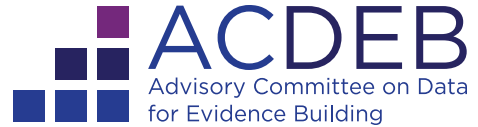


Advisory Committee on Data for Evidence Building: Year 1 Report

October 29, 2021





October 29, 2021

Shalanda Young, Acting Director of the Office of Management and Budget (OMB)

Dear Acting Director:

On behalf of the Advisory Committee on Data for Evidence Building and pursuant to the Foundations for Evidence-Based Policymaking Act of 2018 (Public Law 115-435), I am pleased to submit the following report detailing our findings and recommendations from the Committee's first year.

The Committee affirms the need for a National Secure Data Service, as asserted by the Commission on Evidence-Based Policymaking. The data service will be an important component of the broader evidence-building ecosystem and will build on the framework established in the Evidence Act. The Committee's Year 1 recommendations are actionable, relevant, and timely, covering high-priority items that support the next steps for implementing the Evidence Act and establishing and operationalizing a successful National Secure Data Service. Furthermore, the Committee will form its Year 2 recommendations based on the recommendations presented in the Year 1 report.

As you know, the Committee began a year ago under the leadership of Dominic Mancini in his role as Acting Chief Statistician of the United States. I was honored to accept the role from him when other responsibilities of OMB required him to step away. I want to thank him for launching the Committee on such able footing.

The Committee is taking a holistic, phased approach to gather information, build findings, weigh options, and make recommendations. During its first year, the Committee focused on fact finding to develop an overarching vision for a data service and lay out the contours of what is required to realize that vision. The Committee spent the first several months gathering information from outside experts and sharing knowledge and experience relevant to fundamental evidence-building issues.

From there, the Committee entered the deliberation phase of its work, organized around five focus areas and related subcommittees—legislation and regulations; governance, transparency, and accountability; technical infrastructure; government data for evidence building; and other services and capacity-building opportunities. These focus areas were designed to address a wide range of opportunities and obstacles for a data service and the evidence-building ecosystem more broadly, leveraging the vast input and insights from Committee members.

Continued

In this deliberative phase, the Committee continued to build its knowledge base by harnessing the expertise of its members and conducting virtual site visits to existing data facilities. The Committee also began to collaboratively synthesize different perspectives and use cases into a coherent understanding of the current state of and future needs for the use of data for evidence building. The Committee members recognize their efforts as a work-in-progress that will continue across the next 12 months.

By building on its Year 1 recommendations and ongoing efforts across federal, state, and local governments as well as the private sector, the Committee will promote the use of data for evidence building and champion the need for and value of a National Secure Data Service. These efforts promise to fundamentally transform data sharing, data linkages, and privacy-enhancing techniques.

I call on the Director of OMB to adopt these Year 1 recommendations and work in conjunction with the Committee to continue this invaluable work during Year 2.

Respectfully yours,



Emilda Rivers

ACDEB Chair on behalf of fellow Committee members

Laila Alequresh

Richard Allen

Otis Brown

Leonard Burman

Charles Cutshall

Shawn Davis

Gregory Fortelny

Nicholas Hart

Christine Heflin

Anna Hui

Barry Johnson

Ted Kaouk

Elisabeth Kovacs

Edward Kwartler

Julia Lane

Christin Lotz

Brian Moyer

Kimberly Murnieks

Amy O'Hara

David Park

Todd Richardson

Matthew Soldner

Kenneth Troske

Mayank Varia

Christina Yancey

Contents

1. Executive Summary	1
1.1. Background and Vision of the Advisory Committee on Data for Evidence Building	1
1.2. The Foundation on Which the Committee Builds	2
1.3. Current Evidence-Building Ecosystem	3
1.4. The Committee’s Charge	4
1.5. Vision for a National Secure Data Service	4
1.6. The Framework for the National Secure Data Service	5
1.7. Resources for the National Secure Data Service	6
2. Committee Recommendations	8
2.1. Recommendation 1 – Evidence Act Regulations	9
2.2. Recommendation 2 – Chief Statistician of the United States	10
2.3. Recommendation 3 – Standard-Setting Procedures	10
2.4. Recommendation 4 – Appropriations Requests	11
2.5. Recommendation 5 – Value-Driven Pilot Program	11
2.6. Recommendation 6 – Privacy-Preserving Technologies Case Studies	13
2.7. Recommendation 7 – Communication	13
3. Areas of Focus	14
3.1. Legislation and Regulations Findings and Recommendations	14
3.2. Governance, Transparency, and Accountability Findings and Recommendations	19
3.3. Technical Infrastructure Findings and Recommendations	26
3.4. Government Data for Evidence Building Findings and Recommendations	31
3.5. Other Services and Capacity-Building Opportunities Findings and Recommendations	35
4. Year 2 Roadmap	39
4.1. Year 2 Planned Approach	39
4.2. Initial Year 2 Focus	39
5. Appendices	40
Appendix A. Committee Charter, Membership, Process, and Meetings	40
Appendix B. Selected Meeting Presentation Summaries	43
Appendix C. Public Comments Summary	51
Appendix D. Project Inventory	56
Appendix E. Virtual Site Visits	58
Appendix F. Case Study Criteria Matrix	87
Appendix G. Crosswalk of Recommendations from the Evidence Commission to ACDEB Focus Areas	92
6. References and Acronyms	94
6.1. References	94
6.2. Acronyms	100

1. Executive Summary

1.1. Background and Vision of the Advisory Committee on Data for Evidence Building

The Advisory Committee on Data for Evidence Building (Committee or ACDEB) was established as part of the [Foundations for Evidence-Based Policymaking Act of 2018](#) (Evidence Act) to review, analyze, and make recommendations to the White House Office of Management and Budget (OMB) Director on how to promote the use of federal data for evidence building with a particular emphasis on evaluating the need for and value of a potential National Secure Data Service (NSDS). Congress recognized that the establishment of a National Secure Data Service would have broad impact for both governmental and non-governmental entities and thus wanted a broad committee established. The Committee, composed of experts from federal, state, and local governments and the private sector, represents multiple areas of expertise such as privacy, technology, and research method. The Committee is unique because of its diversity and wealth of expertise.

Over the past 12 months, the Committee has engaged in extensive fact finding by:

- examining the recommendations of the Commission on Evidence-Based Policymaking (Evidence Commission) and the implications of their partial implementation through the Evidence Act;
- leveraging the expertise of its members;
- hearing from researchers, government leaders, other experts, and the public;
- conducting virtual field trips to existing data facilities; and
- beginning to collaboratively synthesize the different perspectives and use cases into a coherent understanding of the current state of and future needs for the use of data for evidence building.

The Committee members recognize their efforts as a work-in-progress that will continue across the next 12 months. This report summarizes the Committee's first-year activities and resulting findings, articulating a vision for an NSDS and the future of data sharing, data linkages, and privacy enhancing techniques across federal agencies and with state and local governments. It lays out recommended actions that can be taken today to build toward that vision while also describing the path that the Committee intends to take across the next year to further develop recommendations for implementing the vision.

1.2. The Foundation on Which the Committee Builds

In 2016, Congress created the [Evidence Commission](#) and, in 2017, the Commission issued 22 recommendations on improving data access, modernizing privacy protections, strengthening evidence-building capacity, and establishing an NSDS as part of the broader evidence environment. The Evidence Commission viewed the establishment of a data service as a means for addressing gaps in current capabilities of the federal government and partners, while also supplementing expertise and capabilities with new frameworks for managing, using, and protecting data. In that spirit, the Evidence Commission offered a series of recommendations that serve as core parameters for a potential design of an NSDS while deferring many specific details for further discussion.

Congress addressed 11 of the Evidence Commission recommendations through the Evidence Act, including several specific provisions related to NSDS authorities and functions. The Act builds on three big ideas: (1) evidence-based policymaking requires systematic planning, (2) effective data use requires high-quality data governance, and (3) access to protected data for evidence building can be safely expanded.

To build on the first big idea, the Evidence Act requires planning, through learning agendas, capacity assessments, and evaluation plans. This planning is coordinated by newly established Evaluation Officers and implemented in concert with newly established Statistical Officials and Chief Data Officers (CDOs). To build on the second idea, the Evidence Act requires data governance and management structures in agencies, led by CDOs. The third big idea, which is the primary focus of the Committee, leverages and builds upon statistical law (i.e., the Confidential Information Protection and Statistical Efficiency Act (CIPSEA)) and current and future statistical agencies. Thus, the Evidence Act leverages the primary existing federal evidence capacity and establishes new authorities, functions, and frameworks, creating the current evidence-building ecosystem.

A key element of the third big idea of the Evidence Act is that safely expanding access to and use of data, including administrative data, for evidence building can substantially improve public policies. For instance, as response rates to traditional government surveys decline and costs to administer them increase, agencies are relying more on and more data that are the result of administrative processes—like records from employers and participation in government programs—to supplement or replace traditional surveys. These data sets, and the underlying metadata, are often the same data that researchers need for evidence building.

The Evidence Act granted new authorities and even greater protections to federal statistical agencies, giving them responsibilities that the Evidence Commission identified as ones for an NSDS. The Evidence Act creates a statistical agency presumption of access to federal government data, which statistical agencies are to protect by implementing a consistent, secure, and comprehensive approach set in regulation, and then make available through a standard application process and tiered access framework.

These requirements are designed to enhance the way the statistical agencies work as a system to facilitate evidence building, just as envisioned by the Commission, even prior to introducing an NSDS coordination function into the ecosystem. Establishment of the Committee within the Evidence Act was also intended as a signal that Congress was interested in further recommendations about improving the national data infrastructure for evidence-building activities, including defining the role and appropriate location and structure of an NSDS within the changing ecosystem.

1.3. Current Evidence-Building Ecosystem

The Committee recognizes that the landscape has changed substantially since the Evidence Commission issued its 2017 recommendations and the Evidence Act passed in 2018. Federal agencies have made real progress under the Evidence Act toward addressing challenges for accessing and using data for evidence building, largely focused on new authorities, roles, and requirements. In addition, advances in privacy-enhancing technologies and secure remote access facilities offer new opportunities for transforming the evidence-building ecosystem.

The evidence-building system is highly decentralized, with over 100 federal agencies or units that engage in statistical activities as well as other entities providing administrative data used for statistical purposes. Of these entities, there are 13 principal statistical agencies, each of which has a primary mission to produce relevant, accurate, and timely information to inform public and private decision-making. The Evidence Act strengthened these statistical agencies as the trusted intermediaries between the data stewards and subjects, on the one hand, and the data users, on the other.

For example, statistical agencies are implementing an Evidence Act requirement for a standard application process to streamline and harmonize researcher access to government data sets for authorized uses. The Commission envisioned this coordinating function as a core role for the National Secure Data Service. Therefore, NSDS would be joining such federal evidence-building efforts already in motion as a new partner.

In addition, within the broader evidence-building ecosystem in the United States, there are thousands of state and local governments who collect administrative data that have vast potential for informing evidence-based decision-making. Often, these data are gathered through federally funded or run programs and thus are, in part, being shared with the federal government today; however, the power of administrative data for evidence building has not yet been realized. Because they are not designed for evidence building, these data often lack the documentation and quality emphasis needed for this purpose. Likewise, despite the potential for evidence building, resources are not provided from the federal or state level down to point-of-origin for data collection to support high-quality information flowing back up to federal, state, and local decision-makers.

Along with data quality, data protection is an ongoing focus in enabling data access and linkages. Statistical agencies currently acquire, protect, and make data accessible for evidence building as part of their statutory missions. These statistical agencies have decades of experience developing, testing, and using traditional confidentiality and security protections. However, they have limited experience with newer methods and cutting-edge technologies that can enable stronger security and privacy protections while providing value for evidence building. These technologies offer a starting point for evaluating solutions for secure access to non-identifiable data for evidence building and would benefit from a coordinated approach to identifying and evaluating them, which the Commission noted is lacking in the evidence-building ecosystem. The Commission recognized the need for a coordinator and recommended this as a role that an NSDS could play, including identifying, researching, and facilitating adoption of innovations to create capacity for the existing statistical agencies.

1.4. The Committee's Charge

The Committee will continue to advance the use of data for evidence building by leveraging what came before, including the work of the Evidence Commission; the vision of the Evidence Act; ongoing efforts across federal, state, and local governments; and advances in privacy-preserving technologies.

The promise of the Evidence Act has only begun to be realized and is not yet fully implemented, particularly as it intersects with the Committee's focus on expanding access to non-public data for evidence building. The evidence-building ecosystem would both benefit from and support the success of an NSDS.

The Committee seeks to understand and define how an NSDS could enhance the existing evidence-building ecosystem in the United States in a way that makes the entire system better. The Committee's recommendations for the National Secure Data Service, and for other needed evidence-building capacity and coordination, aim to advance the vision for how an NSDS fits alongside the federal evidence-building system and other key federal and non-federal actors to facilitate data access, enable data linkages, and develop privacy-enhancing techniques in support of increasing data availability for evidence building across the entire evidence-building ecosystem.

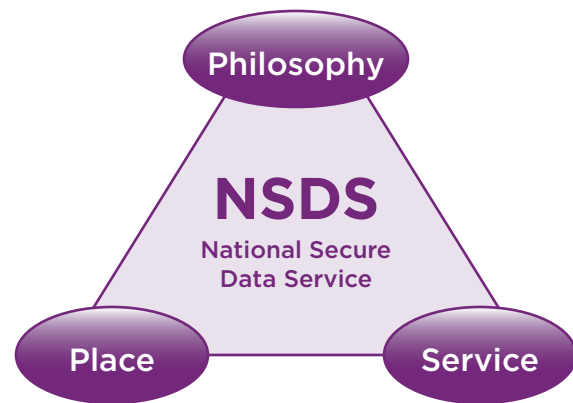
1.5. Vision for a National Secure Data Service

The Advisory Committee on Data for Evidence Building members concur with the Evidence Commission that the United States needs to establish a National Secure Data Service. Four years have elapsed since the Evidence Commission provided its recommendations to Congress and the President, and during that time the need for establishing a data service and a coordinated federal capacity for data sharing, linkage, and protection has only increased. Better and more useful data are central to meeting the challenges the nation faces today, from battling a pandemic and recovering from its effects, to identifying and addressing inequities that impede collective success.

The NSDS will be an important component of the broader evidence-building ecosystem, whose goal is to develop evidence for decision-making. As such, the NSDS will build on the framework established in the Evidence Act and advanced through the [Federal Data Strategy](#), including operating under CIPSEA and complying with associated privacy and confidentiality requirements. Therefore, the establishment and success of the NSDS, and meeting the goals of the Evidence Act, depend heavily on the full implementation of the Evidence Act and the Federal Data Strategy. These provisions and the related activities strengthen the evidence-building ecosystem in which the NSDS will operate by establishing essential authorities, responsibilities, and frameworks.

1.6. The Framework for the National Secure Data Service

Building on the recommendations of the Evidence Commission and work since 2017 on how to establish a data service, the *Advisory Committee members acknowledge that a National Secure Data Service should be a philosophy, a service, and a place*. Each concept in this framework is further described below, and it is recognized that the data service should continue to adapt and “evolve” to offer more capabilities over time beyond what may be possible during an initial phase of operations.



Philosophy. As a philosophy, the NSDS will produce value for the American public by facilitating evidence building. Federal agencies, with the NSDS as a coordinator, will operate as a unified evidence-building system, partnering with state and local governments and non-governmental organizations. As such, the NSDS will coordinate with its federal, state, and local government partners to elevate evidence building that is relevant and timely for these key stakeholders, especially when their data and other resources are involved.

Service. As a service, the NSDS will provide coordination and capacity-building services for data users, data providers, and related communities of practice. The NSDS will coordinate evidence-building efforts that cut across entities, including federal, state, and local governments as well as non-governmental organizations, to use data of many types in evidence building. The NSDS will do this by facilitating linkage of, secure access to, and analysis of non-public data. The NSDS will use its coordinating role to educate data providers and data users and identify and connect interested parties pursuing related work, including facilitating approaches to meet related federal, state, and local evidence needs.

The NSDS will exhibit and champion privacy and confidentiality best practices, including promoting transparency; educating the public and other stakeholders about the value and benefits of using their data and how it is protected; and facilitating research, development, and adoption of practices and methods that enhance privacy and confidentiality.

The NSDS will identify and support high-value data and evidence-building activities. Additionally, NSDS will foster and promote data standardization to enable more efficient and high-quality linkage, access, and analysis.

Finally, the NSDS will be an instrument for innovation across the evidence-building ecosystem and will continually evaluate opportunities for ongoing improvement, including reducing burden for the public, minimizing delays for accessing timely data, and increasing transparency of the outcomes of research and evidence-building activities.

Place. As a place, NSDS will be a legally recognized entity that functions within the larger ecosystem, with hardware, software, and administrative infrastructure and capacity that allows it to meet its mission. The NSDS could be structured as a quasi- or non-governmental entity that is sponsored by a federal statistical agency with the following key attributes: (1) transparency and trust; (2) accessibility; (3) independence; (4) legal responsibilities to acquire, protect, and link data that support and enhance the value of data for evidence building; (5) scalable functionality; (6) sustainability; (7) oversight and accountability; and (8) intergovernmental support.

1.7. Resources for the National Secure Data Service

The NSDS will not succeed without proportionate investment in data producers at all levels, including federal, state, and local partners. NSDS will support capacity building (including skills and knowledge development) at the data producer level so that data can be continuously improved as they are used and analyzed. Ongoing engagement is essential to understand the limitations and advantages of different data sources, types, and methods for analysis and evidence building.

The Advisory Committee recognizes that, for a data service to be successful, legislation may be needed to provide appropriate authority, scope, and funding for a National Secure Data Service. The Committee will continue to explore how a data service can be implemented under current law.

Acknowledgments

The Office of the Under Secretary for Economic Affairs (OUSEA) at the Department of Commerce and, by extension, the Bureau of Economic Analysis (BEA) and the Census Bureau, is responsible for administering the Committee. In addition, the Committee received ongoing support from the National Center for Science and Engineering Statistics (NCSES) and the Office of Management and Budget (OMB). The Year 1 Report offers a window into this team's tireless efforts and commitment to excellence in supporting the Committee. The Committee would like to express its gratitude to the following individuals who provided invaluable assistance during the Committee's first year:

Avi Alpert, AA Strategy Consulting

Jeannine Aversa, BEA

Ryan Byrnes, BEA

Peter Fisk, BEA

Danielle Helta, BEA

Lucas Hitt, formerly BEA

Alyssa Holdren, BEA

Francise Jackson, NCSES

Colby Johnson, BEA

Gianna Marrone, BEA

Shelly Martinez, OMB

Kathryn McNamara, OMB

David Mendez, BEA

Lonna Morrow, BEA

James Plante, retired BEA

Kenneth Pond, BEA

Dondi Staunton, BEA

Duke Tran, BEA

Elizabeth "Meagan" Tydings, Census Bureau

Ryan Wist, OUSEA

2. Committee Recommendations

This section presents recommendations that have been developed, reviewed, and approved by the full Committee for immediate action by OMB. The Committee’s recommendations are actionable, relevant, and timely, covering high-priority items that support the next steps for implementing the Evidence Act and establishing and operationalizing a successful NSDS. The Committee will build its Year 2 recommendations on the recommendations presented here.

In addition, the report includes recommendations that have been approved by subcommittees and presented to the Committee at large. The focus area recommendations align with the overall vision, and the Committee will fully integrate and synthesize these recommendations in Year 2. The “Recommendation Link” boxes throughout this section show connections among the Committee recommendations, the focus area recommendations, and the Evidence Act. For more information and the findings that support these recommendations, see the “Executive Summary,” “Areas of Focus,” and “Year 2 Roadmap.”

2.1. Recommendation 1 – Evidence Act Regulations

The OMB Director should take immediate steps to promulgate draft guidance and regulations required under the Evidence Act, including reviewing and incorporating the Committee’s preliminary advice provided in this report and engaging with the Committee as part of the comment process. These expected policies include:

- Notice of Proposed Rulemaking under CIPSEA Sec. 3581 for the Presumption of Accessibility for statistical agencies and units.
- Notice of Proposed Rulemaking Under CIPSEA Sec. 3582 for expanding access to CIPSEA data assets, including data sensitivity considerations.
- Notice of Proposed Rulemaking under CIPSEA Sec. 3563 on responsibilities for statistical agencies and public trust.
- Implementation guidance for the OPEN Government Data Act (i.e., Title II of the Evidence Act) , including how agencies should implement “open data by default” and data inventories. This guidance should also provide increased clarity on interagency and intergovernmental data sharing responsibilities and expectations that comply with all relevant federal and state laws.

This guidance and regulations will provide necessary frameworks to inform the development of the NSDS and support evidence-based decision-making. The Committee will make additional recommendations to advance the implementation of the Evidence Act and to establish an NSDS in Year 2. For more information on tie-ins to the Evidence Act, see “Recommendation Link: The Evidence Act.”

Recommendation Link: The Evidence Act

As described in the “Executive Summary,” the Committee’s efforts build on the groundwork of the Evidence Act. Table 2A presents subcommittee recommendations that link directly to provisions of the Evidence Act. The Committee will continue to connect to the Evidence Act in Year 2.

Table 2A. Subcommittee Recommendations and the Evidence Act

Evidence Act	Recommendation
Presumption of Accessibility (Sec. 3581)	Legislation and Regulations Recommendations 1, 2, 3 Governance Recommendations 1 and 4
Expanding access to CIPSEA data assets (Sec. 3582)	Governance Recommendations 2 and 4 Technical Infrastructure Recommendation 1 Government Data Recommendation 4
Standard Application Process (Sec. 3583)	Other Services Recommendation 1
Responsibilities for statistical agencies and public trust (Sec. 3563)	Governance Recommendations 1 and 3

2.2. Recommendation 2 – Chief Statistician of the United States

The OMB Director should take immediate steps to designate a full-time Chief Statistician of the United States and elevate this position within OMB’s organizational structure, in alignment with Evidence Commission recommendation 5-4.

2.3. Recommendation 3 – Standard-Setting Procedures

The OMB Director and Chief Statistician of the United States should leverage existing authority under the Paperwork Reduction Act to establish a clear procedure for public and stakeholder engagement on future data standards for intergovernmental, interagency, or intra-agency data sets.

These procedures will be foundational to the development and implementation of standards and standardization for both the NSDS and stakeholders in the broader evidence-building ecosystem. The Committee will make additional recommendations on standards in Year 2. For more information on the importance of standards, see “Recommendation Link: Standards.”

Recommendation Link: Standards

The importance of standards and standardization has emerged as a common theme that cuts across the Committee’s focus areas and includes nuances around data quality, common metadata, data definitions, systems integration, technical interoperability, legal frameworks, data sharing agreements, reporting requirements, and best practices. Committee Recommendation 3 sets the stage for progress in setting different kinds of needed standards. Table 2B presents subcommittee recommendations related to developing and implementing standards. The Committee will continue to explore the issue of standards in Year 2.

Table 2B. Subcommittee Recommendations Related to Standards

Topic	Recommendation
Data standards	Legislation and Regulations Recommendation 1 Governance Recommendation 4 Technical Infrastructure Recommendation 1 Government Data Recommendation 1
Legal standards	Legislation and Regulations Recommendation 1 Governance Recommendation 4

2.4. Recommendation 4 – Appropriations Requests

The OMB Director should include specific requests for increased funding to support implementation of the Federal Data Strategy priorities and Evidence Act requirements in the fiscal year (FY) 2023 President’s Budget request to Congress.

2.5. Recommendation 5 – Value-Driven Pilot Program

The Committee recommends that the Chief Statistician of the United States, in concert with the Interagency Council on Statistical Policy, establishes a pilot program, with funding as needed, that demonstrates the value of streamlining data sharing and increasing coordination, specifically with projects that highlight cross-functional, cross-agency, and cross-governmental topics.

The pilot program should evaluate ways to ensure private information is protected while expanding research access. Projects should include federal agencies, states, and localities that already have well developed data systems and involve people who have experience in data sharing between federal, state, and local governments and have addressed such issues in the past. The program could start by building on efforts already under way on unemployment insurance data, education and workforce, and health. The Committee will use this program to help inform its recommendations in Year 2. For more information on the role of recent, ongoing, and planned evidence-building projects, see “Recommendation Link: Evidence-Building Projects.”

Recommendation Link: Evidence-Building Projects

The Committee recognizes that a National Secure Data Service will be well-situated to prioritize coordination and leadership in cross-functional, cross-agency, and cross-governmental topics. This includes supporting projects that demonstrate the value of data and evidence building as well as advancing the use of privacy-protecting techniques. Along these lines, ACDEB is exploring a variety of projects to inform its recommendations. Examples include:

Equity. With the public’s current interest in addressing disparities and reducing inequities in programs and policies, a data service could support key descriptive analytics and evaluations relevant to this priority. This could include, for example, applying data assets on income/earnings, race/ethnicity, gender, sexual orientation, or other characteristics for analysis that link existing data assets rather than collecting new sensitive data elements from the American people.

Box continues

Recommendation Link: Evidence-Building Projects *(Continued)*

COVID-19 pandemic. The global coronavirus pandemic highlighted a vast number of possibilities where better data linkage infrastructure (and data generally) could benefit decision-making for public health and economic recovery, including making it easier to reach vulnerable populations with economic assistance. A data service could supplement existing capabilities in public health and economic development agencies. This includes providing researchers a front door for access to sensitive or restricted data (as is being developed by the federal statistical system through the Standard Application Process); supporting agency needs at the federal, state, and local levels; and facilitating the use of industry and third-party data linkage.

Education and workforce. Unprecedented changes in labor markets have led to fundamental changes in skill demands. Both sets of changes underscore the need to strengthen the connection between employment services, post-secondary programs, and workforce outcomes. Building these links will help individuals decide what education paths best meet their needs and will encourage high-return investments in skills that yield long-run economic security and mobility.

Workforce/unemployment insurance (UI). Workforce data in the United States is highly decentralized, including for the UI system. In light of policy changes over the last 2 years that included pandemic benefits and extended unemployment, workforce/UI offers a compelling opportunity to pilot analytics projects with high-value, high-salience data assets.

Privacy technologies. Privacy-enhancing technologies like multiparty computation, validation servers with synthetic data, and other encrypted approaches continue to face limited acceptance given the dearth of applications. A data service could support a broad research agenda around pilot projects and demonstrations of scalable applications for privacy-enhancing technologies, including devising best practices for public administrators, lawyers, and other partners in particular projects.

Committee Recommendations 5 and 6 acknowledge the importance of evidence-building projects to inform the Committee’s recommendations and the development of the NSDS. Table 2C presents subcommittee recommendations related to evidence projects; however, as described above, the Committee will continue to explore a variety of project ideas in Year 2. For more information, see “Year 2 Roadmap” and “Appendix D. Project Inventory.”

Table 2C. Subcommittee Recommendations Related to Pilot Projects

Topic	Recommendation
Value-driven projects	Government Data Recommendation 4
Privacy-preserving technologies pilots	Legislation and Regulations Recommendation 4 Technical Infrastructure Recommendation 2

2.6. Recommendation 6 – Privacy-Preserving Technologies Case Studies

The Committee recommends that the United States Chief Statistician, in concert with the Interagency Council on Statistical Policy, publishes case studies where privacy-preserving technologies were employed, in alignment with Evidence Commission recommendation 3-2.

These case studies should identify legal barriers to be overcome and technical requirements needed to foster the widespread use of these technologies. The case studies should inform future coordination between federal, state, and local governments and help develop a framework for expanding engagements beyond the federal evidence-building system, including communicating methodological advancements. The Committee will use this report on case studies to help inform its recommendations in Year 2. For more information on the role of evidence-building projects, see “Recommendation Link: Evidence-Building Projects.”

2.7. Recommendation 7 – Communication

The Committee recommends that the Chief Statistician of the United States, in concert with the Interagency Council on Statistical Policy, develops a comprehensive communication and education strategy, including highlighting the value of data access, linkage, and analysis; the importance of collaboration; and the safety of data protection methods. The strategy should be developed with input from key stakeholders and should have a wide range of interests in mind, including those of the public; federal, state, and local policymakers in executive and legislative roles; data providers; researchers and other evidence-building partners; and data, transparency, and privacy advocates.

The Committee will provide the initial outline for this communication strategy as part of its recommendations in Year 2. For more information on the importance of communication and the Committee’s work to develop a communications strategy, see “Recommendation Link: Communication.”

Recommendation Link: Communication

Communication is a key ingredient to the success of a National Secure Data Service. Table 2D presents subcommittee recommendations related to communication. The Committee will continue to explore communications strategies, methods, and mechanisms in Year 2.

Table 2D. Subcommittee Recommendations Related to Communication

Topic	Recommendation
Transparency and trust	Governance Recommendation 2
Use assessments and communications	Technical Infrastructure Recommendation 5
Communication strategy	Other Services Recommendation 2

3. Areas of Focus

3.1. Legislation and Regulations Findings and Recommendations

This section presents recommendations on legislation and regulations, as developed, reviewed, and approved by the related subcommittee. The focus area recommendations align with the overall vision, and the Committee will fully integrate and synthesize these recommendations in Year 2. For more information, see “Areas of Focus” and “Year 2 Roadmap.”

Introduction

The legal and regulatory framework to support the use of data in government is essential for implementing ethical practices and confidentiality protections and for providing clear and consistent direction to agency staff, partners, researchers, evaluators, and analysts. The ACDEB Legislation and Regulations Subcommittee in Year 1 largely built on the foundational recommendations of the Evidence Commission. The subcommittee extended the work through fact finding related to implementation of the major titles of the Evidence Act as well as to the statistical, data governance, data analysis, and evaluation communities in government.

The subcommittee viewed its scope as focusing on four primary areas:

- ***Emphasize the impact of and clarify the need for using existing legal authorities*** that are not fully employed because regulations and guidance have not yet been developed, including residual authorities under the Evidence Act (e.g., presumption of accessibility, OPEN Government Data Act, and standard data sensitivity assessment).
- ***Identify additional guidance or regulation*** that may be needed to promote successful data sharing policies, potentially including standardized memoranda of understanding (MOUs).
- ***Weigh Evidence Commission-suggested additional authorities and statutory modifications*** based on current environment and capabilities, particularly authorities needed to support implementation of an NSDS.
- ***Consider additional proposals to ensure resourcing, support intergovernmental cooperation, promote evidence-building system efficiency,*** and advance priorities from other subcommittees.

Findings

Using Existing Legal Authorities

Implementation of the Evidence Act is ongoing across the federal government, with effective and enthusiastic implementation in many agencies. However, key guidance and regulatory actions required by law or needed for implementation have not yet been completed by OMB. Several items are currently listed in the OMB Regulatory Agenda but have not been issued as drafts according to the published timelines. The Committee has the opportunity to inform OMB policy development with its recommendations.

During multiple meetings, the Committee made references to the leadership role of the Chief Statistician of the United States for the federal evidence-building system. In fact, under the requirements of the Evidence Act, the chair of the Committee is the Chief Statistician. Despite this position being vacant, the Committee has made great progress, first under the leadership of the OMB Office of Information and Regulatory Affairs Deputy Administrator and then later the Director of the National Center for Science and Engineering Statistics at the National Science Foundation. Beyond the duties and responsibilities for chairing this Committee, the Chief Statistician should be instrumental to leading and coordinating an effective federal data ecosystem. The Chief Statistician also plays a key role in supporting government-wide data standards that, in some cases, are long overdue for updates. The role was last filled by a full-time civil servant on January 4, 2020. The extended vacancy may be contributing to the delay in publishing guidance on Evidence Act implementation.

Additional Guidance and Regulations

During the Advisory Committee's fact finding, multiple participants emphasized the challenges in coordinating legal agreements and processes for data sharing, even when law clearly permits sharing. The use of data for evidence building could be facilitated, if processes for sharing data were clear and efficient. Federal, state, and local governments and other partners would all benefit. In addition, clear guidance would also increase agency confidence in following the "open by default" direction of the OPEN Government Data Act (i.e., Title II of the Evidence Act) and support consistent interpretation of existing legal requirements and flexibilities.

Authorities and Statutory Modifications

Of the Evidence Commission's 22 recommendations from 2017, about half were addressed in the Evidence Act while other recommendations require further action. Based on discussions from the Committee, public feedback, site visits, and subcommittee discussions, many of the remaining recommendations that address legal authorities are still relevant for further action by Congress and the Executive Branch.

Resources for Data Infrastructure and Evidence-Building Activities

The Evidence Commission’s Recommendation 5-5 acknowledged a need for allocating resources to support envisioned evidence-building activities. For the most part, agencies have not received supplemental appropriations or necessary funding flexibilities to support implementation of the Evidence Act or the Federal Data Strategy. While some agencies have reallocated funding or identified other creative funding solutions for data management, statistical programs, data infrastructure, evaluation, and data analysis, other agencies are still in need of resources, including both funding and staffing. OMB units charged with supporting data management and evidence building also need to be staffed to support their expanded functions.

Resources are particularly needed for curating data.gov. Curation and quality control are critical functions for the site to provide value to its users. Likewise, resources are needed at the agency level to organize data sets, ensure data are useful and interoperable, and process data access requests.

Recommendations: Additional Guidance and Regulations

Legislation and Regulations Recommendation 1 - Model Memoranda of Understanding (MOUs)

The OMB Director, working in consultation with the Interagency Council on Statistical Policy, the Chief Data Officer Council, and the Federal Privacy Council, should facilitate the creation of updated model language for MOUs (or a shared standard for data sharing) to facilitate data sharing and linkage projects and reduce the number of MOUs required to conduct regular evaluations of programs and policies.

Legislation and Regulations Recommendation 2 - Systems of Record Notices

The OMB Director, working in consultation with the Federal Privacy Council and other relevant officials, should develop standard language for modifying Systems of Record Notices under the Privacy Act to facilitate secondary uses of administrative data when permissible by law and necessary for evidence-building activities.

Legislation and Regulations Recommendation 3 - Guidance

The OMB Director should issue a detailed memorandum to agency heads clarifying expectations that agencies use existing authorities and flexibilities to facilitate data sharing and use.

Cases-In-Point: Realizing Current Law's Vision

Federal Statistical Research Data Centers (FSRDCs)

FSRDCs address Evidence Act requirements by expanding data access and supporting transparency. FSRDCs expand access to data for evidence building through secure physical and virtual enclaves, a standard application, and by facilitating data linkages. The network offers broad U.S. coverage, multiple access modes (both in-person and virtual environments), and proposal development support. Through the FSRDC network, authorized users can link data from multiple sources, for example, linking data for households and employers from the Longitudinal Employer-Household Dynamics (LEHD) program and linking user-provided data to statistical agency data. FSRDCs abide by stringent transparency requirements, including identifying who is using the data and for what purpose, publishing the number of ongoing projects and by what agency, and developing an inventory of project metadata. The network recently hired a Knowledge Transfer Officer to help with these endeavors.

United States Department of Agriculture (USDA) Enterprise Data Analytics Platform and Toolset (EDAPT)

EDAPT enables statistical agency data access. The USDA data strategy and infrastructure help bridge the gap between administrative and statistical agencies by creating an environment where data, tools, and computing power can be shared on a common platform to resolve major historical challenges. An example of this is the Integrated Modeling and Geospatial Estimation System (IMAGES) project where the USDA Chief Data Officer and EDAPT worked with USDA's National Agriculture Statistics Service (NASS) to produce timelier indications of planted and prevent plant acreages. EDAPT provided the IT infrastructure, tools, and cloud computing capacity NASS was lacking to use all available, useful data simultaneously. As a result, IMAGES has increased the use and value of data on hand, increased the coverage of agricultural production, resulted in process efficiencies, enhanced analytical capabilities, and added value to the workforce.

U.S. Chamber of Commerce Foundation's Jobs and Employment Data Exchange (JEDx)

JEDx is seeking to leverage existing policies to improve the quality of workforce data. There is OMB guidance that directs agencies to participate in standard-setting bodies that transcend government, making it feasible for external stakeholders to inform federal agencies on those definitions. Most employers use human resources vendors like ADP, so these vendors are key participants who can standardize the definitions in their systems, allowing employers to report to vendors, and then those vendors can submit data to meet multiple reporting and compliance requirements.

For comprehensive descriptions of these programs, see Appendix E.

Recommendations: Authorities and Statutory Modifications

Legislation and Regulations Recommendation 4 - Evidence Commission Proposals

The OMB Director should develop legislative proposals for Congress or regulatory actions to consider in implementing the remaining Evidence Commission recommendations, as recommended by the Committee, including the following:

- Recommendation 2-6 from the Evidence Commission encouraged expanded access to income and earnings data already acquired by federal agencies to facilitate evidence-building activities. Using legislation to expand access to the National Directory of New Hires for research and evaluation is one low-burden, high-value strategy for advancing access to priority data assets.
- Recommendation 3-2 from the Evidence Commission encouraged innovation for privacy-preserving technologies. A legislative proposal to test and pilot emerging approaches such as multi-party computation would support the scaling and future adoption of new approaches for protecting data, which could include consideration of safe harbor provisions for those testing new approaches.

The Committee will continue to evaluate the Evidence Commission report and make recommendations on related legislative action in Year 2. For more information, see “Year 2 Roadmap.”

Recommendations: Resources for Data Infrastructure and Evidence-Building Activities

Legislation and Regulations Recommendation 5 - Funding Flexibilities

In addition to any direct appropriations, the OMB Director should propose legislative flexibilities for facilitating funding set-asides for data infrastructure and analysis activities, recognizing these activities are core functions of government.

Year 2 Focus

The Committee will continue evaluating potential legal and regulatory requirements for an NSDS based on ACDEB discourse, pilots, and public input. This will include further exploration of more detailed Evidence Commission recommendations where fact finding is incomplete or ongoing (e.g., Title 13 and Title 26 authorities, state sharing expectations for federally funded programs, etc.). The Committee will also further identify specific areas where changes to legislative language would support the success of the NSDS.

3.2. Governance, Transparency, and Accountability Findings and Recommendations

This section presents recommendations on governance, transparency, and accountability, as developed, reviewed, and approved by the related subcommittee. The focus area recommendations align with the overall vision, and the Committee will fully integrate and synthesize these recommendations in Year 2. For more information, see “Areas of Focus” and “Year 2 Roadmap.”

Introduction

A governance structure for a secure data service that is national in scope is required to execute on the Committee’s vision for the National Secure Data Service (NSDS). The ACDEB Governance Subcommittee in Year 1 evaluated potential approaches as recommended by the past efforts of the Evidence Commission and subsequent research.

The Committee’s initial findings and recommendations lay out the frame for a governance approach that supports the Committee’s vision of an NSDS. The NSDS could be structured as a quasi- or non-governmental entity that is sponsored by a federal statistical agency with the following key attributes: (1) transparency and trust; (2) accessibility; (3) independence; (4) legal responsibilities to acquire, protect, and link data that support and enhance the value of data for evidence building; (5) scalable functionality; (6) sustainability; (7) oversight and accountability; and (8) intergovernmental support. Across the next year the Committee will develop more detailed findings and specific recommendations to operationalize that approach.

Findings

The Evidence Commission envisioned the NSDS as an independent statistical agency housed within the Department of Commerce. A recent influential report, [Hart and Potok](#) (2020), which took stock of the substantial institutional changes and the changing data landscape since the Commission report, identified key attributes and functions the NSDS should include. Building on the Hart-Potok report and other inputs, the subcommittee identified the following attributes and functions that are critical for a successful NSDS:

1. Transparency and trust
2. Accessibility
3. Independence
4. Legal responsibilities to acquire, protect, and link data that support and enhance the value of data for evidence building
5. Scalable functionality
6. Sustainability
7. Oversight and accountability
8. Intergovernmental support

Each of these items is discussed further below in relation to governance issues.

Transparency and Trust

To maintain the public's trust, the NSDS must be transparent about what data are being used and in which projects, provide fair access to qualified researchers, share its policies, and take steps to inform the public about the value it is providing. In addition, the NSDS requires strong security and privacy protection policies. Protection of privacy should include cybersecurity protections, procedures for maintaining the confidentiality of data, and appropriate use of innovative cutting-edge methods to prevent re-identification of individuals in sensitive data sets. The NSDS would need to implement tiered access to sensitive data that conforms to OMB guidance that was mandated by the Evidence Act as well as the privacy regimes of the statistical agencies originating the data.

In addition, the NSDS should leverage ongoing efforts to develop a Standard Application Process that includes a publicly accessible portal for sharing information about current and past evidence-building projects and the value they are providing to the public. The portal should include information on policies and procedures, links to data inventories such as data.gov, and how to apply to access data. Other information to assist researchers could include consistent metadata, commentary on the quality of data and fitness for various uses, and information on other similar research projects possibly using artificial intelligence and/or machine learning approaches.

Accessibility

Accessibility is a key accountability principle. The NSDS will fail to live up to its mission if it exists merely in the realm of case study, accessible only to the savviest of data users. Accessibility embraces plain language, diverse stakeholder engagement in governance, and ongoing assessments of timeliness in data availability, access, and use. This includes continuous improvement to data suitability across tiers of access and its fitness for various uses.

The NSDS must explicitly promote equitable access through its processes and protocols. As the access involves tiers of controls related to the sensitivity of the data, so too should the corresponding burden on users exist on a continuum. Transparency to burden of access includes a clear and cost-neutral fee structure. Other, non-monetary clearance processes and protocols should also follow a corresponding stepwise security framework. Further, for non-technical audiences, the NSDS should offer some accessible utility, such as maps, graphs, and other visual representations of underlying data, perhaps derived as artifacts from research using the NSDS.

Independence

To be successful, the NSDS must set strategic priorities and operate apart from political influence, while still coordinating closely with federal agencies. It would need to support objective analyses and ensure that data are used only for approved statistical purposes. The NSDS would be able to support government-wide, cross-cutting agency priorities, rather than serving primarily the mission of one agency. An advisory board and the Interagency Council on Statistical Policy should have a role in setting priorities.

Legal Responsibility To Acquire, Protect, and Link Data

The NSDS’s legal authority must cascade from the Evidence Act’s reauthorization of the Confidential Information and Statistical Efficiency Act (CIPSEA). This authority allows statistical agencies, and their designated agents, to access data for evidence-building purposes. In addition, the NSDS needs to comply with supporting statutes and regulations, which can vary widely by agency, so the oversight and audit mechanism of the NSDS needs to assure compliance is being monitored and there are accountability feedback loops with the statistical system and the contract management process. Further, as new regulatory and/or statutory adjustments are potentially needed in the future, these feedback loops must also include formal documentation to OMB about the barriers and opportunities for future development.

NSDS could lead efforts to standardize how agencies adopt frameworks for data security and privacy. NSDS could also support the use of standardized legal agreements or Memoranda of Understanding (MOUs) for data sharing between federal agencies, facilitated by forthcoming OMB regulations on the presumption of accessibility under the Evidence Act that will apply government-wide. The amount of time it currently takes for agencies to negotiate data sharing agreements can slow down projects by years and is a major barrier to evidence-building activities.

Scalable Functionality

Scalability requires that the policy and programmatic oversight capabilities are sufficiently staffed to prioritize the fulfillment of the NSDS mission, as well as the IT architecture enabling cost-effective expansion without substantial capital investment as demand grows. This includes dedicated policy and legal resources to draft and administer regulations. In addition, qualified staff need to be scaled up in a timely way as work demands increase.

It will be crucial for the NSDS to hire and retain personnel who understand data and computer science, research needs, IT architecture, and cybersecurity. It will also be important to put in place processes that prevent the NSDS from being a bottleneck due to increased demand without increased capacity. The approval process for research projects, for example, should be a streamlined process that gets answers back to researchers in a reasonable timeframe and allows them to trace the progress of their proposals through this process. This will require a management structure with feedback loops from agencies and researchers. It also requires a restructuring of existing processes that have serial rather than parallel or group procedures for project approval.

Sustainability

Stable funding is necessary to assure continuity, oversight, and the ability to meet future demands. The NSDS should receive funding from fees charged to agencies and direct appropriations. Fees should also be assessed to outside researchers; however, these fees should not be prohibitive for researchers who may not have large funding grants, particularly at the start of their careers. The NSDS should work closely with users to establish a schedule of fees and develop requests for appropriations that are well documented and can be audited by the Government Accountability Office and other relevant parties.

An important governance element is the ability to pivot to another governance or operational model or approach if the current one is not meeting its mission. Having some funding coming from appropriations allows a measure of congressional oversight that would be more difficult to achieve if all funding came from fees or from outside government.

Oversight and Accountability

To ensure accountability to its diverse stakeholders, the NSDS would require a steering or advisory board with representation from various stakeholder groups, including the public, privacy advocates, and researchers. There should be representation from communities that are being studied with personal data they have provided to the federal or state governments. This board must be empowered to provide guidance and oversight on the NSDS's policies and procedures.

Intergovernmental Support

Sub-national data are critical when assessing the outcomes achieved by federal programs. The NSDS would need the ability to coordinate across governmental boundaries, bringing in state and local governments as research partners for federal agencies. In addition, the NSDS may need to comply with state rules on accessing state data or establish standards to which state and local partners will need to adapt their statutes and regulations. Therefore, it would be important to have state representation on the advisory board.

Case-In-Point: Governance Facilitates Evidence-Building Programs

The Coleridge Initiative's Midwest Collaborative (MWC)

Regional collaboratives of state agencies are mitigating the legal and privacy barriers to sharing state data across state lines by using new technologies and demonstrating the value of better data and evidence for policy. MWC's focus is practical in nature; states invest millions of dollars in education and training to create high-wage jobs for their citizens yet the returns to those investments can only be seen if the graduates stay in the state. Initial examples, made possible by new secure platform technologies, have been very informative. The lessons learned made it clear that an organizational structure could empower states to determine common regional metrics and produce regional dashboards and portals as springboards for evidence. There are now limitless possibilities for developing common research agendas and linking needed data across state lines.

Developing the governance to coordinate the use of state-driven data models in the formulation of a national evidence-based agenda is the clear next step. In that context, MWC is establishing a formal governance structure guided by a set of explicit goals:

- Facilitate inter-state collaboration on data
- Define a state-led data analytics infrastructure
- Establish a professional development curriculum
- Build production-level technical capacity
- Develop process for collective use of research and evaluation data
- Inform and shape the national agenda

Experience has shown a focus on the value proposition effectively strengthens the rooting of evidence-based practices at local levels and establishes the rationale for the Collaborative. In addition, it was particularly important for the Collaborative to develop a single agreement specifying conditions for states to opt in or out of certain streamlined data protections and formal bylaws detailing operations and oversight. Finally, it was important to include two organizations with established subject matter expertise to serve in a consultative role as well as offering mature national experience supporting its vision for networking with other regional collaboratives on national agenda.

As such, MWC has four components in its governance leadership structure: a policy council, data stewardship board, administering organization, and platform organization. State representatives from the policy council and stewardship board serve on the Executive Committee that exercises final approval on all policy recommendations and project proposals. The Administering organization and Platform organization serve in a supportive, advisory role.

Box continues

Case-In-Point: Governance Facilitates Evidence-Building Programs *(Continued)*

U.S. Chamber of Commerce Foundation’s Jobs and Employment Data Exchange (JEDx)

JEDx is using a transparent and open process to engage experts from around the public and private sectors. The foundation has formed an advisory committee consisting of over 50 public and private members representing government, employers, human resources (HR) technology companies, and other stakeholders to inform the JEDx program. In addition, the foundation is establishing a National Leadership Team to provide guidance in developing the public-private approach for improving federal and state reporting, starting with state UI reporting and the use of data for public-private workforce analytics.

For comprehensive descriptions of these programs, see Appendix E.

Recommendations

Governance Recommendation 1 - Guiding Governance Framework

The governance structure of the NSDS should address the eight identified elements of a successful entity.

Governance Recommendation 2 - Transparency and Trust

A governance structure for the data service should be organized to support radical transparency. The first step in this is to build a strong program of data stewardship that has the goals of providing maximum access to data while assuring maximum protections for privacy and confidentiality and preventing data breaches. The governance structure must promote transparency by: (1) demonstrating accountability to the public about how the data are being used and the evidence being built, (2) documenting secure data linkage methods, (3) facilitating secure data access and analysis, and (4) advancing the use of privacy enhancements while maintaining interoperability of the high-value data.

Governance Recommendation 3 - Independence

The governance structure should assure the independence of the NSDS but maintain strong accountability for efficient, high-quality operations and responsiveness to cross-agency policy priorities.

Governance Recommendation 4 - Legal and Regulatory Authority

The governance structure should enable the NSDS to maintain standard legal agreements, metadata, and other elements where lack of standard approaches has led to delays and barriers to evidence building.

Year 2 Focus

The Committee will continue to develop the details of governance and present more comprehensive recommendations across the Committee’s second year and in its final report.

With a specific focus in Year 2 to advance measurable progress, the Committee will identify administrative obstacles that can be resolved by a group or agency positioned to have a “whole of government” view of relevant issues, including the development of a robust catalog or repository of standards documents, inter-agency agreements, and data sharing policies and procedures. These activities would be pursued as an intermediary step toward establishing common data and metadata standards; common language related to data collection, storage, and sharing (e.g., research participant consent forms, Systems of Records Notices, data sharing agreements); and broadly accepted policies and procedures for data sharing and linkage.

A particular area of focus will include identifying the best means for the NSDS to support a primary coordinating function with a research function. This will address how to coordinate across federal, state, and local entities while facilitating and promoting research and advances. Additionally, the Committee will further develop potential mechanisms to incentivize reporting on federally sponsored programs.

3.3. Technical Infrastructure Findings and Recommendations

This section presents recommendations on technical infrastructure, as developed, reviewed, and approved by the related subcommittee. The focus area recommendations align with the overall vision, and the Committee will fully integrate and synthesize these recommendations in Year 2. For more information, see “Areas of Focus” and “Year 2 Roadmap.”

Introduction

For data to inform evidence building, data sets need to be findable, definable, and extractable, while protecting the privacy of data subjects. The ACDEB Technical Infrastructure Subcommittee is exploring necessary structural and functional characteristics for this to happen across the broad data ecosystem that includes the NSDS, federal evidence-building system, state and local governments, and research communities that interact with it.

Technical infrastructure in the NSDS and the data ecosystem at large need to support data contributors and users. The NSDS itself must integrate with state and local government consortia, existing university-based research networks, and the federal evidence-building system. Given the variety of data needed for evidence building, this ecosystem must provide tiered access. For more information, see “Governance, Transparency, and Accountability Findings and Recommendations.”

Findings

The Committee recognizes that the focus of data in evidence building extends beyond federal statistical agencies and even the federal government. Therefore, the federal statistical ecosystem, including the NSDS, should strive to provide interoperability across platforms for data gatherers and users both within and external to the federal government. Shared data standards are needed to enable meaningful data linkages that lead to helpful data sets for users with a goal of reducing the data cleaning and reformatting needed to bring data together.

The NSDS must assist potential data partners to increase their capacity for organizing and sharing administrative and program data. This includes:

- Infrastructure for accessing data across what may be stove-piped divisions or agencies within a department or state
- Subject matter expertise to create or improve metadata where there may be none suitable for supporting research or analytics and to implement data standards
- Record linkage capacity when large administrative data systems are used to augment data collected from other sources
- Infrastructure to contribute directly to federal, state, or local dashboards or reporting without having to physically share data
- Infrastructure to securely transmit potentially large data to the NSDS or other data users’ infrastructure under appropriate conditions
- Concierge staff to facilitate data sharing

Technical infrastructure must address privacy issues for data providers as well as data subjects. Agencies sharing data within the ecosystem need an NSDS that coordinates access to best practices when data need to be moved securely, and to build capacity to link and analyze data in situ to produce aggregate outcomes without unnecessarily increasing disclosure risk for private data.

The United Nations Global Working Group Task Team on Privacy Preservation Techniques has recently recommended the use of these technologies by government statistical agencies. Recent efforts to conduct privacy-preserving linkages and secure multiparty computation have been featured in demonstration projects (e.g., with the [National Center for Health Statistics](#) and the [National Center for Education Statistics](#)). Innovation in this area, however, has been fragmented, illustrating both the need and promise for an NSDS that coordinates efforts and expedites research and development.

NSDS must also coordinate the use of disclosure avoidance approaches that aligns with the data being analyzed and published to protect privacy of data subjects. While nearly all of the major technology companies use differential privacy internally in some of their products and several of them even have open-source libraries (e.g., Amazon, Apple, Google, Microsoft), not all data sources or analyses will require formally private outputs. NSDS can coordinate indexes of methods and standards offering suitable privacy protections, depending on the tier of access needed.

Recommendations

Technical Infrastructure Recommendation 1 - Secure and Efficient Data Access

NSDS must coordinate secure and efficient data access. As such, the NSDS should provide services to move data for evidence building safely, easily, and efficiently to users with various tiers of access, including transfers by an intermediary or through secure enclaves. The NSDS must pursue development of data standards that facilitate data interoperability. The NSDS must ensure that data are accompanied with standard metadata, providing technical assistance and tools to expedite metadata production as needed.

Technical Infrastructure Recommendation 2 - Privacy-Preserving Technologies

NSDS must use and promote privacy-preserving technologies. The case studies described in Committee Recommendation 6 will form the basis for additional recommendations on the NSDS's role in enabling regular, ongoing research and development in privacy-preserving methods. Areas of focus will include secure multiparty computation and fully homomorphic encryption, privacy preserving record linkage, tools to automate metadata production and provenance tracking, trusted execution enclaves, and producing synthetic data with validation servers. Once the NSDS is established, it will support experiments and competitions aimed at developing new methods that increase secure access to data.

Technical Infrastructure Recommendation 3 - Data Linkage Capacity

The NSDS must expand data linkage capacity. The NSDS should build on existing work, including the Federal Chief Data Officer (CDO) Council efforts to index data inventories and metadata repositories and the ongoing development of the Standard Application Process. In addition, an aspect of the case studies outlined in Committee Recommendation 6 should be evaluating the availability and formats of tokens for linkages.

Once established, the NSDS data concierge service should coordinate with the Federal CDO Council and statistical agencies to advance efforts to index data inventories and metadata repositories. The data concierge service should also coordinate with state and local officials seeking linkage services. These efforts will support data linkage, one of the central roles for the NSDS.

In addition, the NSDS must build capacity to link data from different sources reliably, whether in situ via Privacy Preserving Record Linkage, by an intermediary, or in an enclave. To do so, the NSDS must promote standards that streamline data linkages (e.g., consistent formatting of Social Security Numbers).

Cases-In-Point: Creating Capacity, Access, and Capability Through Technology

USDA Enterprise Data Analytics Platform and Toolset (EDAPT)

One of EDAPT's key successes is creating technical capacity for its customers across the department. To address USDA's varied needs, the department brought new technologies online that enabled the department to move from descriptive to predictive analysis and to address even bigger challenges. Aspects of this toolkit include:

- **Capabilities and governance.** The department introduced advanced analytical techniques, like artificial intelligence, machine learning, and natural language process.
- **Data management.** USDA implemented a governance process for cataloging data and standardizing analytics tools.
- **Open Data Platform.** The department implemented USDA's Open Data Platform, enabling the department to publish dashboards that provide the public and third-party authenticated users with the ability to draw data-driven insights, as well as download data about USDA programs.

Box continues

Cases-In-Point: Creating Capacity, Access, and Capability Through Technology *(Continued)*

The Coleridge Initiative's Administrative Data Research Facility (ADRF) and Applied Data Analytics Program

The ADRF, a 2018 Government Innovation Award winner, is a secure cloud-based computing platform designed to promote collaboration, facilitate documentation, and provide information about data use to the agencies that own the data. The platform was established by the Census Bureau with funding identified, in part, to inform decision-making of the Evidence Commission. It has enabled secure access to over 150 confidential data sets owned by more than 75 different agencies at all levels of government. The platform's data processing and management capabilities include data ingestion, data documentation, data analytics tools, and data stewardship. All access is exclusively via web browser over secure connections, and the system leverages cloud-based services for efficiency, reusability, and cost transparency. The Coleridge Initiative's Applied Data Analytics Training programs have trained over 800 government employees on how to work with confidential data (within the ADRF) to produce evidence-based research that informs policy.

National Center for Health Statistics (NCHS) Data Modernization Activities

NCHS deploys a tiered access model to expand access and use of data assets which includes:

- ***Open data assets.*** NCHS provides access to public-use data files.
- ***Web-based query system.*** Users can create tabular data views via CDC's WONDER portal.
- ***Physical data enclaves.*** NCHS operates four data centers and is part of the national FSRDC network.
- ***Virtual data enclave (VDE).*** As a forthcoming addition to the tiered access framework, VDE supports implementation of the Evidence Act, adds another access tier, disseminates data to wider audiences, eliminates barriers that exist with physical location, decreases costs for researchers, and expands access to confidential data to a whole new class of researchers.

For comprehensive descriptions of these programs, see Appendix E. Cases-In-Point: Creating Capacity, Access, and Capability Through Technology

Technical Infrastructure Recommendation 4 - Innovation Sandbox

NSDS should feature a sandbox for testing new and innovative technologies and software for multiple data access tiers, data protection protocols, and data analysis. The NSDS should be a neutral ground between agencies where secure testing of new data linkages, privacy-preserving technologies, and model approaches can occur.

This sandbox should support the development and refinement of disclosure limitation techniques and multiple access modes, in line with existing efforts by the Interagency Council on Statistical Policy and the Federal Committee on Statistical Methodology.

Technical Infrastructure Recommendation 5 - Fitness for Use Assessments and Communications

NSDS must support fitness for use assessments and communications. To accomplish this, there must be an investment in the NSDS infrastructure that allows for both assessing and publicly communicating the quality of data prior to and after evidence building. In terms of data capacity and readiness, huge variation exists across agencies at the federal, state, and local levels. NSDS must meet data providers where they are, providing a secure environment and tools to assess and monitor data quality, and to report out on data fitness for use and planned uses.

Consistently communicating the quality and utility of data is critical to build social license, and to have public support for the use of administrative data in evidence building. As the Evidence Commission noted, radical transparency will be critical for NSDS success. The NSDS should build on the significant work already underway by statistical agencies in this space.

Year 2 Focus

Following on these initial recommendations, the Committee is going to dive deeper into potential interoperability standards and technologies with a goal to issue more specific recommendations across the next year. The Committee will explore potential tools and methods for increasing provider and user capacity and capability, particularly privacy-preserving technologies and automating data quality checks, documentation, and harmonization for data providers. The Committee anticipates soliciting public input and exploring the development of communities of practice to support and inform the efforts.

The Committee will also identify recommendations for the specific functional and technological requirements for executing on its vision for the NSDS, as discussed elsewhere in this report, reviewing past studies and work under way in the government and the private sector.

While privacy-preserving techniques are not applicable in all scenarios, the Committee intends to explore current feasibility, highlighting current uses (e.g., [Boston wage gap study](#) and a [recent pilot](#) with the National Center for Education Statistics). The U.S. federal government already invests heavily on basic research in privacy-preserving techniques through grant programs by the National Science Foundation, Defense Advanced Research Projects Agency, Intelligence Advanced Research Projects Activity, and other agencies. Statistics applications involving human-scale data (data sets of hundreds of millions of rows) are largely feasible with secure multiparty computation today. However, some operations are still challenging, including scaling up to “big data” or to a substantial number of computing parties, or neural networks with many layers or deep learning algorithms. The NSDS must advance privacy-preserving technologies where currently feasible and support the evolving science to further innovate evidence building on private data.

3.4. Government Data for Evidence Building Findings and Recommendations

This section presents recommendations on government data for evidence building, as developed, reviewed, and approved by the related subcommittee. The focus area recommendations align with the overall vision, and the Committee will fully integrate and synthesize these recommendations in Year 2. For more information, see “Areas of Focus” and “Year 2 Roadmap.”

Introduction

Data for evidence building must be high quality and available when users need it. The ACDEB Government Data for Evidence Building Subcommittee in Year 1 explored ways the federal government in general, and NSDS in particular, can help ensure quality data are available for evidence building.

Data quality starts at the data source. Often the data source is outside the federal government, at the state and local level. Thus, to ensure that the federal government has the high-quality data it needs, federal agencies should support building capacity and capability among state and local entities.

Data for evaluation often needs to be granular by county or Census tract to see the impact of programming. For performance measurement, it also needs to be current and timely to monitor progress.

Additionally, evidence building is enhanced when the data are available to staff at all levels of government. So, while the federal government must work with state and local governments to provide quality data, it also needs to ensure that state and local officials have access to the data once they go up to the federal databases. Beyond government users, the business and academic research community needs timely access to high-quality data to develop insights and analytic tools that benefit citizens, policymakers, and the private sector.

Findings

State and local data management operates under distinct regulatory and legal regimes. Until recently, only a few visionaries have been interested in harmonizing these regimes. What’s more, federal agencies use inconsistent timelines and requirements for their data that are collected by state and local entities. This increases the burden of data gathering and is an impediment to combining and linking data sets. Differences as simple as inconsistent fiscal and program year cycles create complex challenges that are not easily addressed.

The Committee’s field trips included demonstrations of ongoing efforts by some states to build data services and improve evidence. This exceptional work has been rewarded with enhanced capability to answer questions about the individual states. However, except for efforts like the Coleridge Initiative and the [Multistate Longitudinal Data Exchange](#) run by the Western Interstate Commission for Higher Education, there is minimal coordination among these efforts.

With the possible exceptions of the [Longitudinal Employer-Household Dynamics](#) (LEHD) database, which includes state unemployment insurance data (though not without significant challenges), and the [vital records program](#) between the National Center for Health Statistics and state vital registrars, coordination between the federal government and the states has been minimal. States continue to struggle to get federal agencies to share data that state and local governments provide to federal agencies, and federal agencies have failed to lead in setting standards that would make the data easier to combine and share.

Recommendations

Government Data Recommendation 1 – Standardization Across Levels of Government

OMB should explore standardizing elements of federal data collection across agencies such as harmonizing reporting timelines and data structures. State and local governments should be incentivized to adopt the federal standards.

Government Data Recommendation 2 – Coordination Across Levels of Government

One important role of the NSDS should be helping to coordinate evidence-building efforts across federal, state, and local government agencies as well as with the research community. As part of this role, the NSDS should coordinate with federal, state, and local government partners to share data on federal programs with the entities who are responsible for administering federal programs and for providing data to the federal government. These data regularly flow “up” to the federal government but do not flow back “down” to the data providers for their own evidence-based decision-making.

Government Data Recommendation 3 – Capacity Building

A focus of the NSDS should be capacity building, particularly helping government users at all levels to develop program expertise and data science skills. This should include data privacy expertise.

Government Data Recommendation 4 – Unemployment Insurance Pilot

Building on the pilot program described in Committee Recommendation 5, the federal government should explore pilots involving the transfer of state data to the federal government to validate capabilities, illuminate current challenges, and illustrate the potential value of data sharing and coordination. These projects could draw on existing federal, state, and local government experiences or, with funding, explore new areas.

Pilots around unemployment insurance data should be a high priority. These pilots could integrate approaches from the physical sciences perspective and involve federal agencies and participating states that are looking to broaden permitted uses of unemployment insurance data provided for the Longitudinal Employer-Household Dynamics product.

Cases-In-Point: Enabling State/Federal Partnership

South Carolina's Revenue and Fiscal Affairs Office (RFA)

As RFA seeks to build a dynamic and engaged culture around using data and evidence to solve issues for South Carolina citizens, it has encountered challenges harnessing the value of federal data. There are many challenges states must overcome to access federal data resources that would provide key insights (e.g., the National Directory of New Hires and federal wage data). To improve data sharing for evidence building, there should be a better two-way flow of data from the states to the federal government and vice versa. This process should be driven by collaboration between state and federal actors rather than reporting requirements, as is currently often the case.

Federal Statistical Research Data Centers (FSRDCs)

FSRDCs have agreements with many states (for example, through the LEHD program) and with agencies who provide state-level data (like Supplemental Nutrition Assistance Program benefits). While the network is working to bring these data into the program, it is complicated because there are many federal and state-level requirements for access and review of outputs. For example, LEHD is a partnership between the Census Bureau and 47 or 48 active partner states. Around 20 of these states allow researchers to access their data if they meet Census Bureau requirements, and the remaining states reserve the right to review the research proposal. Researchers regularly gain access to data from about 30 states. The program is working to strengthen these relationships and improve access to state data sets. The Standard Application Process currently being developed could also be expanded to include data from the states or any other provider.

Coleridge Initiative's Applied Data Analytics Program

The Coleridge Initiative has partnered with almost 30 states and the Department of Labor's Employment and Training Administration in a scalable training program that uses unemployment insurance claims data to produce unemployment to reemployment dashboards. Coleridge has also partnered with the Department of Health and Human Services and eight states in a scalable training program to produce new measures of employment for welfare recipients.

Box continues

Cases-In-Point: Enabling State/Federal Partnership *(Continued)*

National Center for Health Statistics (NCHS) Data Modernization Activities

NCHS leverages federal, state, and local partnerships to improve data quality and timeliness. The National Vital Statistics System (NVSS) is the oldest and most successful example of inter-governmental data sharing in the public health realm and is the mechanism by which NCHS collects and disseminates the nation’s vital statistics, including birth certificates, death certificates, and fetal death reports. NVSS is a decentralized, cooperative system comprising 57 jurisdictions: 50 states, the District of Columbia, New York City, and 5 territories. NCHS provides coordination for the program, including standard certificates and forms, instruction and coding manuals, training and instructional materials, and model law to govern data collection. NCHS purchases vital records from the jurisdictions through the Vital Statistics Cooperative Program contracts. There has been some funding over the last decade to help develop state systems, including work on systems interoperability and data standards. More recently, the CARES Act provided \$77 million for NVSS modernization—all 57 jurisdictions will receive funding to work toward required activities. These awards will help the recipients move closer to providing data to NCHS using common standards.

For comprehensive descriptions of these programs, see Appendix E.

Year 2 Focus

The Committee will identify specific areas where additional OMB guidance would be helpful without increasing unfunded burdens on state and local data collectors. Additionally, the Committee is working to identify best practices and model approaches that can be published and disseminated across federal, state, and local levels.

The Committee is working to identify case studies of ongoing research designed to answer questions that are important to multiple levels of government and use both federal and state data. The Committee will continue to explore the challenges and opportunities of sharing data sets on income, employment, education, and health across levels of government and linking those data sets to business data, like sales or investment metrics, that are used in calculating economic indicators.

3.5. Other Services and Capacity-Building Opportunities Findings and Recommendations

This section presents recommendations on other services and capacity-building opportunities, as developed, reviewed, and approved by the related subcommittee. The focus area recommendations align with the overall vision, and the Committee will fully integrate and synthesize these recommendations in Year 2. For more information, see “Areas of Focus” and “Year 2 Roadmap.”

Introduction

The NSDS must be useful and used for it to fulfill the promise envisioned by the Evidence Commission, the Evidence Act, and the members of this Committee. The ACDEB Other Services and Capacity-Building Opportunities Subcommittee in Year 1 considered what may be needed to ensure NSDS users—be they federal, state, local, or other authorized entities—have a thorough understanding of the utility of an NSDS and then can make the best possible use of the data service’s potential for secure and privacy-protecting evidence building, regardless of their existing analytic capacity.

The subcommittee focused on two foundational components necessary to achieve that goal:

- **Awareness:** Approaches to communicating about the NSDS with citizens, policymakers at all levels of government, and researchers focused on the data service’s potential value proposition for each group; and
- **Capability:** Providing technical assistance to qualified researchers, data stewards, and others in federal, state, and local government so they can fully leverage the NSDS to build evidence.

Findings

The subcommittee found that potential users of an NSDS are likely to face two key barriers to making the best use of its services:

- Awareness of the opportunities the NSDS brings to improving policymaking at all levels of government, and
- The technical capability and resources needed to navigate and use the available services.

The subcommittee recognizes that to address these barriers, the NSDS must plan and implement a variety of tailored approaches to communicating about the NSDS with citizens, policymakers at all levels of government, and researchers. These approaches will need to facilitate two-way communication to ensure that stakeholders are aware of the data service’s value and that the NSDS can gather and act on stakeholder input.

Cases-In-Point: Increasing Partners' Capacity

Federal Statistical Research Data Centers (FSRDCs)

FSRDCs are an example of a successful partnership to expand the federal government's capacity to facilitate external researcher access to data for statistical purposes. Federal statistical agencies collaborate with host organizations—including universities, non-profit research institutions, and government agencies—to enable approved researchers to access confidential data from multiple agencies through a network of secure data enclaves. Partnerships have been a key element of their success. Currently, the network includes 100+ universities and research institutions, 31 physical data enclaves each with its own Executive Director, 7 statistical agency partners, and nearly 500 research projects.

Regular and frequent communication is key to maintaining these relationships. The program management office meets with Executive Directors every week and stays in regular contact with agency partners through liaisons with each agency. In addition, the Executive Committee meets monthly. Meetings with these partners and stakeholder groups are a mechanism for two-way communication to discuss ongoing activities and share needs and concerns.

South Carolina's Revenue and Fiscal Affairs Office (RFA)

The RFA increases capacity for contributors by concentrating data assets and skills, developing a data culture, and providing them focused expertise. The RFA helps develop a data culture to champion research and data use for evaluation. RFA efforts cover a broad spectrum from supporting basic research to identifying best practices—these activities seek to create a robust environment for asking questions and using data to answer them. Finally, the RFA builds subject matter knowledge to improve data, processes, and results. RFA statisticians oversee sets of agencies, learn about programs, and become internal data experts. Staff uses this knowledge to complete quality control checks, identify issues, and discuss improvements to data and processes with the source agencies. The more substantive knowledge the staff has, the better the exchanges with the source agencies and the results.

The Coleridge Initiative's Applied Data Analytics Program

The Applied Data Analytics program is a project-focused learning approach designed to train government employees and public policy analysts on how to tackle important policy problems by applying modern data analysis tools to their own confidential data. Agency staff are trained through direct use of their data to answer real, present policy questions they face and to develop practical tools after the training ends. A primary goal of the training is also to facilitate the establishment of a community of practice, by building relationships among and between agencies, universities, and non-profits. Since 2017, this program has partnered with over a dozen top universities and organizations to provide professional development training to over 750 participants across more than 250 organizations.

For comprehensive descriptions of these programs, see Appendix E.

Additionally, the NSDS will need to consider ways to provide training, coaching, and technical support and assistance to users. The technical assistance could include:

- Completing administrative processes required to gain access to administrative data,
- Incorporating (or integrating) those data into the NSDS ecosystem,
- Analyzing linked or linkable data in secure, privacy-preserving ways, and
- Protecting privacy, including tools to analyze risk associated with releasing de-identified or synthetic versions of confidential data.

Recommendations

Other Services Recommendation 1 - Data Concierge

The NSDS should include a data concierge service to facilitate technical assistance and help identify data sources for users. Data concierges would serve as “librarians” who have a comprehensive awareness of the universe of data available through the NSDS. Concierges would also be familiar with associated metadata that positions them to assist researchers in determining whether a given data set is likely to address their proposed research questions. Concierges would be supported by agency-based subject matter experts who can ensure that potential users are connected to the data most appropriate for their needs and who can provide more detailed support in data use.

The concierge service could also have a role in matchmaking stakeholders who have questions researchable with the NSDS but who are lacking analytic capacity or other necessary resources to external researchers who could support their work. Concierges could also build communities of practice around specific topics, data sets, analytic techniques, or other areas of common interest.

Other Services Recommendation 2 - Communication Strategy

Building on the communication strategy described in Committee Recommendation 7, NSDS should maintain, update, and execute a comprehensive communication strategy around the benefits of a robust NSDS. The strategy should be responsive to the interests of wide range of stakeholders including: the public; federal, state, and local policymakers in executive and legislative roles; data providers; researchers and other evidence-building partners; and data, transparency, and privacy advocates.

For each group, potential areas of emphasis could include:

- ***Policymakers:*** Demonstrating that evidence about “what works” and how to improve agency operations is critical to meeting the needs of the citizens and maximizing taxpayers’ return on investment. Making it easier to build evidence to improve public policy—while protecting privacy—is a primary goal of an NSDS. The technical assistance services that an NSDS could provide can make it possible for a wide range of entities, including those that do not have their own analytic capacity, to benefit.

- **Data providers:** Fostering improved data quality at all levels, starting at collection, and encouraging the development of widely accepted data and metadata standards that make linking easier.
- **Evidence-building partners:** Emphasizing the NSDS's role in supporting agency evidence-building needs, such as issues identified in their evidence-building plans.
- **Research community:** Sharing results derived from expanded access to data and statistics generated from those data on privacy and the availability of those data for future research (i.e., through privacy budgets).
- **General audience:** Promoting the benefits to data providers and society that arise from facilitating greater access to data across all levels of government, and the specific role of the NSDS in both promoting the use of data and protecting privacy.

Year 2 Focus

The Committee will develop detailed recommendations for implementing the data concierge service by exploring existing models in use within the federal government and elsewhere, such as [National Institutes of Health librarians](#) or the model within [FSRDCs](#), detailing best practices, lessons learned, and their current value proposition.

Additionally, the Committee will expand on the elements of a communication strategy including identifying best practices from other similar efforts, exploring what structures may be required to facilitate ongoing communication, and detailing what communication capabilities, potentially including a communication function, should look like within the NSDS.

4. Year 2 Roadmap

4.1. Year 2 Planned Approach

The Committee’s first year was focused on fact finding to develop an overarching vision for an NSDS and laying out the contours of what is required to realize that vision from the perspectives of five focus areas—legislation and regulations; governance, transparency, and accountability; technical infrastructure; government data for evidence building; and other services and capacity-building opportunities. The Year 2 agenda will focus on the needed steps to operationalize the NSDS and that vision.

To accomplish this, the Committee will take a holistic approach to synthesizing and integrating the subcommittee recommendations from this report and building on those recommendations. Next steps include investigating specific topics more deeply, documenting findings in those areas, and drafting actionable recommendations. Subcommittees will report out findings and recommendations in the Committee’s public meetings. The full Committee will discuss the material and will decide whether to issue the recommendation or request further investigation. Recommendations may be released by the Committee on a flow basis throughout the year, as well as being captured in the Committee’s final report, to guide swift action and implementation.

In addition, the Committee will seek public input on its Year 1 recommendations and overarching NSDS vision. This input will help guide the Year 2 agenda and inform the specific findings and recommendations the Committee makes moving forward.

4.2. Initial Year 2 Focus

In addition to integrating and synthesizing recommendations across the focus areas, subcommittees intend to explore specific topics more fully over the next 12 months. For more information, see “Areas of Focus.”

5. Appendices

Appendix A. Committee Charter, Membership, Process, and Meetings

This appendix provides an overview of the Committee’s charter, membership, process, and meetings. Additional information related to the Committee can be found on [ACDEB’s website](#). Specific items are described below.

Charter

The Committee was established under the Evidence Act to review, analyze, and make recommendations to the OMB Director on how to promote the use of data for evidence building. ACDEB is an officially chartered committee under the Federal Advisory Committee Act (FACA). As such, the [Committee’s charter](#) provides information on its authority, objectives and scope of activities, description of duties, agency or official to whom the Committee reports, support, estimated annual operating costs and staff years, Designated Federal Officer, estimated number and frequency of meetings, duration, termination, membership and designation, subcommittees, and recordkeeping.

Membership

The Committee is composed of 27 members, representing diverse perspectives and a wealth of expertise from federal, state, and local government as well as the private sector (including the privacy community).

A full list of Committee members and their bios are available on the [ACDEB website](#).

Process

The Committee’s work falls into two main phases: (1) knowledge sharing and (2) deliberations and recommendations formation. Table A1 provides an overview of ACDEB milestones.

Table A1. ACDEB Milestones

October 2020 – April 2021	May 2021 – October 2021	October 2021 – October 2022
Knowledge Sharing	Deliberations / Recommendation Formation	
<ul style="list-style-type: none"> • Overview: Commission on Evidence-Based Policymaking and the Evidence Act • Relevant Technologies • International, State, and Local Perspectives • Performance Officers and Evaluations Officers Perspectives • Federal Statistical System Perspectives and Experience • Chief Data Officers Perspectives • External Researchers Perspectives • Privacy and Confidentiality Concepts and Technologies • Physical Sciences Data Challenges • Hart-Potok Framework 	<ul style="list-style-type: none"> • Subcommittee Kickoff • Continue Information Gathering • Begin Implementation • Develop Year 2 Plan • Due: Year 1 Report 	<ul style="list-style-type: none"> • Adopt Year 2 Plan • Continue Information Gathering • Monitor Implementation • Finalize Recommendations • Due: Year 2 Report

Information gathering. ACDEB launched its work in October 2020 and spent the first 7 months sharing knowledge and experience relevant to fundamental evidence-building issues. During this information-gathering phase, Committee members and outside experts shared presentations at public meetings. For key takeaways from selected presentations, see Appendix B. In addition, the Committee solicited feedback from the public through a request for comment in the Federal Register. Appendix C provides a summary of responses to this request. The full set of [public comments](#) is available on the ACDEB website.

Deliberations. From there, the Committee began to build on this knowledge base as it entered the deliberative chapter of its work. The Committee is taking a multi-phase, holistic approach that breaks discussion into five focus areas and related subcommittees: (1) legislation and regulations; (2) governance, transparency, and accountability; (3) technical infrastructure; (4) government data for evidence building (with an emphasis on administrative data); and (5) other services/capacity-building opportunities. These focus areas were designed to address a wide range of opportunities and obstacles for a National Secure Data Service and the evidence-building ecosystem more broadly, using the vast input and insights of Committee members.

In addition, the Committee established a coordinating committee to ensure consistency across focus areas and minimize duplications of effort. This group comprises a cross-section of members from the different focus areas and with different expertise and experiences.

Per FACA requirements, subcommittees present their findings and recommendations to the full Committee at its public meetings. Table A2 lists subcommittee co-chairs and members.

Table A2. ACDEB Subcommittees

Subcommittee	Members
Coordinating Committee	Emilda Rivers (chair), Laila Alequresh, Charles Cutshall, Nicholas Hart, Anna Hui, Julia Lane, Amy O’Hara, Matthew Soldner
Legislation and Regulations	Nicholas Hart (co-chair), Christine Heflin (co-chair), Gregory Fortelny, Ted Kaouk, Edward Kwartler, Christin Lotz, Todd Richardson, Mayank Varia
Governance, Transparency, and Accountability	Charles Cutshall (co-chair), Julia Lane (co-chair), Otis Brown, Shawn Davis, Gregory Fortelny, Edward Kwartler, Brian Moyer, Kimberly Murnieks, Christina Yancey
Technical Infrastructure	Amy O’Hara (co-chair), David Park (co-chair), Otis Brown, Barry Johnson, Ted Kaouk, Elisabeth Kovacs, Mayank Varia, Christina Yancey
Government Data for Evidence Building	Anna Hui (co-chair), Kenneth Troske (co-chair), Laila Alequresh, Richard Allen, Leonard Burman, Christine Heflin, Elisabeth Kovacs, Christin Lotz, Brian Moyer
Other Services/Capacity-Building Opportunities	Kimberly Murnieks (co-chair), Matthew Soldner (co-chair), Richard Allen, Leonard Burman, Shawn Davis, Barry Johnson, David Park, Todd Richardson

Tools. The Committee developed and leveraged several tools to help inform its work, including a resource library, a project inventory, virtual site visits, a case study criteria matrix, and a crosswalk to the recommendations from the Commission report. For more information on these tools, see Appendices D, E, F, and G.

For more information on the Committee’s process and work plan, see presentations from the April 2021 public meeting on the “Work Plan Approach” and “Next Steps: ACDEB Focus Areas and Committee Discussion,” from the July 2021 meeting on “Decision-making Standards,” from the August 2021 meeting on “Year 1 Report Approach,” and from the September 2021 meeting on “Taking Stock: The Vision, Decision-Making Process, and Year 1 Report Expectations.”

Meetings

The content and structure of the Committee’s public meetings mirrors its process. The first seven meetings focused on information gathering, and the remaining meetings in Year 1 centered on subcommittee reports and Committee discussion around process and possible recommendations.

Meeting agendas, presentations, and minutes are available on the [ACDEB website](#).

Appendix B. Selected Meeting Presentation Summaries

November 2020 ACDEB Meeting

Amy O'Hara: Relevant Technologies

- Must determine how to develop trust and transparency, with both data subjects and data contributors. Technology can help improve transparency in order to increase trust with data providers. At the same time, we need to increase public trust and transparency about data users, results, and benefits.
- Outlined four ways to think about technology (getting data to/in NSDS, protecting the data, connecting users to NSDS, getting results out of NSDS), all of which can be improved by combining a variety of technology solutions. Different technologies need to be included in a mesh instead of technology silos.
- In order to use technology, we need to have clarity across participating agencies about what their laws say and how they are interpreting these laws. This is especially related to what constitutes disclosure, who is an authorized user, what barriers are needed, and where can data be disclosed.

December 2020 ACDEB Meeting

Julia Lane: The Use of Data for Evidence-making: International Lessons Learned

- Summary is not available. Presentation can be found on ACDEB website with questions directed to Julia Lane.

Anna Hui, Elisabeth Kovacs, Christin Lotz, Kimberly Murnieks: State Perspectives

- Compared four states (Missouri, South Carolina, Tennessee, and Ohio) regarding how they look at evidence, how they are organized for evidence and data, and their state-specific concerns.
- Identified three main categories for state-level challenges and issues:
 - ***Lack of incentives:*** Including no standardized reporting framework at the federal level, reporting at different times of the year (FY, program year, calendar year, etc.), lack of federal inter-agency data sharing, legacy data systems, and the need for additional resources to encourage states to integrate state data systems.
 - ***Culture of government data:*** Government programs are linear, there are protective instincts of data ownership by agencies, federal laws/regulations that deter data sharing, and balancing compliance with results.
 - ***Social impact:*** Public distrust of data protections, fear of government managing people through data, and related privacy and security concerns.
- One key recommendation is the need for a common schema or data structure for state and federal governments.

Laila Aleqresh, David Park: Local Perspectives

- **Local challenges:** Identified challenges at the local level including reporting requirements, granular comparison data, the ease of locating information, and employee data skills/expertise.
- **Partnerships:** Need full partnerships between local, state, federal, and other government entities for data gathering and sharing. This contrasts with the current system where local government relationships are too often one way and top down.
- **Data standards:** When considering data standards, local governments and local government staff may not currently have the capacity or workforce skill sets to do all desired data standards work.
- **Workforce skills building:** Need to provide funding for local government consulting assistance and clearinghouse to provide help and build required data skill sets.
- Highlighted that any solutions must be accessible, equitable, and intuitive.
- Cities are often the data generators, but many federal systems don't center on the user experience (cities, researchers, etc.). Architecture needs to be re-designed to better enable data uploads and improve data access and transparency for analytics.

January 2021 ACDEB Meeting

Todd Richardson, Christina Yancey, Matthew Soldner, Christine Heflin: Perspectives from Federal Evaluation and Performance Improvement Officers on Administrative Data Needs

- Federal legislation will be required to meet performance and evaluation needs, including updating the CIPSEA companion legislation to enable sharing combined census and tax data.
- Need to broaden the standardization of data sharing agreements across government entities (including states). Need to both include the value to the agencies as well as the benefits to the states. Broaden the standardization of data sharing agreements across government entities.
- Should identify a group or agency in government to guide the standardization of interagency and intergovernmental agreements and data harmonization. Suggest that OMB should take the lead on this with ICSP and start a repository to capture the data sharing agreements that exist today to assess the commonalities and differences between them.
- Many of the barriers to program evaluation are administrative. These barriers can be smoothed through the development of standard operating procedures, common data-sharing agreements, and common forms that protect privacy and support reuse. Once the administrative barriers are solved, we can then move on to the technical barriers.
- Should test procedures on matching activities that leverage high-value data sets with known/knowable statutory barriers (e.g., LEHD, IRS, NDNH).

Brian Moyer, Barry Johnson, Emilda Rivers: The U.S. Federal Statistical System

- Identified what agencies and subagencies participate in the federal statistical system and its core components, its legislative and regulatory mandates, its primary work and activities, examples of its work, and its priorities and pilot projects as it looks to the future.
- The federal statistical system is neither singular nor static—it is a decentralized structure that is constantly evolving to meet the needs of its customers. Guiding the statistical system is a robust legal and regulatory framework including OMB statistical directives, Confidential Information Protection and Statistical Efficiency Act (CIPSEA) and Implementation Guidance, and the Evidence Act and Federal Data Strategy.
- The federal statistical system provides an existing framework, structure, and ecosystem that has proven to be well-suited to collecting, curating, and disseminating high-quality data and protecting privacy. The Committee must draw upon this as we think about enacting an NSDS and make sure that we keep the existing strengths and capacities of the system while addressing its limitations. We may need to update standards, guidelines, processes, etc., but we do not need to start from scratch.
- Title III of the Evidence Act makes statistical agencies major actors in expanding access to protected data by strengthening the technical autonomy of agencies, providing new authority for agencies to acquire data for the purpose of evidence building, requiring agencies to make data access in multiple access tiers based on data sensitivity, and requiring agencies to adopt a standard application process that will serve as a “single front door” for those who want to apply for access to restricted data for statistical purposes.
- A few pilots or key initiatives highlighted during the presentation:
 - Federal Statistical Research Data Centers (FSRDCs) are an example of how to potentially make data available to researchers within the federal system. The FSRDCs face challenges such as a lack of funding for technology, staff, and support but should be considered as part of a national tiered data access model.
 - The Standard Application Process was developed as a common “front door” and a first step toward achieving standard processes, standard forms, metadata, etc.
 - The Federal Committee on Statistical Methodology’s [Data Protection Toolkit](#) and [framework](#) for evaluating and communicating data quality are two examples of how statistical agencies are applying their specialized skills and experiences to advance expanded, privacy-protecting access to data.
 - The statistical agencies’ response to COVID-19 showed the innovation and nimbleness of the statistical system in meeting customer needs.

February 2021 ACDEB Meeting

Ted Kaouk, Greg Fortelny, Richard Allen: Chief Data Officers Presentation

- Discussion looked at the role and function of CDOs and the CDO Council in the federal government, as well as the importance of data sharing.
- Provided case studies in data sharing and integration from the Departments of Agriculture and Education and the Environmental Protection Agency.
- Identified challenges in data sharing: While significant attention has already been placed on technological solutions, especially those catering to the secure transfer or integration of data across organizational boundaries, less attention seems to be placed on people and processes. Until these components are resolved, data sharing will continue to be an ad hoc and costly exercise inhibiting participation. Individual agencies (intra-agency) must take steps to ensure their data management practices lend itself to secure sharing. Government (inter-agency) must take proactive steps to reduce costs and increase perceived value of secure data sharing.
- Looked at what is needed including data standards, data quality, data governance, data management resources, awareness of data, improvements to accessing data, coordination and collaboration, data stewardship, data training, and data security.
- Identified the following critical key elements to building an effective system: communication, inviting participation, building common understanding, promoting data sharing, focusing on programmatic needs not just data needs, management for data services and integration, and a focus on people and building community.

Leonard Burman: Synthetic Data and Validation Server: Safely Expanding Researcher Access to Sensitive Data

- Looked at emerging technologies, including fully synthetic data, validation servers, and secure multiparty computation, that enable increased data release while protecting privacy.
- New technologies offer the prospect of making data protection easier, faster, and safer for agencies, while making it easier for researchers to access and use the data, but these technologies are still in an early stage of development.
- Identified several of the challenges, including:
 - Traditional statistical disclosure limitation methods are labor intensive and imperfect.
 - Some agencies do not have the resources to implement them; some data sets and statistics are released with minimal protection.
 - Methods such as adding noise, rank swapping, micro aggregation, and dropping or combining sensitive variables to address perceived threats require a lot of staff time and extensive vetting to balance risk against utility.
- Identified opportunities, such as:
 - New methods could allow more data to be released, more quickly, with a strong privacy guarantee.

- Privacy protection could be largely automated, reducing the staff time needed to implement.
- More users could access the data more easily and perform more statistical analyses.
- The data service could spur innovation by promoting R&D of new methods, improving and expanding existing methods, and helping agencies adopt best practices.
- Emphasized that the value proposition is key. These technologies must work for agencies and data users. The challenge is to implement technologies that produce timely data sets, can be managed by agencies with limited resources, and that researchers will want and be able to use to produce high-quality evidence.

Kenneth Troske: Using Administrative Data to for Policy Research

- Using administrative data, if done well, can produce evidence much quicker and at much lower cost.
- Issues that need to be confronted with using administrative data are the following: it is difficult to conduct a nationwide study, it is often difficult to get data from all states, data is often inconsistent and requires significant time to get into a common format, researchers often have to work through an intermediary, and needed data (such as UI data) are less available.
- Potential ways to move forward are the following:
 - Provide a centralized mechanism for researchers and policymakers to access federal and state-level administrative data (National Secure Data Service). This should provide metadata and a standard format for data, clear guidelines on who can apply for data access, a common application for access, and the most advanced technology and methods available for protecting privacy.
 - Enhance the interactions between program administrators and the research community.
 - Increase interaction and coordination between federal and state policymakers, particularly on programs that are a joint federal-state cooperation.
 - Have a discussion on what we mean by “protect privacy” and what the goals are.

March 2021 ACDEB Meeting

Charles Cutshall: Privacy & Confidentiality Concepts

- ACDEB should build on previous work in government data privacy law and set a foundation for future thinking on how to expand access while respecting privacy and maintaining trust in government.
- Should consider that privacy is contextual and not absolute. It exists on a continuum and has multiple tiers for which different groups can be given different levels of privileges. The use of tiered access for different purposes, with different value and different security risk, can help in balancing the trade-offs.

Shawn Davis: Privacy and Confidentiality: Concepts & Challenges

- Main CEP Report Privacy Recommendations
 - Amend Privacy Act and CIPSEA to require departments to conduct risk assessments for public releases of de-identified confidential data.
 - Provide secure and restricted access to confidential data.
 - Adopt cutting-edge technology for data security, integrity, and confidentiality.
- Maintaining data security can be seen as a balancing act between confidentiality, data integrity, and data availability.
- There are different types of direct and indirect identifiers, as well as disparate agency interpretations of de-identification.
- The risk of re-identification increases when geolocation or a combination of other indirect identifiers remain within a single data set or during linkage across data sets.
- Transparency is key to ensure public trust regarding how confidential data are used.
- A proper application process, tiered secure access, auditing, and enforcement are needed to reduce the risk that confidential data are used outside of disclosed statistical purposes.

Mayank Varia: Cryptographically Protected Computing

- Secure multiparty computation (SMC) allows parties to aggregate data sets without transferring or revealing private data to each other or a third party.
- The [Boston wage gap study](#) and other pilot deployments of SMC show that the technology is ready to use in some (but not all) scenarios that involve aggregating data sets for evidence purposes. SMC can add value and privacy by enabling evidence gathering when the underlying data cannot be shared.
- Multiple law and policy questions need to be answered to use this solution more broadly with government data. This includes definitional questions (do encoded pieces count as personal data?), process questions (does computing over encodings constitute disclosures?), and liability questions (who should be blamed if there is an error?).

Leonard Burman: Privacy and Confidentiality Technologies

- Protecting confidentiality in public data sets has never been more challenging or more important. Using synthetic data plus validation server allows expanded access to administrative data with more robust privacy protections than using traditional disclosure control methods.
- Looked at the option of [using fully synthetic data plus validation server](#) to allow wider research access to IRS tax data with stronger privacy guarantee and lighter demands on IRS staff.
- Challenges include identifying an appropriate privacy standard; developing and implementing privacy protections consistent with that standard; measuring and allocating the “privacy budget;” educating researchers about the privacy budget; building a useful, general program interface for researchers; and ensuring reasonable processing time.

Edward Kwartler: Data Ethics

- Recommendations must consider data ethics, defined as responsible and sustainable data collection, processes, and storage practices as well as the ethical use of data.
- AI and machine learning have great power to do good, but they must be used in an ethical, explainable, and equitable manner. This is particularly true in the application of AI and machine learning for data-based decision making.
- Committee recommendations must emphasize that any outcome must be the following:
 - Explainable—consistent with defined process, and don't hide the societal benefit behind bureaucratic process, hard to navigate UI, or ambiguous TOS.
 - Equitable—this is the people's data and value should be broadly maximized.
 - Accessible—need a data visualization layer to provide an overall view of the data without regard to location, format, or source.

April 2021 ACDEB Meeting

Otis Brown: Data Challenges in the Physical Sciences

- Provided two views from a federal government perspective and one from a user perspective about open scientific data and the challenges of accessing and sharing it.
- Discussed Title II of the Evidence Act and its impact on the 23 CFO agencies: data is open by default, agencies must have sharing plans, and must incorporate user input into their plans, etc.
- Project using CDC data to improve decisions and health outcomes emphasizes that the capability to do this (and similar) projects is a combination of the technological infrastructure and the people who are able to operate it.
- The [NOAA effort](#) required integrating data from federal, state, and private sources to capture the cost associated with events. The key challenges faced by NOAA included the lack of homogeneity in data standards and access to multi-agency and private data holdings which may be something a National Secure Data Service (NSDS) could help address.
- [First Street's efforts](#) to build the first publicly available flood risk assessment for 140 million properties in the United States highlighted several challenges, including finding a secure mechanism for accessing and sharing property-level data, individual flood claims, or flood adaptation.

Nicholas Hart: Design Considerations for Implementing a National Secure Data Service

- The [Hart-Potok Report](#) reviewed the changing landscape and legal framework since enactment of the Foundations for Evidence-Based Policymaking Act.
- Identified eight key attributes for a national data service that should be weighed when determining the structure, location, and design of NSDS: transparency and trust, legal authority to protect privacy and confidentiality, independence, legal authority to collect data from agencies, scalable functionality, sustainability, oversight and accountability, and intergovernmental support.

- Considered and weighed four potential options for where to stand up an NSDS based on the attributes: new statistical agency within Commerce, re-tasked agency within Commerce, new public-private partnership in conjunction with NSF, or university-based
 - Recommended standing up NSDS as an FFRDC at the NSF that has CIPSEA designation and provides a roadmap for key considerations specific to NSF.
- Four questions that the Committee should consider when developing NSDS recommendations:
 - Value proposition of the NSDS.
 - Organizing within the existing ecosystem to ensure NSDS fits in the current and future ecosystem, including the role of FSRDCs.
 - Sustainable resources and financing, including what level of resources is needed and how to approach reimbursable arrangements.
 - Oversight and measurement needed to ensure accountability, including oversight mechanisms, transparency requirements, and archiving policies.

Appendix C. Public Comments Summary

The Committee issued a general solicitation of comments from the public to offer researchers, evaluators, contractors, government entities, and other interested parties the opportunity to inform the Committee's work. The request for public comments was issued in the Federal Register on December 15, 2020, and comments were collected through February 9, 2021. Committee members, in coordination with ACDEB support staff, compiled the comments and presented an [analysis of the results](#) at the public meeting in April 2021. The full set of [public comments](#) is available on the ACDEB website. The questions and a summary of the responses is provided below.

Questions for Public Comment

Central Questions

1. What are the main challenges faced by national, state/provincial, or local governments that are trying to build a basis for evidence-based policy? Briefly describe the bottlenecks and pain-points they face in the evidence-based decision-making process.
2. What are examples of high-impact data uses for evidence-based policymaking that successfully effected change, reduced costs, or improved the welfare of citizens?
3. Which frameworks, policies, practices, or methods show promise in overcoming challenges experienced by governments in their evidence building?
4. The Commission on Evidence-Based Policymaking [recommended](#) the creation of a National Secure Data Service. Do you agree with this recommendation, and if so, what should be the essential features of a National Secure Data Service?
5. How can federal agencies protect individual and organizational privacy when using data for evidence building? Recommend specific actions the Office of Management and Budget and/or other federal agencies can take when using data for evidence building, as well as suggested changes to federal laws, policies, and procedures.

Secure Data Access

6. If created, how should a data service be structured to best facilitate: (1) research and development of secure data access and confidentiality technologies and methods, (2) and agency adoption of those technologies and techniques?
7. Government agencies have argued that secure data access has value because it: (1) improves service delivery, (2) improves efficiency (lowers costs), (3) produces metrics for performance measurement, and (4) produces new learnings/insights from the data. Which of these propositions do you agree holds value and why? Do you have examples that demonstrate these benefits? Do you have other examples of the value of secure data access?

Data Services to Federal, State, and Local Agencies and the Public

8. What are the most pressing data needs of state and local decision-makers and how would making data accessible from federal agencies help meet those needs? To share data, what guarantees do data owners (or data controllers) need regarding privacy, data stewardship, and retention?
9. What are the key problems and use cases where collaborative work between federal, state, and local authorities' data analysis can inform decisions? What are key decision support tools? How would greater communication about data and tools benefit expanded evidence building?

Infrastructure for Meeting Public and Evidence-Building Needs

10. What basic public data services are essential for a data service to address existing capacity gaps and needs? What infrastructure or incentives can the federal government create that locals and states cannot?

Key Takeaways

- Widespread support for the Commission's recommendations, including affirmation for NSDS with suggestions for structure/functions.
- A solid collection of use cases, frameworks, and other examples of data sharing systems/mechanisms like: (1) linking data from 28 Florida community colleges to develop performance funding metrics, (2) using insights from a Securities and Exchange Commission (SEC) investor testing initiative to inform evidence-building policies across government, and (3) applying principles of the FSRDC framework to the NSDS; comments provide an extensive inventory of items for further investigation.
- A healthy mix of descriptive vs. prescriptive comments; while most commenters offered general observations and examples from their own perspectives, many commenters also made specific suggestions and recommendations like how to increase federal and state and local evidence-building capacity, how to structure the NSDS, and how to protect privacy when using statistical data for evidence building; comments provide a list of options for consideration.
- In-depth discussions of challenges—covering both wide-reaching difficulties with evidence for decision making and specific hurdles for the NSDS; most common challenges listed were data access, data availability, and resource constraints.
- Highlight several methods for building capacity and expertise; partnership was the most mentioned mechanism, followed by data literacy and user training.

Overview

Number of Commenters/Comments

- 35 commenters (including the submission of several research papers)
- 141 separate comments directly addressing the 10 questions from the Federal Register notice

Table C1. Types of Organizations

Organization type	Number	Average comments
Non-profit institutions, including trade organizations	12	3
Other private firms, including advocacy groups and consultants	9	3
Academic institutions or researchers	6	5
Federal government	3	4
State and local government	2	7
Private citizens (former OMB official, longtime data practitioner)	2	8
Public-private partnership (FSRDC)	1	10

Table C2. Comments by Federal Register Notice Question

Question summary	Comments
1. Challenges faced by national, state, or local governments with evidence-based policymaking (building systems and decision-making process)	21
2. High-impact data uses for evidence-based policymaking	13
3. Promising frameworks, policies, practices, or methods in overcoming challenges	16
4. Support for NSDS recommendation and essential features	17
5. Protecting privacy when using data for evidence building; specific actions when using data for evidence building	11
6. Structure of NSDS to facilitate R&D of secure data access and confidentiality technologies and agency adoption of technologies and techniques	12
7. Value proposition of secure data access (benefits and examples)	11
8. Most pressing needs of state and local decision-makers and how a federal system could help; data owner guarantees on privacy, stewardship, and retention	12
9. Key problems/use cases where collaboration between federal, state, and local authorities can inform decisions; key decision support tools; impact of communication about data and tools to expand evidence building	16
10. Essential public data services and infrastructure that the federal government can create that locals and states cannot	12

Focus Areas

- **Broader evidence community**, including current and former government officials from the Census Bureau, the U.S. Department of Transportation, OMB, and SEC, as well as organizations and researchers like the American Economic Association, Center for Open Data Enterprise, Data Coalition, Mathematica, Project Evident, Rotunda Solutions, Robert George from Chapin Hall at the University of Chicago, and professors from the University of Kentucky
- **State and local government**, including representatives from state and local governments, such as Washington State and State Chief Data Officers, as well as groups who represent and support them, like the Western Governors' Association, Actionable Intelligence for Social Policy at the University of Pennsylvania, and the Inter-University Consortium for Political and Social Research
- **Data equity**, including the Center for Democracy and Technology, the Annie E. Casey Foundation and partners, and the Leadership Conference on Civil and Human Rights
- **Education**, including the American Educational Research Association, Data Quality Campaign, Lumina Foundation, and the National Student Clearinghouse
- **Health care**, including Datavant, Inc. and Trewon Technologies
- **Indigenous Peoples**, including UCLA professors and research papers on culturally responsive Indigenous evaluation (CRIE)

What's Missing?

Privacy: While many commenters acknowledged the challenges with protecting privacy, only a handful of commenters scratched the surface of emerging technology solutions, and there was only one set of comments (from Galois Inc., University of Boston, and Barnard College) from the perspective of technology researchers and providers and privacy advocates.

Comparisons with the Commission on Evidence-Based Policymaking

Reminders: (1) ACDEB's work is an extension of the Evidence Commission's efforts and (2) the Evidence Commission enacted an extensive, multi-pronged communications approach to gather input to inform its recommendations.

- The Evidence Commission received over 350 responses to its FRN request for comments; in addition, the Evidence Commission's fact-finding and deliberative processes included public meetings, public hearings, meetings with selected organizations, a survey of federal offices, and other public input received through email correspondence.
- Over one-third of the ACDEB commenters also submitted comments to the Evidence Commission and/or participated in the Commission's targeted communication efforts.

Possible Future Steps for Analysis

- Inventory use cases, frameworks, and other examples of data sharing systems/mechanisms and sort into categories for further analysis.
- Flag descriptive vs. prescriptive comments and sort suggestions into categories for consideration.
- Provide additional information to ACDEB subcommittees, as requested.

Appendix D. Project Inventory

The Committee is populating a Project Inventory to inform its recommendations about the development of a National Secure Data Service. The inventory includes recent, ongoing, and planned projects that can offer lessons learned, highlight opportunities to build on existing efforts, and help the Committee identify where there is a need for new, targeted pilots.

Projects

ACDEB is especially interested in highlighting projects that: (1) call out solutions to existing challenges to evidence-based decision-making and (2) provide timely, relevant tie-ins to broader federal priorities.

- Challenges could be related to legislative barriers, lack of regulations or guidance, administrative practices, data access or sharing, data quality, transparency, accountability, stakeholder engagement, statistical system coordination, skills gaps, IT systems, or technical capacity.
- Broader priorities could include the Biden Administration’s focus areas (including those in Executive Orders or memoranda), implementation items under the Evidence Act or the Federal Data Strategy, or recommendations from the Evidence Commission.

Partners

In addition to projects submitted by ACDEB members or featured at the Committee’s public meetings, the Committee reached out to the members of the Interagency Council on Statistical Policy (ICSP) to gather project ideas from the federal statistical system.

A National Secure Data Service would join the statistical system as an additional actor and could significantly affect how agencies acquire, curate, link, and make data accessible for statistical and evidence-building purposes. Therefore, it is essential that the work of the statistical system informs ACDEB’s efforts. The goal was to include at least one project from each statistical agency in the inventory. Examples of statistical system projects include the following:

- Cross-cutting projects, like the Standard Application Process and the Data Protection Toolkit
- Efforts from Statistical Officials who are also ACDEB members, such as National Center for Health Statistics work with state administrative COVID-19 data
- Initiatives funded at the Census Bureau and the Bureau of Economic Analysis specifically to inform the creation of a federal data service

The Committee provided a version of the inventory to ICSP members and to the Chief Data Officers Council to highlight use cases that demonstrate how federal agencies are sharing administrative data.

Progress and Plans

The Committee used the inventory to identify topics for virtual site visits and case studies to feature in this report. Moving forward, the Committee will continue to update the inventory and will leverage this tool to guide the Year 2 work plan. The latest version of the [Project Inventory](#) is available on the ACDEB website. For more information on ACDEB virtual site visits, see Appendix E.

Appendix E. Virtual Site Visits

The Committee sponsored a series of virtual site visits to help inform the development of a National Secure Data Service. These field trips provided examples of recent, ongoing, and planned projects that offered lessons learned, highlighted opportunities to build on existing efforts, and helped the Committee identify where there may be a need for new, targeted pilots.

The goal of these site visits was to: (1) provide an overview of the framework, system, or project; (2) spotlight existing challenges and possible solutions to linking data for evidence-based decision-making; and (3) offer timely, relevant tie-ins to broader federal priorities.

- Challenges could be related to legislative barriers, lack of regulations or guidance, administrative practices, data access or sharing, data quality, transparency, accountability, stakeholder engagement, statistical system coordination, skills gaps, IT systems, or technical capacity.
- Broader priorities could include the Biden Administration’s focus areas (including those in Executive Orders or memoranda), implementation items under the Evidence Act or the Federal Data Strategy, or recommendations from the Commission on Evidence-based Policymaking.

The Committee held these site visits about weekly starting in mid-August and running through September. Each field trip was 60–90 minutes, including time for the presentation and Committee discussion. The ACDEB Designated Federal Officer attended each site visit and monitored attendance levels to ensure compliance with all requirements of the Federal Advisory Committee Act. The Committee used information from these sessions to inform this report and to help plot next steps.

See table E1 for an overview of each site visit. The Committee would like to thank all hosts, presenters, and support staff who made these events a reality. This appendix also includes more in-depth summaries for the site visits.

Table E1. Virtual Site Visits Overview

Virtual Site Visit	Host(s), Speaker(s), and Support Staff
U.S. Department of Agriculture (USDA) Enterprise Data Analytics Platform and Toolset (EDAPT) (August 16, 2021)	Host: Ted Kaouk, U.S. Office of Personnel Management Chief Data Officer (CDO) (formerly USDA CDO) and ACDEB member Speakers (USDA): Ted Kaouk, Chris Alvares, Jacqueline Ponti-Lazaruk, Kasey Martin, Linda Young, Chris Nelson, Cyndy Parr, Lynn Overman
South Carolina Revenue and Fiscal Affairs (RFA) Office and the Integrated Data System (August 25, 2021)	Host: Elisabeth Kovacs, Deputy Director of Workforce Development at the South Carolina Department of Commerce and ACDEB member Speaker: W. David Patterson (RFA)
Coleridge Initiative (Coleridge) (August 30, 2021)	Host: Julia Lane, Coleridge co-founder and Director and ACDEB member Speakers: Julia Lane, George Putnam (State of Illinois), Deshawn Preston (United Negro College Fund), Jessica Cunningham (KYStats), Ismail Coskun (Coleridge), Nancy Potok (former Chief Statistician of the United States), Nathan Barrett (Coleridge)
Federal Statistical Research Data Centers (FSRDCs) (September 16, 2021)	Host: Bill Beach (Bureau of Labor Statistics), Barbara Downs (Census Bureau), Maggie Levenstein (University of Michigan), Cathy Fitch (University of Minnesota) Speaker: Barbara Downs, FSRDC Program Director
U.S. Chamber of Commerce Foundation (Chamber Foundation) Jobs and Employment Data Exchange (September 21, 2021)	Host: Jason Tyszko, Vice President of the Center for Education and Workforce at the U.S. Chamber of Commerce Foundation Speakers (Chamber Foundation): Jason Tyszko, Robert Sheets Support (Chamber Foundation): Sarah Castro
National Center for Health Statistics (NCHS) Modernization Activities (September 29, 2021)	Host: Brian Moyer, NCHS Director and ACDEB member Presenters (NCHS): Brian Moyer, James Carver, Travis Hoppe, Irma Arispe, Lisa Mirel, Neil Russell, Paul Sutton Support (NCHS): Kiana Morris, Leslie Rivas, Lisa Wagner, Meagan Walters

Virtual Site Visit 1. USDA Enterprise Data Analytics Platform and Toolset (EDAPT)

Background

The U.S. Department of Agriculture (USDA), with its diverse mission of 19 agencies, including two principal statistical agencies, is a microcosm of the federal evidence ecosystem. Leveraging the department’s vast data sets as a strategic asset is critical to achieving ambitious administration goals in addressing climate change, increasing nutrition security, ensuring thriving rural communities and economies, increasing opportunity for agricultural products at home and abroad, and advancing equity across federal programs and services.

The challenge. In 2017, data-driven decision-making and evidence building was difficult across the sprawling USDA, with its 29 agencies and staff offices and nearly 100,000 employees relying on data trapped in hundreds of unconnected silos throughout the organization. When a leader at USDA asked a simple logistical question—like “How many vehicles does the agency have, and how many are underutilized?”—the answer might take weeks, requiring manual data calls across multiple agencies and offices.

This challenge extended to nearly every part of the department’s administration, including human resources (HR), finance, operations, and more. The situation was no different for programs delivering support to citizens, such as farm loans and disaster assistance: gaining fast, data-driven insights for actual mission delivery was nearly difficult or impossible. This was problematic, because making better decisions ultimately relies on the ability to assemble many types of data, from many disparate sources, and then transform those data into something actionable. Yet this substantial challenge presented an even greater opportunity.

The solution. Over the past 3 years, USDA has established the Chief Data Officer (CDO) role with the responsibilities required under the Evidence Act and created the Assistant CDO position in each USDA mission area. Centralized analytics teams support these leaders to perform and advance the use of analytics. The department has also developed enterprise-wide data dashboards to improve decision-making while reducing manual data collection and has launched the EDAPT.

EDAPT provides a standardized, centrally available set of tools and connected data sources to enable a broad range of analytics from descriptive methods to advanced predictive techniques and natural language processing. EDAPT eliminates technology as a barrier to data sharing and analytics within and across agencies and integrates data analysis for more than 150 sources from every corner of USDA as well as outside the department. These initial investments have enabled partnerships across department, mission area, program area, Statistical Official, CDO, and Evaluation Officer functions that substantially advance the department’s evidence-building capacity both internally and externally.

The benefits—pandemic response. With these key technical and workforce capabilities in place, the department was better able to respond when the need for data was exacerbated during the COVID-19 pandemic.

USDA worked with the Federal Chief Data Officer Council to identify common agency data needs and to develop dashboards that could provide relevant information fast. Within hours, all efforts shifted to COVID-19. USDA had the first COVID-19 dashboard operating 2 weeks later, overlaying case figures from Johns Hopkins University over existing HR data, so that USDA could understand where hot spots were occurring and their risk to its 100,000-person workforce. The new dashboard has since evolved to incorporate additional data to support data-driven workforce safety decisions and has been shared with multiple other federal agencies for reuse through the Federal CDO Council. In just a few weeks, these tools were delivered to nearly 5,000 leaders across USDA, who now use the dashboard to make informed, risk-based decisions in a time of crisis—an accomplishment that would have been nearly impossible just a few years earlier.

The enterprise analytics team has also built additional tools that directly support USDA's program delivery to better serve American citizens during the pandemic. The Food and Nutrition Service (FNS), for example, uses a new dashboard for the [Meals for Kids](#) program, which assists families in finding free meals in their areas while schools are closed. Another dashboard helps USDA track the spending of nearly \$19 billion appropriated to it by the CARES Act for the Coronavirus Food Assistance Program, which provides direct relief to American farmers and ranchers who have been adversely affected by the pandemic.

The benefits—value delivered. From more efficiently spending taxpayer dollars to keeping its workforce safe during a pandemic, USDA's dashboard project has evolved into a full-fledged [USDA Data Strategy](#) that has delivered value throughout the entire department. USDA leaders can get rapid answers to vital questions, and as a result, the department's customers can get better service. This transformation has been recognized externally, with the project being named [FedScoop Magazine's Federal Government Innovation of the Year](#) in 2018.

More than that, the team's ability to respond and pivot as needed on an ongoing basis means that USDA can stay on top of whatever emerges, like COVID-19. And USDA can do it quickly: In this most recent crisis, a process that had previously taken 3 months took 2 weeks. The entire mindset at USDA has shifted, where leaders now expect to have the tools and insights to tackle challenging tasks in near real-time.

The project was the first in the federal government to provide a comprehensive suite of administrative dashboards and analytics tools accessed on a single platform, and several other agencies have recently followed suit, with many CDOs now sharing these ideas through collaborative forums such as the CDO Council.

Why has EDAPT been successful?

Through its recent modernization efforts, including the development of EDAPT, USDA has expanded access to statistical data for evidence building within the department. Key elements of this success include the following:

Culture—nurtures USDA’s culture as a data-driven organization. While its recent data and analytics innovations have delivered immense value to USDA, the most longstanding benefit may be the growth of the department’s culture as a data-driven organization that supports evidence building internally and externally.

- **Overcoming skepticism.** In the early days, internal skepticism was the most substantial obstacle: employees had become accustomed to slow answers to questions requiring data as part of the solution. Over the course of this initiative, the team was able to flip this paradigm on its head—fundamentally raising the bar when it comes to data and analytics.
- **Building buy-in.** With employee buy-in, a new, data-driven decision culture has taken root at USDA, removing manually burdensome processes and enabling employees to focus on their mission of delivering vital programs that support America’s people, farmers, ranchers, and producers.
- **Fostering collaboration.** To address the need for greater integration and to foster a more collaborative, data-driven environment, USDA established two communities of practice (CoPs) under the leadership of Assistant CDOs, bringing together individuals with an interest or established skill sets in data visualization and analytics. These CoPs share best practices and tips as well as examples of advanced dashboard development and data usage, encouraging questions and conversations among groups and effectively raising the collective awareness and knowledge of staff across the department.

Value—provides value throughout the department and to the American public. The USDA CDO worked closely with departmental partners to overhaul and rethink how the department uses its massive amounts of data in new and innovative ways. Highlights of this process include the following:

- **The department.** Starting in November 2017, the team interviewed administrative leaders to understand the most pressing questions for USDA leadership. Rather than traditional requirements documents, these conversations leveraged design thinking strategies and engagement sessions. Once those pressing questions were identified, the team began 3-month “sprints” for each administrative function: to understand what data were needed to improve operations, what data existed, and how reliable they were. Turning that data into automated, real-time insights for USDA leaders required building extensive new infrastructure—standing up what would become the EDAPT. In just one year, the team delivered over [120 distinct dashboards](#) that illuminate critical data across the department.
- **The public.** With a solid foundation that aided leadership’s decision-making throughout the department, USDA shifted its focus in the second year of the project to providing the same kinds of solutions for some of the agency’s specific citizen-facing programs. The team worked with each of USDA’s eight mission areas, each with diverse programs and needs, to develop a foundational set of tools that spanned nearly every facet—from employee attrition to fighting forest fires to understanding the impact of research. For example:

- State conservationists at Farm Production and Conservation have access to a real-time scorecard with more than a dozen key metrics spanning both program outcomes and operational management to help more efficiently administer government resources in support of conservation.
- Forest supervisors in the Forest Service have access to an integrated view of timber sales, fuels treatments, budget, fleet, and HR data for each individual forest to help them better manage natural and government resources sustainably and efficiently.
- Scientists and operational field leaders across the Office of Food Safety (OFS) have access to more timely data around the status of laboratory sampling as well as key public health indicators for every food processing establishment OFS regulates.
- For more information on the value of these efforts as well as use cases, see “Benefits—pandemic response” and “Benefits—value delivered” above and “Legal authority” below.

Governance and management—operationalizes an organizational vision to address data leadership and skills gaps and effectively scale emerging capabilities. The increased expectations about what was possible at USDA drove the department to establish entirely new data functions across the organization, including creating the role of Assistant CDO within each mission area to lead data strategy and governance activities and provide leadership in the following areas:

- Establish consistency and governance in managing data across the mission area so that data can be more easily used and shared within proper security parameters
- Oversee centralized analytics teams and provide analytics capacity to address cross-cutting questions or issues within the mission area (for more information, see “Subject matter expertise” below)
- Enable program areas to conduct program-specific analytics with common tools
- Ensure mission area alignment with the USDA data strategy

Expertise—builds capacity within each mission area to leverage data as a strategic asset. USDA has developed an organizational model for centralized analytics teams in each of its mission areas. These teams, led by the Assistant CDOs, support their respective mission areas by answering key cross-cutting questions with data while fostering analytics training and development more broadly by performing the following activities:

- Identify and help solve major cross-cutting strategic questions using data analysis and advanced data analytics techniques and methods (including advances in data science such as machine learning, neural networks, and other forms of artificial intelligence)
- Create analytics products such as data visualizations, scenario analysis tools, and prescriptive or predictive models to draw insight from across mission area data sets—structured, semi-structured data, and unstructured (for example, text)—for day-to-day use by business leaders
- Provide ad hoc analytics services to various parts of the mission area
- Cultivate a data-driven organization through the development and enablement of the workforce

Technical capacity—develops and implements new advanced capabilities to facilitate collaboration and data sharing across the department. To address USDA’s varied needs, the department brought new technologies online that enabled the department to move from descriptive to predictive analysis and to address even bigger challenges. Aspects of this expanded toolkit include the following:

- **Capabilities and governance.** The department introduced advanced analytical techniques, like artificial intelligence, machine learning, and natural language process.
- **Data management.** USDA implemented a governance process for cataloging data and standardizing analytics tools.
- **Open Data Platform.** The department implemented USDA’s Open Data Platform, enabling the department to publish dashboards that provide the public and third-party authenticated users with the ability to draw data-driven insights, as well as download data about USDA programs.

Policy—enables data access for statistical agencies. The USDA data strategy and infrastructure are also helping to bridge the gap between administrative and statistical agencies by creating an environment where data, tools, and computing power can be shared on a common platform to resolve major historical challenges. The Integrated Modeling and Geospatial Estimation System (IMAGES) project demonstrates the challenges, gaps, solutions, and benefits around leveraging the presumption of accessibility, including the following:

- **The challenge.** In 2019, major crop-producing areas in Illinois and the surrounding states experienced excess soil moisture due to above normal precipitation events and below normal temperatures. This caused historic delays in planting, late plantings, and increased prevent plant (acreage that farmers planned to plant to a specific crop but could not). Ultimately, there was a large difference from farmers’ March planting intentions to their final June planted acreages, differences substantiated by follow-on surveys conducted by USDA’s National Agricultural Statistics Service (NASS). As insurance against future challenges, NASS began to explore novel technologies, more data sources, and new analytical methods to produce timelier indications of planted and prevent plant acreages.
- **The gap.** Although NASS has highly skilled geographers, mathematical statisticians, and data scientists dedicated to advanced research methods, NASS was lacking the IT infrastructure, tools, and cloud computing capacity to use all available, useful data simultaneously. NASS was struggling to modernize its estimation methods and unable to develop and scale new national and state-level data models that leverage its own geospatial data products with emerging, timelier, and more granular satellite imagery data.
- **The solution.** NASS and USDA’s Statistical Official partnered with the CDO and the EDAPT team to address this issue. In FY 2021, NASS worked closely with the CDO to confirm requirements, modernize, and launch its analytics capacities in EDAPT through the IMAGES project. The CDO and NASS worked to fully enable NASS data scientists to use the platform for a variety of analytics workloads.

- **The benefits.** These efforts have resulted in several direct and indirect benefits for NASS and USDA. In its research phase, IMAGES has increased the use and value of data on hand, increased the coverage of agricultural production, resulted in process efficiencies, enhanced analytical capabilities, and added value to the workforce. As far as the increased value of data on hand, IMAGES is expanding the use of collected data, both current and historical, providing added value to some products and resulting in new ones. EDAPT allows IMAGES to utilize all available non-survey data, including the following:
 - Administrative data, such as Farm Service Agency (FSA) Form 578 (signup records), Risk Management Agency, and Agricultural Marketing Service data
 - Geospatial data, including historic NASS Cropland Data Layers (CDLs) enhanced with FSA Common Land Unit (CLU) data
 - Economic data, such as recent and historic corn and soybean prices data as well as futures data
 - Environmental data, such as weather, climate, moisture, and soil type data

What challenges and opportunities remain?

While USDA has made great strides in recent years, there are opportunities to make even more progress, including the following:

Additional applications. Using all available data is responsive to survey respondents concerns about burden, perceived duplication of data requests, and the rising cost of survey data collections. For example, IMAGES represents a fuller utilization of available data and resources while committing to provide new, timelier, and more frequent insights. IMAGES may improve the accuracy of NASS surveys and censuses or allow reporting on a finer spatial or temporal scale. By layering geospatial data and other data, NASS is increasing its coverage of agricultural production.

Scaling up beyond the department. What started as a project at USDA and moved to becoming a data strategy at the department offers insights in how to advance all agencies' missions and improve data-driven accountability across the federal government. By building capacity to share and analyze data internally first, USDA gained support from agencies for an enterprise approach to data management, analytics, skills development, and evidence building. Lessons learned from this work can be applied at the state and local government level as well.

Policy. These efforts position the department to better support the broader goals of the Evidence Act, the Federal Data Strategy, and a future National Secure Data Service. This includes informing CDO strategies for modernizing systems and reducing duplicative efforts, sharing data to meet agency needs and make the presumption of accessibility a reality, and delivering program resources more effectively. Success often depends on different policies, so it is critical to have clear authorities supported by official guidance.

Virtual Site Visit 2. The South Carolina Revenue and Fiscal Affairs (RFA) Office and the Integrated Data System

Background

The State of South Carolina’s Integrated Data System is one of the most comprehensive health care data systems in the nation, encompassing nearly 20 state agencies and more than 60 hospitals. The South Carolina [Revenue and Fiscal Affairs Office](#) (RFA) plays an integral role in this system by providing independent research, analysis, and resources to facilitate informed policy decisions and administration of services.

RFA’s processes are built on both internal governance and compliance bodies working in conjunction with an array of external stakeholders and data owners, state-level executive and legislative leadership, and external requestors. Authority to receive data is granted by statute and proviso. This legal framework covers all state agencies, as well as acute care hospitals and emergency departments, and requires that data providers/owners execute MOUs with RFA covering the content, transmission, and permissible uses of their data.

The agency uses existing data to statistically link records across various sources thereby making a far broader range of analytic questions potentially answerable. The process requires that each data owner authorizes every data use, and RFA maintains a comprehensive accounting of all uses and disclosures. The process works through the concentration of technical expertise within RFA, while maintaining data owner autonomy through a federated decision-making process regarding data use and release.

This data sharing process has been in existence for more than 20 years. Thanks to the impact of information technology, RFA annually aids more than 150 integrated data projects serving internal and external customers.

Why has RFA been successful?

South Carolina’s Integrated Data System has emerged over time through an organic process built on trust, good stewardship, and technical competence. This system is one of the most comprehensive state-level data systems in the country and is frequently cited as an example of a “mature” integrated data system. Key elements of this success include the following:

Neutrality—houses data in a neutral setting. RFA is an independent office with a targeted mission. As the data integrator, RFA is the servant of the data holders, the researchers who ask and generate the questions, and the program administrators who serve clients and seek to improve outcomes, enhance operations, and realize efficiencies. As such, most of the agency’s work is funded through contracts with other state agencies.

Value—produces value for contributors by concentrating data assets and skills. Every data element held by RFA can be linked and combined to serve a multitude of purposes from addressing a specific research question, to supporting evaluation, to monitoring operations, and tracking outcomes. While there are three key pieces of legislation and more than a dozen others that require coordination among agencies for data sharing and project coordination, the system works well because of the value it produces for all parties involved.

Culture—develops a data culture to champion research and data use for evaluation. RFA efforts cover a broad spectrum from supporting basic research to identifying best practices—these activities seek to create a robust environment for asking questions and using data to answer them.

Access—provides equal access for all users, including state agencies and outside data requestors. The system is open to all comers and maintains consulting services to assist with and track project applications. RFA works with agencies and researchers to produce a wide variety of outputs, including analytic products like data linkage, de-identified data sets, and dashboards with query functions on its public website as well as custom software applications for its agency partners.

Expertise—builds subject matter knowledge to improve data, processes, and results. RFA statisticians oversee sets of agencies, learn about programs, and become internal data experts. Staff uses this knowledge to complete quality control checks, identify issues, and discuss improvements to data and processes with the source agencies. The more substantive knowledge the staff has, the better the exchanges with the source agencies and the results.

Governance and management—relies on transparent processes from data acquisition to destruction.

- Data are acquired via secure transport under terms of MOUs, with mutually agreed specifics in terms of timing and content.
- Data are checked for quality, staged into a probabilistic record linking process, and geocoded. An internal randomly generated ID number allows for later linkage.
- Identifiers are removed and placed into storage, and data files are moved to storage associated with specific statistical units.
- As requests for de-identified and/or aggregate data are received, the project team and compliance officer reviews request and produces an application for data use and any associated data use agreements. These are reviewed by data owners, and a decision is rendered.
- Data are linked to create a project data set containing no identifiers and only the variables required by the project. Any aggregations and statistical analyses are performed.
- Data set is transmitted to the requestor by encrypted transport or deployed into software application for use.
- Data releases are approved by data owners or multi-stakeholder councils and committees.
- Project management monitors use and follows up, as needed, based on the terms of the data use agreement, including properly retiring systems and destroying data.

Security and privacy—prioritizes data security and privacy. The public has overarching and compelling concerns with security, and data breaches can undermine trust in agencies that hold data on behalf of citizens. RFA takes a holistic approach to protecting security and privacy, including enforcing specific provisions—like HIPAA and FERPA, complying with more general areas of regulatory compliance, and employing substantial security and monitoring resources. RFA was one of the first state offices in the country to have an internal Privacy Officer and strives for achieve best practices in this space.

What challenges and opportunities remain?

RFA seeks to build a dynamic and engaged culture around using data and evidence to solve issues for South Carolina citizens. This goal, however, has not been fully fleshed out or institutionalized. Opportunities include the following:

Fully utilize RFA resources. The potential of the Integrated Data System and the staff who supports it has not been fully realized.

- ***Agency continuity.*** Different agency directors have different priorities. It is often difficult to get traction around a set of problems, so they can be addressed on an ongoing basis. This limits the ability of agencies to collect data and develop evidence to inform decision-making.
- ***Grant funding.*** Data use and evaluation are often funded by grants. While grants can seed these opportunities, this funding mechanism tends not to drive long-term cultural shifts. Often, strides are made that evaporate once the funding goes away.

Expand efforts to institutionalize multi-agency collaboration. Many problems cut across agencies and require collaboration from many stakeholders to solve them. Stakeholders should expand efforts to institutionalize multi-agency collaboration. Recent examples of such efforts include the following:

- ***South Carolina Early Learning Extension.*** The program is a collaboration between several South Carolina agencies—the Department of Education, Department of Social Services, First Steps to School Readiness, and RFA. The goal of the program is to extend South Carolina’s statewide longitudinal data system to incorporate early childhood programs and enable access to data that can inform stakeholders about the benefits of these programs.
- ***Coordinating Council for Workforce Development.*** The council was formed to improve coordination efforts around workforce development by state agencies involved in educating and training the state’s workforce. Membership includes the Secretary of Commerce and nine other members from state agencies or organizations involved in education and training.
- ***Big Data for Health Science Activities.*** The University of South Carolina has developed a data center that uses big data and advanced analytics techniques to address health care problems. The center has convened a set of stakeholders, including the Hospital Association, the Medical Association, the Health Sciences of South Carolina, and the Department of Health, around emergent health care issues in the state. The university uses some of the more advanced quantitative techniques that are necessary for the discovery of problems and real solutions. The program has already received several grants from the National Institutes of Health, primarily around HIV and COVID-19.

Harness the value of federal data. There are many challenges states must overcome to access federal data resources that would provide key insights (for example, the [National Directory of New Hires](#) and federal wage data). To improve data sharing for evidence building, there should be a better two-way flow of data from the states to the federal government and vice versa. This process should be driven by collaboration between state and federal actors rather than reporting requirements, as is currently often the case.

Virtual Site Visit 3. The Coleridge Initiative

Background

The [Coleridge Initiative](#) (Coleridge) is a not-for-profit organization that works with federal, state, and local governments to use their confidential data more effectively for public decision-making. Originally established in 2016 at New York University, Coleridge has worked with over 800 government agency staff from more than 250 agencies to develop over 130 pilot projects in their Applied Data Analytics training programs.

Coleridge’s multi-faceted efforts include three components vital to supporting secure data linkage, access, and analysis: the [Administrative Data Research Facility](#) (ADRF), the [Applied Data Analytics training program](#), and the development of “[Rich Context](#)” methods for discovering how public data are used.

Administrative Data Research Facility (ADRF). The ADRF is a secure cloud-based computing platform that allows agencies to safely host and analyze de-identified data sets. The platform was designed to promote collaboration, facilitate documentation, and provide information about data use to agency data owners. The ADRF was established by the Census Bureau with funding identified, in part, to inform the decision-making of the Evidence Commission.

The ADRF has enabled secure access to over 150 confidential data sets owned by more than 75 different agencies at all levels of government. The facility is FedRAMP moderate certified and is listed on the FedRAMP Marketplace. The ADRF has received authorization to operate from the Census Bureau, the U.S. Department of Agriculture, and the National Science Foundation and has an Authorization to Use from the U.S. Department of Health and Human Services. The platform won a [2018 Government Innovation Award](#).

The platform’s data processing and management capabilities include data ingestion, data documentation, data analytics tools, and data stewardship. All access to the ADRF is exclusively via web browser over secure connections, and the system leverages cloud-based services for efficiency, reusability, and cost transparency.

Applied Data Analytics training program. The Applied Data Analytics training program is a project-focused learning approach designed to train government employees and public policy analysts on how to tackle important policy problems by applying modern data analysis tools to their own confidential data. Agency staff are trained through direct use of their data to answer real, present policy questions they face and to develop practical tools after the training ends. A primary goal of the program is also to facilitate the establishment of a community of practice, by building relationships among and between agencies, universities, and non-profits.

Since 2017, this program has partnered with over a dozen top universities and organizations to provide professional development training to over 800 participants across more than 250 organizations.

Rich Context. Coleridge received funding from several major philanthropic foundations and federal agencies to develop machine learning and natural language processing tools to discover how public data are used and by whom. The resulting information can be made available in an Application Programming Interface so that participating agencies can easily provide the public with information about data use, engage with their stakeholders to develop new uses, and create data scorecards with drilldown capabilities. Agencies can also produce data inventories that respond to the mandates in Title II of the Evidence Act.

Coleridge worked with Kaggle (an online community of data scientists and machine learning practitioners) to launch a “[Show US the Data](#)” competition in March 2021. The competition attracted 1,610 data science teams who worked for 3 months to develop models to find data sets used in scientific research publications. From the [winning models](#), Coleridge is building the following three proofs of concept:

- Data usage scorecards that summarize usage for data sets of key interest to participating agencies
- Automated Data Inventories that provide an overview of how data have been used, and by which experts, as well as the topics for which they have been used
- Evidence basis for topics of key interest to agencies by providing lists of data that have been used to study those topics

Why has the Coleridge Initiative been successful?

Value—demonstrates value through a process that is driven by agencies and states. The Applied Data Analytics training program is structured to leverage the ADRF to produce evidence that is relevant, timely, actionable, and scalable. The [Unemployment to Reemployment](#) portal, an interactive tool for exploring the experiences of claimants over time through examining weekly unemployment insurance claims data, is a prime example of how this state-driven process yields value for participants. The portal, which was originally developed by the state of Illinois and has since been adopted by several other states throughout the country is useful because the information it provides is:

- **Relevant.** The portal allows users to analyze real-time labor market information critical for local pandemic response.
- **Timely.** The portal features evidence that is both high frequency (weekly unemployment statistics) and current (updated within a day of the weekly release).
- **Actionable.** Results provide intelligence for statewide and local stakeholders, including governors’ offices (policymaking), state agencies (program administration), and local workforce boards (strategic resource allocation).
- **Scalable.** Through a common data model and standardized code, six states are currently building on the approach of the Unemployment to Reemployment portal.

Access—provides secure data access to meet evidence-based policy goals. Coleridge’s ADRF provides secure data access based on the principles of the “Five Safes”:

- **Safe projects.** The ADRF only includes agency approved projects that have been proposed and agreed upon by agency project and data set stewards. Approved projects require signed agreements and have isolated workspaces within the ADRF, separate from other projects and with controlled access to workspace resources through individual and group memberships.
- **Safe people.** The system of access implemented by Coleridge allows only approved analysts and researchers to access a given project workspace and related resources. All individuals with access to a project on the ADRF are required to complete an on-boarding process in which they must sign data use agreements and terms of use, as well as complete data security training on privacy, confidentiality, and disclosure avoidance. Data resources are explicitly granted based on project requirements and are strictly in a read-only mode to ensure integrity of the data source.
- **Safe settings.** The ADRF is designed to provide secure methods of data transfer for agency microdata, specifically data that include personally identifiable information (PII) in the data set. Only agency-identified and -authorized personnel are invited to perform data transfers. Data transfers happen in a secure environment and are restricted to upload applications only. Additional security protocols include regular vulnerability scanning and third-party penetration testing.
- **Safe data.** The ADRF ensures safe data through two primary mechanisms: the data hashing application and the data stewardship application. For more information, see “Governance and management—Data stewardship” below.
- **Safe exports.** Coleridge prevents the unauthorized removal of any information from within the secure ADRF environment. Researchers must request the export of their work through a formal process of disclosure review. Once an export request is initiated, the request must go through internal and external reviews for disclosure avoidance before data are released to the researcher. Strict standards must be met to suppress or aggregate output that could enable re-identification. Coleridge staff maintains a log of export requests for auditing purposes and to evaluate subsequent requests.

Capacity building—equips agencies and states to harness confidential data and cutting-edge tools. The Applied Data Analytics training program uses a robust “training the trainer” model. The “Leveraging Big Data to Achieve Equity” training program held between July and October 2021 exemplifies how this model works:

- **The challenge.** Many Historically Black Colleges and Universities (HBCUs) and Hispanic-Serving Institutions (HSIs) are not “Level 1” research institutions, and there is a need to build out research activities within these institutions.
- **The objectives.** The “Leveraging Big Data” Applied Data Analytics training program has two complementary objectives: (1) Provide HBCUs, HSIs, and their faculty with guided training and capacity building to enable production of critical research with confidential microdata, and (2) enhance the research capacity of HBCUs and HSIs to inform the advancement of Black and brown populations in postsecondary education across the college to career pipeline.

- **The model.** The training program includes participants from five HBCUs and five HSIs. Three successive cohorts are trained in three successive classes with different levels of engagement. These three unique training program cohorts play different roles as far as receiving training, involvement in curriculum development and facilitation, and leading training program project teams. By the third training, participants are responsible for leading lectures and discussions, overseeing breakout groups, and supporting the development of research projects through final presentations.
- **The data.** Training participants have access to a broad range of data sets through the ADRF, including working with data from National Center for Science and Engineering Statistics' Survey of Earned Doctorates and Survey of Doctoral Recipients, as well as administrative records from universities. Participants use these data sets to explore questions like: *How does federal funding for PhD candidates differ by research field, race, and sex? Where does funding come from? How to design career pathways with funding?*

Partnerships—harnesses multi-state collaborative efforts. The [Midwest Collaborative](#) is a coalition of state workforce and education agencies working in partnership with Coleridge and regional university partners to design a system that enables individual states to answer critical questions that are relevant to societal well-being. State partners set the priorities, contribute data, and establish the agenda for research and data products. Founding states include Iowa, Illinois, Indiana, Kentucky, Michigan, Missouri, New Jersey, Ohio, and Tennessee. The Midwest Collaborative's work to produce the [Multi-State Postsecondary Dashboard](#) demonstrates the possibilities of cross-state collaboration.

- **The program.** In 2020, Coleridge, together with the Ohio State University, delivered an Applied Data Analytics training program designed to use education and workforce data to enhance state data analysis capabilities.
- **The product.** Analysts who took the program, which included participants from Kentucky, Ohio, Tennessee, and New Jersey, worked on projects that evolved into the Multi-State Postsecondary Dashboard. The two primary goals of the dashboard are to demonstrate value of: (1) combining administrative data from multiple states and (2) standardizing data models in reproducing analyses across state lines. The information presented in the dashboard will inform state decision-making in education and workforce policy by providing employment outcome measures from multiple states' workforce data.
- **The possibilities.** Moving forward, the Midwest Collaborative is looking to generate more granular breakdowns, including academic outcomes by major, student demographics, firm-based characteristics (e.g., who is employing graduates and how graduates are moving across state lines), and postsecondary non-completer outcomes. All of this supports the bigger goal of building states' abilities to evaluate cross-state education and workforce outcomes through research and analysis.
- **The promise.** States in other regions, notably the South and the East, are also forming regional collaboratives to jointly develop new products for evidence building and share knowledge across state lines.

Governance and management—provides transparent processes for governing and managing data assets.

- **Governance.** Features of ADRF governance include an enterprise data catalog, management with Rich Context methods, and disclosure review for exports.
 - **Enterprise data catalog.** The system includes an enterprise data catalog with technical metadata.
 - **Rich Context.** The ADRF applies artificial intelligence and natural language processing to manage data assets and build data inventories. For more information, see “Rich Context” above.
 - **Disclosure review.** Coleridge applies a rigorous set of internal and external reviews before exports are released to researchers. For more information, see “Access—Safe exports” above.
- **Data stewardship.** The Data Stewardship Application within the ADRF is positioned primarily as the management and monitoring console for ADRF projects and data stewards. It provides detailed insight on project workspace configurations, ADRF user activity, ADRF user onboarding status, and overall costs of an ADRF project. The application focuses on four primary pillars of information: users, projects, data sets, and agreements.

Policy—builds the capacity of federal agencies to meet the requirements of the Evidence Act. The Evidence Act requires that agencies use data and evidence to inform their work, build measures to inform the public of data use, and provide ways for the public to request specific data assets be prioritized. Coleridge is partnering with federal agencies to leverage Rich Context methods and to develop data usage scorecards and automated data inventories.

What challenges and opportunities remain?

While the ADRF, the Applied Data Analytics training program, and Rich Context methods have produced significant value over the last 5 years, challenges and opportunities remain, including the following:

- **Expanding scope.** Agencies are interested in increasing the value that they provide to their citizens by including data about health, welfare, and criminal justice in the ADRF.
- **Expanding understanding of how data are used.** Rich Context algorithms have the potential to go beyond just looking at peer reviewed journals. One focus is making data inventories and scorecards more comprehensive by expanding the type of text documents to include government reports, Federal Register notices, and public media. A second focus is to build a better understanding of the data ecosystem by documenting which data are used together to address key social problems.
- **Expanding training programs.** The Applied Data Analytics training programs can be expanded to include both more basic and more advanced certificates. The partnerships with universities that have been the hallmark of the programs can be formalized to establish the equivalent of an agricultural extension program for public service.

- **Coordination.** There is an important role for central coordination to harness ongoing efforts for data collaboration across federal, state, and local governments—as well as other non-profit and private-sector organizations. Participation should be driven by value to agencies and program administrators, not by mandates from a central governing body. Instead, a more fluid coordinating entity could work with a network of data providers. The purpose of this role is coordination—not to serve as a single “place” for linking data records. There is certainly potential for a “collaborative of collaboratives.”

Virtual Site Visit 4. Federal Statistical Research Data Centers (FSRDCs)

Background

The U.S. Census Bureau, in partnership with federal statistical agencies and research institutions, operates a network of 31 [FSRDCs](#) in 19 states and the District of Columbia to provide secure access to a range of federal restricted-use microdata for statistical purposes.

FSRDCs are an example of a successful partnership to expand the federal government’s capacity to facilitate external researcher access to data for statistical purposes. Federal statistical agencies collaborate with host organizations—including universities, non-profit research institutions, and government agencies—to enable approved researchers to access confidential data from multiple agencies through a network of secure data enclaves.

The FSRDC network, and demand for its services, has grown significantly over the last 25-plus years. The first remote research data center opened in Boston in 1994, and the newest location is slated to open in Florida in 2021. With this addition, the FSRDC network will have more than doubled in size since 2014. Over the years, the FSRDC network has supported thousands of different projects. There are currently over 900 active researchers working on around 475 active research projects. Most of these projects, about 98 percent, are using data from the Census Bureau and the National Center for Health Statistics.

Recently, the FSRDC network has been exploring ways to provide secure data access through remote access and cloud environments and has participated with the interagency working group developing a standard application process that streamlines project application and approval processes.

Why has the FSRDC network been successful?

Through its statistical and institutional partners, the FSRDC network safely and securely increases access to confidential statistical data for evidence building. Key elements of this success include the following:

Partnerships—brings together federal statistical agencies, research institutions, and researchers to facilitate research for statistical purposes. Currently, the network includes 100+ universities and research institutions, 31 physical data enclaves each with its own Executive Director, 7 statistical agency partners, and nearly 500 research projects. An Executive Committee and a program management office at the Census Bureau oversee and support the FSRDC network. Regular and frequent communication is key to maintaining these relationships. The program management office meets with Executive Directors every week and stays in regular contact with agency partners through liaisons with each agency. In addition, the Executive Committee meets monthly. Meetings with these partners and stakeholder groups are a mechanism for two-way communication to discuss ongoing activities and share needs and concerns.

Access—provides secure access to statistical data in support of participating agency missions. Access modes include secure, networked physical locations as well as remote access using telework technology. FSRDCs secure access—and trust in the system—is built on the principles of the Five Safes.

- **Safe data.** FSRDCs assess data to identify where there is a disclosure risk in the data itself.
- **Safe projects.** Data centers rigorously review proposals, including considering implications for disclosure. There must be a demonstrated need for access to restricted-use microdata.
- **Safe people.** All researchers must be granted Census Bureau Special Sworn Status, which includes an oath of confidentiality for life. In addition, approved users must abide by the legal authority of the data sets—Title 13 (Census Bureau), Title 15 (NCHS), Title 26 (IRS), the Privacy Act, and the Confidential Information Protection Statistical Efficiency Act. All researchers take regular training to ensure compliance with statutes, regulations, and data-sharing agreements.
- **Safe settings.** FSRDCs are secure Census Bureau facilities within host institutions. Each center has a Census Bureau employee on site who serves as a data concierge. Only authorized personnel are permitted to access the facility. The computing environment is secured with access granted at the project level, and there are no unapproved transfers across projects.
- **Safe output.** There is a mandatory disclosure avoidance review for all output, printing is strictly controlled, and no written materials may leave the data center.

Policy—addresses Evidence Act requirements.

- **Expands access.** FSRDCs expand access to data for evidence building through secure physical and virtual enclaves, a standard application, and data linkages. The network offers broad U.S. coverage, multiple access modes (both in-person and virtual environments), and proposal development support. In addition, FSRDCs will be part of a Standard Application Process that will provide a single portal for users to apply for access to restricted data from multiple agencies. Through the FSRDC network, authorized users can link data from multiple sources, for example, linking data for households and employers from the Longitudinal Employer-Household Dynamics (LEHD) program and linking user-provided data to statistical agency data.
- **Supports transparency.** FSRDCs abide by stringent transparency requirements, including identifying who is using the data and for what purpose, publishing the number of ongoing projects and by what agency, and developing an inventory of project metadata. Over the past year, the network has made notably progress in this area, including hiring a Knowledge Transfer Officer to help with these endeavors. For more information on active and completed projects, see the FSRDC [“Projects”](#) webpage.

Value—enables relevant research that yields critical insights. Ongoing projects and feedback from users help improve programs, identify gaps, and create new knowledge for decision-making. Recent examples include the following:

- [Gollust and Karaca-Mandic](#) linked data from the American Community Survey to researcher-provided data to show that increased advertising is associated with declines in un-insurance.
- [Phillips and Herkenhoff, et al.](#) used data from LEHD and business surveys and researcher-provided data to demonstrate that increased credit limits allow unemployed individuals more time to find a job, and once these individuals were working, their earnings were higher and work was done at more productive firms.
- [Babina and Ouimet, et al.](#) leveraged data from Census Business Surveys and the Bureau of Labor Statistics National Compensation Survey to examine the impact on workers and businesses when workers face differential outside employment opportunities, investigating how decisions by workers impact future firm and employee outcomes.
- [Kniffin](#) used data from Census LEHD, NCSSES Survey of Earned Doctorates, and UMETRICS to explore the short, medium, and long-term effects of interdisciplinary research vs. traditional discipline-based research on career outcomes.

Governance and management—establishes clear procedures that support partner institutions and project results.

- **Governance.** FSRDC’s processes are governed by both a program management office that oversees administrative, budget, and technical operations as well as an Executive Committee that provides overall program direction.
- **Administrative support.** FSRDCs rely on several types of agreements, including joint statistical agreements between the Census Bureau and each location that hosts a data center, inter-agency agreements to support partner agency-approved projects, and memoranda of understanding with other federal agencies for other types of data sharing. In addition, over the last few years, the FSRDC network has developed a set of processes to support multi-agency projects. These processes include proposal development review that ensures statutory compliance; project management procedures, including reconciling access restrictions and time limits; and disclosure avoidance review, including managing staff clearances.
- **Technical support.** The program management office tracks all activities throughout the project lifecycle, including modifications and outputs. The office also manages the IT infrastructure, enforces physical and remote IT security protocols, and ensures that researchers comply with all certification, training, and monitoring requirements.
- **Resources.** To support a research data center, the Census Bureau and the institutional partner invest an equitable level of resources. The Census Bureau provides the IT infrastructure and program management, and the host institutions provide the physical space, the Executive Director who acts as the liaison between the FSRDC program and the university, and funding for the Census Bureau staff member who is on site. Over the last three decades, tens of millions of dollars have been put into supporting and expanding the FSRDC network.

Finding the right balance of resources requires continuous adjustment. As the program has evolved, the network has fine-tuned things like the IT infrastructure, staffing, and the layouts of the data centers to accommodate more researchers.

What challenges and opportunities remain?

While the FSRDC network has made great strides in recent years, there are still challenges to overcome, including the following:

Costs. While there are no costs associated with applying for access, there are several potential costs for researchers, including the following:

- ***Special Sworn Status.*** Each researcher must apply for Special Sworn Status from the Census Bureau that is necessary to access data through the FSRDC network. This can cost \$1,800 per person if researchers are doing a non-Census data project.
- ***Data preparation and linkage.*** Agencies may charge fees for cleaning and packaging their data sets. For example, data linkages can cost \$20,000 per file at the Census Bureau if a researcher wants to bring in their own data.
- ***Facilities.*** Each FSRDC location may charge its own fees, and costs vary by location and researcher affiliation (researcher networks often offer discounts for their members).

Timeframe from proposal to access. The time needed to develop a proposal depends on the researcher and coordination with the agency contact. Once submitted, most agencies complete their reviews of the proposal within 3 months. Once approved, the Census Bureau starts the Special Sworn Status process, which can take another 3 months. Overall, the process takes about 6–12 months, and Special Sworn Status is good for 5 to 7 years.

Access to state data. FSRDCs have agreements with many states (for example, through the LEHD program) and with agencies who provide state-level data (like Supplemental Nutrition Assistance Program benefits). While the network is working to bring these data into the program, it is complicated because there are many federal and state-level requirements for access and review of outputs. For example, LEHD is a partnership between the Census Bureau and 47 or 48 active partner states. Around 20 of these states allow researchers to access their data if they meet Census Bureau requirements, and the remaining states reserve the right to review the research proposal. Researchers regularly gain access to data from about 30 states. The program is working to strengthen these relationships and improve access to state data sets. The Standard Application Process currently being developed could also be expanded to include data from the states or any other provider.

Virtual Site Visit 5. The U.S. Chamber of Commerce Foundation Jobs and Employment Data Exchange (JEDx)

Background

The U.S. Chamber of Commerce Foundation's [Jobs and Employment Data Exchange \(JEDx\)](#) is a data standards-based approach for how employers can produce enhanced and more timely data on both jobs and employment. JEDx is a unique opportunity to modernize America's workforce through a national public-private partnership and data trust.

One of the public's biggest knowledge gaps is accurate, timely, and trusted data for the dynamic U.S. economy. As the United States emerges from a historic economic downturn and seeks to put Americans back to work, there is a critical need for: (1) improved labor market information and (2) enhanced employment data for evidence-based policymaking and the administration of government programs (for example, unemployment insurance, or UI).

Through JEDx, the Chamber of Commerce Foundation has assembled a unique coalition of state and national public and private sector partners, stakeholders, and leaders that stands ready to close this gap on a national scale. JEDx will begin testing data and use cases as early as 2022 with the goal of forming a public-private data trust by 2024.

The vision of JEDx is to streamline and improve how employers report data to government agencies, produce better longitudinal data about jobs and employment to power new workforce analytics while protecting privacy, and empower Americans with data and trusted records to verify their work history as well as their eligibility for government benefits.

Why has JEDx been successful?

While JEDx has only recently entered the design and test phase of its work, the groundwork laid during the planning phase and the lessons learned from its demonstration projects will help determine the ultimate success of the program. Key elements for success include the following:

Partnerships—explores how states, technology partners, and employers can improve data and evidence for decision-making. This includes implementing standards and sharing employment data, potentially through a public-private data trust and shared services. Data sharing under this partnership would not come from mandates but from agreements among peers—this is a new take on an old problem.

Value—reduces costs and creates higher value for stakeholders. Potential benefits to key stakeholders include the following:

- **Workers and learners.** JEDx stands to offer better information on job opportunities and requirements and more current and trusted records for use when applying for jobs, education opportunities, and government benefits.
- **Education and workforce partners.** JEDx could improve information on in-demand jobs, skills, and credentials and provide better career guidance for learners.

- **Government agencies.** The program has the potential to enhance information on labor markets and employment outcomes to support evidence-based policymaking and improve the administration of government programs and benefits, including UI. For example, the Biden Administration is very focused on competitiveness—understanding the characteristics of the workforce is key to this goal.
- **Employers.** JEDx could reduce federal and state reporting burden and offer better communication of in-demand jobs and skills to job seekers and programs that prepare people for jobs.

Governance and management—harnesses the power of a data trust, data standards, and advisory groups.

- **Data sharing agreement.** JEDx uses an agreement among parties to share, access, and use data for agreed upon and permitted uses.
- **Data standards.** JEDx is built on a standardized format for how data are organized, so they can be shared, compared, and discovered. Data standards allow for data to be organized and compared in citizens' daily lives.
- **Advisory groups.** The Chamber of Commerce Foundation is using a transparent and open process to engage experts from around the public and private sectors. The foundation has formed an advisory committee consisting of over 50 public and private members representing government, employers, human resources (HR) technology companies, and other stakeholders to inform the JEDx program. In addition, the foundation is establishing a National Leadership Team to provide guidance in developing the public-private approach for improving federal and state reporting, starting with state UI reporting and the use of data for public-private workforce analytics.

Policy—leverages existing policies to improve the quality of workforce data. There is OMB guidance that directs agencies to participate in standard-setting bodies that transcend government, making it feasible for external stakeholders to inform federal agencies on those definitions. Most employers use HR vendors like ADP, so the HR vendors are key participants who can standardize the definitions in their systems, allowing employers to report to vendors, and then those vendors can submit data to meet multiple reporting and compliance requirements.

What challenges and opportunities remain?

As the JEDx program designs and tests its system, pilots demonstration projects, and moves toward implementation, a variety of challenges and opportunities remain, including the following:

Broader public-private approach. JEDx will consider both public and private workforce analytics that include statistical and evidence-based policy applications.

Empowering individuals. The program must lay the foundation for empowering individuals with their data and trusted records to use in accessing government programs. The National Leadership Team will also explore ways to engage everyday Americans in the decision-making process for JEDx; this will be an important part of the governance structure.

Data standards. JEDX will identify opportunities for government agencies to use public-private data standards to inform and improve data collection and use. For example, program staff are working with stakeholders to evaluate different employment relationships and definitions—there are many differences in how states address the self-employed. Another consideration is how to geocode this work.

Reduce costs and improve quality. JEDx projects will explore new roles for employers and HR service providers in organizing and sharing data to reduce reporting costs and improve data quality.

Privacy and security. The program will set the rules of the road for privacy and security, including establishing criteria for data sharing, authorized users, and approved users; the National Leadership Team is looking at these issues, including exploring technology and governance.

Virtual Site Visit 6. The National Center for Health Statistics (NCHS) Data Modernization Activities

Background

The NCHS collects and disseminates statistical data to guide public health and inform health policy decisions. NCHS data:

- Inform and measure the impact of public policies and programs, including data on COVID-19, opioid overdose deaths, and health insurance coverage
- Identify health disparities and use of health care by race, ethnicity, socioeconomic status, and geographic region
- Document the health status of the U.S. population and monitor health indicators
- Track access to and use of the health care system to support health services research
- Track morbidity and mortality rates/causes throughout the United States

Collaborating with other public and private health partners, NCHS uses a variety of data collection mechanisms to obtain accurate information from multiple sources. Sources of data collection included the following:

- Birth and death certificates
- Patient medical records, including electronic health records
- Personal interviews (in households and by phone)
- Standardized physical examinations and laboratory tests
- Health care facilities and providers

The Data Modernization Initiative has been a key part of business at NCHS for the past decade, through a focus on improvements in vital statistics data collection, analysis, and release. Beginning in FY 2020, Congress began to appropriate additional resources for data modernization at the Centers for Disease Control and Prevention (CDC), with some funding awarded to NCHS. Data modernization at NCHS focuses on improved data access, collection of new data sources, and embarking on new data analyses, including the use of data science techniques and increased data linkages. These data modernization efforts have increased partnership opportunities across CDC, state and local health departments, and other federal and non-federal organizations.

Why has NCHS been successful?

NCHS data modernization efforts have greatly improved the availability of health care data for evidence building. Key elements of this success include the following:

Culture—offers new paradigm for data sharing across the health statistics ecosystem. The goal of the Data Modernization Initiative is to move from siloed and brittle public health systems to connected, resilient, adaptable, and sustainable systems. The initiative helps address known problems, yielding improved data for evidence building like better and faster hot spot analysis, identifying vaccine

desserts, tracking diseases at the neighborhood level, integrating climate data, and performing interventions and evaluating them in real time to course correct.

Access—deploys a tiered access model to expand access and use of data assets. This model includes the following tiers:

- **Open data assets.** NCHS provides access to public-use data files.
- **Web-based query system.** Users can create tabular data views via CDC’s [WONDER](#) portal.
- **Physical data enclaves.** NCHS operates four data centers and is part of the national FSRDC network (for more information, see “Virtual Field Trip Featuring Federal Research Data Centers”). Physical data enclaves enable vetted researchers to access confidential data in a controlled, secure environment.
- **Virtual data enclave (VDE).** VDE is under development and will provide a tiered access framework that has many benefits: supports implementation of the Evidence Act, adds another access tier, disseminates data to wider audiences, eliminates barriers that exist with physical location, decreases costs for researchers, expands access to confidential data to a whole new class of researchers.

Data standards—leverages metadata and document identifiers for data sharing and discovery. Key elements include the following:

- **Open, accessible, and high-quality data.** NCHS efforts to provide open data and improve data quality build on statutes, official guidance, executive orders, and recognized frameworks, including Title II of the Evidence Act; OMB memos on open data, transparency, and restoring trust in government; and the Federal Committee on Statistical Methodology’s data quality framework.
- **Data, metadata, and customized tags.** The agency expanded access to NCHS assets on [data.cdc.gov](#), including providing consistent metadata standards to match the U.S. Department of Health and Human Services (HHS) and CDC’s Enterprise Data Catalog system. This managed vocabulary reflects both the nature of NCHS data and its relation to other public health resources.
- **Consistent presentation.** NCHS presents its data assets through [data.cdc.gov](#) with standard fields and descriptions.
- **Reporting and tracking outputs.** The agency is updating its citations to make it easier to report and track outputs. All new NCHS reports include a document object identifier that creates a machine-readable citation for tracking, reporting, and finding relevant work. NCHS is also developing the means to cite provisional and real-time survey results.

Partnerships—leverages federal, state, and local partnerships to improve data quality and timeliness. The [National Vital Statistics System](#) (NVSS) is the oldest and most successful example of inter-governmental data sharing in the public health realm and is the mechanism by which NCHS collects and disseminates the nation’s vital statistics, including birth certificates, death certificates, and fetal death reports. Key aspects of the program include the following:

- **Jurisdictional partners.** NVSS is a decentralized, cooperative system comprising 57 jurisdictions: 50 states, the District of Columbia, New York City, and 5 territories.

- **Coordination.** NCHS provides coordination for the program, including standard certificates and forms, instruction and coding manuals, training and instructional materials, and model law to govern data collection.
- **Modernization.** The system has developed over time, driven by modernization work and challenges, including the development of electronic birth and death registration systems (beginning in the 1990s), birth and death certificate revisions (2003), the Good to Great partnership (formed in 2010), and the Vital Statistics Rapid Release Program (starting in 2014). The COVID-19 pandemic has accelerated these modernization activities.
- **Resources.** NCHS purchases vital records from the jurisdictions through the Vital Statistics Cooperative Program contracts. There has been some funding over the last decade to help develop state systems, including work on systems interoperability and data standards. More recently, the CARES Act provided \$77 million for NVSS modernization—all 57 jurisdictions will receive funding to work toward required activities. These awards will help the recipients move closer to providing data to NCHS using common standards. More technically mature recipients may use the funding to address their own needs and challenges working cooperatively with NCHS.
- **Results.** The system produces annual data files and publications, including public use and restricted use files, inputs to CDC’s WONDER query tool, National Vital Statistics Reports, Data Briefs, and Health E-Stats. In addition, thanks to investments and modernization efforts, NVSS releases timelier and higher-frequency statistics and special analyses, such as quarterly versions of regular files and reports, weekly pneumonia and influenza mortality (since 2014–15 flu season), monthly drug overdose mortality counts (since August 2017), and provisional death counts by week with geographic and demographic detail (since April 2020).

Data linkage—evaluates methods for new linkage opportunities. The [NCHS data linkage program](#) is over 40 years old and has grown over time. Key aspects of the program include the following:

- **Functions.** This cross-cutting program: (1) links health survey data with data collected from vital and administrative records to support high-quality research and program evaluation, (2) provides a standardized, efficient approach for creating official NCHS files, (3) maximizes the value of NCHS population-based surveys by allowing for analyses that are not possible with each data source alone, and (4) models approaches that can be applied inside and outside HHS, including enhancing processes for linking data files (that is, linkage algorithms), improving methods for evaluating data quality, and exploring innovative methods for maintaining secure access to linked data (for more information, see “Privacy and Security” below).
- **Legal authority.** The data linkage program supports multiple federal data initiatives, including CDC’s strategic framework, the Federal Data Strategy Action Plan, OMB directives on the use of administrative data for statistical research, and the Evidence Act.
- **Recent accomplishments.** NCHS implemented new probabilistic linkage algorithms that incorporate machine learning techniques to improve linkage accuracy and efficiency; released updated linked files, including data sets on mortality, Medicare, and housing; expanded

linkage activities to include additional sources of survey data; strengthened partnerships for new linkages that address important policy questions (for example, with Veterans Affairs and Medicaid data); and provided analyses of linked data to demonstrate potential uses.

Privacy and security—explores privacy-protecting approaches and maintains physical and virtual security.

- **Emerging privacy technologies.** NCHS is [exploring privacy preserving record linkage](#) (PPRL) to expand data linkages and maintain high data quality. The agency recently ran a case study using PPRL to replicate results from a known linkage, showing very similar results with PPRL as with unencrypted identifiers. NCHS also piloted methods in synthetic data creation for public use files to improve accessibility and reduce disclosure risk.
- **Physical and virtual security.** Security for NCHS physical enclaves mirrors protocols and procedures of the FSRDC network (for more information see “Virtual Field Trip Featuring Federal Statistical Research Data Centers”). VDE security involves several similar dimensions, including a data use agreement between the researcher’s employer and NCHS, hosting data on secure servers in Atlanta CDC headquarters, CIPSEA status for CDC employees and contractors, restrictions on download or copy/paste of data or outputs, disclosure reviews of output before release, and monitoring logs for researcher activity.

What challenges and opportunities remain?

While NCHS modernization efforts have produced great benefits over the last decade, challenges and opportunities remain, including the following:

Consistency, data standards, and systems interoperability. A major aspect of NVSS modernization activities is consistency across data sets, and this work is being further supported by recent CARES Act funding to all 57 partner jurisdictions. NCHS will continue to advance these efforts to achieve consistency and standardization across data flows, producing richer opportunities for data linkages and analyses.

Technical support. NCHS will continue to explore ways to provide resources to states and help them build capacity to capture and validate data before it rolls up to the federal level. Having a robust quality control/quality assurance program at the jurisdictional level is critical. NCHS uses several quality control methods, including completing a rapid review of data to identify anomalies, outliers, and incomplete data and leveraging a targeted webservice for front-end validation. NCHS is looking to increase training resources to support these activities.

Privacy. NCHS will continue to pilot and test novel approaches for securely expanding access to health data for evidence building. As its work with PPRL and synthetic data has shown, there are both output and input disclosure considerations. For instance, PPRL handles input privacy but not necessarily output privacy; this technology enables data to link to more sources but does not mean that more data can be released securely without disclosure risks.

Upscaling virtual data enclaves. NCHS has received conditional approval for its VDE program; however, the agency will need to address challenges with data hosting, funding, and schedule.

- **Data hosting.** The CIPSEA statute, as written, can make it challenging to host data assets on the cloud. Further complicating the issue is that data need to reside on CDC’s cloud provider, but NCHS does not hold the contract for that service. One possible solution would be for the federal government to develop an enterprise-wide VDE for federal statistical agencies that is “certified” ready to host CIPSEA protected data.
- **Funding.** NCHS has received kickoff funding for FY 2021 and second-year funding through May 2022. The agency will consider future VDE costs and how to cover those costs moving forward. For example, can research data center fees cover VDE costs?
- **Schedule.** NCHS will need to determine how and when the VDE could move to CDC’s cloud provider.

Data sharing for timely decision-making on both sides of the data stream. Successful evidence building involves two-way data sharing between NCHS and the states. The agency will continue to improve both the data flowing from states to the federal government as well as data flowing back down to the states to better inform their own decision-making.

Increasing geographic granularity. Currently, NVSS information is available for all counties through restricted access files, and the WONDER platform allows access but enforces suppression rules. Subcounty data are available; however, there are concerns with data quality. NCHS will begin collecting address information and then geocoding that information and returning it to jurisdictions as a service, making it possible to tabulate statistics at finer geographic levels.

Appendix F. Case Study Criteria Matrix

This appendix presents a framework for evaluating possible case studies to inform the Committee’s work, describes the components of this matrix, and presents two examples. The [Case Study Criteria Matrix](#) was presented and discussed at the July 2021 public meeting.

Table F1 presents the matrix. The **ROWS** indicate major **challenges** to evidence-based decision-making (that is, the roadblocks that need to be addressed to bring data together successfully). The **COLUMNS** represent mechanisms for showing **external validity** by connecting to “authoritative” sources. “Good” case studies “live” at the intersection of the rows and columns by addressing existing challenges AND providing timely, relevant responses to outside priorities.

Case studies can include past, present, and proposed projects. There are benefits for evaluating all three types of projects. For example, past projects provide lessons learned, ongoing projects present opportunities to build on existing efforts, and future projects offer a platform for the Committee to propose new, targeted initiatives. The Committee is also building a Project Inventory; for more information, see Appendix D.

Table F1. Case Study Criteria Matrix

	Administration’s Priority Areas, including Executive Orders	Evidence-Building Implementation Items, including Evidence Act and Federal Data Strategy	Commission on Evidence-Based Policymaking Recommendations (not included in Evidence Act)
Policy			
Administrative			
Data Governance and Management			
Stakeholder Engagement			
Technical			

Table F2 describes the five challenge “buckets” in the case study criteria matrix and provides examples of the types of questions a relevant case study should address.

Table F2. Case Study Criteria: Challenges

Barrier	Description
Policy	<p>Challenges related to legislation, regulations, and guidance</p> <p><i>Key questions:</i> What are the remaining legislative barriers to data access and sharing? How to address them? What additional guidance is needed from OMB to move forward? What statistical system practices can/must (under law/policy) be imported into NSDS (especially relevant to Evidence Act implementation under way now)? How can the federal government improve guidance for data providers at the state/local level to achieve higher quality data inputs?</p>
Administrative	<p>Challenges related to administrative practices</p> <p><i>Key questions:</i> What statistical system practices could be improved to ensure the system works well together? How could standard operating procedures, common data sharing agreements (MOUs), and common consent forms improve access to and use of data for evidence building?</p>
Data Governance and Management	<p>Challenges related to tiered access, metadata existence, data quality, transparency, and accountability</p> <p><i>Key questions:</i> Who can be an authorized user? What are the authorized uses? How can NSDS support existing efforts like the Standard Application Process? How to provide standardized metadata? What is the appropriate data quality framework for NSDS output/inputs? How to measure the use of data and gauge NSDS success? How to provide transparency for data controllers and subjects?</p>
Stakeholder Engagement	<p>Challenges related to internal and external stakeholder needs, statistical system coordination, and capacity building</p> <p><i>Key questions:</i> How to address the needs of government at all levels (federal, state, local)—both to facilitate the transfer of quality data into NSDS and to ensure the system supports evidence-based policymaking, evaluation, and monitoring? How to address the privacy concerns of stakeholders and communicate protections appropriately? How to get all 13 statistical agencies to cooperate and participate? How to build workforce and user capacity? How to improve data literacy? What training and tools are needed?</p>
Technical	<p>Challenges related to IT systems, evidence production, R&D, and methodological capacity</p> <p><i>Key questions:</i> What are the relevant technologies and methods for data acquisition, linkage, and access? How to get data to/in NSDS? How to protect the data? How to connect users to NSDS? How to get results out of NSDS? What role does NSDS have to play in R&D? How to build capacity for innovation, so that technologies are not out-of-date in a few years?</p>

Table F3 describes the authoritative sources in the case study criteria matrix that can be used to demonstrate external validity.

Table F3. Case Study Criteria: External Validity

Source	Description
Administration’s Priority Areas, including Executive Orders	Supports the main objectives of the Biden Administration, including those outlined in Executive Orders and memoranda on topics like scientific integrity, equity, COVID-19, economic recovery, and climate change
Evidence-Building Implementation Items, including Evidence Act and Federal Data Strategy	Furthers the implementation of the Evidence Act (by title) and the Federal Data Strategy
Commission on Evidence-Based Policymaking Recommendations (not included in Evidence Act)	Affirms, expands, and recasts recommendations of the Commission on Evidence-Based Policymaking, as appropriate, to meet the needs of the current evidence-building ecosystem

Example 1. The Federal Committee on Statistical Methodology's Data Protection Toolkit

Status: Present (Phase 1 completed, Phase 2 under development)

Description: Provides information and resources for Statistical Officials, Chief Data Officers, Privacy Officials, and statistical agency staff on disclosure risk assessment, disclosure risk mitigation, controlled access to sensitive data, assessing data accuracy, and governance

Table F4. Case Study Criteria Matrix for the Data Protection Toolkit

	Administration's Priority Areas, including Executive Orders	Evidence-Building Implementation Items, including Evidence Act and Federal Data Strategy	Commission on Evidence-Based Policymaking Recommendations (not included in Evidence Act)
<p>Policy: Provides a central resource for guidance, tools, and templates to help agencies expand access to data assets while avoiding the unintentional release of information that could be used to re-identify individual people or entities</p>	<p>Restoring trust in government: Outlines how statistical agencies are keeping their promises to the American public by protecting private information when releasing data publicly</p> <p>Equity: Highlights ways to provide additional detailed data on small populations while protecting the confidentiality of these more easily identifiable data</p> <p>Data-driven response to COVID-19: Provides an inventory of resources that can be used to inform a review of and improvements to the federal government's existing approaches to open data that are open to anyone at any time from any agency</p>	<p>Action 15 of the Federal Data Strategy: Updates FCSM Working Paper No. 22 to address emerging privacy risks and improvements to Statistical Disclosure Limitation techniques since the 2005 revision</p> <p>Evidence Act Title II: Contains 200+ resources and tools that facilitate open data practices while managing data privacy risks</p> <p>Evidence Act Title III: Provides guidance on the use of tiered access aimed at expanding access to restricted data in a safe and appropriate way</p>	
<p>Administrative: Includes templates (e.g., disclosure review board charter, data use agreements) for assessing, managing, and mitigating the re-identification risk of individuals or enterprises in U.S. federal data products</p>		<p>Action 15 of the Federal Data Strategy: Can be used by agencies to develop and implement cost-effective data protection programs</p>	<p>Rec. 3-1: Offers model for a comprehensive risk assessment and disclosure review board approval before public release of de-identified data</p>
<p>Data Governance and Management: Provides an integrated repository of recommendations and guidelines for the selection and use of effective statistical disclosure limitation methods and resources and tools to assist agencies in designing, improving, and implementing the data access and protection aspects of data governance processes</p>	<p>Restoring trust in government: Equips agencies to enhance public trust by applying proper access and protection approaches; supports transparency by cataloging data protection strategies, including agency policies and charters</p> <p>Equity: Emphasizes need for more granular data on small populations and highlights the role of data governance and management techniques around confidentiality and disclosure avoidance to meet that need</p>	<p>Evidence Act Title II: Provides data governance boards with the right tools to support strong data protection programs, as departments enhance their data inventories and make data as open as possible</p> <p>Evidence Act Title III: Draws parameters around the management of tiered access and includes an inventory of tiered access modes used by federal agencies, including restricted use enclaves and automated data tools</p>	
<p>Stakeholder Engagement: Engages federal agencies with compiling, disseminating, and adopting best practices for data protection and tiered data access</p>	<p>Restoring trust in government: Training strategy establishes process for capacity-building and buy-in within the federal community</p>	<p>Evidence Act Titles II and III: Engages statistical agencies by compiling open data practices and models for tiered access across the statistical system</p>	
<p>Technical: Provides a rich library of automated data protection tools, including software packages and tiered access frameworks</p>	<p>Restoring trust in government: Presents technological solutions for increasing access to federal data assets, while protecting confidentiality; by developing this resource, the federal government is demonstrating commitment to increasing access to federal data assets while simultaneously strengthening privacy and confidentiality promises now and in the future</p>	<p>Evidence Act Title III: Highlights risk assessment tools, controlled access technologies, and emerging methods that can be used to expand secure access to statistical assets</p>	<p>Rec. 4-1: Catalogues existing technical expertise and infrastructure that can be built on and enhanced to ensure sufficient capacity in secure record linkage and data access for evidence building</p>

Example 2. U.S. Department of Agriculture (USDA) Enterprise Data Analytics Platform and Toolset (EDAPT)

Status: Present or ongoing

Description: Over the last 3 years, USDA has made great strides in leveraging the department’s vast data sets as a strategic asset and advancing evidence-building activities. This includes establishing the Chief Data Officer (CDO) and Assistant CDO roles, empowering centralized analytics teams, developing enterprise-wide data dashboards, and launching the EDAPT to integrate data analysis and enable a broad range of analytics.

Table F5. Case Study Criteria Matrix for USDA EDAPT

	Administration’s Priority Areas, including Executive Orders	Evidence-Building Implementation Items, including Evidence Act and Federal Data Strategy	Commission on Evidence-Based Policymaking Recommendations (not included in Evidence Act)
Policy: Provides resources and tools to help USDA expand access to data assets	Restoring trust in government: Supports a full-fledged USDA Data Strategy that has delivered value throughout the department and improved its customer service	Federal Data Strategy (identify priority questions and data sets): EDAPT is the established central repository for data sets needed to answer priority business questions from department leadership	
Administrative: Provides a common platform to feed data-driven decision-making and evidence building across the spectrum of administrative processes, including HR, finance, and operations; provides first-of-its-kind comprehensive suite of administrative dashboards and analytic tools accessed on a single platform	Restoring trust in government: Improves programs delivering support to citizens, such as farm loans and disaster assistance, by gaining fast, data-driven insights for mission delivery	Evidence Act Title I (systematic planning): Features an organizational model for centralized analytics teams to answer key cross-cutting questions Evidence Act Title III (expanding access for statistical purposes): Helps bridge the gap between administrative and statistical agencies by creating an environment where data, tools, and computing power can be shared on a common platform to resolve major historical challenges	Rec. 5-4 (sufficient resources): Aligns administrative processes to support evidence building within USDA; drives a new, data-driven decision culture by removing manually burdensome processes and enabling employees to focus on their mission of delivering programs that support America’s people, farmers, ranchers, and producers
Data Governance and Management: Features a governance process for cataloging data and standardizing analytics tools that enable collaboration and data sharing across the organization	Data-driven response to COVID-19: Thanks to a clear governance process and availability of standard tools, USDA was better able to respond when the need for data was exacerbated during the pandemic	Evidence Act Title II (coordination and open data planning): Establishes new data functions across the organization, creating the role of Assistant Chief Data Officer within each mission area to lead data strategy and governance activities; Open Data Platform enables USDA to publish dashboards that provide the public and third-party authenticated users with the ability to draw data-driven insights, as well as download important data about USDA programs	Rec. 5-3 (evidence building across departments): As a microcosm of the evidence ecosystem, provides insights into coordinating evidence building across departments; establishes consistency and governance in managing data across mission areas so that data can be more easily shared and used
Stakeholder Engagement: Enables partnerships across department, mission area, program area, Statistical Official, CDO, and Evaluation Officer functions that substantially advance the department’s evidence-building capacity both internally and externally; fosters data analytics training and development through centralized analytics teams and communities of practice	Restoring trust in government: By sharing its experiences and ideas through collaborative forums, such as the CDO Council, USDA offers insights in how to advance all agencies’ missions and improve data-driven accountability across the federal government	Evidence Act Title II (coordination): Leverages communities of practice to share best practices, as well as examples of dashboard development and data usage, effectively raising knowledge levels of departmental staff Evidence Act Title III (expanding data access): Responds to the data providers’ concerns about survey burden, perceived duplication of data requests, and rising cost of survey data collections by expanding access to and use of available data, including administrative data, geospatial data, economic data, and environmental data	Rec. 5-3 (interagency Councils): See “Stakeholder engagement: Restoring trust in government”
Technical: Eliminates technology as a barrier to data sharing and analytics within and across agencies; integrates data analysis for more than 150 sources from every corner of USDA as well as outside the department; delivers over 120 distinct dashboards that illuminate critical data	Data driven response to COVID-19: Features COVID-19 dashboards designed to identify risk to its 100,000-person workforce and enables nearly 5,000 leaders across USDA to make informed, risk-based decisions in a time of crisis	Evidence Act Title II (coordination, data management): Turns data into automated, real-time insights for USDA leadership; enables program areas to conduct program-specific analytics with common tools Evidence Act Title II (coordination, Open Data planning): Co-locates the tools, diverse data, and advanced computing capacity that enable agencies to develop complex machine learning models utilizing all available data in new ways	Rec. 5-3 (evidence building across departments): Identifies and helps solve major cross-cutting strategic questions using data analysis and advanced data analytics techniques and methods (including advances in data science such as machine learning, neural networks, and other forms of artificial intelligence)

Appendix G. Crosswalk of Recommendations from the Evidence Commission to ACDEB Focus Areas

This appendix lists each recommendation from the Evidence Commission, tags its status related to the Evidence Act, and flags the possible ACDEB focus areas of interest. The crosswalk was discussed at the May 2021 public meeting. Several of the focus areas used this crosswalk to guide their discussions, and the Committee made specific recommendations in this report on items the Evidence Act did not address. The Committee will continue to consider the Evidence Commission recommendations and identify areas where additional recommendations should be made in Year 2.

Table G1. Evidence Commission Recommendations to ACDEB Focus Area Crosswalk

Recommendations and Notes	Addressed by Evidence Act? Y = Yes, N = No, P = Partially	Legislation and Regulations	Governance, Transparency, and Accountability	Technical Infrastructure	Government Data for Evidence Building	Other Services/ Capacity-Building Opportunities
Rec. 2-1: The Congress and the President should enact legislation establishing the National Secure Data Service (NSDS) to facilitate data access for evidence building while ensuring transparency and privacy. The NSDS should model best practices for secure record linkage and drive implementation of innovative privacy-enhancing technologies.	P		✓	✓	✓	✓
Rec. 2-2: NSDS should be a service, not a data clearinghouse or warehouse. The NSDS should facilitate temporary data linkages in support of distinct authorized projects.	P		✓	✓	✓	✓
Rec. 2-3: In establishing NSDS, the Congress and the President should amend the Privacy Act and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) to require new stringent privacy qualifications as a precondition for NSDS to acquire and combine survey and administrative data for solely statistical purposes. At the same time, the Congress should consider additional statutory changes to enable ongoing statistical production. For CIPSEA, see 44 USC sec 3581—presumption of accessibility.	P	✓				
Rec. 2-4: The Congress and the President should review and amend, as appropriate, statutes such as Title 13 of the U.S. Code to allow statistical uses of survey and administrative data for evidence building within the CIPSEA secure environment.	N	✓				
Rec. 2-5: The Congress and the President should consider repealing current bans and limiting future bans on the collection and use of data for evidence building.	N	✓				
Rec. 2-6: The Congress and the President should enact statutory or other changes to ensure that state-collected administrative data on quarterly earnings are available for solely statistical purposes. The data should be made available through a single federal source for solely statistical purposes.	N	✓	✓	✓	✓	✓
Rec. 2-7: The President should direct federal departments that acquire state-collected administrative data to make them available for statistical purposes. When there is substantial federal investment in a program, federal departments should, consistent with applicable law, direct states to provide the data necessary to support evidence building, such as complete administrative data when samples are already provided.	N	✓	✓	✓	✓	✓
Rec. 2-8: OMB should promulgate a single, streamlined process for researchers external to the government to apply, become qualified, and gain approval for access to government data that are not publicly available. Approval would remain subject to any restrictions appropriate to the data in question. For more information, see USC 44 sec 3583—Standard Application Process; actionable soon.	Y	✓	✓	✓	✓	
Rec. 3-1: The Congress and the President should amend the Privacy Act and CIPSEA to require federal departments to conduct a comprehensive risk assessment on de-identified confidential data intended for public release. De-identified confidential data subject to the Privacy Act and CIPSEA should only be made available after a disclosure review board: (1) approves the release and (2) publicly provides the risk assessment and a description of steps taken to mitigate risk. For CIPSEA, see 44 USC sec 3582—expanding secure access.	P	✓	✓			✓
Rec. 3-2: The President should direct federal departments, in coordination with NSDS, to adopt state-of-the art database, cryptography, privacy-preserving, and privacy-enhancing technologies for confidential data used for evidence building.	N			✓		

Table continues

Table G1. Evidence Commission Recommendations to ACDEB Focus Area Crosswalk (Continued)

Recommendations and Notes	Addressed by Evidence Act? Y = Yes, N = No, P = Partially	Legislation and Regulations	Governance, Transparency, and Accountability	Technical Infrastructure	Government Data for Evidence Building	Other Services/ Capacity-Building Opportunities
Rec. 3-3: The President should direct federal departments to assign a senior official responsible for coordinating access to and stewardship of department's resources for evidence building in collaboration with senior department information technology, privacy, and other leaders. A Principal Statistical Agency (PSA) head, or other appropriately qualified senior official, should serve this function. For more information, see the Evidence Act—creation of Statistical Officials and CDOs; additional action may be needed.	Y	✓				
Rec. 3-4: The Congress and the President should enact legislation to codify relevant portions of OMB Statistical Policy Directive No. 1 to protect public trust by ensuring that data acquired under pledge of confidentiality are kept confidential and used exclusively for statistical purposes. For more information, see USC 44 sec 3582—statistical agency responsibilities; additional action may be needed.	Y	✓				
Rec. 4-1: NSDS should be established as a separate entity in the Department of Commerce that builds upon and enhances existing expertise and infrastructure in the federal government, especially at the Census Bureau, to ensure sufficient capacity in secure record linkage and data access for evidence building.	P		✓	✓	✓	✓
Rec. 4-2: NSDS should establish a Steering Committee that includes representatives of the public, federal departments, state agencies, and academia.	P		✓			
Rec. 4-3: To ensure exemplary transparency and accountability for the federal government's use of data for evidence building, NSDS should maintain a searchable inventory of approved projects using confidential data and undergo regular auditing of compliance with rules governing privacy, confidentiality, and access (that is, a "transparency portal").	N		✓	✓		
Rec. 4-4: NSDS should have specific administrative and implementation flexibilities including the ability to leverage public-private partnerships and to collect and retain user fees.	N		✓			
Rec. 4-5: OMB should increase efforts to make information available on existing federal data sets including data inventories, metadata, and data documentation in a searchable format. For more information, see the Evidence Act—Federal Data Catalog and agency data inventories; OMB working on guidance.	P	✓	✓			
Rec. 5-1: The President should direct federal departments to increase capacity for evidence building through the identification or establishment of a Chief Evaluation Officer, in addition to needed authorities to build a high performing evidence-building workforce. For more information, see the Evidence Act—Evaluation Officers, OMB evaluation policy, evaluation job series, EO council, capacity assessments, etc.; additional action needed?	Y	✓				
Rec. 5-2: The Congress and the President should direct federal departments to develop multi-year learning agendas that support the generation and use of evidence. For more information, see the Evidence Act—learning agendas; additional action may be needed.	Y	✓				
Rec. 5-3: The Congress and the President should direct OMB to coordinate the federal government's evidence-building activities across departments, including through any reorganization or consolidation within OMB that may be necessary and by bolstering the visibility and role of interagency councils. Note: Council activity is under way.	N	✓				
Rec. 5-4: The Congress and the President should align administrative processes to support evidence building, in particular by streamlining approval processes for new data collections and using existing flexibilities in procurement policy.	N	✓			✓	✓
Rec. 5-5: The Congress and the President should ensure sufficient resources to support evidence-building activities about federal government programs and policies.	N	✓			✓	✓

6. References and Acronyms

6.1. References

- “2018 Government Innovation Awards: Data Mashups at Government Scale,” Government Computer News, November 1, 2018 ; https://gcn.com/articles/2018/11/01/psi_administrative-data-research-facility.aspx (accessed October 25, 2021).
- A Framework for Data Quality, Federal Committee on Statistical Methodology (FCSM), September 2020; https://nces.ed.gov/FCSM/pdf/FCSM.20.04_A_Framework_for_Data_Quality.pdf (accessed October 25, 2021).
- ACDEB charter; <https://www.bea.gov/system/files/2021-01/84587-DOC-2021-Charter-ACDEB-1.13.2021.pdf> (accessed October 25, 2021).
- ACDEB project inventory; <https://www.bea.gov/system/files/2021-10/acdeb-project-inventory.xlsx> (accessed October 27, 2021).
- ACDEB public comments; <https://www.bea.gov/system/files/2021-02/ACDEB-Public-Comments.pdf> (accessed October 25, 2021).
- Advisory Committee on Data for Evidence Building (ACDEB) website; <https://www.bea.gov/evidence> (accessed October 25, 2021).
- Babina, Tania, Paige Ouimet, and Rebecca Zarutskie, Going Entrepreneurial? IPOs, Employee Reallocation, and New Firm Creation, Finance and Economics Discussion Series Working Papers, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, March 26, 2017; <https://www.federalreserve.gov/econres/feds/going-entrepreneurial-ipos-and-new-firm-creation.htm> (accessed October 25, 2021).
- Boston Women’s Workforce Council Gender/Racial Wage Gap Measures; <https://thebwcc.org/mpc> (accessed October 25, 2021).

- Bowen, Claire, Victoria L. Bryant, Len Burman, Surachai Khitatrakun, Graham MacDonald, Robert McClelland, Philip Stallworth, Kyle Ueyama, Aaron R. Williams, and Noah Zwiefel, A Synthetic Supplemental Public-Use File of Low-Income Information Return Data: Methodology, Utility, And Privacy Implications, Tax Policy Center, Urban Institute & Brookings Institution, July 9, 2020; <https://www.taxpolicycenter.org/publications/synthetic-supplemental-public-use-file-low-income-information-return-data-methodology> (accessed October 25, 2021).
- CARES Act (Public Law 116-136, March 27, 2020); <https://www.govinfo.gov/content/pkg/PLAW-116publ136/pdf/PLAW-116publ136.pdf> (accessed October 25, 2021).
- CDC Wide-ranging ONline Data for Epidemiologic Research (WONDER); <https://wonder.cdc.gov/> (accessed October 25, 2021).
- Chief Data Officer Council; <https://www.cdo.gov/> (accessed October 25, 2021).
- Coleridge Initiative; <https://coleridgeinitiative.org/> (accessed October 25, 2021).
- Coleridge Initiative – Show US the Data, Kaggle Competition, June 2021; <https://www.kaggle.com/c/coleridgeinitiative-show-us-the-data/> (accessed October 25, 2021).
- Coleridge Initiative’s Administrative Data Research Facility; <https://coleridgeinitiative.org/adrf/> (accessed October 25, 2021).
- Coleridge Initiative’s Applied Data Analytics Training; <https://coleridgeinitiative.org/training/> (accessed October 25, 2021).
- Coleridge Initiative’s Rich Context Project; <https://coleridgeinitiative.org/projects-and-research/rich-context/> (accessed October 25, 2021).
- Commission on Evidence-Based Policymaking, The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking, September 2017; <https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2019/03/Full-Report-The-Promise-of-Evidence-Based-Policymaking-Report-of-the-Comission-on-Evidence-based-Policymaking.pdf> (accessed October 25, 2021).
- [Data.gov](https://data.gov) (accessed October 25, 2021).
- Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g) ; <https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title20-section1232g&num=0&edition=prelim> (accessed October 25, 2021).
- FCSM’s Data Protection Toolkit; <https://nces.ed.gov/fcsm/dpt> (accessed October 25, 2021).
- Federal Data Strategy 2020 Action Plan; <https://strategy.data.gov/2020/action-plan/> (accessed October 25, 2021).

- Federal Data Strategy 2021 Action Plan; <https://strategy.data.gov/2021/action-plan/> (accessed October 25, 2021).
- Federal Privacy Council; <https://www.fpc.gov/> (accessed October 25, 2021).
- FedRAMP Marketplace; <https://marketplace.fedramp.gov/#!/products> (accessed October 25, 2021).
- Foundations for Evidence-Based Policymaking Act of 2018 (Public Law 115–435, January 14, 2019); <https://www.govinfo.gov/app/details/PLAW-115publ435/summary> (accessed October 25, 2021).
- FSRDC Institutional Partners (“data concierges”); <https://www.census.gov/about/adrm/fsrdc/about/partners.html> (accessed October 25, 2021).
- FSRDC network; <https://www.census.gov/about/adrm/fsrdc.html> (accessed October 25, 2021).
- FSRDC “Projects” webpage; <https://www.census.gov/about/adrm/fsrdc/about/ongoing-projects.html> (accessed October 25, 2021).
- Gollust, Sarah and Pinar Karaca-Mandic, 2020. “Geographic Variation in ACA-Related Media Messages and Health Insurance Enrollment” CES Technical Notes Series 20-02, Center for Economic Studies, U.S. Census Bureau; <https://ideas.repec.org/p/cen/tnotes/20-02.html> (accessed October 25, 2021).
- Hart-Potok report – Nick Hart and Nancy Potok, Modernizing U.S. Data Infrastructure: Design Considerations for Implementing a National Secure Data Service to Improve Statistics and Evidence Building, July 2020; <https://www.datafoundation.org/modernizing-us-data-infrastructure-2020> (accessed October 25, 2021).
- Health Insurance Portability and Accountability Act of 1996 (HIPAA) (Public Law 104-191, August 21, 1996); <https://www.govinfo.gov/app/details/PLAW-104publ191/summary> (accessed October 25, 2021).
- Interagency Committee on Standards Policy (ICSP); <https://www.nist.gov/standardsgov/interagency-committee-standards-policy-icsp> (accessed October 25, 2021).
- Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA); <https://www.govinfo.gov/content/pkg/FR-2007-06-15/pdf/E7-11542.pdf> (accessed October 25, 2021).
- Kearns, Ed, First Street Foundation, Communicating Climate Change Risk Through the Use of U.S. Government Open Data, Committee Public Meeting, April 23, 2021; <https://www.bea.gov/system/files/2021-04/Kearns-Apr23-2021.pdf> (accessed October 25, 2021).

- Kniffin, Kevin M., Andrew S. Hanks, Xuechao Qian, Bo Wang & Bruce A. Weinberg, Dissertators with Distantly Related Foci Face Divergent Near-Term Outcomes, National Bureau of Economic Research Working Papers, September 2020; <https://www.nber.org/papers/w27825> (accessed October 25, 2021).
- Longitudinal Employer-Household Dynamics; <https://lehd.ces.census.gov/> (accessed October 25, 2021).
- Midwest Collaborative: Data for Evidence-Based Policy, March 4-6, 2019; https://coleridgeinitiative.org/wp-content/uploads/2020/04/Midwest_Spring_Summary_Report.pdf (accessed October 25, 2021).
- Modernizing the National Vital Statistics System; <https://www.cdc.gov/nchs/nvss/modernization.htm> (accessed October 25, 2021).
- Multi-State Postsecondary Dashboard, Coleridge Initiative; <https://coleridgeinitiative.org/projects-and-research/multi-state-post-secondary-dashboard/> (accessed October 25, 2021).
- National Directory of New Hires; <https://www.acf.hhs.gov/css/training-technical-assistance/overview-national-directory-new-hires> (accessed October 25, 2021).
- National Institutes of Health librarians; <https://www.nihlibrary.nih.gov/get-help/find-your-librarian> (accessed October 25, 2021).
- NCHS Data Linkage; <https://www.cdc.gov/nchs/data-linkage/index.htm> (accessed October 25, 2021).
- NCHS National Vital Statistics System (NVSS); <https://www.cdc.gov/nchs/nvss/index.htm> (accessed October 25, 2021).
- OMB, “Federal Data Strategy – A Framework for Consistency” (memorandum M-19-18); <https://www.whitehouse.gov/wp-content/uploads/2019/06/M-19-18.pdf> (accessed October 25, 2021).
- OMB Regulatory Agenda; <https://www.reginfo.gov/public/do/eAgendaMain> (accessed October 25, 2021).
- OMB Statistical Policy Directives; <https://www.whitehouse.gov/omb/information-regulatory-affairs/statistical-programs-standards/> (accessed October 25, 2021).
- “On a Mission to Make Federal Data Sets More Useful and Accessible,” Scholarly Publishing and Academic Resources Coalition (SPARC), August 3, 2021; <https://sparcopen.org/news/2021/on-a-mission-to-make-federal-data-sets-more-useful-and-accessible/> (accessed October 25, 2021).
- Paperwork Reduction Act (Public Law 104-13, May 22, 1995); <https://www.reginfo.gov/public/reginfo/pra.pdf> (accessed October 25, 2021).

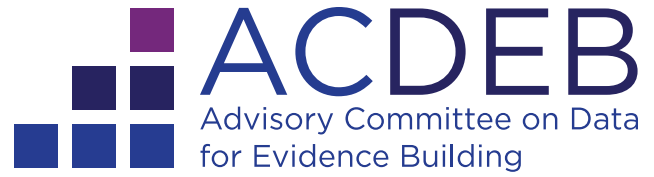
- Phillips, Gordon, and Kyle Herkenhoff, The Impact of Consumer Credit Access on Earnings Mobility, Entrepreneurship and Income Inequality, National Bureau of Economic Research project, August 1, 2018 – July 31, 2022; <https://www.nber.org/programs-projects/projects-and-centers/7512-impact-consumer-credit-access-earnings-mobility-entrepreneurship-and-income-inequality> (accessed October 25, 2021).
- Privacy Act, 5 USC § 552a(b)(4) (1974); <https://www.justice.gov/archives/opcl/page/file/844481/download> (accessed October 25, 2021).
- Privacy Preservation Techniques Task Team of the UN Committee of Experts on Big Data and Data Science for Official Statistics; <https://unstats.un.org/bigdata/task-teams/privacy/index.cshtml#:~:text=The%20Privacy%20Preserving%20Techniques%20Task,platform%2C%20specifically%20around%20supporting%20privacy> (accessed October 25, 2021).
- ResearchDatagov; <https://www.icpsr.umich.edu/web/pages/appfed/index.html> (accessed October 25, 2021).
- Resnick, Dean, Christine Cox, and Lisa Mirel. (2021). Using synthetic data to replace linkage derived elements: a case study. Health Services and Outcomes Research Methodology. 21. 1-18, September 2021; <https://link.springer.com/article/10.1007/s10742-021-00241-z> (accessed October 25, 2021).
- Smith, Adam B., NOAA National Centers for Environmental Information (NCEI) Center for Weather and Climate, The Use of Federal Data for Costing Weather and Climate Extreme Events: Data Challenges, Needs and Opportunities, Committee Public Meeting, April 23, 2021; <https://www.bea.gov/system/files/2021-04/Smith-The-Use-of-Federal-Data-for-Costing-Weather-and-Climate-Extreme-Events-Data-Challenges-Needs-and-Opportunities.pdf> (accessed October 25, 2021).
- South Carolina Big Data for Health Science Activities; <https://bigdata.sc.edu/> (accessed October 25, 2021).
- South Carolina Coordinating Council for Workforce Development; <https://www.sccommerce.com/sc-coordinating-council-workforce-development> (accessed October 25, 2021).
- South Carolina Early Learning Extension; <https://earlychildhoodsc.org/media/t4li3cvn/scele-overview.pdf> (accessed October 25, 2021).
- South Carolina Revenue and Fiscal Affairs Office; <https://rfa.sc.gov/> (accessed October 25, 2021).
- Straus, Stephanie, “A Federal Government Privacy-Preserving Technology Demonstration,” McCourt School of Public Policy at Georgetown University, June 29, 2021; <https://mccourt.georgetown.edu/news/a-federal-government-privacy-preserving-technology-demonstration/> (accessed October 25, 2021).

- System of Record Notices; <https://www.fpc.gov/elements-of-federal-privacy-program/system-of-record-notice/> (accessed October 25, 2021).
- Title 13 – Census; <https://uscode.house.gov/view.xhtml?path=/prelim@title13&edition=prelim> (accessed October 25, 2021).
- Title 26 – Internal Revenue Code; <https://uscode.house.gov/view.xhtml?path=/prelim@title26&edition=prelim> (accessed October 25, 2021).
- Unemployment to Reemployment Portal, Coleridge Initiative; <https://coleridgeinitiative.org/projects-and-research/unemployment-to-reemployment-portal/> (accessed October 25, 2021).
- U.S. Chamber of Commerce Foundation’s Jobs and Employment Data Exchange (JEDx); <https://www.uschamberfoundation.org/JEDx> (accessed October 25, 2021).
- USDA Data Strategy; <https://www.ocio.usda.gov/sites/default/files/USDA%20Data%20Strategy.pdf> (accessed October 25, 2021).
- USDA’s Meals for Kids program; <https://www.fns.usda.gov/meals4kids> (accessed October 25, 2021).
- “USDA CXO Dashboards Declared ‘Innovation of the Year,’” USDA Press Release 0238.18, November 2, 2018; <https://www.usda.gov/media/press-releases/2018/11/02/usda-cxo-dashboards-declared-innovation-year> (accessed October 25, 2021).
- Western Interstate Commission for Higher Education Multistate Longitudinal Data Exchange; <https://www.wiche.edu/key-initiatives/multistate-longitudinal-data-exchange/> (accessed October 25, 2021).

6.2. Acronyms

Acronym	Definition
ACDEB	Advisory Committee on Data for Evidence Building
ADRF	Administrative Data Research Facility
BEA	Bureau of Economic Analysis
CARES	Coronavirus Aid, Relief, and Economic Security
CDC	Center for Disease Control
CDL	Cropland Data Layers
CDO	Chief Data Officer
CEP	Commission on Evidence-Based Policymaking
CIPSEA	Confidential Information Protection and Statistical Efficiency Act
CLU	Common Land Unit
CoP	Community of Practice
CRIE	Culturally Responsive Indigenous Evaluation
EDAPT	Enterprise Data Analytics Platform and Toolset
FCSM	Federal Committee on Statistical Methodology
FDS	Federal Data Strategy
FedRAMP	Federal Risk and Authorization Management Program
FERPA	Family Educational Rights and Privacy Act
FNS	Food and Nutrition Service
FSA	Farm Service Agency
FSRDC	Federal Statistical Research Data Center
FY	Fiscal Year
HBCU	Historically Black Colleges and Universities
HHS	Department of Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act of 1996
HR	Human Resources
HSI	Hispanic-Serving Institution

Acronym	Definition
ICSP	Interagency Council on Statistical Policy
IMAGES	Integrated Modeling and Geospatial Estimation System
IRS	Internal Revenue Service
JEDx	Jobs and Employment Data Exchange
LEHD	Longitudinal Employer-Household Dynamics
MOU	Memorandum of Understanding
MWC	Midwest Collaborative
NASS	National Agricultural Statistics Service
NCES	National Center for Education Statistics
NCHS	National Center for Health Statistics
NDNH	National Directory of New Hires
NIH	National Institutes of Health
NOAA	National Oceanic and Atmospheric Administration
NSDS	National Secure Data Service
NSF	National Science Foundation
NVSS	National Vital Statistics System
OFS	Office of Food Safety
OMB	Office of Management and Budget
OPEN	Open, Public, Electronic, and Necessary
PII	Personally Identifiable Information
PPRL	Privacy Preserving Record Linkage
PSA	Principal Statistical Agency
RFA	Revenue and Fiscal Affairs Office
SAP	Standard Application Process
SEC	Securities and Exchange Commission
SMC	Secure Multiparty Computation
UI	Unemployment Insurance
UN	United Nations
USDA	United States Department of Agriculture
VDE	Virtual Data Enclave
WONDER	Wide-ranging Online Data for Epidemiologic Research



Advisory Committee on Data for Evidence Building: Year 1 Report

[Visit the ACDEB website](#)

Contact the Committee at evidence@bea.gov
with questions or comments

