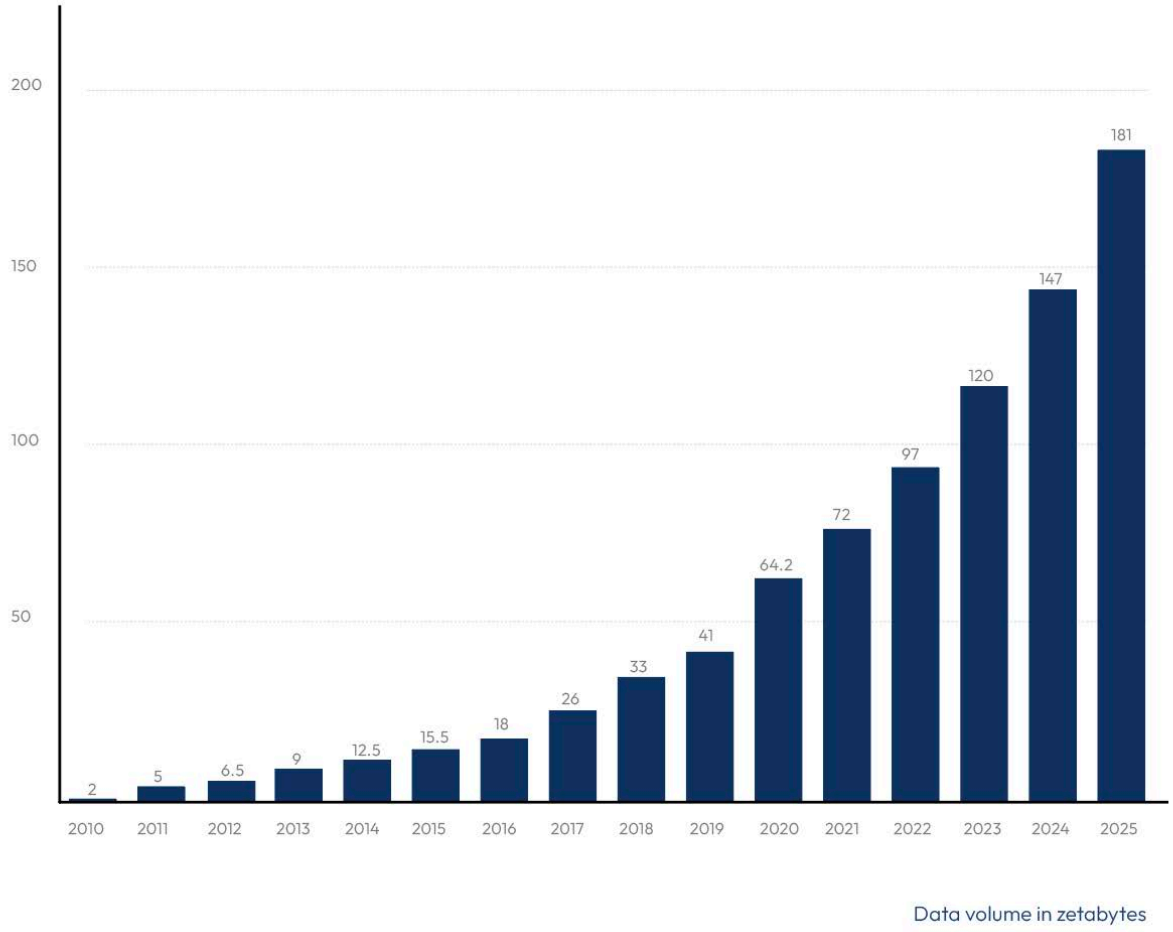
[3] There is no universal definition of "data." This report takes the broadest view of data to include all the information generated every day about people, objects, environments, and systems. It can include speech, text, imagery, behaviors and actions, sounds, locations, and much more that has been recorded in any form. Pub. L. 115–435, Foundations for Evidence-Based Policymaking Act of 2018 §202 (2019); National Data Strategy, UK Government (2020).

[4] The total amount of data globally created and consumed reached 64.2 zettabytes in 2020. By 2025, annual global data creation is projected to grow to more than 180 zettabytes. Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide From 2010 to 2020, With Forecasts From 2021 to 2025, Statista (2022). A zettabyte is equal to a trillion gigabytes.

[5] The inherent value in data is unlocked when enough relevant and quality data is combined and analyzed to improve human decision making and outcomes across government, the private sector, academia, and civil society. Data enables us to identify novel patterns, investigate causes and impacts, and generate original insights about almost anything in our world. The speed and scale of these insights can be amplified when supported by data-enabled artificial intelligence (AI). For AI applications that serve the greatest public purpose, access to quality data is key. Bhaskar Chakravorti, et al., Which Countries Are Leading the Data Economy?, Harvard Business Review (2019). The availability of quality data is essential for training many machine learning systems, with products and services rapidly moving from pattern recognition and insight generation to more sophisticated forecasting techniques and, thus, better decision recommendations.

More and More Data[6]

Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025.



[6] Petroc Taylor, Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2020, With Forecasts from 2021 to 2025, Statista (2022).

Data for U.S. Government:

- Data in the hands of the government supports evidence-based public policy and the provision of better public services. Governments can filter through and leverage vast amounts of data to glean contextual insights that can be used to guide policy that supports economic and national security interests. For example, public datasets combined with private datasets can inform industrial strategy. Specifically, data sharing is essential to inform government policy surrounding supply chain resilience, as vulnerabilities that exist several tiers into the supply chain are often only visible to industry.[7] Indeed, in the microelectronics sector, the U.S.-EU Trade and Technology Council recognizes the potential for an early warning system for supply chain shortages based on data sharing.[8] Data also shapes the USG by allowing it to collect and analyze data on program impacts and then iterate for better, improved services[9] suited to user needs; optimize operations; rely on a robust enterprise data infrastructure that allows for secure, real-time cross-department sharing of data; and, reshape the social contract by being transparent about how public dollars are spent.[10]

Data for Economy:

- In industry's hands, data is the basis for many new products and services, driving productivity and resource efficiency gains across all sectors of the economy. Large, diverse datasets are an essential resource for start-ups and small and medium-sized enterprises (SMEs) if they can gain access

[7] Testimony of John VerWey before the U.S.-China Economic and Security Review Commission, U.S.-China Competition in Global Supply Chains (2022).
[8] FACT SHEET: U.S.-EU Trade and Technology Council Establishes Economic and Technology Policies & Initiatives, The White House (2022).
[9] Early in the COVID-19 pandemic, technology companies made available vast amounts of anonymized aggregated human mobility data and analytics. The COVID-19 Mobility Data Network (CMDN) was a voluntary collaboration created in March 2020 in order to "establish routine analytic pipelines between tech companies and policy makers, providing meaningful policy-relevant information supported with scientific evidence and methodological rigor." Annual Report 2020: COVID-19 Mobility Data Network, Crisis Ready at 7 (2020). This kind of data provided otherwise unavailable insights into the interactions of population mobility and disease spread, informing reopening strategies among other government decisions. Serina Change, et al., Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening, Nature 589 at 82–87 (2021). As a final illustration of the potential value of non-USG data for USG use, NASA's Western Water Applications Office is using geospatial data from Climate Engine and Google to help improve how water is managed in the arid western United States. Leveraging Google Geospatial AI to Prepare for Climate Resilience, Google Cloud (2021); also see Western Water Applications Office, National Aeronautics and Space Administration (last accessed 2022).
[10] Oliver Wise, Data is About Much More Than Decision-Making, StateScoop (2018).

to such data.[11] Data will also fuel wider implementation of transformative practices such as the use of digital twins in manufacturing. Digital twins create a virtual replica of a physical product, process, or system. The replica can, for example, predict when a machine will fail, based on data analysis, which allows for increased productivity through predictive maintenance.[12] Data generated through digital twins can also be used to accelerate the training of machine learning AI systems, which require vast, quality datasets.[13] Data also can generate value as a commodity that can be monetized. While it can be difficult to make a precise valuation for data,[14] there is a significant and growing market in data aggregation, consolidation, and sale. For example, "data broker market size was valued at $257.16 billion in 2021 and the total data broker revenue is expected to grow at 4.5% from 2022 to 2029, reaching nearly $365.71 billion."[15]

Data for Public Good:

- Leveraging data is also critical for realizing its value for public good. Data promises to transform societies and our everyday lives by giving us a better understanding of our world and environment. As just one example, data enables medical caregivers to recognize genetic predisposition to diseases, identify illnesses faster and more precisely, and respond to them with personalized therapeutic strategies that can save lives and reduce healthcare costs.[16]

[11] Small- and medium-sized enterprises are often disadvantaged in the data economy. "Incumbents in the data economy appear to be earning large rents, as reflected by high reported profits and equity market valuations, and many digital markets currently feature high degrees of concentration. This may reflect a practice of hoarding data on their customers, creating a barrier to entry that is stifling competition from smaller firms in some cases." Yan Carriere-Swallow & V. Haksar, The Economics and Implications of Data: An Integrated Perspective, International Monetary Fund at 32 (2019).

[12] Marc Hamilton, Supercomputing Superpowers: NVIDIA Brings Digital Twin Simulation to HPC Data Center Operators, NVIDIA (2022). Idaho National Laboratory Demonstrates First Digital Twin of a Simulated Microreactor, U.S. Department of Energy (2022).

[13] Kosmas Alexopoulos, et al., Digital Twin-Driven Supervised Machine Learning for the Development of Artificial Intelligence Applications in Manufacturing, International Journal of Computer Integrated Manufacturing (2020).

[14] "Despite rhetorical consensus on the value of data for business, individuals and society, organizations struggle to measure it." Articulating Value from Data, World Economic Forum at 4 (2021).

[15] Data Broker Market: Global Industry Forecast (2022-2029) by Data Category, Data Type, Pricing Model, End Use Sector, and Region, Maximize Market Research (2022).

[16] Roy Adams, et al., Prospective, Multi-Site Study of Patient Outcomes After Implementation of the TREWS Machine Learning-Based Early Warning System for Sepsis, Nature (2022) (Demonstrating that using the machine learning-based TREWS resulted in earlier detection of sepsis and lower mortality rates); Junaid Bajwa, et al., Artificial Intelligence in Healthcare: Transforming the Practice of Medicine, Future Healthcare Journal (2021) (Showing that AI

The U.S. government is key for improving data accessibility throughout the entire U.S. data ecosystem, which is why this report makes recommendations for USG action. As data holders, USG agencies can make their relevant datasets more accessible to other parts of the USG, the private sector, academia, and civil society. As rule-maker, the USG creates restrictions, protections, and incentives to shape behaviors. As a trusted convener, the USG can bring together diverse groups to overcome barriers to sharing for the benefit of all.

The Problem: The United States Is Not Leveraging the Nation's Data Assets in the Global Competition

When compared to other advanced economies like the European Union (EU) and China, the United States has not effectively organized as a whole nation to fully leverage USG and non-USG data (private sector, academia, and civil society) assets to develop a robust and resilient domestic data ecosystem for economic and societal benefits.

While the United States is home to the world's leading technology firms and universities with enormous data assets, U.S. laws, policies, and regulatory systems have lagged far behind the exponential growth of data and its expanding role in governance systems, the economy, and the personal lives of citizens. For example, the United States lacks a comprehensive federal data privacy framework. While the United States generates enormous volumes of data, and may use it in narrow silos, very little of the real potential of data is used.[17]

The lack of a holistic U.S. approach to data also poses a growing threat to U.S. interests outside its borders, leaving nations to adopt data governance and localization policies hostile to U.S. strategic and economic interests in the absence of a clear U.S. data policy posture. The United States needs whole-of-nation data strategies and policies that leverage national data resources across government, industry, academia, and civil society, to gain global competitive advantages. Such broad strategies and policies would address multiple issues, many of which are beyond the scope of this report. A comprehensive overall national data approach would need to address domestic issues such as upholding democratic values beyond data privacy; coordinating data efforts across USG, industry, academia, and civil society to support national interests; and assessing the need for incentives like digital intellectual property (IP) or IP-type rights protections. Such strategies and policies also must grapple with international issues such as ensuring data policy alignment with partners and allies, establishing additional digital trade agreements, and adopting mechanisms directed to data threats from China and other adversaries.

tools can perform many image-based diagnoses as well or better than human experts.); see also Big Data: A Tool for Inclusion or Exclusion?, U.S. Federal Trade Commission at 7 (2016).

[17] "Building a better world through data is crippled by our using only a small fraction of the existing data. We deserve more from the data we provide and pay for. The sequestering of private and public data hurts society by making both private enterprise and government far less effective than they could be. Sharing data across sectors can help us better tackle societal problems and grow the economy." Robert M. Groves & Adam Neufeld, Accelerating the Sharing of Data Across Sectors to Advance the Common Good, Georgetown University at 23 (2017). Statistics about data volume generation are sometimes not correlated with the real value of the data. For example, the enormous volume of data created and used on streaming video every day in the United States says very little about the societal, economic, and national security data opportunities being advanced across society.

The absence of domestic governance frameworks impedes alignment with partners and allies to establish agreements on trustworthy data flows,[18] which are fundamental to setting the conditions for democracies to leverage their combined data resources more effectively in the global competition. Without a common approach to data, the United States and fellow democracies are at risk of ceding leadership in global data governance.

As U.S. partners and rivals race ahead to define the terms of the digital future, the United States' window of opportunity to lead is closing. Over the past few years, the EU, the world's second-largest democratic market economy and close U.S. partner, has established digital policies and data governance that other nations could adopt, especially given the lack of a U.S. response. In 2020, the EU developed an agenda under the von der Leyen Commission to achieve "digital sovereignty," described as "Europe's ability to act independently in the digital world."[19] This push for digital sovereignty relies in part on protecting European citizens' data, but also harnessing it for economic prosperity – within European borders. The EU has made significant headway to craft an approach to data governance with a heavy emphasis on regulation. The EU's regulatory frameworks, including the General Data Protection Regulation (GDPR) that went into force in 2018, the 2022 Digital Services Act package, and the draft European Data Act, are establishing precedents that impact the global collection and use of data.[20] The longer the United States delays establishing its own data governance approaches, the more the divergence between America's policy void and the EU's highly regulated approach will grow. This will leave a vacuum for other nations to fill with governance models that have global influence and do not align with U.S. interests and values or allowing for fragmented global data governance.

A lack of U.S. strategic vision and tech governance policies also hinders transatlantic digital trade and data transfer agreements. Two previous data transfer agreements were judicially invalidated by the European Court of Justice (ECJ) for insufficient protections under GDPR requirements,[21] and interrupted digital trade for many companies. Beyond the EU, it prevents the

[18] Distant U.S. ambitions such as treaties or conventions directed towards bilateral and multilateral data flows depend on the United States having in place privacy and security frameworks. Faced with another crisis, like the COVID-19 pandemic that has cross-border implications, the United States should be in a position to be able to quickly set up these types of international data flows.

[19] Tambiama Madiega, Digital Sovereignty for Europe, European Parliamentary Research Service (2020).

[20] The EU championed the first large scale and comprehensive data governance regulation in 2016 with the General Data Protection Rule (GDPR). This regulation had a considerable global impact, serving as a model for other countries' data regulations — and impacting companies considerably with far-reaching compliance needs. One of its biggest provisions is the "right to be forgotten," which grants individuals the right to ask companies to erase their data. Additionally, the GDPR had implications for international data flow agreements and digital trade agreements, as challenges arose for not complying with GDPR rules. More recently, the EU passed their Digital Services Act, closely followed by a UK equivalent — the Online Safety Bill — that includes provisions for users to be able to request access to their data collected by online services they are using. It has specific requirements for large platforms, principally targeting U.S. and Chinese online giants. The EU also started tackling non-personal data – also called "industrial data" – with their draft Data Act, seeking to create rules to encourage internal data sharing, but also implement fair and transparent B2B and Business-to-Consumer data sharing rules. EU's Digital Markets Act, DSA's sister act in the Digital Services Act package, restricts large online companies categorized as "gatekeepers" from aggregating users data without their consent, including data obtained by third parties and for online advertising purposes "legitimate interests" as a basis for combining of cross-use personal data or for sign in end users to other gatekeeper services in order to combine personal data.

[21] Following the 2013 Snowden revelations, the European Court of Justice (ECJ) ruled in Schrems I that the U.S.-EU Safe Harbor Framework did not meet adequacy requirements of the fundamental right to privacy and a fair trial under the Charter of Fundamental Rights of the European Union. C-362/14, Maximillian Schrems v. Data Protection

United States from increasing alignment on and ties with data issues with other allies and partners. The U.S. Government should be pursuing agreements on digital trade and cross-border data flows with like-minded partners[22] to advance its cyberspace vision and expand commercial opportunities for U.S. workers and firms.[23] The United States has examples of digital trade that it should build on: it has signed a few digital trade agreements, such as the U.S.-Japan Digital Trade Agreement and the digital chapter in the U.S.-Mexico-Canada (USMCA) trade agreement.[24] These offer potential models for future agreements with other economies, such as Australia, Chile, Colombia, Korea, New Zealand, Peru, Taiwan, the UK, and members of ASEAN.[25] The United States can also learn from allies that are further ahead in developing principled democratic approaches to data. Tokyo's "Data Free Flow with Trust" (DFFT) concept, for example, offers principles to bolster international cooperation on data flows to boost economic growth while offering data protections that align with democratic values.[26]

The People's Republic of China (PRC) is developing a comprehensive data governance regime that allows for state control over data and is moving ahead of the United States in establishing data strategies, laws, regulations, and policies to benefit its interests and shape global standards. Domestically, the PRC is bolstering government control over all data assets in China and restricting data flows out of China. Party officials last year issued one of the world's strictest pieces of data privacy legislation aimed at curbing data collection by PRC tech companies and reeling in a growing black market for data inside China, but leaving room for expanded government surveillance and data access.[27] Additionally, Beijing's economists have framed data as a "factor of production" – on par with land, labor, and capital – and Chinese Communist Party (CCP) leaders have created policies and regulatory frameworks designed to harness data for

---

Commissioner, European Court of Justice (2015). The US and EU agreed to another framework shortly thereafter – Privacy Shield – which the ECJ invalidated in the Schrems II decision. C-311/18, Data Protection Commissioner v. Facebook Ireland Limited, Maximillian Schrems, European Court of Justice (2020). Earlier this year, the TTC paved the way to a new Transatlantic Data Privacy Framework, which is reaching finalization. See FACT SHEET: United States and European Commission Announce Trans-Atlantic Data Privacy Framework, The White House (2022); FACT SHEET: President Biden Signs Executive Order to Implement the European Union-U.S. Data Privacy Framework, The White House (2022).

[22] Nigel Cory, U.S. Options to Engage on Digital Trade and Economic Issues in the Asia-Pacific, Information Technology & Innovation Foundation (2022).

[23] The Digital Trade Revolution: How U.S. Workers and Companies Can Benefit from a Digital Trade Agreement, U.S. Chamber of Commerce at 4 (2022).

[24] See Digital Trade & E-Commerce FTA Chapters, Office of the United States Trade Representative (last accessed 2022).

[25] The Digital Trade Revolution: How U.S. Workers and Companies Can Benefit from a Digital Trade Agreement, U.S. Chamber of Commerce (2022).

[26] The DFFT concept, championed by the late former Prime Minister Shinzo Abe during Japan's G20 presidency in 2019, is moving into the implementation phase and includes principles that would enable cross-border data flows to bolster digital trade and power economic growth. DFFT "maps a multi-dimensional architecture for international cooperation on data flows, between governments, as well as involving business, with recommendations to increase levels of governance trust and build openness through trade rules and other tools." See Data Free Flow with Trust (DFFT): Paths toward Free and Trusted Data Flows, World Economic Forum (2020). For a strategic rationale for the United States to develop a coherent national approach to data and strengthen alignment with allies and partners to strengthen leverage vis-a-vis Beijing in the global competition to control the terms of the data revolution, see Matthew Pottinger & David Feith, The Most Powerful Data Broker in the World Is Winning the War Against the U.S., New York Times (2021) (arguing that DFFT offers a blueprint for democratic allies to "work together to promote data sharing among themselves while limiting flows to China").

[27] Eva Xiao, China Passes One of the World's Strictest Privacy Laws, Wall Street Journal (2021).

economic benefit.[28] Data is seen as an amplifier of other factors of production in addition to having intrinsic value in a digital economy.[29] Beijing likewise is pursuing a vision of "cyber sovereignty"[30] – a concept that dates back to a 2010 PRC government white paper.[31] However, the PRC vision is one that severely restricts international data sharing and asserts control over all aspects of the digital ecosystem, representing a stark contrast from the vision promoted by the United States and many of its allies and partners for an "open, free, global, interoperable, reliable, and secure Internet."[32] As United States Secretary of Commerce Gina Raimondo recognizes, "[t]hey have firewalled their data economy from the rest of the world."[33] China also has been strengthening its leverage over international data collection.[34]

The PRC's growing influence over global data infrastructure and corresponding virtual networks, technical standards, and governance regimes increases the prospect that, if the United States fails to implement effective data policies and strategies,[35] an autocracy will take the lead in the

[28] Over the past decade, Beijing has been asserting stronger government control over data through a comprehensive regulatory framework and has rewritten the Party's version of Marxist economic theory to designate data as a distinct "factor of production," in addition to traditional factors like land, labor, and capital – illustrating its view that data is not just a byproduct of technology, but also the lifeblood of the digital economy. Qiheng Chen, China Wants to Put Data to Work as an Economic Resource - But How?, DigiChina (2022); Restoring the Sources of Techno-Economic Advantage, Special Competitive Studies Project at 11 (2022); Aynne Kokas, Trafficking Data: How China Is Winning the Battle for Digital Sovereignty, Oxford at 65-75 (2022).

[29] "[A]ccording to the Chinese Academy of Information and Communications Technology, a key distinction between data and the traditional production factors is in the multiplier effect—that data can amplify other factors of production such a s labor and capital and produce even more significant economic gains." Lindsay Gorman, China's Data Ambitions, National Bureau of Asian Research (2021).

[30] The PRC's Cybersecurity and Data Security Laws impose restrictions on cross-border data flows and establish data localization requirements designed to limit foreign access to PRC domestic data holdings. Aynne Kokas, Trafficking Data: How China Is Winning the Battle for Digital Sovereignty, Oxford at 65-67 (2022) (noting that these laws formalize the PRC government's access to all data generated in China and require that "critical information" be kept in PRC government-run servers). See also Adam Segal, China's Internet Conference: Xi Jinping's Message to Washington, Council on Foreign Relations (2015).

[31] The Internet in China, Information Office of the State Council of the People's Republic of China (2010) (archived by the National Security Archive, George Washington University).

[32] See A Declaration for the Future of the Internet, The White House (2022).

[33] Remarks by U.S. Secretary of Commerce Gina Raimondo on the U.S. Competitiveness and the China Challenge, U.S. Department of Commerce (2022).

[34] A lack of systemic data governance in the United States, combined with the growth of PRC technology platforms, such as TikTok and WeChat, in the U.S. market, leaves Americans vulnerable to having their data trafficked [or exploited] in ways that empower Beijing, since PRC laws enshrine the government's ability to access corporate data both domestically and internationally, as well as to disinformation elevated by opaque algorithms. Aynne Kokas, Trafficking Data: How China Is Winning the Battle for Digital Sovereignty, Oxford at 2-4 (2022). Kokas argues that "the movement of data from tech firms in the United States to China threatens digital sovereignty around the world," and that U.S. laissez-faire approaches to regulating data flows have provided an opening for the PRC government to build on U.S. tech firms' "long tradition of exploiting the public for commercial gain" to move U.S. citizens' data across borders without citizens' consent and use it to advance PRC state objectives. U.S. tech firms, meanwhile, are blocked out of the China market unless they submit to "formal centralized oversight of all corporate data as a condition of their presence in the Chinese market." In other words, the situation is highly asymmetric to China's strategic and economic benefit, and a failure to establish sufficient laws, regulations, and protections is essentially unilateral disarmament by the United States.

[35] For example, U.S. data strategies should focus on U.S. business and personal data transferred and stored outside our borders. The U.S. Government, in collaboration with industry partners, should develop policy concepts and technology solutions that safeguard U.S. business and personal data from being improperly collected and exploited by technology platforms operated by the PRC or other countries of concern. Policymakers could consider previously drafted provisions, such as Section 202 of the proposed American Data Privacy and Protection Act (ADPPA) that

digital revolution and gain control over the markets, information flows, and geopolitical power that are emerging from it. The PRC's digital influence to date has already enabled it to extend its domestic levers of social control to populations overseas, which could deepen as Beijing pursues a mutually reinforcing set of "technology spheres of influence" to project power abroad.[36]

requires covered entities to disclose whether data is made accessible to the PRC, Russia, Iran, or North Korea. See H.R.8152, American Data Privacy and Protection Act §202 (2022). If existing executive branch authorities prove insufficient to address the threat posed to U.S. data by adversary nations' platforms, new legal authorities should be developed. The U.S. approach should be guided by at least two criteria: national security and reciprocity. National security criteria should include factors such as the ownership, control, and management of the tech platforms, the ability of third parties to audit the platform, and the scope and sensitivity of the data being collected by the platform. Additional criteria should be developed to address the lack of reciprocity between U.S. and PRC data regulations as a barrier to market access – an obstacle that ultimately hurts American economic and technological competitiveness. Independent of any congressional action, steps can be taken to impose data security requirements on PRC tech platforms that impose or restrict conditions on the flow of U.S.-origin data back to China that could be exploited for national security purposes or used for PRC's own technology development.

[36] Emily de la Bruyere, et al., China's Digital Ambitions: A Global Strategy to Supplant the Liberal Order, National Bureau of Asian Research (2022); Samantha Hoffman, Engineering Global Consent: The Chinese Communist Party's Data-Driven Power Expansion, Australian Strategic Policy Institute (2019) (explaining how the party-state regime in Beijing engages in data collection on a massive scale in order to shape global sentiment in ways that favor the interests of the Chinese Communist Party over those of the state or individuals and requires a constant expansion of the PRC's technology-enabled authoritarianism overseas). See also Harnessing the New Geometry of Innovation, Special Competitive Studies Project at 10-14 (2022).

Approaches to Data Governance

Areas of Governance	U.S.	EU	PRC
 Domestic Data Privacy Laws	• **No comprehensive federal data privacy protection law for non-USG data.** • Sector-specific data privacy regulatory frameworks (e.g., HIPAA, FCRA, FERPA, GLBA, ECPA, COPPA, and VPPA). • Patchwork of states passing data privacy laws – CA, CO, CT, UT, IL, VA.	• **2018 General Data Protection Regulation (GDPR).** Model for other countries' data regulations. • **2022 Digital Services Act (DSA).** Increases accountability and transparency in how Internet platforms manage content. • **2022 Digital Markets Act (DMA).** Sister act of the DSA that regulates large online companies.	• **2021 Personal Information Protection Law.** Requires firms operating and handling data inside China to abide by certain data handling, storage, and protection requirements.
 Domestic Laws and Strategies for Leveraging Data for National Interests	• **2018 OPEN Government Data Act.** Makes USG data open and accessible through data.gov. • **Federal Data Strategy (FDS).** Principles and practices for making the most of USG data.	• **2018 EU Digital Strategy.** Relies on the concept of digital sovereignty. • **2022 EU Data Act.** Tackles non-personal data – also called "industrial data." • **2022 Data Governance Act.** Strengthens EU's digital single market's governance.	• Beijing has designated data as a distinct "factor of production," emphasizing that data is the lifeblood of the digital economy. • **14th Five Year Plan for National Informatization (2021).** Party policy document that calls for maximizing the utilization of data for economic development. • **2021 Data Security Law.** Endorses the idea of "data exchanges" where firms can trade and exchange public data deemed non-sensitive.
 International Data Sharing	• The United States has negotiated a number of individual data sharing agreements with allies and partners such as Japan and the EU, no comprehensive strategy toward international data sharing.	• The EU push for "digital sovereignty" in its Digital Strategy, is paving way towards containing European data within EU borders. • **U.S.-EU data agreement.** Safe Harbour and Privacy Shield were judicially invalidated for insufficient protections under GDPR requirements.	• **2010 White Paper on "the Internet in China."** Introduces concept where the CCP has control over all information and data flows in and through the country. • **2017 Cyber Security Law.** Codified views on Internet sovereignty and established top-level legal principles. • **2018 Measures for the Administration of Scientific Data.** Places restrictions on PRC and foreign scientific data flows and joint R&D projects from being shared outside China. • **2021 Data Security Law.** Establishes requirements for data localization inside China and security vetting of bulk data before being sent outside of the PRC.

The Way Forward to Increasing U.S. Data Accessibility

In parallel to developing broad national strategies and policies to fully leverage U.S. data assets (both USG and non-USG data), the USG must implement near-term actions to make quality data accessible in and outside the USG, while ensuring privacy protections. Improving data accessibility is a tremendous challenge. Barriers to unlocking the full potential of data held by USG and non-USG entities include bureaucratic friction, privacy and security concerns, IP and legal risks, lack of infrastructure and funding, misaligned incentives, and missing leadership prioritization. We must overcome these challenges with a sense of urgency to maintain geopolitical competitiveness.

SCSP recommendations for increasing the accessibility of national data assets as a lever in the global competition are organized into three action areas:

1. **Protect and promote public trust in the U.S. data ecosystem;**
2. **Accelerate USG data accessibility by USG and non-USG entities (academic, private sector, and civil society); and**
3. **Facilitate non-USG data accessibility by USG and other non-USG entities.**

Three Areas of Action

Data Privacy

- Pass comprehensive Federal data privacy protections laws
- Incentivize the creation and use of privacy-enhancing technologies
- Apply data governance principles to USG and regulated data

USG Data
to USG/non-USG recipients

Executive Actions

- Establish national level CDO in OMB
- Mandate OMB complete Evidence Act guidance, issue a 2023 FDS action plan, and require agencies meet milestones
- Mandate development/deployment of data.gov V2 based on stakeholder feedback

Legislative Actions

- Extend lifetime of CDO Council
- Fully authorize and fund the development and implementation of the NAIRR

Non-USG Data
to USG/non-USG recipients

Executive Actions

- Convene stakeholder engagements in each of the six priority technology areas highlighted by the Mid-Decade Challenges to National Competitiveness report to assess high potential areas for increasing appropriate data accessibility
- Require CDOs to stand up narrowly-scoped PPP to facilitate data accessibility to solve specific challenges
- Require CDO Council, or national-level CDO if established, to analyze infrastructure, capabilities, policies, and funding needed to ensure agencies and their respective CDOs are enabled to be sophisticated consumers of non-USG data

Action Area 1: Protect and Promote Public Trust in the U.S. Data Ecosystem

Data privacy measures are foundational for the U.S. public to trust the U.S. data ecosystem.[37] Public awareness of the data privacy and protection practices of an entity that holds their data, such as knowing how their personal data will be treated, leads to greater trust in the entity and its use of their data.[38]

The U.S. public is well-attuned to the proliferation of USG and non-USGs entities collecting their data and the resulting privacy risks. According to a 2019 Pew Research Center Study, "roughly six-in-ten U.S. adults say they do not think it is possible to go through daily life without having data collected about them by companies or the government."[39] In parallel, public trust in data practices is challenged as reflected by "some 81% of the public say[ing] that the potential risks they face because of data collection by companies outweigh the benefits, and 66% say[ing] the same about government data collection."[40] With increasing regularity, news reports confirm Americans' perception – data about an individual, combined with data about others, has been used for making inferences about or influencing a group, even for nefarious purposes.[41]

To combat this skepticism, trustworthiness in data practices must be established.[42] If the United States is to remain a technology leader, the U.S. Government must lead with nationwide data policy and governance solutions that protect individuals' right to privacy. U.S. policies must require networks, products, and services that rely on data to be trustworthy. Foremost, data collection, storage, use, and sharing practices must protect the right to privacy.

[37] A discussion of public trust in the data ecosystem would not be complete without also addressing cybersecurity. The Cyberspace Solarium Commission was a Congressionally mandated commission charged to "develop a consensus on a strategic approach to defending the United States in cyberspace against cyber attacks of significant consequences." Pub. L. 115-232, John S. McCain National Defense Authorization Act for Fiscal Year 2019 §1652 (2018). As the Cyberspace Solarium Commission report observes, "Data security is a necessary first step for data privacy, because if the security of data is not guaranteed, its privacy cannot be either." Cyberspace Solarium Commission Report, U.S. Cyberspace Solarium Commission at 94 (2020). The Cyberspace Solarium Commission proposes a strategy of layered cyber deterrence and provides 80 recommendations to implement the strategy. In addition, cyber security protections of USG data are mandated by the Federal Information Security Modernization Act, which sets forth base security requirements for Federal agencies' information technology systems and state agencies that administer federal programs–with the chief goal of data protection. See Federal Information Security Modernization Act, U.S. Cybersecurity & Infrastructure Security Agency (last accessed 2022).

[38] Jim Boehm, et al., Why Digital Trust Truly Matters, McKinsey & Company (2022).

[39] Brooke Auxier, et al., Americans and Privacy: Concerned, Confused, and Feeling Lack of Control Over Their Personal Information, Pew Research Center (2019).

[40] Brooke Auxier, et al., Americans and Privacy: Concerned, Confused, and Feeling Lack of Control Over Their Personal Information, Pew Research Center (2019).

[41] Nicholas Confessori, Cambridge Analytica and Facebook: The Scandal and Fallout So Far, New York Times (2018). Cambridge Analytica harvested tens of millions of Facebook user profiles to target U.S. voters in advance of the 2016 election is an example of activities undermining public trust.

[42] Nadia Hewett, Responsible Data Collection Could Inspire Consumer Trust - Here's How, Forbes (2021).

Privacy rights are enshrined in federal laws, however, they are largely directed at the USG.[43] While there are notable laws that address sector-specific data,[44] there is no comprehensive U.S. law, regulation, or policy governing data privacy of all types of data in the non-government sector. And even where there are laws and regulations, there are gaps, such as in the digital health space, leaving Americans' data subject to exploitation.[45] This void in governing the privacy of nonpublic data, particularly consumer data, and its implications for society, is significantly exacerbated by the digital revolution and the ever-increasing production and use of data.

With no federal data privacy standard or regulatory structure, states have taken the lead, leaving companies to face an uneven patchwork of privacy compliance requirements. California, for example, has enacted sweeping state-level privacy legislation,[46] and many other states such as Colorado, Connecticut, Utah, and Virginia have passed their own consumer data privacy laws.[47] While the right to access, delete, and port personal information is a universal feature of these state laws, along with the right to opt-out of the sale of personal information, that is where the similarities end.[48] As the patchwork of laws persist, many Americans are left vulnerable and businesses must navigate regulatory uncertainty.

[43] The Privacy Act of 1974, for example, established rules for Federal government collection, storage, use, and disclosure of personal information. The Privacy Act also granted individuals the right to request personal records, to request certain changes such as inaccuracy in personal records, and to be protected from unwarranted invasion of privacy. 5 USC § 552a (1974).

[44] Thorin Klosowski, The State of Consumer Data Privacy Laws in the US (And Why It Matters), New York Times (2021) ("The United States doesn't have a singular law that covers the privacy of all types of data. Instead, it has a mix of laws that go by acronyms like HIPAA, FCRA, FERPA, GLBA, ECPA, COPPA, and VPPA"). See also, for example, the Health Information Portability and Accountability Act (HIPAA) for the healthcare sector and Gramm-Leach-Bliley Act (GLBA) for the financial sector. Pub. L. 104-191, Health Information Portability and Accountability Act of 1996 (1996); Pub. L. 106-102, Gramm-Leach-Bliley Act (1999). HIPAA covers communications between an individual and "covered entities," which include doctors, hospitals, pharmacies, insurers, and other similar entities, but does not, for example, protect Fitbit data. GLBA requires consumer financial products, like loan services or investment-advice services, to explain how they share data, as well as the customer's right to opt out, but does not restrict how companies can use collected data if such usage is disclosed beforehand.

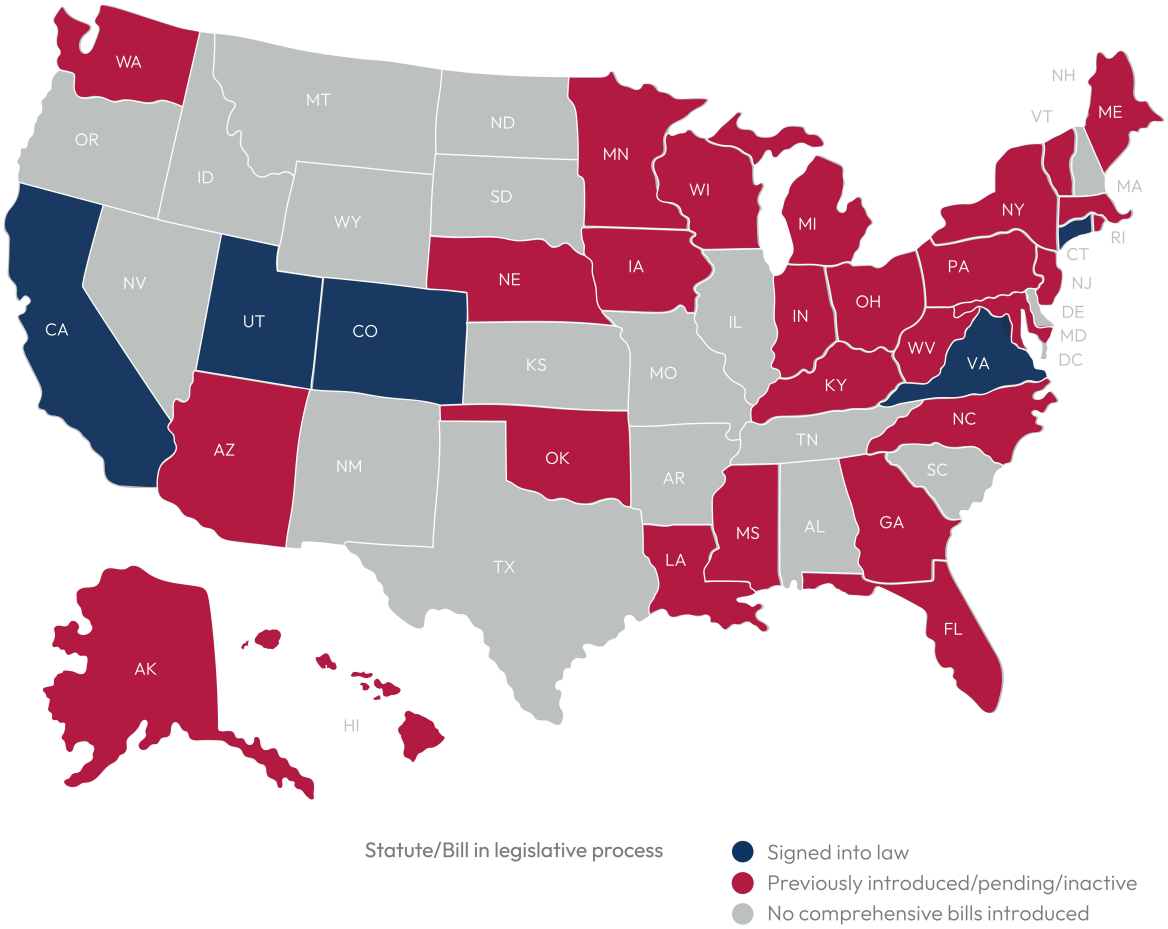
[45] Tatum Hunter & Jeremy B. Merrill, Health Apps Share Your Concerns With Advertisers. HIPAA Can't Stop It, Washington Post (2022); Thorin Klosowski, The State of Consumer Data Privacy Laws in the US (And Why It Matters), New York Times (2021).

[46] See, for example, California Consumer Privacy Act of 2018, California (2018) (giving consumers the right to request a business to disclose the categories and specific pieces of personal information that the business has collected about the consumers as well as the source of that information and business purpose for collecting the information); Data Broker Registration, California (2019) (defining a "data broker" as a "business that knowingly collects and sells to third parties the personal information of a consumer with whom the business does not have a direct relationship," and requiring data brokers to register with, and provide certain information to, the Attorney General); Privacy Rights for California Minors in the Digital World Act, California (2013) (allowing minors to remove, or to request and obtain removal of, content or information posted on an Internet Web site, online service, online application, or mobile application).

[47] State Laws Related to Digital Privacy, National Conference of State Legislatures (2022). See also Data Privacy, U.S. Chamber of Commerce (last accessed 2022). In 2008, Illinois became the first state to enact a biometric data privacy law. The law establishes requirements for any organization operating in Illinois that is using or storing biometric identifiers. It provides a private right of action for recovering statutory damages when they do not comply with the requirements. Biometric Information Privacy Act, Illinois (2008).

[48] U.S. State Privacy Legislation Tracker, IAPP (2022).

U.S. State Privacy Legislation Tracker 2022[49]



The varied regulatory and policy landscape among state governments presents an arduous compliance environment with cost implications for the private sector, academia, and civil society. To put these costs in concrete terms, the Information Technology & Innovation Foundation (ITIF) estimates that the patchwork of state-enacted privacy laws will impose out-of-state costs of $98 to $112 billion annually with small businesses bearing $20 to $23 billion annually. In the absence of comprehensive federal privacy law, the ITIF projects that these out-of-state costs would exceed $1 trillion over a 10-year period.[50]

[49] Anokhy Desai, U.S. State Privacy Legislation Tracker, International Association of Privacy Professionals (2022).
[50] Daniel Castro, et al., The Looming Cost of a Patchwork of State Privacy Laws, Information Technology & Innovation Foundation (2022).

Pass Comprehensive Federal Data Privacy Protections Into Law

The United States must take action to protect the privacy of individuals' data. Congress should pass comprehensive data privacy legislation that establishes a sufficient level of protection and sets reasonable, transparent, consistent standards by which data actors must abide. Legislation[51] should:

1. Give individuals the right to access, correct, delete, or port their personal data;
2. Promote data minimization and prohibit organizations from collecting, storing, using, or transferring data beyond what is reasonably necessary;
3. Provide transparency requirements for how organizations manage data, including providing reasonable notice of how data is collected and used;
4. Provide the ability to object[52] to the transfer of personal data;
5. Provide the ability to object to targeted advertising;
6. Prohibit discrimination against those individuals who exercise their privacy rights;
7. Close covered party loopholes in existing agency sector-specific privacy laws;[53]
8. Provide reasonable controls over third-party data brokers;
9. Establish stronger privacy protections for minors, including prohibiting certain marketing toward children;
10. Provide clear definitions of the types of data covered and for which entities the law would apply; and
11. Provide clear mandates for the Federal Trade Commission (FTC) to enforce the established standards.[54]

During the 117th Congress, legislators proposed comprehensive federal data privacy protections in the form of the American Data Privacy and Protection Act (ADPPA). This bi-partisan legislation

[51] Such principles are consistent with tenets of the Blueprint for an AI Bill of Rights including rights to Data Privacy and Notice and Explanation. Blueprint for an AI Bill of Rights: Making Automated Systems Work For the American People, The White House at 6 (2022).
[52] Objection could come in the form of "opt-in" or "opt-out' requirements.
[53] For example, personal health data provided by individuals to health applications not determined to be a covered entity under HIPAA are not subject to the same privacy and notifications protections as health data provided to a HIPAA-covered entity. See The Access Right, Health Apps, & APIs, U.S. Department Health & Human Services (2021). The ADPPA, as reported from the House Energy and Commerce Committee, would not expand HIPAA coverage to such health applications; however, it would subject these applications to limitations on collection and transfer of such "sensitive covered data." See H.R.8152, American Data Privacy and Protection Act (2022).
[54] In its March 2020 report, the Cyberspace Solarium Commission (CSC) recommended passage of comprehensive data privacy legislation that includes: national minimum common standards for the collection, retention, analysis, and third-party sharing of personal data; definitions of personal data, to include that which can be linked, directly or indirectly, to individuals or households; thresholds for what entities are covered by the legislation; timelines for deleting, correcting, or porting personal data upon request by the appropriate persons; and a clear mandate for the FTC to enforce the standards with civil penalties. Cyberspace Solarium Commission Report, U.S. Cyberspace Solarium Commission at 93 (2020).

attempts to address all of these principles to varying degrees.[55] Different from prior attempts at national privacy legislation, the ADPPA is a "comprehensive law" that many privacy experts assert is "distinctly stronger" and broader than state laws such as the California Consumer Privacy Act. The ADPPA would also have required transparency by covered entities about covered data that is collected and transferred to, processed by, or hosted in other adversarial countries like China, Russian, North Korea, and Iran.[56] Congress must consider, debate, and pass comprehensive legislation, like the ADPPA, that encompasses these principles and accounts for or mitigates national security risks. Overall, the United States can no longer wait to act. Further delay on establishing this foundational element of a comprehensive approach to data privacy puts the United States at risk of even greater social harms and lost economic opportunity.

Incentivize the Creation and Use of Privacy Enhancing Technologies

Absent Congressional action, there are steps the USG can take, in partnership with the private sector, to increase data privacy protections. One important step is to incentivize the creation, integration, and adoption of privacy enhancing technologies (PETs) and approaches.[57] While there is no one consensus definition of a "PET,"[58] the term generally refers to technology that mitigates threats to privacy in data use and sharing. Techniques include federated learning,[59]

[55] Jonathan M. Gaffney, et al., Overview of the American Data Privacy and Protection Act, H.R.8152, Congressional Research Service (2022). See also Daniel Castro, Review of the Proposed "American Data Privacy and Protection Act," Part 1: State Preemption and Private Right of Action, Information Technology & Innovation Foundation (2022); Daniel Castro, Review of the Proposed "American Data Privacy and Protection Act," Part 2: The Good and the Bad, Information Technology & Innovation Foundation (2022).

[56] See Cameron Kerry, Will California Be The Death of National Privacy Legislation?, Brookings Institution (2022); Brandon Pugh, Will California Derail National Push to Protect Data Privacy?, R Street Institute (2022).

[57] "Privacy-enhancing technologies (or PETs for short) allow us to analyze data while protecting people's personal information and company's confidential business information." Privacy-Enhancing Technologies – A Day with PETs, Deloitte (last accessed 2022). There are a variety of kinds of PETs, but they all provide ways to limit access to sensitive data while still enabling processing of the data.

[58] "There is no single definition or standard for what constitutes a PET, though the term is typically used to refer to technologies or approaches that can help mitigate privacy and security risks. … Leading academic researchers define PETs as a 'wide array of technical means for protecting users' privacy', while industry stakeholders use the term to refer to various technical means for protecting privacy by providing anonymity, pseudonymity, unlinkability and unobservability of data subjects. Policymakers typically use the term 'PETs' to refer to technological tools or methods that help to achieve compliance with privacy or data protection legislation or requirements, often in combination with organisational measures, including information security-related policies and procedures, personnel management and access controls, recordkeeping, and audits, among others." Elizabeth Renieris, Why PETs (Privacy-Enhancing Technologies) May Not Always Be Our Friends, Ada Lovelace Institute (2021).

[59] "Federated learning is an emerging approach allowing the training of machine learning models on decentralised data for privacy or practical reasons. A central server coordinates a network of nodes, each of which has training data. The nodes each train a local model, and it is that model which is shared with the central server. In other words, data is protected at the device level." Protecting Privacy in Practice: The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis, The Royal Society at 50 (2019).

differential privacy,[60] homomorphic encryption,[61] and secure multi-party computation.[62] The last few years have seen a significant amount of venture capital investment in PET developers.[63] PETs can provide researchers and businesses the ability to analyze sensitive data "without ever having access to the data itself."[64] PETs could revolutionize research by allowing new collaborations among entities and individuals who would otherwise not have access to vast but sensitive datasets. Such embedding of data protection and privacy would support data governance – minimizing misuse, maximizing innovation, and enabling peer-to-peer data sharing agreements.

The White House has recently given attention to the importance of PETs. In July 2022, the White House's Office of Science and Technology Policy (OSTP) issued a request for public comment on privacy-preserving data sharing, seeking "to better understand how to accelerate the responsible development and adoption of PETs in a manner that maximizes the benefit to individuals and society, including increasing equity for underserved or marginalized groups and

[60] Unlike many PETs which address privacy during computation, differential privacy addresses privacy in disclosure of the dataset or result by ensuring that it does not give "much more information about a particular individual than if that individual had not been included in the dataset." It, thus, mitigates "the risk of revealing whether a specific individual or organisation is present in a dataset or output" and can be applied at different phases of the data analysis lifecycle. "Differentially private mechanisms can, in particular, provide secure public access to private datasets and protect data whilst disclosing derived information." Protecting Privacy in Practice: The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis, The Royal Society at 41-43 (2019).

[61] "Homomorphic encryption is a form of encryption that allows certain computations on encrypted data, generating an encrypted result which, when decrypted, matches the result of the same operations performed on the data before encryption. … Homomorphic encryption can be used to analyze data in circumstances where all or part of the computational environment is not trusted, and sensitive data should not be accessible. … Compared with computing on unencrypted data, homomorphic encryption is extremely computationally expensive and has lower throughput. Encryption can entail a substantial increase in data size, which can cause a major bandwidth problem." Protecting Privacy in Practice: The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis, The Royal Society at 31-33 (2019).

[62] "Secure multi-party computation (MPC) is a subfield of cryptography concerned with enabling private distributed computations. MPC protocols allow computation or analysis on combined data without the different parties revealing their own private input. In particular, it may be used when two or more parties want to carry out analyses on their combined data but, for legal or other reasons, they cannot share data with one another. For example, MPC can allow bidders to identify who has won an auction without revealing anything about the actual bids. … Whilst secure multi-party computation has been applied in a limited number of 'products', research and development is ongoing and other applications are at a 'proof of concept' stage." Protecting Privacy in Practice: The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis, The Royal Society at 38-41 (2019).

[63] The Investors' View on Privacy-Enhancing Technologies, Kisaco Research (last accessed 2022) ("The PETs industry is projected to grow from approximately 2.4 billion USD IN 2022 to 26 billion USD by 2029."). This is a reflection of opportunities in start-ups responding to the growing demand signal for PETs. Gartner Identifies Top Security and Risk Management Trends for 2021, Gartner (2021) ("Trend 6: Privacy-enhancing computation techniques are emerging that protect data while it's being used — as opposed to while it's at rest or in motion — to enable secure data processing, sharing, cross-border transfers and analytics, even in untrusted environments. … Gartner predicts that by 2025, 50% of large organizations will adopt privacy-enhancing computation for processing data in untrusted environments or multiparty data analytics use cases."). Indeed, privacy budgets across varying sized organizations are growing. Forged by the Pandemic: The Age of Privacy, Cisco at 9 (2021). A Future of Privacy report describes trends in the Privacy Tech marketplace. Buyers increasingly want enterprise-wide solutions and integrated technologies over specific narrow PET products. Some vendors are moving to either collaborate and integrate or provide fully integrated solutions themselves. New FPF Report Highlights Privacy Tech Sector Evolving From Compliance Tools to Platforms For Risk Management and Data Utilization, Future of Privacy Forum (2021).

[64] Alexander Macgillivray & Tess deBlanc-Knowles, Advancing a Vision for Privacy-Enhancing Technologies, The White House (2022).

promoting trust in data processing and information technologies."[65] OSTP has indicated the feedback will be used to develop a national strategy[66] and identify actions – to include investment for research and education – to advance and adopt privacy-preserving data sharing technologies.[67] The government is also seeking to incentivize innovation in PETs. In July 2022, the White House announced a $1.6 million, joint U.S.-U.K. prize challenge for the development of privacy-preserving technologies.[68]

To realize the full potential of PETs, more research and development is needed. The USG should continue to prioritize and fund not only the open source advancement of individual PETs,[69] but more importantly, their integration into demonstration sandboxes; and within this technology infrastructure, establish safe harbor laws and policies to lower friction and risks for industry, academia, government, civil society, and international partners to participate in these sandboxes with the intent to scale successful demonstrations.[70] To this end, the USG should partner with the private sector in these pilots in order to bring regulatory clarity[71] along with trustworthiness stemming from demonstrated auditability and oversight, which is inherently challenging given the complexity of integrated PET solutions (e.g., due to their typical use of strong encryption and anonymity).[72]

[65] 87 Fed. Reg. 35250, Request for Information on Advancing Privacy-Enhancing Technologies, The White House, Office of Science and Technology Policy (2022).
[66] Aaron Boyd, White House Developing National Strategy to Increase Data Collection as Privacy Tech Improves, Nextgov (2022).
[67] 87 Fed. Reg. 35250, Request for Information on Advancing Privacy-Enhancing Technologies, Office of Science and Technology Policy (2022).
[68] The U.S.-UK prize challenge encourages innovators to "develop privacy-preserving federated learning solutions that enable artificial intelligence models to be trained on sensitive data without organizations having to reveal, share, or combine their raw data." U.S. and U.K. Launch Innovation Prize Challenges in Privacy-Enhancing Technologies to Tackle Financial Crime and Public Health Emergencies, The White House (2022). The technologies are to focus on targeting financial crime and pandemic response. In November 2022, 12 prize winning-papers were selected; the teams now compete through building the solutions envisions in their technical papers. Winners Announced in First Phase of U.S.-UK Privacy-Enhancing Technologies Prize Challenges, U.S. National Science Foundation (2022); U.S. PETs Prize Challenge, DrivenData (last accessed 2022).
[69] As recommended in the NSCAI Final Report, the USG should "[a]ssure privacy protection in data use for AI development and operation through advancements in anonymity techniques and technologies such as multi-party federated learning." Final Report, National Security Commission on Artificial Intelligence at 190 (2021).
[70] The NSCAI Final Report also recommends that "The United States should work with key allies and partners to establish the Multilateral AI Research Institute (MAIRI). MAIRI will facilitate joint efforts to develop technologies that advance responsible, human-centric, and privacy-preserving AI/ machine learning (ML) that better societies and allow allies to pool their talents and resources." Final Report, National Security Commission on Artificial Intelligence at 249 (2021). The Final Report identifies a set of priorities for the initial research agenda for MAIRI, which includes "Privacy-preserving AI/ML technologies, including technologies like federated learning and on-device prediction that enable remote execution, encrypted computation through multi-party computation and homomorphic encryption, and differential privacy." Id. at 538 (2021). See also Andrew Trask, AIME Presentation: Privacy-preserving AI - July 14, 2022, U.S. National Institute of Standards and Technology at 54:30 minutes (2022).
[71] Sebastiao Barros Vale, Event Report: FPF Side Event and Workshop on Privacy Enhancing Technologies (PETs) at the 2022 Global Privacy Assembly (Gpa), Future of Privacy Forum (2022).
[72] Elizabeth Renieris, Why PETs (Privacy-Enhancing Technologies) May Not Always Be Our Friends, Ada Lovelace Institute (2021).

Apply Data Governance Principles

In addition, applying the four principles for AI governance described in Mid-Decade Challenges to National Competitiveness to data governance would further engender public trust in the U.S. data ecosystem.[73]

Govern by Use and Sector: First, organize governance by the agency sector of the intended ultimate use of the data. The risks and opportunities presented by data collection are tied to the context in which data are used. Currently, the United States is pursuing agency sector-specific efforts to regulate AI by adapting existing regulatory frameworks and agencies to address new issues introduced by the adoption of AI, and the use of data in these sectors often overlaps with the ways AI is used.[74] This principle will not cover all data governance issues given regulators sometimes have authority over the use and not the data itself.

Empower and Modernize Existing Regulators: Second, empower and modernize existing regulatory bodies. The United States should rely on its existing constellation of agency sector-specific regulators,[75] which can be equipped to address new regulatory needs raised by data collection and use. Existing regulatory bodies have the sector expertise that allows for tailoring rules, ensuring that the data governance complements existing governance, and assessing impacts. However, we must identify the resources these agencies currently lack to address regulatory challenges posed by data collection and use.

Focus on High-Consequence Uses:[76] Third, focus governance on high-consequence use cases, both beneficial and harmful. Because it is impractical to govern every instance of data collection and use, regulation should focus attention on the most high-consequence instances. The United States needs a framework for categorizing data use cases. Categories should, at a minimum, account for use cases that 1) have the potential to cause significant harm to individuals or

[73] Mid-Decade Challenges to National Competitiveness, Special Competitive Studies Project at 87 (2022).

[74] Sector-specific regulation adapts agencies' existing regulatory frameworks to address new issues introduced by the adoption of AI. Examples include the Food and Drug Administration (FDA) for machine learning (ML) as a medical device and good ML manufacturing processes, the aviation community for how AI in safety-critical avionics should be addressed, and the FTC applying its current regulatory authorities to new commercial uses of AI DOT/FAA/TC-16/4. See Artificial Intelligence and Machine Learning in Software as a Medical Device, U.S. Food and Drug Administration (2021); Good Machine Learning Practice for Medical Device Development: Guiding Principles, U.S. Food and Drug Administration (2021); Verification of Adaptive Systems, U.S. Federal Aviation Administration (2016); Christoph Torens, et al., Guidelines and Regulatory Framework for Machine Learning in Aviation, AIAA SCITECH 2022 Forum (2021); Elisa Jillson, Aiming for Truth, Fairness, and Equity in Your Company's Use of AI, U.S. Federal Trade Commission (2021).

[75] For example, the Food and Drug Administration has established regulatory guidance for aspects of data for medical device approval and surveillance and for pharmaceutical manufacturing. See e.g., Data Integrity and Compliance With Drug Current Good Manufacturing Practice, U.S. Food and Drug Administration (2018); FDA Data Standards Advisory Board, U.S. Food and Drug Administration (2022). The FTC has regulatory authority in several aspects of data collection including control over what information websites can collect from children, consumer privacy, the use of credit reports, and compliance with the Gramm-Leach-Bliley Act that requires financial institutions to explain their information-sharing practices to their customers. Privacy and Security Enforcement, U.S. Federal Trade Commission (last accessed 2022).

[76] There are challenges in adapting these governance principles to data collection. It is challenging, if not impossible, to anticipate the various ways that data may be shared and combined, including in ways that are high-consequence only in combination. This makes a focus on high-consequence use cases more about governing the use of data and the outcomes of such use.

communities, such as widespread discrimination or privacy violations, and 2) have high potential for positive impact that is inhibited by a lack of governance, such as providing for equitable outcomes.

Strengthen Non-Regulatory Mechanisms: Fourth, strengthen non-regulatory data governance. In addition to its regulatory guardrails, the United States should strengthen and nurture its robust non-regulatory ecosystem as it relates to data collection and use. Civil society participation in governance is an American strength, and non-regulatory mechanisms draw on this by exerting power through incentives and public opinion. Illustrations include technically-based investigative journalism (for example the ProPublica investigation into algorithmic decision support in criminal justice)[77] and independent third-party audits (for example, an audit by university researchers of a health care cost estimation system that affected millions of patients).[78]

[77] Julia Angwin, et al., Machine Bias, ProPublica (2016).
[78] Heidi Ledford, Millions of Black People Affected by Racial Bias in Health-Care Algorithms, Nature (2019). Other examples are voluntary standards and best practices, self-governance, advocacy, philanthropy, policy research, legal recourse, government contracting requirements, government funding, incentives, waivers, exemptions, Congressional public hearings and investigations to inform potential legislation, and government-issued policy guidance or frameworks. Mid-Decade Challenges to National Competitiveness, Special Competitive Studies Project at 89, fn. 16 (2022).

Data Gray Zones:

There are also "gray zones" where these data governance principles may fall short. One challenge is data collected for one purpose and under a specific governance regime can be repurposed for different uses that are not governed.[79] Another "gray zone" challenge is that these governance principles do not address privacy and rights abuses due to third-party data brokers and the use of combined datasets to identify sensitive data about individuals.[80] Concerns about privacy violations by data brokers are a very important part of data privacy concerns for regulators and for legislation.[81] Addressing these gray zones will require evolving data privacy frameworks and a focus on data aggregation and off-purpose use and resale. While regulation and privacy laws can evolve to address some of the gray zone challenges, the data inference threat is largely a consequence of the use of various data sets rather than the access to or sharing of them. In this way, addressing inappropriate inference use is more like the management of data breaches and the regulation of uses of data that may violate specific prohibitions in a sector (e.g., loan approval or making an employment offer). This will require existing agency sector regulators to monitor for signals of such abuses and to establish easily accessible mechanisms for the reporting of suspected inappropriate use and information sharing, as is done for cybersecurity.

[79] For example, the primary legal structure governing the use of personal health information (PHI) is the Health Insurance Portability & Accountability Act of 1996. Fitness and health wearables and supporting online apps are generally not required to comply with HIPAA when they are used for personal, self-health tracking and do not directly transmit PHI to an electronic health record. This leaves a considerable amount of personal health-related data potentially unprotected for resale, aggregation, and application for other purposes. Data Privacy When Using Wearable Health and Fitness Devices, Maryland Health Care Commission (2022).

[80] Sandra Wachter & Brent Mittelstadt, A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI, Columbia Business Law Review (2018); Zeynep Tufekci, Think You're Discreet Online? Think Again, New York Times (2019).

[81] On August 11 2022, the FTC announced that it is exploring new rulemaking to address what it refers to as commercial surveillance. "Commercial surveillance is the business of collecting, analyzing, and profiting from information about people. …Companies use algorithms and automated systems to analyze the information they collect. And they make money by selling information through the massive, opaque market for consumer data, using it to place behavioral ads, or leveraging it to sell more products." Agency Seeks Public Comment on Harms from Business of Collecting, Analyzing, and Monetizing Information About People, U.S. Federal Trade Commission (2022). The proposed American Data Privacy and Protection Act (ADPPA) also address data brokers "The ADPPA requires data brokers to register with the FTC. Under the bill, the FTC will establish and maintain an online, searchable, central public registry of all registered data brokers, and a 'Do Not Collect' registry, which will allow individuals to request that all data brokers delete their data within 30 days. The ADPPA will also enable third-party audits of how data brokers share information with others." American Data Privacy and Protection Act – Could a Federal Privacy Law be on the Horizon?, Bass, Berry, & Sims (2022).

Action Area 2: Improve Accessibility of USG Held Data by USG and non-USG entities

The USG collects and maintains diverse data sets related to issues and sectors such as agriculture, housing, and waterway navigation, among many others.[82] While not all data for any purpose should be made available to anyone, maximizing the accessibility of USG data supports more effective policy implementation, economic growth, transparency and trust, and social good.

Congress and the Executive Branch recognize the importance of data and have put in place requirements to increase USG data accessibility and use. The two most notable mandates are the 2018 Open, Public, Electronic and Necessary (OPEN) Government Data Act (Title II of the Foundations for Evidence-Based Policymaking Act of 2018)[83] and the Office of Management and Budget (OMB)'s 2019 Federal Data Strategy (FDS).[84] The OPEN Government Data Act, for example, requires that agencies make data assets open by default when not prohibited by law and to the extent practicable.[85] The OPEN Government Data Act mandate places open data access and management at the heart of making USG data widely available. The Act also requires the Government Services Administration (GSA) to maintain a single public interface online for a Federal Data Catalog.[86]

In 2019, OMB committed to issuing guidance to facilitate agency compliance with the OPEN Government Data Act.[87] However, OMB has yet to release agency guidance for "Open Data Access and Management."[88] Also in 2019, OMB established a Federal Data Strategy, and

[82] Highlights, Data.gov (last accessed 2022).

[83] With the goals of enhancing transparency and harnessing the innovative and economic value of USG data, the OPEN Government Data Act builds off the 2013 Executive Order, Making Open and Machine Readable the New Default for Government Information, and mandates that agencies make public and nonpublic data assets open by default at no cost to the public. Pub. L. 115-435, Foundations for Evidence-Based Policymaking Act of 2018, § 202 (2018), see also Executive Order -- Making Open and Machine Readable the New Default for Government Information, The White House (2013).

[84] M-19-18, Federal Data Strategy - A Framework for Consistency, U.S. Office of Management and Budget (2019). The FDS sets forth a strategy to extract value from USG data. It lays out data "principles" and practices and indicates that annual action plans will be issued that identify and prioritize practice-related steps for a given year, along with target timeframes and responsible entities." An important example of a practice is "Leverage Partnerships: Create and sustain partnerships that facilitate innovation with commercial, academic, and other partners to advance agency mission and maximize economic opportunities, intellectual value, and the public good." In addition, OMB issued a Federal Data Strategy 2021 Action Plan that outlines detailed steps for creating a federal space for the virtual exchange of information along with guidelines and suggestions on timelines for achieving certain goals. Federal Data Strategy: 2021 Action Plan, U.S. Office of Management and Budget (2021).

[85] H.R.1770 - OPEN Government Data Act, § 3562(b) (2018). The Act further mandates that agencies develop an open data plan that prohibits the dissemination and accidental disclosure of nonpublic data assets. H.R.1770 - OPEN Government Data Act, § 3564(a)(2)(F) (2018). The Act also requires each agency to develop and maintain a comprehensive data inventory. Some agencies have followed through with collecting and making their data more accessible. The Act further requires every agency to establish the role of Chief Data Officer (CDO) (Section § 3520. Chief Data Officers in H.R.4174 - Foundations for Evidence-Based Policymaking Act of 2018 (2018)), and requires OMB to establish a CDO Council and to issue guidance to agencies for Evidence Act compliance. Section 3520A. Chief Data Officer Council in H.R. 4174 - Foundations for Evidence-Based Policymaking Act of 2018 (2018).

[86] GSA hosts this data catalog on Data.gov.

[87] M-19-23, Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance, U.S. Office of Management and Budget (2019).

[88] In Phase II guidance, OMB plans to include information for agencies on implementing the provisions in Title II of the Evidence Act - the OPEN Government Data Act - to (1) develop and maintain comprehensive data inventories, and (2) fulfill their responsibilities to make data open by default. M-19-23, Phase 1 Implementation of the Foundations for

committed to providing annual FDS Action Plans that "identify and prioritize practice-related steps for a given year, along with target timeframes and responsible entities" in support of advancing the FDS.[89] In 2022, the Office of Science and Technology Policy (OSTP) issued the Blueprint for an AI Bill of Rights[90] and updated a 2013 OSTP Memorandum that helped reshape the landscape for data and research by sharing results from federally funded research freely and openly with the public and the scientific community.[91]

Despite the above actions, at least four barriers have prevented making better use of USG data: 1) inadequate guidance and implementation, 2) inadequate resources (funding, infrastructure, and staffing), 3) ineffective mechanisms for data accessibility, and 4) inadequate access for diverse stakeholders. The USG should strengthen the accessibility of its data by both USG and non-USG entities.[92] In many cases, the steps the USG should take to make its data accessible for

Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance, U.S. Office of Management and Budget at 4 (2019). As of the publication of this paper, OMB has not yet issued official Phase II guidance.

[89] M-19-18, Federal Data Strategy - A Framework for Consistency, U.S. Office of Management and Budget (2019). OMB's issuance of the FDS was in response to the 2018 President's Management Agenda and laid out a new Cross-Agency Priority (CAP) Goal: Leveraging Data as a Strategic Asset to develop and implement a comprehensive Federal Data Strategy. Leveraging Data as a Strategic Asset, Performance.gov (2021).

[90] Blueprint for an AI Bill of Rights, The White House (2022). The Blueprint makes data collection and use a top focus throughout, recognizing the relationships between AI and data. The Blueprint contains considerable discussion of data privacy issues including concerns about surveillance as well as rights to controlling the collection, use, and reuse of personal data. The Blueprint importantly identifies a type of data that presents specific challenges: "Data that is derived from other data through the use of algorithms, such as data derived or inferred from prior model outputs." Id. at 20. It notes that derived data should be viewed as potentially high-risk inputs that may lead to feedback loops, compounded harm, or inaccurate results. The Blueprint addresses the critical need for a comprehensive statutory or regulatory framework governing the rights of the public when it comes to personal data. Although the Technical Companion to the Blueprint includes a great deal of detail on recommendations for collecting, using, and sharing data, none of the recommendations are mandated.

[91] Alondra Nelson, Ensuring Free, Immediate, and Equitable Access to Federally Funded Research, The White House (2022) (amending the 2013 guidance to refine requirements for making Scientific data resulting from federally funded research widely available by default). The 2022 guidance also requires that by default all peer-reviewed scholarly publications authored or coauthored by individuals or institutions resulting from federally funded research are made freely available and publicly accessible by default in agency-designated repositories without any embargo or delay after publication. See also John P. Holdren, Increasing Access to the Results of Federally Funded Scientific Research, The White House (2013).

[92] There are some public efforts to facilitate USG data sharing both within USG and external to USG. The National Oceanic and Atmospheric Administration (NOAA) Open Data Dissemination Program provides public access to NOAA's open data on commercial cloud platforms through public-private partnerships. About the NOAA Open Data Dissemination Program, U.S. National Oceanic and Atmospheric Administration (2022). USAFacts, a not-for-profit, supports the collection, analysis, and dissemination of USG data. About USAFacts, USAFacts (last accessed 2022) ("USAFacts is a not-for-profit, nonpartisan civic initiative making government data easy for all Americans to access and understand. We provide accessible analysis on US spending and outcomes in order to ground public debates in facts."). Several USG data sets are made available on Amazon's AWS Data Exchange, including the American Community Survey, an ongoing survey that provides information about jobs and occupations, educational attainment, veterans, whether people own or rent their homes, and other topics. The data is aggregated at the census block group level. See AWS Data Exchange, AWS (last accessed 2022); ACS - Sociodemographics (USA, Census Block Groups, 2019), AWS Marketplace (last accessed 2022). The FDA provides the OpenFDA Application Programming Interface that serves public FDA data such as recall enforcement reports and adverse events about drugs, devices, and foods. About the openFDA API, U.S. Food and Drug Administration (last accessed 2022).

use across agencies and for non-USG use are the same. The USG has massive amounts of data that are useful for itself, as well as to organizations outside the federal government.[93]

Inadequate Guidance and Implementation: First, OMB has yet to issue statutorily-required implementation guidance to agencies on making data open by default and comprehensive data inventories in support of the OPEN Data Act objectives.[94] OMB also has not issued an FDS Action Plan for 2022. The 2021 Action Plan, issued two months before the end of 2021, states that delineated milestones were only "aspirational" and further acknowledges that some milestone dates in the 2020 action plan "were unachievable because of a lack of published guidance."[95]

Inadequate Resources: Second, a 2022 survey of USG Chief Data Officers indicates that key challenges to using data towards agency missions included lack of direct funding, a limited workforce, and data governance challenges.[96] If agencies are not funded or staffed to prepare their data for sharing, then they are not properly resourced nor incentivized beyond meeting their own mission.

Ineffective Mechanisms for Data Accessibility: Third, stakeholders have indicated the USG portal through which agencies are required to make their data accessible, data.gov, is ineffective. A December 2021 Government Administration Office survey of current users of data.gov (from the private sector, state and local governments, and nonprofits) indicated that the portal had limited usefulness. In fact, stakeholders often went directly to various other data sources because "[data.gov] data sets were difficult to discover or not organized in a useful way."[97]

Inadequate Diversity in Research Organizations: Fourth, an important aspect of increasing the accessibility of USG data is making sure more people can take advantage of available data. For the United States to maintain leadership in key technology areas and compete globally, there must be whole-of-nation participation where research organizations and businesses of all sizes have access to the data and compute power to create the applications of the future. However, researchers and developers sitting beyond the walls of well-resourced universities, large companies, and national laboratories often lack access to large-scale datasets and advanced

[93] Ellen Hughes-Cromwick & Julia Coronado, The Value of US Government Data to US Business Decisions, Journal of Economic Perspectives at 145 (2019) ("The value of government data is difficult to measure, but it is clearly a substantial strategic asset for the US business sector. Such data are used by a wide range of companies from auto producers to digital platform companies, and for purposes that include production and investment decisions, marketing and inventory management, and long-range strategic planning…While companies are generating ever-increasing amounts of big data from their own operations, it is often the combination of proprietary data with comprehensive government data that provide critical context and allow for maximum strategic benefit (a public good externality).").

[94] GAO-22-104574, Open Data: Additional Action Required for Full Public Access, U.S. Government Accountability Office (2021).

[95] Federal Data Strategy: 2021 Action Plan, U.S. Office of Management and Budget at ii, 5 (2021).

[96] CDO Survey Analysis, Federal CDO Council at 63 (2022).

[97] GAO-22-104574, Open Data: Additional Action Required for Full Public Access, U.S. Government Accountability Office at 29 (2021).

compute power.[98] This leaves small U.S. businesses and, even more so, small academic institutions shut out of the U.S. innovation ecosystem. Leaving out these voices decreases the diversity of researchers, especially those from underserved and underrepresented communities,[99] and narrows the scope of research topics to those largely focused "on private profit, rather than public benefit."[100]

Executive Implementation Actions

The White House should signal that improving USG data accessibility is a priority by issuing an Executive Order (E.O.). An E.O. should:

1. Establish a national-level Chief Data Officer in the Office of Management and Budget that drives implementation of a government-centric data strategy across departments and agencies. Roles and responsibilities would include execution of the OPEN Government Data Act. The national-level Chief Data Officer should also be given the authority and resources necessary to verify the status of efforts to implement the FDS Action Plan, and other data-related Federal policies and regulations;
2. Require OMB and the CDO Council working with agencies to issue a revised FDS every three years to reflect the ever-changing data environment;
3. Mandate that OMB complete the Evidence Act compliance guidance, issue a 2023 FDS Action Plan early in 2023 to reflect all that has passed since the last Action Plan issued in 2021. Further require that agency milestones in all issued action plans be met, and highlight priority near-term areas in FDS action plans;
4. Mandate the development and deployment of a data.gov Version 2.0 based on stakeholder feedback. At a minimum, data.gov Version 2.0 should prioritize making datasets available, and ensuring processes to keep them complete and up to date, in formats that aid uptake and create opportunities for monetizing USG data. Specifically, data.gov Version 2.0 should:
    a. Create an environment for Public-Private-Partnerships (PPPs) to leverage industry innovation and the potential to commercialize USG datasets similar to existing successful models;[101]
    b. Prioritize datasets for critical application domains, functions, and requirements in order to create a lower cost, higher benefit model, which can be built upon and broadened over time;

[98] Envisioning a National Artificial Intelligence Research Resource: Preliminary Findings and Recommendations, National AI Research Resource Task Force (2022); Daniel E. Ho, et al. Building a National AI Research Resource: A Blueprint for a National Research Cloud, Stanford Institute Human-Centered Artificial Intelligence (2021).
[99] Envisioning a National Artificial Intelligence Research Resource: Preliminary Findings and Recommendations, National AI Research Resource Task Force (2022).
[100] Daniel E. Ho, et al. Building a National AI Research Resource: Blueprint for a National Research Cloud, Stanford Institute Human-Centered Artificial Intelligence at 20 (2021).
[101] An example of a successful PPP that leverages USG data through commercialization is the National Oceanic and Atmospheric Administration (NOAA) Open Data Dissemination Project, which "provides public access to NOAA's open data on commercial cloud platforms through public-private partnerships." NOAA Open Data Dissemination Project, U.S. National Oceanic and Atmospheric Administration (2022).

c. Require all data registered on data.gov Version 2.0 to adhere to standards that ensure quality and interoperability with emphasis on relevant industry and academic data standards; and

d. Require Application Processing Interfaces (APIs) that facilitate SMEs and academic institutions in obtaining and utilizing quality datasets;

5. Mandate that agency senior leadership prioritize data accessibility through agency participation in data.gov Version 2.0; and
6. Mandate that the CDO Council assess agency funding needs and, if required, agencies prioritize this need in their annual budget requests.[102]

Legislative Implementation Actions

Congress should pass legislation to:

1. Extend the lifetime of the CDO Council, clarify its roles and responsibilities, and establish the national CDO as the Chair of the CDO Council;[103]
2. Congress should follow the recommendations of the National AI Research Resource (NAIRR) Task Force, the NSCAI, and the university and research community and fully authorize and fund the development and implementation of the NAIRR, as recommended by NSCAI;[104] and
3. Prioritize funding for agencies to address their data needs.

[102] The E.O. should further encourage individual departments/agencies to shift around their own resources to advance these priorities. Alternative funding mechanisms include expanding the mandate of the federal Technology Modernization Fund to include data projects with funding; and expanding the Presidential Innovation Fellows (PIFs) program to make more fellows available to implement FDS.

[103] The CDO Council is a valuable coordinating body and critical to implementing agency data implementation plans, like the FDS. However, the CDO Council is set to sunset in 2025. Phil Goldstein, Federal CDOs Seek More Guidance on Government's Data Strategy, FedTech (2021). The clarified roles and responsibilities of the CDO Council should include, at a minimum, 1) assessing whether FDS implementation gaps are related to FDS prioritization or implementation, and 2) authorities to verify the implementation status of other USG policies that impact data acquisition, sharing, and use practices.

[104] To democratize access to data and compute power and bring greater diversity in research participation, the National Security Commission on Artificial Intelligence recommended establishing a National AI Research Resource coordinated by the USG as a public-private partnership to provide verified participants with access to compute resources and AI-ready USG and non-USG datasets. Final Report, National Security Commission on Artificial Intelligence (2021). In response, Congress directed the National Science Foundation (NSF) and the White House Office of Science and Technology Policy to create a Task Force to explore the feasibility of, and develop a roadmap for, a NAIRR. Pub. L. 116-283, William (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021, § 5106 (2020). In its preliminary report, the NAIRR Task Force emphasizes the critical need to provide a national cloud resource and curated datasets to a diverse set of stakeholders to fuel AI research and development. Envisioning a National Artificial Intelligence Research Resource (NAIRR): Preliminary Findings and Recommendations, NAIRR Task Force (2022). The NAIRR Task Force further provided its vision of providing such support to tens of thousands of users and supplied an interim roadmap for the development and sustainment of the NAIRR. The Task Force's final report is expected in the coming months.

Action Area 3: Facilitating flows of non-USG data to USG and non-USG entities

The USG has the opportunity to accelerate a whole-of-nation data ecosystem by ensuring appropriate governance and facilitation of data access and use of non-USG data (private sector, academia, and civil society) by USG and non-USG entities.

The private sector has begun to recognize the importance of democratizing access to data across organizations, including between the public and private sectors.[105] However, a variety of studies have identified at least four barriers that inhibit the private sector from sharing data among other private sector organizations and with the government when the business model is not commoditizing data as a service or product: 1) privacy concerns, 2) inadequate security provisions, 3) inadequate incentives, and 4) inadequate IP protections.

Privacy concerns: Organizations outside of the USG are often reluctant to share their data because of concerns about uncertain legal and regulatory constraints on data privacy (e.g., the exchange of personal health information between covered and noncovered entities)[106] and because of concerns over hindered ability to ensure privacy of their data, which requires stakeholders to trust another organization's enforcement of privacy controls.[107]

Inadequate security provisions: Organizations express concerns over losing control of their data and having to trust another organization's enforcement of cybersecurity controls.[108]

Inadequate incentives: In many instances, there is no business case for investing in sharing data when there is not an obvious return on investment to justify the costs and risks.[109] This is especially

[105] Open Data Campaign, Microsoft (last accessed 2022) ("We believe everyone can benefit from opening, sharing and collaborating around data to make better decisions, improve efficiency and even help tackle some of the world's most pressing societal challenges."); Open Data Policy Lab (last accessed 2022) ("Governments play a critical role in providing data that is vital to addressing today's most pressing problems and improving people's lives'... governments at all levels can benefit from functional access to private-sector data to address important public policy challenges,").

[106] Daniel M. Walker, et al., Approaches for Overcoming Barriers to Cross-Sector Data Sharing, The American Journal of Managed Care (2022) ("Interviewees expressed concerns regarding what specifically is allowable under HIPAA and whether [specific potential sharing partners] could be HIPAA compliant, ultimately resulting in their reluctance to enter into a [Business Associate Agreement.]").

[107] Robert Groves & Adam Neufeld, Accelerating the Sharing of Data Across Sectors to Advance the Common Good, Georgetown University at 12 (2017) ("Privacy risks must be weighed when sharing or linking data, and will sometimes be a significant obstacle to sharing.").

[108] Enhancing Access to and Sharing of Data, Organisation for Economic Cooperation and Development (2019) ("Evidence confirms that risks of confidentiality breach, for instance, have led users to be more reluctant to share their data, including providing personal data, and in some cases to use digital services at all."); Robert Groves & Adam Neufeld, Accelerating the Sharing of Data Across Sectors to Advance the Common Good, Georgetown University at 13 (2017) ("Despite trust in accepting the government's intentions, they may fear that it does not have the capability to fulfill its commitments to data security."); Sam Tawfik, The Top 5 Barriers to Data Sharing and How to Overcome Them, Immuta (2021) ("CDOs deemed data security risk assessments a top barrier for sharing data, particularly externally.").

[109] Robert Groves & Adam Neufeld, Accelerating the Sharing of Data Across Sectors to Advance the Common Good, Georgetown University at 11 (2017) ("Neither governments nor companies generally have an incentive to prioritize the hard work of sharing granular data with other sectors. The benefits of providing data to other sectors typically do not fall to the data owner, at least in the short term, making it hard to justify taking energy and resources away from achieving the core societal mission of the government agency and the profit-making mission of companies.");

true when externally sourced data, that may be accessible through collective sharing agreements, is difficult to integrate before it can be used.

Inadequate IP or IP-type protections: Private sector organizations are concerned about data sharing that results in providing IP or proprietary advantage to competitors.[110]

Additional barriers specific to USG uptake of non-USG data: In a survey of USG CDOs, roughly 80 percent of large agency USG CDOs indicated "additional directed funding" is required for their agency's CDO role to be successful.[111] USG CDOs also often lack adequate guidance on priorities and milestones, hindering progress on USG data accessibility.[112] There is often inadequate USG infrastructure (e.g., capacity, data management tools, standardization) and capabilities (e.g., required skill sets in staff) to uptake, integrate, and analyze data.[113] With respect to the uptake of for-sale data as a service or product,[114] barriers include challenges with USG acquiring and sharing commercial data across agencies and resources to acquire the data.[115]

Society will be better served with improved data sharing across and within sectors and domains—between entities that might have competing interests or face barriers to sharing. The USG can play a critical role in facilitating data access and sharing across the private sector, academia, civil society, and USG agencies either as a trusted broker or by enabling an independent third-party.[116] There are some bright spots demonstrating the potential. For example:

Enhancing Access to and Sharing of Data, Organisation for Economic Cooperation and Development (2019) ("But data quality may not only affect the ability and the cost to re-use data. It can also prevent stakeholders from participating in data-sharing arrangements. According to some studies, uncertainties about data quality may explain, for instance, why open data repositories are used at far lower rates than most scholars and practicing data curators would expect.").

[110] Enhancing Access to and Sharing of Data, Organisation for Economic Cooperation and Development (2019) ("[Intellectual Property Rights] and other legitimate commercial and non-commercial interests need to be protected, otherwise incentives to contribute data and to invest in data-driven innovation may be undermined, in addition to the risks of direct and indirect harm to right holders, including data subjects. Evidence confirms that risks of confidentiality breach, for instance, have led users to be more reluctant to share their data."). Indeed, China understands the importance of companies being able to protect their data IP. China's National IP Administration is planning pilot projects for the protection of data IP in its high-tech areas at the center of its emerging data economy. Trivium Tech Daily, Trivium China (Dec. 6, 2022).

[111] CDO Survey Analysis, Federal CDO Council at 49 (2022).

[112] As noted earlier, OMB has not issued an FDS Action Plan for 2022. The 2021 Action Plan, issued two months before the end of 2021, states that delineated milestones were only "aspirational," and further acknowledges that some milestone dates in the 2020 action plan "were unachievable because of a lack of published guidance." Federal Data Strategy 2021 Action Plan, U.S. Office of Management and Budget at ii, 5 (2021).

[113] SCSP discussion with a Senior Government official (November 2022).

[114] Some organizations sell data as a commodity in the form of a product or service (e.g., companies who sell data they have collected or third-party data brokers). For example, Definitive Healthcare sells a wide range of aggregated data. "From financial and quality metrics to affiliation and technology data, HospitalView gives you an unparalleled level of intelligence on every hospital and IDN in the United States. With more than 9,300 distinct facility profiles packed with contextualized data curated from nearly 40 different public, private and proprietary sources, you'll expand the breadth and depth of your knowledge." HospitalView, Definitive Healthcare (last accessed 2022). As another example, Bloomberg Data Products provides a range of financial data. "Access reference, pricing, regulatory and alternative data with extensive history." Bloomberg Data, Bloomberg (last accessed 2022). These entities are already incentivized to monetize value from data.

[115] SCSP discussion with a Senior Government official (November 2022).

[116] Robert Groves & Adam Neufeld, Accelerating the Sharing of Data Across Sectors to Advance the Common Good, Georgetown University at 23 (2017).

- Aviation Safety Information Analysis & Sharing (ASIAS) is a collaboration between the Federal Aviation Administration (FAA) and the aviation community. ASIAS data resources include internal FAA datasets, airline proprietary safety data, publicly available information, manufacturers' data, and weather data.[117] The structure of data sharing agreements and the operational governance of ASIAS have overcome the reluctance of private sector organizations to share proprietary data with direct competitors in order to realize the collective benefit from unlocking aggregate analyses on safety benchmarking and improvements for aviation without compromising the security of their confidential business models.[118]

- The National Cancer Institute's Cancer Research Data Commons (CRDC) integrates and makes available a wide range of datasets from various NIH funded academic and private sector research projects, provides a mechanism for other organizations to contribute data and to search the collective data, and provides compute and analytic resources for the research community.[119] The CRDC has been cited as a key resource in many refereed research publications,[120] demonstrating the value in unlocking shared access to data from USG, the research community, and the private sector.

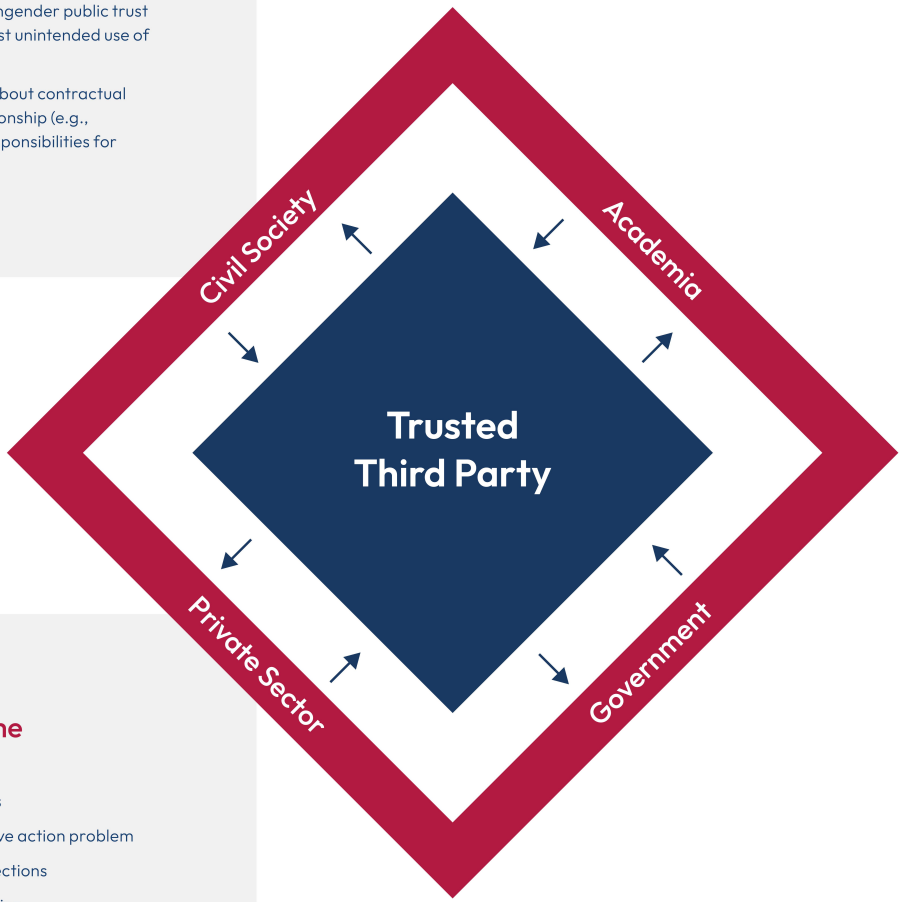
- The Identity Theft Tax Refund Fraud Information Sharing and Analysis Center (the IDTTRF-ISAC) was established to "provide a secure platform via a sustainable public/private partnership, to facilitate information sharing consistent with applicable law, and analytics necessary to detect, prevent, and deter activities related to stolen identity refund fraud."[121] The IDTTRF-ISAC is an independent partnership among the Internal Revenue Service (IRS), industry, and states. The technology and analytical services for the IDTTRF-ISAC are provided by a Trusted Third Party (TTP) and sponsored by the IRS.[122]

[117] Overview, Aviation Safety Information Analysis & Sharing (last accessed 2022) ("ASIAS's resources include both public and non-public aviation data. Public [USG] data sources include, but are not limited to, air traffic management data related to traffic, weather, and procedures. Non-public sources include de-identified data from air traffic controllers and aircraft operators, including digital flight data and safety reports submitted by flight crews and maintenance personnel.").
[118] Federal Aviation Administration Report to Congress: Report on the Status of Aviation Safety Information Analysis and Sharing (ASIAS) Capability Acceleration, U.S. Federal Aviation Administration (2020) ("MITRE/CAASD fulfilled the role of the [ASIAS] trusted third party to help facilitate the sharing of proprietary data and to ensure associated protections for the data. Data protection and de-identification protocols had to be established before participants were willing to share sensitive data.").
[119] Cancer Research Data Commons, U.S. National Institutes of Health (last accessed 2022).
[120] Selected Publications, U.S. National Institutes of Health (last accessed 2022).
[121] 2021 Annual Report, Identity Theft Tax Refund Fraud Information Sharing and Analysis Center at 2 (2021).
[122] 2021 Annual Report, Identity Theft Tax Refund Fraud Information Sharing and Analysis Center at 5 (2021) ("TTPs are critical in ISACs, in that they facilitate information sharing among entities that wouldn't otherwise do so. The platform serves as the centralized information-sharing vehicle for the ISAC and includes controls to help ensure that sharing occurs in a manner that is consistent with applicable laws.").

Data-Sharing Public-Private Partnerships

Common Elements

- Address a discrete critical problem or opportunity and a justification for each participant to engage.
- PPP is independent and trusted by all participants, with clear guardrails on data accessibility, sharing, and use.
- Strong focus on ensuring privacy and security in the operations of the PPP to engender public trust in adequate protections against unintended use of the shared data.
- Participants establish clarity about contractual and legal issues for their relationship (e.g., treatment of IP protection, responsibilities for avoiding conflicts of interest).



Barriers to Overcome

- Privacy concerns
- Inadequate security provisions
- Inadequate incentives/collective action problem
- Inadequate IP or IP-type protections
- Technical and human capabilities

There are many different kinds of PPPs, but the model here focuses on data sharing.[123] These PPPs often use trusted third parties for data access and sharing to unlock the opportunities of aggregating private and public sectors data in a controlled and trusted manner for the collective good of all participants. Each PPP is shaped by the mission and participants, but there are common elements in successful data sharing PPPs:

- They are formed to address a discrete critical problem or opportunity, which includes a sense of urgency and a justification for each participant to engage.[124]
- The PPP is viewed as independent and trusted by all participants, with clear guardrails on data accessibility, sharing, and use.[125]
- There is a strong focus on ensuring privacy and security in the operations of the PPP, to engender public trust that there are adequate protections against unintended use of the shared data.[126]
- The partners establish clarity about contractual and legal issues for their relationship (e.g., the treatment of IP or IP-type protection and data rights, policies for avoiding conflicts of interest, the role of an independent trusted third party if one is established, avoiding antitrust/anti-competition issues, funding and authorization of the PPP).[127]

A recent example of the USG successfully catalyzing outcome-oriented stakeholder participation with a sense of urgency occurred in response to the COVID-19 pandemic. The White House and individual agencies convened a variety of stakeholders and rapidly met specific outcomes and agreements.[128]

---

[123] Ted Senkrecht, Tales & Tips from the Trenches: Extend the Impact of Enterprise Data through Partnerships, MITRE (2021) ("Data-sharing PPPs involve multi-party collaboration around information sharing and analysis to take action on complex problems without boundaries. These PPPs are predicated on shared decision-making, shared resourcing, and shared benefit to the partners and the public.").

[124] Ted Senkrecht, Tales & Tips from the Trenches: Extend the Impact of Enterprise Data through Partnerships, MITRE (2021) ([Data sharing PPPs have a] "common mission - Partners are driven by a sense of urgency and the realization that their own interests are served by working together on a shared goal.").

[125] Robert Groves & Adam Neufeld, Accelerating the Sharing of Data Across Sectors to Advance the Common Good, Georgetown University at 19 (2017) ("[A] key barrier to sharing data across sectors is trust. Companies want to make sure the government does not use the data in ways that hurt their business interests, and the government taking control of private sector data imposes security and legal risks. The public's privacy concerns also seem to diminish when the government is not the one combining data.").

[126] Robert Groves & Adam Neufeld, Accelerating the Sharing of Data Across Sectors to Advance the Common Good, Georgetown University at 20-21 (2017) ("The technology for protecting privacy has evolved substantially since organizations simply deleted names, addresses, and Social Security numbers from spreadsheets. Technical approaches (such as query tools, synthetic data, and multiparty shared computing) and mathematical methods (such as differential privacy) now allow for far more sophisticated ways to reduce the risk of re-identifying people. … Decisions on technical infrastructure are intertwined with privacy, cybersecurity, and confidentiality concerns.").

[127] Ted Senkrecht, Tales & Tips from the Trenches: Extend the Impact of Enterprise Data through Partnerships, MITRE (2021).

[128] See e.g., Executive Order on Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats, The White House (2021) (explicitly assigning roles and responsibilities across the Executive Branch to achieve data sharing and analysis outcomes, including how to de-identify COVID-19-related data). The National Center for Advancing Translational Sciences' (NCATS) initiative, National COVID Cohort Collaborative (N3C), is a centralized data analytics platform which, in a matter of months, brought together clinical, laboratory, and diagnostic data from medical research sites across the country. A key enabler of NC3 was the rapid agreement on a common data format that enabled the different ways contributing hospitals store patient data to be converted into

Executive Implementation Actions

The USG should look for the most promising opportunities to apply the outcome-directed stakeholder participation approach that was successfully leveraged to catalyze the sharing of data in response to the COVID-19 pandemic. Such an approach can be used to maximize the value of national data assets in areas of critical importance to U.S. national competitiveness. Trusted intermediaries can responsibly bring together data from a range of private and public entities, link and analyze the data into actionable information, and share both the insights as well as the underlying data (as appropriate) with all parties.

Aligned with the conclusions in SCSP's report on Mid-Decade Challenges to National Competitiveness, the USG should prioritize data-sharing PPPs to advance the six priority technology areas identified by SCSP: AI, novel computing paradigms, next generation communications networks, biotechnology, energy generation and storage, and areas of technological convergence like advanced manufacturing.[129] Within each of these areas, cross-domain and application data sharing would, for example, help reduce redundancy in data collection and analytics, guide the direction of efforts across the full life cycle of the technology (i.e., from identifying research and development (R&D) gaps to scaling production to fielding operational use-cases), generate better industry analytics to inform business decision and government policy making, and assist in collective tracking and mitigation of geopolitical adversarial actions.

The White House should signal that improving whole-of-nation data accessibility through the establishment of PPPs[130] is a policy priority for the Administration by issuing an Executive Order (E.O.). An E.O. should:

1. Charge appropriate senior leaders to oversee stakeholder engagements[131] in each of the six priority technology areas highlighted by the Mid-Decade Challenges to National Competitiveness report: AI, novel computing paradigms, next generation communications networks, biotechnology, energy generation and storage, and areas of technological convergence like advanced manufacturing.[132] The diverse stakeholders will assess high potential areas for increasing appropriate data accessibility in their respective priority technology areas;

---

the common format to enable combined "apples-to-apples" analyses. See Announcement: Access to the COVID-19 Data Analytics Platform Is Open, U.S. National Institutes of Health (2022),

[129] Mid-Decade Challenges to National Competitiveness, Special Competitive Studies Project at 32 (2022).

[130] PPPs focused on data sharing to address discrete challenges set the conditions for improved sharing of both USG and non-USG data, while also creating new value for USG and non-USG participants.

[131] Diverse participants from the USG, private sector, civil society, and academia will be needed to identify the most promising opportunities and partners, propose initial steps and requirements, and generate buy-in among key stakeholders.

[132] This implementation action is aligned to the Federal Data Strategy: M-19-18, Federal Data Strategy - A Framework for Consistency, U.S. Office of Management and Budget at 7 (2019) ("Leverage Partnerships: Create and sustain partnerships that facilitate innovation with commercial, academic, and other partners to advance agency mission and maximize economic opportunities, intellectual value, and the public good.").

2. Require appropriate CDOs, once final ideas are selected, to stand up narrowly-scoped PPPs to facilitate data accessibility to solve specific challenges. The PPPs should be equipped with agreements and controls surrounding privacy, data security, democratizing access for SMEs and researchers, ensuring incentives, addressing anti-competitive concerns, and proprietary IP or IP-type protections tailored to the relevant PPP; and
3. Require CDO Council, or national-level CDO if stood up, to analyze infrastructure, capabilities, policies,[133] and funding needed to ensure agencies and their respective CDOs are sophisticated consumers of non-USG data.

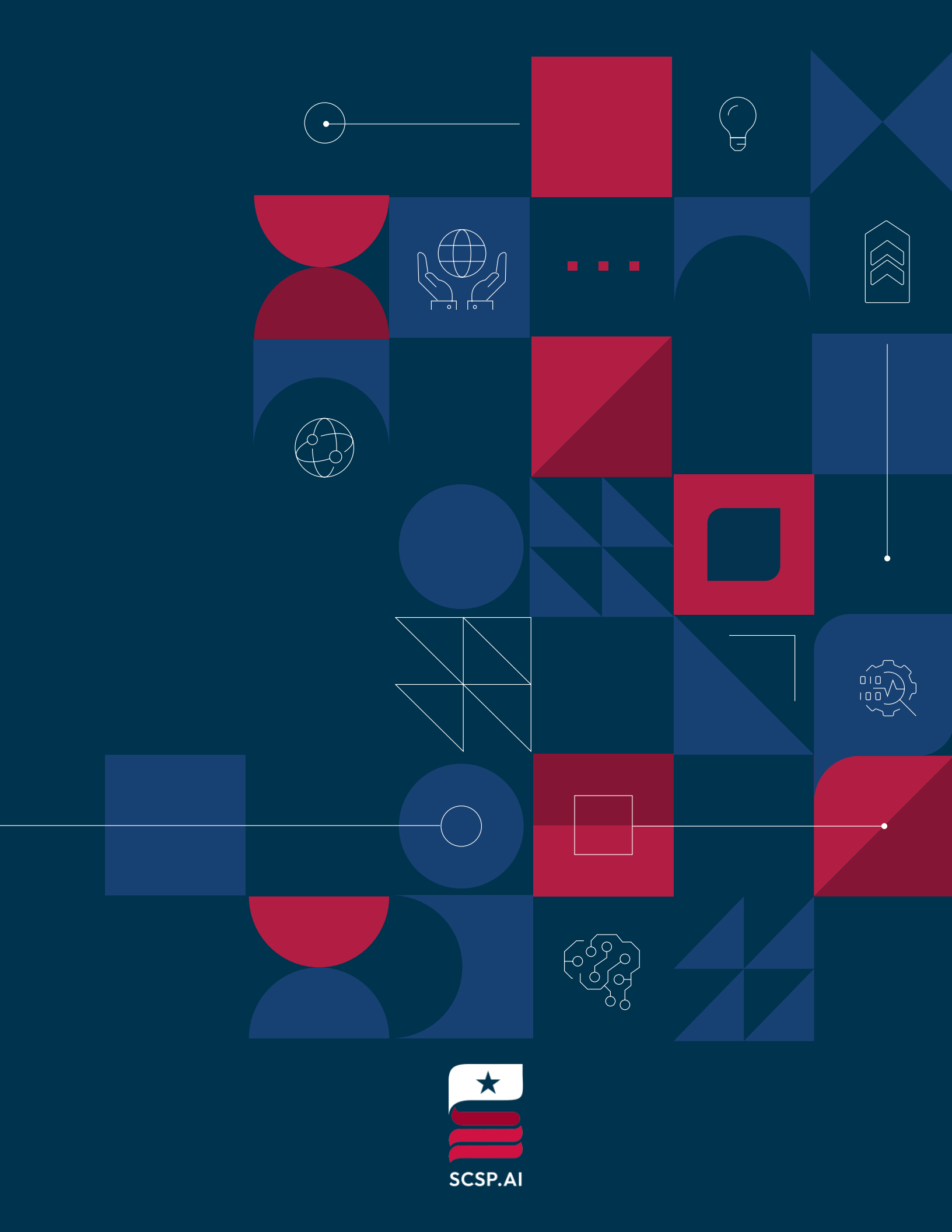
Conclusion

Data-driven technologies will continue to increasingly shape every aspect of our economy, security, and personal lives. Data supports improved technology-driven decision making, research advances, and innovation implementation across emerging technologies. The USG has a critical role to play in establishing a democratic framework for data governance, promoting accessibility and sharing of rich data sets held by the USG and non-USG entities, and demonstrating a democratic vision for data governance globally. These actions are essential to leveraging data for national competitiveness by improving government decision making, cultivating data-driven innovation in AI and other emerging technologies, and addressing societal challenges while protecting individual rights in accordance with laws and values.

To fully maximize U.S. data assets, the USG must implement specific yet far-reaching strategies and policies such as those presented in this report. Positioning itself to have a robust and resilient data ecosystem that protects democratic values like data privacy will enable the United States to align with its partners and allies on data efforts and help put the United States in a strategically advantageous position with respect to China.

The United States needs to start by ensuring that its data assets are appropriately accessible while also ensuring sufficient data privacy protections. For USG data, the path is known. Laws and policy directives already require greater data accessibility. High-level leadership prioritization is needed to overcome bureaucratic friction and accelerate policy implementation. For non-USG data, the USG can convene and promote the role of trusted brokers to set the conditions for greater data access and use through public-private partnerships. Priority efforts should go to emerging technology areas critical for national competitiveness including AI, novel computing paradigms, next generation communications networks, biotechnology, energy generation and storage, and areas of technological convergence like advanced manufacturing.

[133] For example, developing strategies and policies with respect to acquiring and sharing non-USG across the USG.



SCSP.AI