National Secure Data Service Demonstration Project: Two-Year Report

I. Background

On August 9, 2022, President Biden signed Public Law 117-67, also known as the CHIPS and Science Act, authorizing numerous programs and activities for federal agencies to support the continued advancement of science, engineering, and technology in the U.S. Section 10375 authorized the establishment of one such program, the National Secure Data Service (NSDS) Demonstration project. The NSDS Demonstration project was authorized to '*develop, refine, and test models to inform the full implementation of the Commission on Evidence-Based Policymaking recommendation for a governmentwide data linkage and access infrastructure for statistical activities conducted for statistical purposes as defined in chapter 35 of title 44, United States Code*[1].' The U.S. National Science Foundation (NSF) was designated as the agency responsible for establishment of the Demonstration project in accordance with the requirements in the legislation, operated directly or via a contract that is managed by the National Center for Science and Engineering Statistics (NCSES), the statistical agency within the NSF. In addition, Section 10375(a) requires consultation with the Office of Management and Budget and the National Artificial Intelligence Research Resource (NAIRR) pilot.

This report is submitted in accordance with Section 10375 (h) and highlights work completed during the first two years of the Demonstration project. The report addresses the following requirements: (1) the current privacy policy landscape; (2) data linkage and NSDS Demonstration activities to date; (3) risk mitigation and removing barriers to an NSDS implementation; and (4) a plan for scaling up the Demonstration project to support implementation of a full National Secure Data Service (NSDS).

II. Policies for Protecting Data

The Federal Statistical System (FSS) is the premier provider of statistical information about the nation, its people, and the economy. Ensuring respondent confidentiality is a central tenet of the FSS. The FSS relies on the public trust in obtaining responses to surveys, promoting data sharing arrangements, and maintaining its reputation. Declining survey response, a proliferation of non-federal data sources of mixed quality, and increased concerns over privacy and algorithmic bias highlight the importance of maintaining that public trust to safeguard the continued collection of critical data within the FSS. The FSS data collected, however, are highly valued due to their expansive coverage of the U.S. population. Sharing these data, therefore, is

[1] PL-117-67, Section 10375(a).

also critical to ensure that evidence-building activities are supported with data that are fit for purpose to fuel data-driven decision-making.[2]

Currently, confidential data within the FSS are protected through a collection of statutes, regulations, and policies designed to ensure that respondent confidentiality and public trust are maintained. These include, but are not limited to, the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), the Health Information Protection and Portability Act (HIPPA), the NSF Act of 1950, Title 13, and Title 26. CIPSEA, in particular, seeks to balance confidentiality with data sharing. The NSDS Demonstration project, thus far, has supported data sharing and/or linkage with data protected by each of these legal authorities. A significant challenge to data sharing is a statutory prohibition of access, even to RSAUs, to some of the government's most useful data. As a result, the effectiveness of an NSDS may be limited for the purposes of data linkage between many agencies' data assets. While regulations and guidance under CIPSEA may provide some additional clarity and support for data sharing, statutory changes may be needed to ensure that data sharing and linkage is possible to support evidence-building activities for certain agency data.

The current NSDS Demonstration project is testing several methods for protecting data that are consistent with federal law and agency policy. Lessons[3] from one of the first NSDS Demonstration projects, an environmental scan of privacy preserving technologies, reinforced the need for legal consultation and clear legal advice when developing data sharing agreements. Other Demonstration projects are relying on privacy preserving record linkage (PPRL), an application of secure multiparty computation, that enables the sharing of data without sharing direct unencrypted personally identifiable information. To do this, data sharing agreements have been established between two federal statistical agencies and between a federal statistical agency and its parent agency. In establishing these agreements, the statutes and mandates that govern each organization's data were acknowledged and addressed by working closely with each agency's Office of General Counsel. As a result, for the cross-statistical agency project, relevant NCSES staff have become designated agents of the National Center for Health Statistics. CIPSEA defines an "agent" as a person "...who is working under the authority of a government entity with which a contract or other agreement is executed by an executive agency to perform exclusively statistical activities under the control of an officer or employee of that agency." 44 U.S.C. 3561(2). Similarly, the agreement between a federal statistical agency and its parent agency involved consulting the System of Records Notice (SORN) to ensure that the proposed data collection activities fell within its scope. Two other Demonstration projects are exploring

[2] 2018 Foundations for Evidence-Based Policymaking Act and *Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good,* National Academy of Science, Engineering, and Medicine 2023

[3] Project description and final report can be found online, Award: Privacy Preserving Technologies Phase 1: Environmental Scan (PPT) – ADC | America's Datahub Consortium (americasdatahub.org).

tiered access[4] with the creation of synthetic data[5]. The synthetic data are being created to promote and apply tiered access models to enhance the utility of data while protecting confidentiality. Synthetic data can provide additional publicly available data for evidence-building, hence increasing overall access to data. For these projects, policies for protecting privacy have also been incorporated into the project framework.

The data utilized in these PPRL and synthetic data projects, with their sponsoring agencies, include:

- Annual Business Survey, U.S. Census Bureau, NCSES
- National Health Interview Survey, National Center for Health Statistics, Centers for Disease Control and Prevention
- Principal Investigator Award Database, U.S. National Science Foundation
- Survey of Earned Doctorates, NCSES
- Survey of Doctoral Recipients, NCSES.

Data security policies and procedures from the above agencies guide the protection of data utilized in the NSDS Demonstration project, ensuring secure use and the protection of personally identifiable information (PII) and business identifiable information (BII), including at rest and in transit. As additional agency data are shared, linked, and analyzed during the Demonstration project, further data protection policies and practices will be implemented, utilizing the initial linkage projects as guides. The projects through the NSDS Demonstration have informed and will serve as a framework for future data sharing. Examples below include key considerations that support this framework.

- *CIPSEA Agent Designation:* CIPSEA requires that all users of confidential data protected under CIPSEA obtain CIPSEA Agent status. CIPSEA defines "agent" as "an individual…who is working under the authority of a government entity with which a contract or other agreement is executed by an executive agency to perform exclusively statistical activities under the control of an officer or employee of that agency."[6] All users must sign a Data Use Agreement and take annual data security training. All project performers that handle CIPSEA-protected data under the NSDS Demonstration project are required to become CIPSEA agents.
- *Data Security Agreements:* The NSDS Demonstration project has developed a data security agreement template that can be applied when working with agency data that are covered under CIPSEA. For example, for all protected data on NCSES-sponsored systems used within the Demonstration project, the data user's organization must

[4] Tiered access is an application of the privacy principle of data minimization, which means giving access to the least amount of data needed to complete an approved project.

[5] A synthetic dataset does not contain the exact records of the original dataset, but instead retains the statistical properties of the original dataset, preserving information useful to researchers and their queries.

[6] 44 U.S.C. 3561(2) (defining "agent").

complete and sign an NCSES Confidentiality Plan and Data Security Procedures (CPDSP) form, based on a template, that explains how the organization will meet all data security requirements. The plan is reviewed internally within NCSES, and the NCSES Confidentiality Officer issues a final determination memo regarding access.

- *Data Sharing Agreements:* Sharing confidential data requires a data sharing agreement that outlines the terms of use and how the data will be secured. The agreement must be signed and fully executed by all data sharing parties. Any changes to data security and/or data sharing procedures require an agreement modification. The NSDS Demonstration project has developed two templates, one for sharing data when data are covered under differing statutory mandates and another for similar statutory mandates.
- *Secure Data Access:* Confidential information may be used only on secure servers by approved data users. For the Demonstration project, all data are used within a secure data enclave with secure partitioned project spaces and strict access controls. The NCSES and NCHS data are accessed within an NCSES-sponsored secure data enclave, while Census Bureau data are being used within a Federal Statistical Research Data Center (FSRDC) or Survey Sponsor Data Center (SSDC). Restricted virtual desktop environments ensure secure access with no access to internet, email, or printing.
- *Special Sworn Status:* The U.S. Census Bureau requires that all users of confidential data protected under Title 13, CIPSEA, Title 26, or another confidentiality pledge obtain Special Sworn Status (SSS) and take annual data security training. For example, the Annual Business Survey contains confidential Census and tax data, so access to the survey data requires SSS. Obtaining SSS requires undergoing a background check and obtaining a security clearance. All users must also sign a Rules of Behavior document, certifying that they understand all security measures. In addition, some data, such as those protected under Title 26, must be accessed in a security facility, such as an FSRDC or SSDC.
- *Disclosure Avoidance:* Confidential data cannot be removed from secure project spaces. All output from data use, such as tables, graphs, or statistics, must be cleared for public release through the agency's disclosure avoidance procedures to protect respondents from the risk of reidentification.

For all federal agencies involved in the Demonstration project, data security is paramount. Similar practices exist across agencies for signing security and data use agreements, taking annual data security training, and accessing data on secure servers and in secure project spaces to ensure that only approved individuals have access to only the data approved for their use. These common elements are part of a framework that will inform a future National Secure Data Service's ability to ensure that legally protected data are kept secure, and that respondent confidentiality is maintained. As part of the NSDS Demonstration project, integrating privacy-preserving technologies into existing policy frameworks is being explored, discussed further in the next section.

III. Completed and Active Data Linkage (and Other Shared Services) Activities and Projects

During its first two years, the NSDS Demonstration project has tested and evaluated potential services that could be utilized and shared government wide to support data access and linkage to further evidence-building and informed decision-making across the federal government as well as among state, local, territorial, and tribal governments. These shared services are being informed through eighteen projects that have been initiated, with several completing and additional projects planned for future years of the Demonstration. The NSDS Demonstration projects are informed by stakeholder engagements and the recommendations of the Commission on Evidence-Based Policymaking (CEP) and Advisory Committee on Data for Evidence Building (ACDEB). This envisioned data service would include infrastructure, tools, customer service, and capacity-building anchored with strong stakeholder engagement and governance to streamline and facilitate data sharing and linkage. Projects established in the first two years of the Demonstration project have sought to inform this shared services vision through an exploration of technology and tools, novel data sharing and linkage, infrastructure, and stakeholder needs to support data sharing and linkage of federal data.

Most of the projects for the Demonstration project have been supported through the America's DataHub Consortium (ADC). In August 2021, NCSES established the America's DataHub Consortium (ADC) by awarding an Other Arrangement (42 U.S.C. § 1870(c)) to Advanced Technology International (ATI), a 501(c)3 public service nonprofit. The ADC was formed as a coordinated research and development program to advance NCSES's statutory role as a central federal clearinghouse for the collection, interpretation, analysis, and dissemination of objective data on science, engineering, technology, and research and development[7]. ATI serves as the consortium management firm for ADC, and in that role, has two primary functions: fostering collaboration and accelerating project awards. As a result, the ADC has played a central role in engaging stakeholders and piloting projects in support of the NSDS Demonstration project.

The following tables each highlight projects within a particular area of contribution to the Demonstration project. Table 1 lists projects that link data to provide insight into secure linkage and how linked data can be harnessed to inform evidence-building. Table 2 lists projects that inform equity and customer service, exploring how to best deliver services to a wide variety of data users and increase their capacity to engage in evidence-building. Table 3 lists projects that inform data access, security, and interoperability, highlighting ways to make more data available while preserving privacy. Table 4 lists projects that inform infrastructure for a National Secure Data Service, exploring data access and use platforms. Table 5 lists projects that provide guidance from experts in visioning and shaping a potential NSDS. The tables list the status of each project. Most are ongoing and four have been completed to date. These projects are critical to developing a National Secure Data Service, with shared services informing capabilities and infrastructure and linkage projects providing critical use cases for how such a service can

[7] PL 111-358, Section 505.

facilitate secure data sharing, linkage, and analysis. Future projects will build on lessons learned from existing activities including piloting services with the ultimate goal of informing a prototype of a National Secure Data Service.

Table 1: Projects that Inform Data Linkage

Title	Description	Status
Utilizing Privacy Preserving Record Linkage to Link Data from Two Federal Statistical Agencies	This project utilizes an application of secure multiparty computation, known as privacy preserving record linkage (PPRL) to link the two data sources without the sharing of direct personal identifiers. This project highlights the ability to deploy an open source PPRL tool in a secure environment with a trusted third party supporting the linkage process. This technique is a potential shared technology that could serve as a critical component in support of secure data linkage within a future National Secure Data Service. This project also involves a data sharing agreement between two federal statistical agencies, with lessons learned from that process that could inform standardization of agreements in the future. These linked data will inform how differing questions on disability could impact policy discussions, specifically in the case of the congressionally mandated report: Women, Minorities, and Persons with Disabilities in Science and Engineering.	Ongoing
Utilizing Privacy Preserving Record Linkage with Parent Agency Data and Statistical Agency Data to Inform Programs and Policies	This project uses a commercial, non-open source PPRL tool to link the two data sources without the sharing of direct personal identifiers and serves as a second use case for deploying a PPRL tool in a secure environment with a trusted third party. It involves a data sharing agreement that could inform standardization of data sharing agreements. This project links survey data from NCSES and NSF's principal investigator award database to inform the learning agenda of NSF and assess equity in awards for recent doctorate recipients.	Ongoing
Foreign-Born Scientists and Engineers and the U.S. Workforce	This project uses clear text record linkage to link data from NCSES survey respondents to employer records to explore the U.S. return on investment on U.S. training of foreign-born scientists and engineers. This project serves as a use case for data linkage to inform a policy issue and assists in highlighting the different linkage methods for possible inclusion as services within an NSDS.	Ongoing

Table 2: Projects that Inform Equity and Customer Service

Title	Description	Status
Models for a Data Concierge Service for a National Secure Data Service	This project explores models for a data concierge service that would serve as a shared resource for individuals seeking access to federal confidential data. The concierge service would coordinate between data providers and data users, assisting data users in answering general questions; identifying confidential data assets that meet their evidence-building needs; providing data linkage support; directing them to subject matter experts and more focused guidance when needed; and helping them develop evidence-building proposals to apply for access to confidential data.	Ongoing
National Secure Data Service Website	This project develops a website to support the NSDS Demonstration project while informing a future website that would support, and potentially serve as the front door, to a National Secure Data Service.	Ongoing
Informing Evidence-Building Capacity among State, Local, Territorial, and Tribal Governments within a National Secure Data Service	This project explores how a future NSDS could support capacity building for research and data science among state, local, territorial, and tribal governments, either through skill building, continuous learning opportunities, and/or access to infrastructure and tools that governments may not have access to. Recommendations will include capacity-building strategies for these stakeholders that could be provided by a data concierge service within a future NSDS.	Ongoing
Challenges to inclusive and equitable evidence-building research in the Federal Statistical Research Data Centers (FSRDCs)	This project informs strategies to increase equity in utilization of the FSRDCs. Through a survey and focus groups, this project explores barriers to FSRDC access by minority-serving institutions, less-resourced colleges and universities, local government agencies, and low-resource nonprofit agencies that serve underrepresented groups and communities, and implications for a future NSDS.	Ongoing

Table 3: Projects that Inform Data Security, Access, and Interoperability

Title	Description	Status
Data Protection Toolkit Use Case Analysis	This project conducted a use case analysis of the Federal Committee on Statistical Methodology's Data Protection Toolkit (DPT) to identify successful use cases and potential enhancements to the Toolkit for enabling access to federal data assets while protecting confidentiality. The Data Protection Toolkit is a government-wide, shared resource in response to the Evidence Act and Federal Data Strategy that builds on well-established expertise within the Federal Statistical System. The Toolkit	Completed

	provides resources, tools, and content about protecting data while increasing access to a variety of users.	
Privacy Preserving Technologies Environmental Scan	This project provided insight into the current landscape of privacy preserving technologies (PPT) for the protection of persons, data, and systems that contribute to the use of confidential data, including both individual-level and business data, for evidence-building and policymaking. PPT includes technologies, techniques, methodologies, approaches, tools, and other like terms that relate to preserving privacy. PPT leaders and innovators in government, academic institutions, and private industry were interviewed and convened in expert panels to discuss the current landscape of PPT and considerations and challenges to implementation of PPT. Key takeaways included the importance of first clearly specifying an evidence-building question and then deciding on the right PPT tool to inform the question at hand, the importance of developing a multidisciplinary team when planning to implement PPTs, bridging information gaps between technology and legal teams and leveraging technology and governance best practices to streamline data sharing agreements.	Completed
Evaluation of Noise Infusion for a Large-Scale Demographic Sample Survey	This project evaluates the use of noise infusion as a possible privacy-preserving method in the use of federally confidential data from a demographic survey. Investigations focus on use cases where noise infusion may be appropriate and use cases where noise infusion may introduce quality issues that reduce confidence in the use of resulting estimates for decision-making.	Ongoing
Creation of Synthetic Data for the and Development and Use of Verification Metrics	The objective of this project is to produce a synthetic data file for public use as part of a tiered access model when working with a dataset that is a census rather than a sample survey, explore the use of synthetic data for evidence-building, and test the use of verification metrics in validating estimates produced from synthetic data.	Ongoing
Creating and Validating Synthetic Data for a Nationally Representative Survey of Businesses	The objective of this project is to test and compare methods for creating synthetic data as part of a tiered access model for survey data; explore the use of synthetic data for evidence-building; and test the use of verification metrics in validating estimates produced from synthetic data.	Ongoing
National Center for Health Statistics: National Vital Statistics System Modernization—New Opportunities for Interoperable Data	This project highlights opportunities for the use of interoperable health data to support timely research and public health surveillance. For the purposes of this project, “data interoperability” encompasses a wide range of related topics—data quality, standards, metadata, definitions, systems, and	Ongoing

	technologies—needed to share information effectively and to support the creation of better evidence for decision-making. The results of this work will inform a future NSDS and best practices when working with state and territory data.	

Table 4: Projects that Inform Infrastructure

Title	Description	Status
Secure Compute Environment Scan	This project summarized the current landscape of secure computing environments within the federal government related to data sharing, privacy, and confidentiality restrictions required by multiple statutes. A secure compute environment is a platform for secure data sharing, linkage, and access. This project was the first step in the upcoming development of an NSDS secure computational environment platform planned as a shared resource for data linking and other analytical activities. The scan included a review of legal requirements under various statutes for data protection and security related to IT infrastructure for the management and use of confidential data.	Completed
Secure Compute Environment Testbed for a National Secure Data Service Demonstration Project	This project builds a secure compute environment as a testbed for research and development for the Demonstration project and pilots a shared IT platform that can be accessed for data sharing, linkage, and analysis. This environment will allow us to test privacy-preserving technologies and will be available for use by federal, state, local, territorial, and tribal governments as well as by other entities that collaborate on projects as part of the Demonstration project.	Ongoing
Federated Data Usage Platform	This project researches and develops a robust and sustainable prototype dashboard to aid the federal data ecosystem in understanding the uses of its data. This project will produce possibilities for a future, state-of-the-art, updatable, and publicly accessible dashboard platform that provides information on usage of federal data. This dashboard would include information about projects that use federal data in order to promote transparency, foster communities of practice, and help inform future data use.	Ongoing

Table 5: Projects that Provide Expert Guidance

Title	Description	Status
Expert Panel for Informing the National Secure Data Service Demonstration Project	This expert panel convenes experts from federal state governments as well as non-profit organizations to offer strategic input on a vision for	Ongoing

	an NSDS, identifying opportunities to collaborate with and leverage other federal and state initiatives, identifying any gaps in the current demonstration activities, and serving as ambassadors to communicate the goals and vision for an NSDS.	
The Quality of National Statistical Data: A Workshop	This workshop assembles a diverse array of experts to inform how different countries have addressed the issues of greater diversity of data sources and methods used to generate national statistics that are more timely, granular, and policy relevant. The workshop also targets how various countries effectively communicate to policymakers and the public, differences in data quality, appropriate data uses, and strengths and tradeoffs across different statistics and data products.	Ongoing
Request for Information: Use Cases to Inform a National Secure Data Service	A request for information (RFI) was submitted for members of the public to submit potential use cases for a National Secure Data Service. The information received will be used to inform use cases to pilot potential shared services within an NSDS. Twenty responses were received from a range of stakeholders including federal agencies, academic institutions, state and local governments, non-profits, and private industry.	Completed

IV. Assessment of the Effectiveness of the Demonstration Project in Mitigating Risks and Removing Barriers to a Sustained Implementation of a National Secure Data Service

Both the year 1 and year 2 reports of the Advisory Committee on Data for Evidence-Building (ACDEB) discussed risks and barriers to data sharing and evidence-building, including legal, cultural, data interoperability, resource and capacity issues, and privacy. During the first two years, the Demonstration project has explored and piloted potential services to address risks and barriers to data sharing to support a future NSDS. These projects, discussed in Section III, include exploring linkage technologies to protect privacy, piloting ways to best serve stakeholders and data users in their evidence-building activities, learning how to expand data access while preserving privacy, and planning the development of a shared IT infrastructure to support an NSDS.

The Demonstration project has identified areas where risks and barriers can be addressed to support the goals of a future, viable NSDS. We plan to continue to explore these issues with ongoing projects and planned future projects. The risks and barriers explored thus far, as well as how mitigation strategies are being explored, are provided below in Table 5.

Table 5: Risks and Barriers with Mitigation Strategies

Risk or Barrier	Mitigation Strategies
Preserving Privacy when Linking Data	Data linkage projects in the Demonstration have shown that use of privacy preserving record linkage (an application of secure multiparty computation) techniques are feasible and can provide a means to securely link de-identified data while balancing privacy and utility. Utilizing a trusted third party to conduct the linkages, after deidentification and encryption, is a viable component of an NSDS infrastructure that would greatly enhance the capacity to link datasets across the federal statistical system, other government agencies, and organizations while protecting privacy. The projects have relied on cryptographic guidelines from NIST (Guideline for Using Cryptographic Standards in the Federal Government: Directives, Mandates and Policies (nist.gov)) as the privacy preserving record linkage tools have been implemented.
Challenges with Data Sharing Agreements	Data sharing agreements have been established between two federal statistical agencies and between a federal statistical agency and its parent agency during this Demonstration period. To establish these agreements, the statutes, mandates, and policies that govern each agency's data were addressed and acknowledged through working closely with the Offices of General Counsel at both agencies. NCSES staff became designated agents of NCHS. For the agreement between the federal statistical agency and the parent agency, the System of Records Notice (SORN) was consulted to ensure that the proposed activities within the project fell within the SORN of the data collection. While these data sharing agreements were successfully executed, significant resources were required to draft and negotiate them. An executed data sharing agreement can also take a minimum of 6 months and sometimes takes longer. This timeframe is not conducive to time-sensitive program or policy issues that require a rapid turnaround. In 2017, the Commission on Evidence-Based Policymaking found that data sharing agreements between two or more agencies can take years to develop and suggested "developing a common MOU template for use in the Federal government." The two projects mentioned above provided insights into the data sharing process and valuable lessons about the resources and processes required to establish agreements. The establishment of templates through these Demonstration projects could reduce some of the negotiation time. The Commission also cited delays caused by inconsistent legal authorities. To mitigate that risk, they called on Congress and the President to review and amend statutes as appropriate to allow statistical use of data for evidence building. Recommendations include amending statutes such as Title 13 to allow statistical uses of survey and administrative data for evidence-building within a secure CIPSEA environment; repealing current and limiting future bans on the collection and use of data for evidence-building; and enacting statutory changes to ensure state-level data on earnings are available for statistical purposes. The full set of recommendations can be found in Section 2 of the 2017 final report.
Need for Additional Data Access Options	Synthetic data are being explored as part of a tiered access approach to data protected by federal statutes. Synthetic data can expand publicly available data while preserving confidentiality and addressing privacy concerns. User workshops have thus far indicated a need and desire to have data prior to submitting a proposal to work in a restricted

	environment and for training purposes. Readily available synthetic data can be used to build models and as training datasets for machine learning. With verification metrics researchers can learn the validity of their estimates from synthetic data and then decide if they should put in a proposal to access restricted data. The Demonstration projects are establishing standard methods for creating synthetic data files, including outlining steps for assessing disclosure risk assessments and validity of the synthetic files when compared to the statistical properties of the underlying datasets. These tools could be utilized within an NSDS environment as a comprehensive toolkit to make creating and disseminating synthetic data more streamlined for data providers.
Challenges Navigating the Federal Data Ecosystem	Models for a data concierge service are being explored to provide customer service to a wide range of stakeholder communities. These models are being informed by discussions with stakeholder communities and data users about their needs and potential use cases for a future NSDS. Engagement with state, local, territorial, and tribal governments is also underway to learn how best to support their evidence-building needs, which could include training, toolkits, and infrastructure. A data usage platform is also being piloted to provide the public with information on uses of federal data to inform their evidence-building needs. This platform will provide key resources to researchers and policymakers while protecting privacy and data security.
Protecting Data while Increasing Sharing	Ways to mitigate risks to privacy and data security through a data protection toolkit were explored with both federal and non-federal data providers. Suggestions for improvement from data providers included providing information on assessing disclosure risk, highlighting unsuccessful disclosure strategies, and improving visibility of resources and navigation within the toolkit. An additional project is exploring using noise infusion with a sample survey and assessing its effects on data quality and disclosure risk.
Promoting Data Interoperability	Interoperability of state vital statistics data is being explored to serve as a model for increasing interoperability of other state and federal data. The National Vital Statistics System is a model of interoperable data that can be used to inform policy based on birth and death statistics. However, there are many other datasets at the state and territory level that could also be used to inform policy. This Demonstration project has laid the foundation for best practices with interoperable data and will inform future initiatives to standardize data that can be then linked among states and between states and the federal government. Common data model repositories that could be available through open-source tools are being explored to promote interoperable data and mitigate risks that would prohibit data sharing and linkage.
Utilizing Privacy Preserving Technologies within the Federal Government	While the field of privacy preserving technology has been expanding rapidly, federal government agencies have been slow to embrace these new technologies. A Demonstration project was conducted to assess the privacy preserving technology landscape, building on the United Nations Privacy Enhancing Technology report. There are both input and output privacy preserving technologies which are at different stages of maturation. Input privacy tools include encryption and cryptographic techniques. Output privacy tools include differential privacy and noise infusion. This project highlighted key areas for next steps to inform a future NSDS, including but not limited to working closely with NIST to

	establish standards to implement privacy preserving technologies, developing communication strategies that provide background on the technology and can be shared with legal teams and governance bodies, and establishing communities of practice across agencies to build and learn from one another. A secure compute environment is being established to assess how federal agencies can use privacy-preserving technologies within the Demonstration project and to pilot a comprehensive IT infrastructure that could support the use of these technologies for data protection.

During the remainder of the Demonstration period, additional services will be explored to remove risks and barriers to secure data sharing, linkage, and use for evidence-building. In addition, ongoing collaboration with the National Artificial Intelligence Research Resource (NAIRR) pilot will provide opportunities to explore how artificial intelligence (AI) can be leveraged within an NSDS and how an NSDS can support research into solutions that include AI. Stakeholder engagement is critical to our efforts in determining the future direction for a potential NSDS. We have held meetings and/or briefings with the following groups:

- Coalition for National Science Funding
- Beeck Center, Georgetown University
- Data Foundation
- Health and Human Services Data Governance Board
- Congressional Research Service
- National Science Foundation Artificial Intelligence and Data Governance Group
- Centers for Disease Control and Prevention Privacy-Preserving Record Linkage Working Group
- Staffers for the House Science, Space, and Technology Committee
- Presidential Council of Advisors on Science and Technology
- Staff for the Hon. Representative Chellie Pingree
- National Science Foundation Enterprise Data Governance and Education Working Group
- Committee on National Statistics
- Chief Data Officers Council Data Sharing Working Group
- State Chief Data Officer Network
- Disability Data Interagency Working Group Resources and Infrastructure Workstream National Science Foundation Social, Behavioral, and Economic Advisory Committee

In addition, we have presented at the following conferences and business meetings:

- Federal Statistical Research Data Centers Annual Business Meeting
- Internal Revenue Service, Statistics of Income Division Business Meeting
- Association of Public Data Users Annual Conference
- Federal Committee on Statistical Methodology Annual Conference
- Joint Statistical Meetings Annual Conference
- International Population Data Linkage Network conference

- Georgetown University Privacy and Public Policy Conference

V. Plan for Scaling-up the Demonstration Project to Facilitate Data Access for Evidence-Building

Scaling-up the Demonstration project will require addressing multiple dimensions to ensure appropriate funding, effective governance, stakeholder engagement, a robust infrastructure exists that can support a variety of services to facilitate data sharing and linkage, and an overall structure and vision for a shared services model that supports data access for evidence-building across the nation. Addressing these issues requires ongoing collaboration with the Interagency Council on Statistical Policy (ICSP), the Office of Management and Budget (OMB), and the NAIRR pilot, as required under Section 10375(a) of the CHIPS and Science Act. Ongoing engagement with stakeholders in federal, state, local, territorial, and tribal governments, policymakers, and non-profit organizations is also critical to ensure that the resulting NSDS can support evidence-building needs across a wide range of communities. It is worth noting that a significant challenge to data sharing will continue to be the lack of a "notwithstanding clause" in CIPSEA that would scope-in agency's title-specific confidential data protections in statute. As a result, the effectiveness of an NSDS may be limited for the purposes of data linkage between many agencies' data assets. Statutory changes may be needed to ensure that data sharing and linkage is possible to support evidence-building activities. The following is a plan for scaling-up the Demonstration project to facilitate data access for evidence-building ultimately informing the feasibility, benefits, and costs of a future NSDS. The following activities will be led by the NSDS Demonstration project in collaboration with stakeholder groups.

- *Explore Funding Models:* The NSDS Demonstration will develop and recommend a funding model for an NSDS. Funding models could include appropriated funds for an NSDS, cost-sharing by participating agencies, and cost-reimbursement models for users.
- *Establish Governance Structure:* The NSDS Demonstration will develop a proposed governance structure for an NSDS, to include determining membership, roles, responsibilities, and processes for decision-making. The governance structure will be informed by existing governance efforts within the NSDS Demonstration project, including ICSP member participation and expert panel input.
- *Engage Stakeholders:* The NSDS Demonstration will continue to engage stakeholders across a wide range of communities to explore ways that an NSDS can meet their evidence-building needs. In addition, an expert panel of federal and state government as well as non-profit stakeholders will be assembled to provide ongoing feedback on projects and potential services.
- *Pilot IT Infrastructure:* The NSDS Demonstration will establish a secure compute environment as a pilot infrastructure for data sharing, linkage, and use of privacy-preserving technologies. NCSES will continue to gather requirements for data access from federal agencies and other data users to ensure systems meet all privacy and security requirements.
- *Specify a Model for an NSDS:* The NSDS Demonstration will inform the development of an overall structure and model for an NSDS. This model will include an overall structure,

including how users would access services, what services would be provided, and how existing shared services for secure data access, such as the Standard Application Process, the Federal Statistical Research Data Centers, and agency secure data enclaves, would be integrated. In addition, where an eventual NSDS will reside will also be addressed as a key decision point.

- *Conduct Ongoing Research and Explore Use Cases:* The NSDS Demonstration will continue to pilot potential new shared services to include in an NSDS as well as use cases that demonstrate the utility of an NSDS in facilitating data access, linkage, and use for evidence-building. These pilots will also identify needs and gaps in existing services to inform potential shared services for an NSDS and determine their value. Focus areas will include privacy preserving technologies to continue expansion of approaches to protect confidentiality while linking and analyzing data; data concierge services and customer service models to inform data use needs; data sharing with state, local, territorial, and tribal governments to uncover and mitigate associated barriers; support for policymakers to address their unique requirements for actionable information; metadata and dataset schema standardization to facilitate data usage and linkage across disparate data sources; and use of artificial intelligence to manage and mitigate risks and provide expansive customer service within an NSDS.

The next three years of the NSDS Demonstration project will be critical for piloting of potential shared services through future projects, exploration of technologies to ensure data protections, establishment of prototype infrastructure, and development of a robust governance framework that includes input from our stakeholder communities. We look forward to continued input from members of Congress during the Demonstration period to shape projects and provide critical feedback on our efforts to inform a potential NSDS.