



THE PROMISE OF EVIDENCE-BASED POLICYMAKING

Report of the Commission on Evidence-Based Policymaking

Transparency

Humility

Data

Privacy

Rigor

Capacity



The Promise of Evidence-Based Policymaking

Report of the Commission on Evidence-Based Policymaking

September 2017

This page intentionally left blank

**KATHARINE G.
ABRAHAM**
CHAIR

RON HASKINS
CO-CHAIR

4600 SILVER HILL ROAD
WASHINGTON, DC 20233



Commission on Evidence-Based Policymaking

WASHINGTON, DC 20233

COMMISSIONERS

SHERRY GLIED
ROBERT M. GROVES
ROBERT HAHN
HILARY HOYNES
JEFFREY B. LIEBMAN
BRUCE D. MEYER
PAUL OHM
NANCY POTOK
KATHLEEN RICE MOSIER
ROBERT J. SHEA
LATANYA SWEENEY
KENNETH R. TROSKE
KIM R. WALLIN

September 7, 2017

The President of the United States
The Speaker of the House
The President of the Senate

Dear Sirs:

On behalf of the Commission on Evidence-Based Policymaking, and pursuant to Public Law 114-140, we are pleased to submit the following report detailing our findings and recommendations. We believe that adoption of the recommendations contained in our report will prove transformational, improving both the privacy protections afforded to the American public and the availability of rigorous evidence to inform policymaking.

Over the course of its work, the Commission heard from numerous expert witnesses and members of the public, received extensive written public comment, and fielded a survey of Federal agencies. This fact-finding phase was followed by several months of deliberations. The recommendations contained in the report have been agreed to unanimously by the 15 Members of the Commission.

The Commission envisions a future in which rigorous evidence is created efficiently, as a routine part of government operations, and used to construct effective public policy. Advances in technology and statistical methodology, coupled with a modern legal framework and a commitment to transparency, make it possible to do this while simultaneously providing stronger protections for the privacy and confidentiality of the people, businesses, and organizations from which the government collects information.

We call on the Congress and the President to act to ensure that this vision becomes a reality.

Respectfully yours,

Katharine Abraham
Katharine G. Abraham, Chair

Ron Haskins
Ron Haskins, Co-Chair

Sherry Glied
Sherry Glied

Robert M. Groves
Robert M. Groves

Robert Hahn
Robert Hahn

Hilary Hoynes
Hilary Hoynes

Jeffrey B. Liebman
Jeffrey B. Liebman

Bruce D. Meyer
Bruce D. Meyer

Paul Ohm
Paul Ohm

Nancy Potok
Nancy Potok

Kathleen Rice Mosier
Kathleen Rice Mosier

Robert J. Shea
Robert J. Shea

Latanya Sweeney
Latanya Sweeney

Kenneth R. Troske
Kenneth R. Troske

Kim R. Wallin
Kim R. Wallin

This page intentionally left blank

Contents

Executive Summary: The Promise of Evidence-Based Policymaking	1
Commission on Evidence-Based Policymaking Recommendations.....	5
Chapter 1. Introduction: Vision for Evidence-Based Policymaking	7
Defining Evidence-Based Policymaking	8
Evolution of Evidence Building in the United States.....	12
Commission’s Charge and Process.....	16
Realizing the Promise of Evidence-Based Policymaking	17
Chapter 2. Secure, Restricted Access to Confidential Data	23
Findings.....	24
Recommendations.....	40
Chapter 3. Enhancing Privacy Protections for Federal Evidence Building	47
Findings.....	50
Recommendations.....	60
Chapter 4. Modernizing America’s Data Infrastructure for Accountability and Privacy	65
Findings	68
Recommendations.....	80
Chapter 5. Strengthening the Evidence-Building Capacity within the Federal Government	87
Findings.....	88
Recommendations.....	101
Chapter 6. Conclusion: Possibilities with More and Better Evidence	105
Acronyms	107
Appendix A: Evidence-Based Policymaking Commission Act of 2016	109
Appendix B: Commissioner Biographies.....	115
Appendix C: Commission’s Fact-Finding and Deliberative Processes.....	119
Appendix D: Examples of Data Productive for Evidence Building.....	125

This page intentionally left blank

Executive Summary:

The Promise of Evidence-Based Policymaking

The American people want a government that functions efficiently and responsibly addresses the problems that face this country. Policymakers must have good information on which to base their decisions about improving the viability and effectiveness of government programs and policies. Today, too little evidence is produced to meet this need.

The Commission on Evidence-Based Policymaking (the “Commission”) envisions a future in which rigorous evidence is created efficiently, as a routine part of government operations, and used to construct effective public policy. The Federal government has already taken important steps towards accomplishing this vision, but much work remains. The growing interest in producing more and higher-quality evidence to support decision-making led the Congress and the President to enact the Evidence-Based Policymaking Commission Act of 2016, creating the Commission.

The Commission was provided just over a year to study and develop a strategy for strengthening government’s evidence-building and policymaking efforts. During the Commission’s fact-finding phase, numerous experts, researchers, government leaders, public and private organizations, and members of the public offered their perspectives on the Commission’s charge.

Based on this collective input, the Commission determined that greater use of existing data is now possible in conjunction with stronger privacy and legal protections, as well as increased transparency and accountability. The Commission believes that improved access to data under more privacy-protective conditions can lead to an increase in both the quantity and the quality of evidence to inform important program and policy decisions.

Traditionally, increasing access to confidential data presumed significantly increasing privacy risk. The Commission rejects that idea. The Commission believes there are steps that can be taken to improve data security and privacy protections beyond what exists today, while increasing the production of evidence. Modern technology

and statistical methods, combined with transparency and a strong legal framework, create the opportunity to use data for evidence building in ways that were not possible in the past. This report describes the Commission’s findings and presents recommendations for fundamental improvements to the Federal government’s evidence-building systems and capabilities. Specifically, the Commission’s report includes recommendations on (1) how the Federal government can provide the infrastructure for secure access to data, (2) the mechanisms to improve privacy protections and transparency about the uses of data for evidence building, and (3) the institutional capacity to support evidence building.

Recommendations for Improving Secure, Private, and Confidential Data Access

There are many barriers to the effective use of government data to generate evidence. Better access to these data holds the potential for substantial gains for society. The Commission’s recommendations recognize that the country’s laws and practices are not currently optimized to support the use of data for evidence building, nor in a manner that best protects privacy. To correct these problems, the Commission makes the following recommendations:

- Establish a National Secure Data Service to facilitate access to data for evidence building while ensuring privacy and transparency in how those data are used. As a state-of-the-art resource for improving government’s capacity to use the data it already collects, the National Secure Data Service will be able to temporarily link existing data and provide secure access to those data for exclusively statistical purposes in connection with approved projects. The National Secure Data Service will do this without creating a data clearinghouse or warehouse.

- Require stringent privacy qualifications for acquiring and combining data for statistical purposes at the National Secure Data Service to ensure that data continue to be effectively protected while improving the government's ability to understand the impacts of programs on a wider range of outcomes. At the same time, consider additional statutory changes to enable ongoing statistical production that, under the same stringent privacy qualifications, may make use of combined data.
- Review and, where needed, revise laws authorizing Federal data collection and use to ensure that limited access to administrative and survey data is possible to return benefits to the public through improved programs and policies, but only under strict privacy controls.
- Ensure state-collected quarterly earnings data are available for statistical purposes, including to support the many evidence-building activities for which earnings are an important outcome.
- Make additional state-collected data about Federal programs available for evidence building. Where appropriate, states that administer programs with substantial Federal investment should in return provide the data necessary for evidence building.
- Develop a uniform process for external researchers to apply and qualify for secure access to confidential government data for evidence-building purposes while protecting privacy by carefully restricting data access to qualified and approved researchers.

Recommendations for Modernizing Privacy Protections for Evidence Building

Enhancements to privacy, coupled with improved methods for secure data access, will revolutionize how government uses and protects the data it collects. Among the Commission's recommendations to achieve this vision are:

- Require comprehensive risk assessments on

de-identified confidential data intended for public release to improve how data are protected and risk is managed.

- Adopt modern privacy-enhancing technologies for confidential data used for evidence building to ensure that government's capabilities to keep data secure and protect confidentiality are constantly improving.
- Assign senior officials the responsibility for stewarding data within government departments. Agencies should improve leadership, coordination, and collaboration when implementing protections for the use of confidential data.
- Codify policies for maintaining integrity and objectivity in Federal statistics to promote continued public trust in the accuracy of information being used to guide government decision-making.

Recommendations for Implementing the National Secure Data Service

The Commission's recommendations for improved data access and strong privacy protections rely heavily on the establishment of the National Secure Data Service. Being able to combine data within a secure environment will be an increasingly vital aspect of the evidence-building community's capacity to meet future demand from policymakers. Increased transparency will enable the public to be informed about how data are being used to improve their government, even as data are being stringently protected. The Commission envisions that the National Secure Data Service will operate an effective and efficient service that can be held accountable by policymakers and the American public. The Commission's recommendations to implement the National Secure Data Service include:

- Build on the infrastructure and expertise already developed in government, including at the U.S. Census Bureau, to ensure that data linkages and access to confidential data for

statistical purposes are conducted in the most secure manner possible.

- Require public input, guidance, and participation in the policies and procedures for data linkage activities through public and stakeholder representation on the National Secure Data Service's steering committee.
- Establish a new transparency and accountability portal for evidence-building activities to ensure the public is notified about how confidential data are used for evidence building and to document routine audits for compliance with rules governing privacy, confidentiality, and data access.
- Innovate continuously on privacy-protective data access approaches with sufficient administrative flexibilities to ensure government can adjust as technology advances.
- Increase efforts to make information available about the government's current data inventories and supply related data documentation to help researchers inside and outside government know which data they need to evaluate programs and policies.

Recommendations for Strengthening Federal Evidence-Building Capacity

More privacy protective approaches and improved access to data alone will not improve the volume and quality of evidence. The evidence-building community also needs sufficient capacity, administrative flexibilities, and appropriate program design to enable a strong and effective evidence-generation system to operate. To strengthen the evidence-building capacity within the Federal government, the Commission makes the following recommendations:

- Identify or establish a Chief Evaluation Officer in each department to coordinate evaluation and policy research and to collaborate with other evidence-building functions within Federal departments.

- Develop learning agendas in Federal departments to support the generation and use of evidence to address the range of policymakers' questions.
- Improve coordination of government-wide evidence building by directing the Office of Management and Budget to facilitate cross-government coordination, and consider how a greater commitment to foundational information policy responsibilities can be achieved, including through any consolidation or reorganization at the Office of Management and Budget that may be necessary.
- Align administrative processes with evidence-building activities, including those relating to the approval of information collections and the procurement of services for evidence building.
- Ensure that sufficient resources to support evidence-building activities are available, including resources to support implementation of the Commission's recommendations.

Conclusion

Generating and using evidence to inform government policymaking and program administration is not a partisan issue. The strategy described in this report offers a non-partisan approach to improving how government staff, private researchers, foundations, non-profits, the business community, and the public interact to make sure government delivers on its promises.

The Commission's recommendations represent a comprehensive strategy for tackling the greatest problems facing evidence building today—data access is limited, privacy-protecting practices are inadequate, and the capacity to generate the evidence needed to support policy decisions is insufficient. The Congress, the President, and the American people are ill-served by this state of affairs. Government must do what it takes to increase the quantity and quality of evidence building. The strategy outlined in the Commission's report simultaneously improves privacy protections and makes better use of data the government already collects to support policymaking. Together with leadership from the President and the Congress

in calling for credible evidence to support policy decisions throughout government, implementation of the Commission's recommendations is an important step in providing the country with an effective government.

Whether deciding on funding allocations, assessing proposed regulations, or understanding

how to improve processes for efficiently providing services, evidence should play an important role in key decisions made by government officials. The Commission proposes modernizing the country's evidence-building capacity to make sure our government's decision-making process is among the best in the world, now and in the future. ■

Commission on Evidence-Based Policymaking Recommendations

Note: Recommendations in this report are numbered sequentially to align with the discussion. For example, 2-1 refers to the first recommendation in Chapter 2.

Improving Secure, Private, and Confidential Data Access

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 the implementation of innovative privacy-enhancing technologies.

REC. 2-2: The NSDS should be a service, not a data clearinghouse or warehouse. The NSDS should facilitate temporary data linkages in support of distinct authorized projects.

REC. 2-3: In establishing the 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 the 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.

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.

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.

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 available through a single Federal source for solely statistical purposes.

REC. 2-7: The President should direct Federal departments that acquire state-collected administrative data to make them available for statistical purposes. Where 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.

REC. 2-8: The Office of Management and Budget should promulgate a single, streamlined process for researchers external to the government to apply, become qualified, and gain approval to access government data that are not publicly available. Approval would remain subject to any restrictions appropriate to the data in question.

Modernizing Privacy Protections for Evidence Building

REC. 3-1: The Congress and the President should amend the Privacy Act and the Confidential Information Protection and Statistical Efficiency Act (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.

REC. 3-2: The President should direct Federal departments, in coordination with the National

Secure Data Service, to adopt state-of-the-art database, cryptography, privacy-preserving, and privacy-enhancing technologies for confidential data used for evidence building.

REC. 3-3: The President should direct Federal departments to assign a senior official the responsibility for coordinating access to and stewardship of the department's data resources for evidence building in collaboration with senior department information technology, privacy, and other leaders. A Principal Statistical Agency head, or other appropriately qualified senior official, should serve this function.

REC. 3-4: The Congress and the President should enact legislation to codify relevant portions of Office of Management and Budget Statistical Policy Directive #1 to protect public trust by ensuring that data acquired under a pledge of confidentiality are kept confidential and used exclusively for statistical purposes.

Implementing the National Secure Data Service

REC. 4-1: The National Secure Data Service (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.

REC. 4-2: The NSDS should establish a Steering Committee that includes representatives of the public, Federal departments, state agencies, and academia.

REC. 4-3: To ensure exemplary transparency and accountability for the Federal government's use of data for evidence building, the 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.

REC. 4-4: The NSDS should have specific administrative and implementation flexibilities including the ability to leverage public-private partnerships and to collect and retain user fees.

REC. 4-5: The Office of Management and Budget should increase efforts to make information available on existing Federal datasets including data inventories, metadata, and data documentation in a searchable format.

Strengthening Federal Evidence-Building Capacity

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.

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.

REC. 5-3: The Congress and the President should direct the Office of Management and Budget (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.

REC. 5-4: The Congress and the President should align administrative processes to support evidence building, in particular by streamlining the approval processes for new data collections and using existing flexibilities in procurement policy.

REC. 5-5: The Congress and the President should ensure sufficient resources to support evidence-building activities about Federal government programs and policies. ■

1

Introduction: Vision for Evidence-Based Policymaking

With the passage and signing of the Evidence-Based Policymaking Commission Act in the spring of 2016, elected leaders issued a bipartisan call to improve the evidence available for making decisions about government programs and policies.¹ (See the box “Charge to the Commission on Evidence-Based Policymaking.”) In an environment of growing partisanship

in the country, it is notable that this legislation was embraced by legislators on both sides of the aisle and enacted without dissent. U.S. House of Representatives Speaker Paul Ryan, a co-sponsor of the Act, described the potential for evidence-based policymaking as a “sea change in how we solve problems.” Likewise, co-sponsor Senator Patty Murray said: “Whether you think we need more government, or less government—you should agree that we should at least have better government.”

1. Evidence-Based Policymaking Commission Act of 2016 (Public Law 114–140, March 30, 2016).

Charge to the Commission on Evidence-Based Policymaking

In the Evidence-Based Policymaking Commission Act of 2016 (see Appendix A), the Congress and the President prescribed a number of duties to the Commission, including the following:

- Study the data inventory, data infrastructure, database security, and statistical protocols related to Federal policymaking. Make recommendations on how data infrastructure, database security, and statistical protocols should be modified.
- Determine the optimal arrangement for which administrative data, survey data, and related statistical data series may be integrated and made available for evidence building while protecting privacy and confidentiality.
- Make recommendations on how best to incorporate evidence building into program design.
- Consider whether a “clearinghouse” for program and survey data should be established and how to create such a “clearinghouse.”

The American people want a government that solves problems. This requires that decision makers have good information to guide their choices about how current programs and policies are working and how they can be improved. While the Federal government has already taken steps towards developing an “evidence culture,” much remains to be done. A particularly important barrier to government’s further progress is lack of access by researchers outside of government and by individuals within government to the data necessary for evidence building, even when those data have already been collected.

While collecting taxes, determining eligibility for government benefits, engaging in economic development, and running programs, government necessarily collects a considerable amount of information. In 2017, the American public will spend nearly 12 billion hours responding to more than 100 billion individual requests for information from the Federal government.² Even though the direct costs of collecting these data are funded by taxpayers, these data are not generally available for producing evidence. Addressing barriers to the use of already collected data is a path to unlocking important insights for addressing society’s greatest challenges.

As the use of existing government data to support policymaking grows, the American public will be concerned about exactly how those data are being used and whether the privacy and confidentiality of individuals and organizations are being protected. Today, data are protected, in part, through pledges of confidentiality, privacy laws, and legal and policy limitations on how they are used, but the government’s approach to data protection has not kept pace with important changes in technology.

Capabilities now exist to improve privacy protections while making better use of already collected administrative data, including recent advances in statistical methodology, computer science, and computational capacity. Growing experience with successful legal models for data stewardship points in the same direction. Government also can dramatically improve transparency about its collection and use of data, improving the Amer-

ican public’s ability to hold the government accountable. Adhering to the highest possible standards with respect to privacy and accountability is an important part of earning the public’s trust. The improvements to privacy and accountability that the Commission envisions can occur simultaneously with providing policymakers the tools to deliver more effective government services.

The Commission envisions a future in which rigorous evidence is created efficiently, as a routine part of government operations, and used to construct effective public policy. While this may sound like a daunting task, the Commission’s vision for the future of evidence-based policymaking in the United States is well within reach (see the box “Examples of the Promise of Evidence-Based Policymaking”). This vision requires that new laws and policies be put into place. When implemented, the Commission is confident that the approaches proposed in this report will greatly improve both the ability to produce evidence in support of better policies and privacy protections for individuals and organizations.

Defining Evidence-Based Policymaking

“Evidence” can be defined broadly as information that aids the generation of a conclusion. Throughout this report, the Commission uses the term in a more specific way—this report uses the shorthand “evidence” to refer to information produced by “statistical activities” with a “statistical purpose” that is potentially useful when evaluating government programs and policies. Following U.S. Office of Management and Budget (OMB) Statistical Policy Directive #1, which in turn follows the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), we define “statistical activities” as “the collection, compilation, processing, analysis, or dissemination of data for the purpose of describing or making estimates concerning the whole, or relevant groups or components within, the economy, society, or the natural environment, including the development of methods or resources that support those activities, such as measurement of methods, statistical classifications, or sampling frames.” A “statistical purpose” is defined as “the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that

2. Office of Information and Regulatory Affairs, Office of Management and Budget, Inventory of Currently Approved Information Collections; <https://www.reginfo.gov/public/do/PRARReport?operation=11> (accessed August 10, 2017).

Examples of the Promise of Evidence-Based Policymaking

Through transactions with the American public, governments collect a considerable amount of data. These administrative data, collected in the first instance to serve routine program operation purposes, also can be used to assess how well programs are achieving their intended goals.¹ Below are examples where administrative data were used to generate evidence that informed government policies.

Permanent Supportive Housing. There is a growing body of research on the impact of providing permanent supportive housing to chronically homeless individuals.² This research demonstrates that an intervention combining long-term housing assistance with supportive services can help chronically homeless individuals maintain stable housing and achieve other positive outcomes, such as improved health outcomes and reduced use of crisis services, including costly emergency room visits or stays in a homeless shelter.³ Cost-effectiveness studies of the intervention also suggest that offering permanent supportive housing to chronically homeless individuals with the highest service needs can reduce taxpayer costs for other components of the

safety net.⁴ These studies were carried out using a combination of survey and administrative data, including administrative data from locally operated Homeless Management Information Systems.

As a direct result of this growing body of evidence, in recent years, the U.S. Department of Housing and Urban Development (HUD) has encouraged and incentivized communities to increase their supply of permanent supportive housing for chronically homeless individuals over the past several years.⁵ Notably, there has been a 27 percent reduction in chronic homelessness nationally between the years 2010 and 2016.⁶

Substance Abuse Education. The Drug Abuse Resistance Education program (DARE), created in 1983, originally aimed to prevent drug use and gang membership for kindergarten through 12th grade students in Los Angeles. In partnership with local law enforcement officers, DARE grew into a national program focused primarily on drug prevention that at its peak was in over 75 percent of the schools in the United States and in more than 50 countries.⁷

More than 30 rigorous evaluations conducted throughout the 1990s and 2000s suggested that the original DARE program did not produce

—continues

1. The Evidence-Based Policymaking Commission Act of 2016 (Public Law 114–140, March 30, 2016) defines “administrative data” as data “(1) held by an agency or contractor or grantee of an agency (including a State or unit of local government); and (2) collected for other than statistical purposes.”

2. The U.S. Department of Housing and Urban Development adopted the Federal definition, which defines a chronically homeless person as “either (1) an unaccompanied homeless individual with a disabling condition who has been continuously homeless for a year or more, or (2) an unaccompanied individual with a disabling condition who has had at least four episodes of homelessness in the past three years.” (See 11 Code of Federal Regulations 91 and 578, 2015.)

3. Dennis P. Culhane and Thomas Byrne, *Ending Chronic Homelessness: Cost-Effective Opportunities for Interagency Collaboration*. Federal Strategic Plan Supplemental Document No. 19 (Washington, D.C.: U.S. Interagency Council on Homelessness, 2010); https://www.usich.gov/resources/uploads/asset_library/DennisCulhane_EndingChronicHomelessness.pdf (accessed August 10, 2017).

4. S. R. Poulin, M. Maguire, S. Metraux, and D. P. Culhane. “Service Use and Costs for Persons Experiencing Chronic Homelessness in Philadelphia: A Population-Based Study,” *Psychiatric Services* 61, no. 11 (2010): 1093–1098; M.E. Larimer, D.K. Malone, M.D. Garner and others, “Health Care and Public Service Use and Costs Before and After Provision of Housing for Chronically Homeless Persons With Severe Alcohol Problems,” *Journal of the American Medical Association*, 301, no. 13 (April 1, 2009): 1349–1357.

5. See HUD’s Fiscal Year 2018 Congressional Justification for the Homeless Assistance Grant Program for more information; <https://portal.hud.gov/hudportal/documents/huddoc?id=22-HomelessAGrants.pdf> (accessed August 10, 2017).

6. HUD, *The 2016 Annual Homeless Assessment Report (AHAR) to Congress; Part I: Point-in-Time Estimates of Homelessness* (Washington, D.C.: U.S. Government Printing Office, 2016).

7. For information about DARE America, go to <http://www.dare.org/about-d-a-r-e> (accessed August 10, 2017).

Examples of the Promise of Evidence-Based Policymaking— continued

substantial reductions in teenage substance abuse over the long-term.⁸ One study carried out in a suburban setting even found that the intervention could contribute to increases in drug use.⁹ In 2001, the Surgeon General summarized the available research and designated DARE as an “ineffective primary prevention program” but also stated “its popularity persists despite numerous well-designed evaluations and meta-analyses that consistently show little or no deterrent effects on substance abuse.”¹⁰

The DARE program partnered with Pennsylvania State University to adopt a new elementary and middle school curriculum called “keepin it REAL.”¹¹ Today, the DARE program focuses on a broader vision of empowering students to respect others and choose to lead lives free from violence, substance abuse, and other dangerous behaviors.¹² Preliminary evidence from the revised curriculum suggests more promising effectiveness at achieving the stated goals related to decision-making.¹³

Workforce Investment. A large portfolio of evidence about workforce investments and job training programs suggests that program

participants can realize improved earnings and employment outcomes, though the evidence is mixed on specific strategies.¹⁴ The U.S. Department of Labor’s Adult Program assists people who are economically disadvantaged facing barriers to employment. The Dislocated Workers Program assists workers who have been laid off or who have been notified that they will be terminated or laid off. Both programs provide a range of training and supportive services. In a study using administrative data in a non-experimental program evaluation, researchers found that participants in the adult program experienced an increase in quarterly earnings relative to a comparison group, while participants in the dislocated workers program actually saw reduced earnings in several quarters.¹⁵ When the workforce investment programs were reauthorized in 2014 through the Workforce Innovation and Opportunity Act (WIOA), the new law provided states additional flexibility to shift funding between the adult and dislocated worker aspects of the program to better target local needs. WIOA included numerous other evidence-informed strategies based on the existing portfolio of evidence.

Implementation of the permanent supportive housing, DARE, and workforce investment programs each were influenced by evidence developed to inform the implementation of Federal policies. With more evidence to inform a range of policy interests and questions, policymakers will have a stronger basis for making decisions in the future.

8. Greg Berman and Aubrey Fox, *Lessons from the Battle Over DARE* (Center for Court Innovation and Bureau of Justice Assistance of the U.S. Department of Justice, 2009); <http://www.courtinnovation.org/sites/default/files/DARE.pdf> (accessed August 10, 2017).

9. Dennis P. Rosenbaum and Gordon S. Hanson, “Assessing the Effects of School-Based Drug Education: A Six Year Multi-Level Analysis of Project DARE,” *Journal of Research in Crime and Delinquency* 35, no. 4 (1998): 381–412.

10. U.S. Department of Health and Human Services, *Youth Violence: A Report of the Surgeon General* (Washington, D.C.: Department of Health and Human Services, 2001): 110.

11. DARE America, “Keepin it Real Elementary School Curriculum,” <https://www.dare.org/keepin-it-real-elementary-school-curriculum> (accessed August 10, 2017).

12. DARE America, “D.A.R.E.’s keepin’ it REAL Elementary and Middle School Curriculums Adhere to Lessons From Prevention Research Principles,” <https://www.dare.org/d-a-r-e-s-keepin-it-real-elementary-and-middle-school-curriculums-adhere-to-lessons-from-prevention-research-principles> (accessed August 10, 2017).

13. Randy Borum and David Allan Verhaagen, *Assessing and Managing Violence Risk in Juvenile* (New York: Guilford, 2006).

14. Office of Management and Budget (OMB), “Using Administrative and Survey Data to Build Evidence,” white paper for the Commission on Evidence-Based Policymaking (Washington, D.C.: OMB, Executive Office of the President, 2016); https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/using_administrative_and_survey_data_to_build_evidence_0.pdf (accessed August 10, 2017).

15. Caroline J. Heinrich, Peter R. Mueser, Kenneth R. Troske, Kyung-Seong Jeon, and Daver C. Kahvecioglu, “Do Public Employment and Training Programs Work?” *IZA Journal of Labor Economics* 2, no. 6 (2013); <https://izajole.springeropen.com/articles/10.1186/2193-8997-2-6> (accessed August 10, 2017).

comprise such groups; and includes the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support such purposes.”³ We return to these definitions in Chapter 2 in the discussion of CIPSEA. The essence of a “statistical activity” with a “statistical purpose” is that the result summarizes information about a group rather than a single individual or organization. For example, a statistical activity could include analyzing a “unit,” such as a state or a grantee, in order to generate average values for all of the individuals included within that unit, such as residents, clients, or firms.

Data can be used for many purposes other than evidence building. These include non-statistical

3. Statistical Policy Directive No. 1, *Federal Register* 79 (December 2, 2014): 71609–71616; <https://www.gpo.gov/fdsys/pkg/FR-2014-12-02/pdf/FR-2014-12-02.pdf>.

purposes such as the use of individual-level information to determine benefit amounts, enforce laws, or otherwise affect the rights or privileges of an individual. The identification of a single individual means the information is not being used for statistical activities and thus, would not be termed “evidence” in the Commission’s definition. Throughout this report, the Commission has been mindful that, consistent with applicable law, efforts to make data available specifically for statistical purposes might also inadvertently put information about individuals at increased risk for use in other ways. The Commission’s proposals attempt to ensure strict structural and institutional separation between statistical and non-statistical uses of data. The Commission’s evidence-building reforms are engineered to make data difficult to repurpose for non-statistical uses.

The Commission defines evidence-based policymaking as the application of evidence to inform

Questions to Answer with More and Better Evidence

The Commission’s fact-finding process produced numerous examples of important questions that individuals who provided input to the Commission reported cannot currently be adequately addressed because of difficulty accessing the right data:

- What effects does the Supplemental Nutrition Assistance Program have on health outcomes?
- Are the earnings of veterans improved by training received while in the military?
- Which transition-to-adulthood experiences make students with disabilities less likely to rely on the Supplemental Security Income program?
- To what extent do eligible active-duty military households participate in antipoverty programs and how does this participation affect their economic self-sufficiency?
- Do Farm Service Agency programs ease credit constraints for farmers?
- What impacts do Federal economic development efforts, such as the Appalachian Regional Commission’s grants, have on the communities they are trying to help?
- What enforcement approaches are most effective in improving clean air regulation compliance?

Each of these questions could be studied using administrative data that the government already collects. Too often, however, the capacity and infrastructure to study pressing questions faced by decision-makers are lacking. The Commission’s vision for evidence-based policymaking would enable each of these important policy questions to be addressed with appropriate information analyzed in a secure and privacy-protected environment, and then used to improve government policies and programs.

decisions in government. For evidence-based policymaking to occur, a supply of evidence must first exist. Thus, the Commission recognizes that evidence-based policymaking requires the generation of evidence, which relies on access to data. As the evidence base becomes stronger, the American public should expect that policymakers increasingly will incorporate new and better evidence into their decisions about the operation of government programs and funding for government services.

Different types of evidence are relevant for policymaking and may involve a variety of methods.⁴ Descriptive statistics provide insights about trends and context. Performance metrics support monitoring of policy outputs and efficiency. Implementation and process studies can identify how well the application of programs and policies aligns with their intended design and goals. Impact evaluations provide insights about whether desired outcomes are achieved. Each of these types of evidence and others are relevant for evidence-based policymaking, and the appropriate approach depends on the policymakers' question (see the box "Questions to Answer with More and Better Evidence").

This report uses the term "evidence-building community," which is meant to describe the collective set of individuals located both inside and outside the Federal government who fulfill a set of roles key to generating evidence for use in policymaking. The evidence-building community includes individuals situated across government and in the business, non-profit, and academic sectors. The community includes individuals who perform statistical activities, such as collecting data to produce national indicators relevant to the country. The community includes researchers who study ways to improve government's programs and policies and evaluators who assess whether those programs and policies are achieving their intended goals. The community also includes individuals who support program administrators with analysis to achieve targeted improvements to their programs and policies.

4. OMB, "Using Administrative and Survey Data to Build Evidence," 2016.

Evolution of Evidence Building in the United States

The nation's founders recognized the importance of information for governance, requiring in the U.S. Constitution a census of the population (see Figure 1).⁵ James Madison argued that collecting more data about the populace could guide congressional decisions about government actions as the young country grew.⁶ Early censuses gathered information about industry, agriculture, and the population. Census questions changed over time, reflecting important societal and governmental information needs of the day.

By the mid-nineteenth century, the Congress had established several permanent units to produce national statistics in specific policy areas. More statistical units were added in subsequent years and formed the basic information infrastructure of the nation's first two centuries. Over the years, the statistics on population size, education, employment, gross domestic product, and others became a routine dashboard on what was happening in society.

Today, the evidence-building community operates under a range of laws, regulations, and policies that evolved over time, a state of affairs that has contributed to a lack of coordination and collaboration across the community. Numerous commissions or committees were convened during the 20th century to recommend improvements for the country's evidence-building system (see online Appendix H). These included recommendations for greater coordination of activities and for enhancements to the protection of privacy.⁷

Information policy setting and coordination across government began in earnest in 1939 when President Franklin D. Roosevelt issued an Executive Order directing the Bureau of the Budget "to plan and promote the improvement, development and coordination of Federal and other statistical

5. U.S. Constitution Article I, Section 2.

6. Annals of Congress, House of Representatives, 1st Congress, 2nd Session "Remarks by James Madison on the Bill for the 1790 Census" (1790): 1145.

7. Janet Norwood, *Organizing to Count: Change in the Federal Statistical System* (Washington, D.C.: Urban Institute Press, 1995).

services.”⁸ The Congress subsequently provided additional authority to conduct central reviews for new data collections in government with the Federal Records Act of 1942.⁹

1960s and 1970s: Expansion of Evidence Building and Privacy Protections

As more data were collected and used by government for implementing programs and for statistical activities, the need for privacy protections became increasingly apparent. In 1973, the U.S. Department of Health, Education, and Welfare developed the Fair Information Practice Principles (FIPPs), recommending that the Congress enact them into what eventually became the Privacy Act of 1974.¹⁰ The principles include transparency, individual participation, purpose specification, data minimization, use limitation, data quality and integrity, security, and accountability and auditing. These principles strive to balance the need for information with privacy protections for the benefit of the American public.

The Privacy Act also codified some U.S. information practices, establishing common requirements related to collecting, maintaining, using, and disseminating government records about individuals. The Privacy Act articulated basic transparency requirements and limitations on how data collected by the government may be disclosed. In 1977, the Privacy Protection Study Commission created by the Privacy Act conveyed two central tenets for evidence building. First, research and statistical uses of data about individuals must exclude any result that would directly affect an individual’s rights, privileges, or benefits. Second, government statistical and non-statistical uses of data should be separated by a bright line, a principle referred to as “functional separation.”¹¹ These

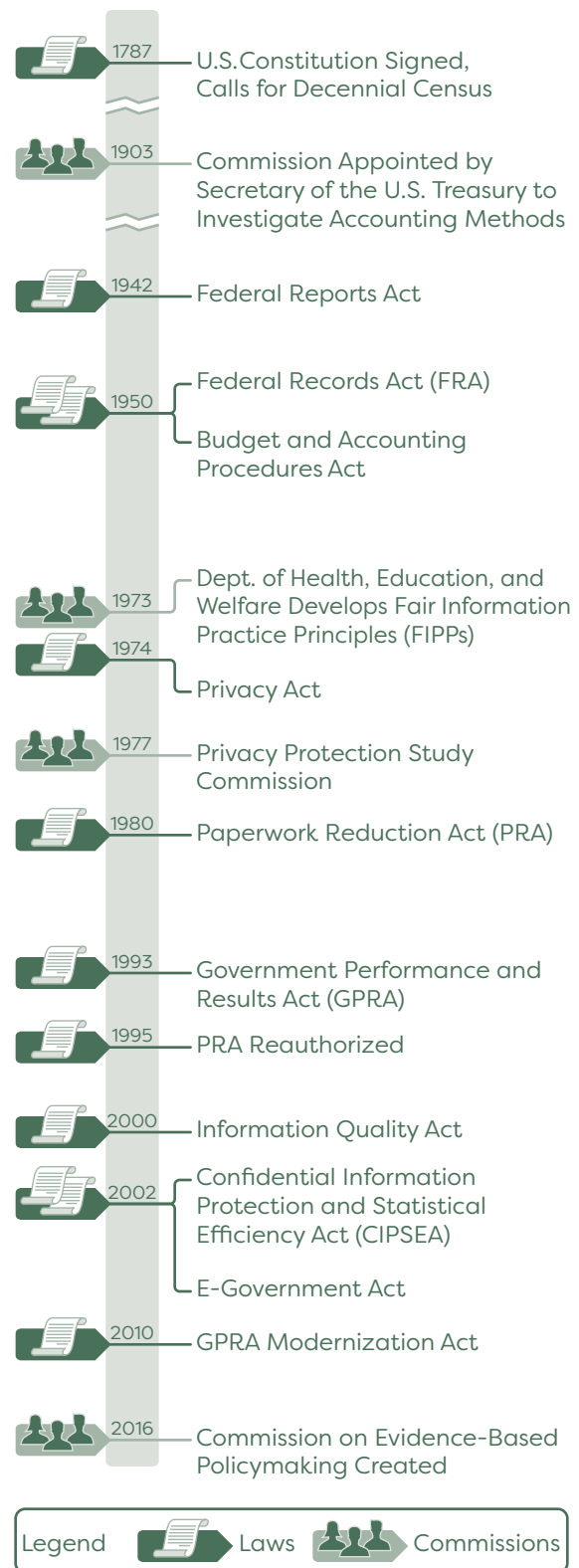
8. Executive Order 824: Establishing the Divisions of the Executive Office of the President and Defining Their Functions and Duties, 3 C.F.R. (September 8, 1939); <https://www.archives.gov/federal-register/codification/executive-order/08248.html> (accessed August 10, 2017).

9. Records Management by Federal Agencies, 44 USC § 3101 et. seq.

10. U.S. Department of Health, Education, and Welfare, *Records, Computers, and the Rights of Citizens: Report of the Secretary’s Advisory Committee on Automated Personal Data Systems* (1973).

11. *Personal Privacy in an Information Society*, The Report of the Privacy Protection Study Commission, July 1977.

Figure 1. Key Milestones in the U.S. Evidence-Building System



*Timeline not to scale

two core concepts carry forward into the nation's privacy protection and statistical laws, and remain key to this Commission's work.

Amidst discussions about improving privacy, approaches to generating and using evidence expanded. The U.S. Department of Defense introduced cost-benefit analysis as part of budgeting activities and the technique gained further traction in the regulatory infrastructure.¹² Similarly, in the 1960s and 1970s, the Department of Health, Education, and Welfare pioneered many of the approaches for executing government's evaluation function in partnership with contractors, grantees, and academics, to meet the need for better information about the impacts of Great Society programs. Many of these efforts created the basic infrastructure and resource mechanisms to support evaluation of education, human services, and health programs.

1980s and 1990s: Improved Evidence Coordination to Meet Demand

As evidence-building activities became more common across government, gaps in coordination mechanisms became more apparent. The Paperwork Reduction Act (PRA) of 1980 specifically recognized the role of the OMB in coordinating government-wide information and statistical policy, which the Congress reaffirmed during a 1995 reauthorization. The PRA specifically encourages data sharing for statistical activities as government policy, both to improve the efficiency and the quality of statistical series and of evidence more broadly.¹³

Within the decentralized Federal evidence-building apparatus, 13 Principal Statistical Agencies (PSAs) in 11 Federal departments provide the core infrastructure to support government's main statistical activities and provide access to statistical datasets for research (see Figure 2).¹⁴ These 13 agencies, along with nearly 100 smaller units embedded within program agencies, collectively

form the Federal Statistical System.¹⁵ The PSAs benefit from explicit statutory authorities and an established culture that supports secure and confidential data collection and stewardship. According to the American Statistical Association, "Federal statistical agencies will play a vital role in evidence-based policymaking and...the historical autonomy of these agencies [is needed] in order to ensure the integrity of their work."¹⁶ The Commission agrees with this assessment. Coordination for the statistical system is supported by the statutory Interagency Council on Statistical Policy and the Federal Committee on Statistical Methodology, and led by the Chief Statistician in OMB.¹⁷ OMB also issues standards and directives to assure the quality and integrity of activities conducted within the system.

With the Government Performance and Results Act of 1993 (GPRA), the Congress also established an expectation for continuous improvement in government by creating government-wide requirements for departments to set goals, measure results, and report progress.¹⁸

2000s: Strengthened Protections for Statistical Activities

In recognition of the need to better protect data collected by the Federal Statistical System, CIPSEA established common legal authorities for statistical activities. Consistent with the recommendations of the President's Reorganization Project for the Federal Statistical System twenty years earlier, CIPSEA provides statutory protections that apply to all 13 PSAs and other designated units for data collected under a pledge of confidentiality for statistical activities.¹⁹

15. OMB, *Statistical Programs of the United States Government* (Washington, D.C.: 2017); https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2017.pdf (accessed August 10, 2017).

16. Clyde Tucker, American Statistical Association, Commission Public Hearing, Washington, D.C., October 21, 2016.

17. The Chief Statistician role and the ICSP are established under the Paperwork Reduction Act, 44 USC § 3504(e) (1995).

18. P. G. Joyce, "The Obama Administration and PBB: Building on the Legacy of Federal Performance-Informed Budgeting?" *Public Administration Review*, 71, no. 3 (2011): 356-367.

19. James T. Bonnen, Theodore G. Clemence, Ivan P. Fellegi, Thomas B. Jabine, Ronald E. Kutscher, Larry K. Roberson, Charles A. Waite, "Improving the Federal Statistical System: Report of the President's Reorganization Project for the Federal Statistical System," *American Statistician*, 35, no. 4 (November 1981, reprint): 184-196.

12. Edward P. Fuchs and James E. Anderson, "The Institutionalization of Cost-Benefit Analysis," *Public Productivity Review* 10, no. 4 (1987): 25-33.

13. Paperwork Reduction Act, 44 USC § 3506 (1995).

14. Throughout this report, the term "departments" is generally used to refer to the 24 Federal agencies subject to the Chief Financial Officers Act of 1990 (Public Law 101-576).

The Privacy Act, CIPSEA, and the PRA offer the legal foundation for coordinating information policy and protecting data used in evidence building, but much work remains to realize the promises embodied in these laws. For example, Federal evaluation and policy research offices are more nascent in their development than the statistical agencies. As such, none of them are recognized under CIPSEA to employ the strong protections that statistical agencies can provide for confidential data.

2010s: Increased Demand for Evidence

Reauthorizations of programs by the Congress during the 2010s provided numerous examples of the increased demand for evidence to inform policy. The GPRM Modernization Act of 2010 reiterated the Congress's interest in better use of evidence for continuous improvement in government.²⁰ The Workforce Innovation and Opportunity Act in 2014 authorized numerous evaluation and multisite projects. The Agricultural Act of 2014 called for testing of innovative approaches to helping individuals gain and retain employment that leads to self-sufficiency.²¹ The Bipartisan Budget Act of 2015 reauthorized the Social Security Disability Insurance program's demonstration authority to test policy improvements.²² The Every Student Succeeds Act in 2015 reauthorized education programs and expanded the use of evidence-based grants.²³ Numerous other laws and draft legislation embody the increasing demand for evidence such as by incorporating requirements about using evidence or supporting innovation in awarding grants.

Within the Executive Branch, calls encouraging more and better evidence also emerged. In 2013, a memorandum to Federal departments specifically encouraged using evidence that is available,

20. "GPRM Modernization Act of 2010" (Public Law 111-352, January 4, 2011); <https://www.congress.gov/111/plaws/publ352/PLAW-111publ352.pdf> (accessed August 10, 2017).

21. "Agricultural Act of 2014" (Public Law 113-79, February 7, 2014); <https://www.congress.gov/113/plaws/publ79/PLAW-113publ79.pdf> (accessed August 10, 2017).

22. "Bipartisan Budget Act of 2015" (Public Law 114-74, Sec. 821, November 2, 2015); <https://www.congress.gov/114/plaws/publ74/PLAW-114publ74.pdf> (accessed August 10, 2017).

23. "Every Student Succeeds Act" (Public Law 114-95, December 10, 2015); <https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf> (accessed August 10, 2017).

Figure 2. Federal Principal Statistical Agencies (PSAs)



producing new evidence to increase knowledge, and continually innovating to improve government performance.²⁴ And in 2017, the President specifically called for generating and using evidence to improve government's effectiveness.²⁵

Partnerships for Policy Research and Evaluation

The capacity within Federal departments to perform policy research and evaluation can be arrayed along a continuum, ranging from a strong centralized approach to dispersed capacity across bureaus and units. Discussing the existing track record within the Federal government, the American Evaluation Association observed that “for the most part, these evaluations have been sporadic, applied inconsistently, and supported inadequately. Training and capacity building for evaluation have been inconsistent across agencies and, in many cases, insufficient to achieve the needed evaluation capacity and sustain it over time.”²⁶ The Commission agrees with this assessment.

Individuals outside of government, such as non-governmental researchers, philanthropic organizations, universities, researchers, and other partners, play an essential role in supporting and extending the evidence-building community within government. Evidence produced by external actors can influence policymakers, encourage policy innovation, and drive the allocation of resources across new and existing programs. Moreover, external actors create a capacity that goes well beyond that of government staff.

Federal departments and universities have initiated collaborations to improve evidence-building activities. For example, the Federal Statistical Research Data Centers operate today as partnerships between statistical agencies and the re-

search organizations hosting these centers. The centers provide researchers with secure access to confidential data, and the statistical agencies and general public benefit from research findings that make important contributions to statistical products and general knowledge.

Commission's Charge and Process

Continuing this history of evidence-building actions in the United States, the Commission on Evidence-Based Policymaking was established by the bipartisan Evidence-Based Policymaking Commission Act of 2016 (Appendix A), jointly sponsored by Speaker of the House Paul Ryan and Senator Patty Murray and signed by President Barack Obama on March 30, 2016. The Act directed the Commission to consider how to strengthen government's evidence-building and policymaking efforts. Specifically, the Act directed the Commission to study how the data that government already collects can be used to improve government programs and policies. The Commission was directed to submit to the Congress and the President a detailed statement of its findings and recommendations.

The Commission recognizes that data are an important building block for the generation of evidence and that many of the greatest gains for evidence building in the near term can be accomplished by addressing the challenge of secure, private, and confidential data access. A theme that runs throughout this report is that access to confidential data for evidence-building purposes should be increased, but only in the context of a modern legal framework providing for strengthened privacy protections and increased transparency. Taking up the privacy challenge, the report recommends specific steps to improve privacy and transparency with regard to evidence-building activities. The report also recognizes the need to strengthen the government's institutional capacity to support the evidence-building system.

Following the Commission's first meeting in July 2016, Commissioners engaged in an eight-month-long fact-finding process to learn about the current state of evidence production and use in the Federal government, as well as the Federal government's policies and practices to protect data confidentiality (see Appendix C). The Commission

24. OMB, “Next Steps in the Evidence and Innovation Agenda” (memorandum M-13-17); <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf> (accessed August 10, 2017).

25. Executive Office of the President, “Building and Using Evidence to Improve Government Effectiveness,” in *Analytical Perspectives: Budget of the United States for Fiscal Year 2018* (March 2017); https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/budget/fy2018/ap_6_evidence.pdf (accessed August 10, 2017).

26. American Evaluation Association (AEA), *An Evaluation Roadmap for a More Effective Government* (Washington, D.C.: AEA, revised October 2013); <http://www.eval.org/d/do/472> (accessed August 10, 2017).

solicited feedback from the public and experts by doing the following:

- Administering a survey to 209 offices of the Federal government that the Commission identified as likely to be generating or using evidence, in order to understand the current or potential capacity of Federal agencies to engage in aspects of evidence-based policymaking (see online Appendix E).
- Convening seven public meetings to hear from 49 invited expert witnesses on a range of issues (see online Appendix F).
- Issuing a Request for Comments in the *Federal Register* and accepted comments by email, which generated over 350 responses from the public (see online Appendix G).
- Holding three open public hearings—in Washington, D.C., Chicago, and San Francisco—during which 39 members of the public presented information directly to Commissioners (see online Appendix G).
- Meeting with more than 40 organizations or experts to solicit additional input.

Following the fact-finding phase of the Commission's work, Commissioners launched a five-month deliberative process to consider all of the gathered input and information. At the outset of its delibera-

tions, the Commission developed a set of five guiding principles that shaped its recommendations (see the box “Guiding Principles for Evidence-Based Policymaking”). The fact-finding and deliberations using these principles culminated in the conclusions presented in this report and a strategy for realizing the promise of evidence-based policymaking. Each of the 15 appointed Commissioners, forming a bipartisan Commission with a range of perspectives on the issues examined, approved this final report and the recommendations herein.

Realizing the Promise of Evidence-Based Policymaking

The Commission believes that enabling improved access to data under more modern privacy-protective conditions will lead to more and better evidence. Improvement in access can occur safely with improved privacy protections. This report maps a path to realizing this vision.

Vision for Improving Secure Access to Confidential Data

During the Commission's fact-finding phase, Commissioners heard about the many barriers to effective use of the data government already collects to generate evidence. The Commission believes that enabling better use of such data will offer substantial gains for society. The Commission also believes that by better coordinating the access to

Guiding Principles for Evidence-Based Policymaking

1. **Privacy.** Individual privacy and confidentiality must be respected in the generation and use of data and evidence.
2. **Rigor.** Evidence should be developed using well-designed and well-implemented methods tailored to the questions being asked.
3. **Transparency.** Those engaged in generating and using data and evidence should operate transparently, providing meaningful channels for public input and comment and ensuring that evidence produced is made publicly available.
4. **Humility.** Care should be taken not to over-generalize from findings that may be specific to a particular study or context.
5. **Capacity.** The capacity to generate and use data and evidence should be integrated within government institutions and adequately funded and staffed.

administrative data, government can improve the security of these data while better protecting individual privacy.

Chapter 2 describes the current flaws in the country's data infrastructure and capacity to use data for evidence building. While the country's data infrastructure and capacity have evolved substantially over the past two and a half centuries, the Federal government, as a whole, has not kept pace with emerging privacy-protecting technologies and protocols to support secure access to data. While there are pockets of innovation, our government has not yet broadly adopted approaches for using data to generate insights that can enhance decision-making that are becoming much more prevalent in the private sector. Some other countries also have more consistently implemented the capabilities needed for developing and using high quality evidence to inform public policy.

“Data are the lifeblood of decision-making and the raw material for accountability. Without high-quality data providing the right information on the right things at the right time; designing, monitoring and evaluating effective policies becomes almost impossible.”

– United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development²⁷

As the amount of data available increases exponentially, improving government's capacity to securely analyze information that crosses the silos of government is an ongoing challenge. The Commission proposes to answer this challenge by building on and reorganizing existing resources within government through the National Secure Data Service (NSDS), establishing a state-of-the-

art capacity for integrating existing data and providing secure data access for exclusively statistical purposes (REC. 2-1). The NSDS will model best practices for secure record linkages, propel implementation of privacy-enhancing technologies, and coordinate a new level of transparency about how government data are used for evidence building. Importantly, the NSDS will facilitate temporary linkages only for authorized projects, and will not serve as a data warehouse (REC. 2-2).

As explained in Chapter 2, the country's laws are not currently optimized either to support the use of data for evidence building across programs or to maximize privacy. Increasing connections between data used in evidence building across topical areas will improve our ability to understand the impacts of programs on a wider range of outcomes. Some existing authorities enable sharing of information across agencies within the Federal government and allow access for non-Federal actors. These approaches could be expanded to enhance the analytical capabilities to cut across additional policy domains.

Inconsistent laws and interpretations of legal authorities for securely sharing and using confidential data have inhibited the ability of the evidence-building community to access the right types of data. The Commission proposes to amend the Privacy Act and CIPSEA to require stringent privacy qualifications for acquiring and combining data for statistical purposes, to ensure that data continue to be effectively protected while improving the government's ability to understand the impacts of programs on a wider range of outcomes (REC. 2-3). In some cases, the purposes for which administrative data may be used are defined narrowly, preventing their use for evidence building. The Commission proposes a review of such statutes to ensure that limitations that preclude the use of administrative data for evidence building are applied only when the Congress and the President deem the limitations still to be necessary (REC. 2-4). In some cases, existing laws specifically prohibit the collection or analysis of information to support evidence building. Again, the Commission calls for a reconsideration of such bans and restraint in the enactment of future bans (REC. 2-5).

Chapter 2 also highlights the value of data collected by states and other jurisdictions about Federal programs and policies, especially data

27. United Nations, Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development. *A World That Counts: Mobilizing the Data Revolution for Sustainable Development* (New York, NY: United Nations, November 2014); <http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf> (accessed August 10, 2017).

on income, wages, and earnings, while noting that these data in most cases are not currently being used productively for evidence building.²⁸ Multiple programs supported with Federal funds and administered by states and other jurisdictions collect income, wage, or earnings data in the course of program operations. Some of these data are provided to the Federal government for specific purposes, but numerous researchers and evaluators reported to the Commission that accessing the data for all states is nearly impossible. The Commission believes that data about earnings are among the most important for enhancing the generation of evidence about Federal government programs. The Commission strongly encourages the Congress and the President to make these data available for statistical purposes (REC 2-6). In addition, many Federal programs rely on states and other jurisdictions to collect data, but those data are not currently available for evaluating programs and policies. An expectation should be established that these data will be available for evidence building (REC. 2-7).

A related issue frequently reported to the Commission is the confusion caused by variation across agencies in the approval processes through which external researchers establish eligibility to access confidential data held by Federal departments. The Commission recognizes that while embedding a certain amount of friction in these processes can enhance privacy—that is, that access should not be too easy—but needless variation can create unintentional barriers that serve no deliberate purpose. To better facilitate access while protecting privacy, the Commission recommends the creation of a single process for external researchers to apply and qualify for access to sensitive government data that are not otherwise publicly available, subject to any restrictions appropriate to the data in question (REC. 2-8).

Vision for Modernizing Privacy Protections for Evidence Building

Many existing privacy protections in the United States are strong, but protections are applied

unevenly across government. Many are not dynamic enough to meet the ever-changing risks associated with the use of data. While the Federal government is subject to many transparency requirements, government could be more open in explaining how individuals' data are used for evidence building.

Chapter 3 discusses privacy as a broad concept encompassing the full range of the Fair Information Practice Principles, which provide a useful framework for the Commission's recommendations. Confidentiality is a distinct but overlapping concept in that it refers specifically to the protection of information about individuals or other entities from unauthorized disclosure.²⁹ Confidentiality is a component of privacy, but protecting and respecting privacy entails substantially more than just ensuring the confidentiality of information. For convenience, the Commission often uses the terms "privacy" or "privacy protection" in the broader sense of the full range of Fair Information Practice Principles, including confidentiality. Therefore, throughout this report, privacy and privacy protection should be interpreted to include the confidentiality of individuals' information, and, where appropriate, the protection of information about businesses and other entities for which the term privacy in a narrower sense may not normally apply.

Chapter 3 describes how risks to privacy have evolved and shows why they must be continually assessed as new information is made more publicly available. Government has an obligation to be open and honest with the American public. The public's trust can be earned only through transparency about these risks. The Commission recommends that, prior to any public release of de-identified confidential data, Federal departments assess the risks that may be associated with the release and the steps taken to mitigate those risks, then make these risk assessments publicly available (REC. 3-1). This recommendation is one of several intended to improve how the Federal government manages privacy risks.

As technologies evolve, the capabilities to keep data secure and protect confidentiality are constantly improving. The Commission calls on government to make better use of state-of-the-art

28. Throughout this report, the phrase "other jurisdictions" should generally be interpreted to mean "Tribes, territories, local governments (such as cities, counties, or districts), or other governing entities."

29. *Personal Privacy in an Information Society*, 1977.

approaches that can be applied for enhancing security and confidentiality for data used in evidence building (REC. 3-2).

With respect to all of these activities, the Commission recognizes a need for leadership within Federal departments. The Commission recommends that a senior official be designated to ensure appropriate collaboration with regard to decisions about data stewardship and the implementation of privacy-protective measures within each department (REC. 3-3). The Commission also recommends that existing policies for maintaining the integrity and objectivity of Federal statistics be codified in law, maintaining the public's trust by ensuring that the confidential data used to produce these statistics continue to be protected and used only for statistical purposes (REC. 3-4).

These enhancements to privacy, coupled with the improved strategy for secure data access outlined in Recommendation 2-1, will advance how government uses and protects the data it already collects, allowing the data to be accessed securely.

Implementing the Vision

A core feature for implementing the Commission's vision is the creation of the NSDS as a new service that builds on and enhances existing Federal government data infrastructure for statistical activities (REC 2-1). The ability to temporarily combine identifiable data within a secure environment, then remove direct identifiers for individual projects is a vital element of the evidence-building community's ability to meet future demand from policymakers. For the NSDS to be a role model for the use of data and held accountable by the American public, it should have five core features. These are described further in Chapter 4.

First, the NSDS should build on the infrastructure and expertise already developed across government, such as at the Census Bureau, to ensure that statistical analyses of confidential data are conducted in the most secure manner possible (REC. 4-1). As the data work of the NSDS proceeds, the public must be afforded opportunities to participate by providing input and guidance on the policies and procedures for conducting these efforts through representation on the NSDS's steering committee (REC. 4-2).

The statistical uses of confidential data can pose risks to privacy, and government must be

held accountable for safely managing those risks. The Commission recommends the creation of a single new Federal transparency portal to ensure that information about approved evidence-building activities that rely on confidential data is readily available. Then, routine audits of government and external users of data should be carried out to ensure that, as promised, data are accessed in the most privacy-protective fashion and for solely statistical purposes (REC. 4-3).

Approaches to statistical analysis of government data that protect privacy will continue to evolve as technologies advance. Government will need to innovate continuously with respect to privacy-protective approaches. The NSDS must play an important role in the process of continuous improvement and be afforded the administrative flexibilities to support innovation and to engage in partnerships that can bring needed expertise to bear (REC. 4-4).

Improving protections and capabilities for better using data also requires new approaches to understanding what data exist about Federal programs and policies. The Commission recommends the creation of a complete inventory of Federal data assets with technical information that will help members of the evidence-building community inside and outside government know which data are available to evaluate programs and policies (REC. 4-5).

Even after the NSDS data capabilities have been established, much work will remain to implement other aspects of the Commission's vision. More privacy protective approaches and improved access to data alone will not improve the volume and quality of evidence. Ultimately, the capacity of the Federal evidence-building community must be strengthened to increase its efficacy.

Chapter 5 emphasizes the need for leadership, coordination, and resources to facilitate efficient evidence-building activities that support continuous learning in the Federal government. Departments need to coordinate their evidence-building functions internally to ensure the activities can be effectively and efficiently implemented. The Commission recommends that Federal departments identify or establish a Chief Evaluation Officer to encourage and coordinate policy research and evaluation (REC. 5-1). Programs can also be more effectively designed to enable evidence building. The Commission recommends that departments

create learning agendas that outline priorities and support the development of all types of evidence to help address the range of policymakers' questions (REC. 5-2).

In addition to coordination within departments, better coordination is also needed across government. OMB and Federal departments must be organized to implement this vision. This may require consolidation or reorganization of certain activities at OMB where information policy and other evidence-building activities are currently spread across the organization (REC. 5-3).

The Commission also heard about administrative processes in government that inadvertently inhibit evidence-building activities. These included processes for reviewing information collections under the PRA and procurement policies. The Commission recommends specific flexibilities

be applied or created to tailor these processes to better facilitate the need for increased evidence building (REC. 5-4). Finally, appropriate resources must be available to support evidence building, including resources to enable the full suite of activities discussed throughout this report, from data stewardship to risk assessments to the evaluations of individual programs (REC. 5-5).

In short, government must improve its existing infrastructure and legal frameworks to enable more and better evidence building. The strategy outlined in this report both improves privacy protections and makes better use of data the government has already collected. The remainder of the report outlines the challenges to increasing the volume, quality, and utility of evidence generated to support policymakers and makes specific recommendations for improvements. ■









This page intentionally left blank

2

Secure, Restricted Access to Confidential Data

Equip the evidence-building community with secure, restricted access to data to facilitate the generation of high quality evidence about government programs and policies.

Recommendations

-  **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 the implementation of innovative privacy-enhancing technologies.
-  **2-2:** The NSDS should be a service, not a data clearinghouse or warehouse. The NSDS should facilitate temporary data linkages in support of distinct authorized projects.
-  **2-3:** In establishing the 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 the 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.
-  **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.
-  **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.
-  **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 available through a single Federal source for solely statistical purposes.
-  **2-7:** The President should direct Federal departments that acquire state-collected administrative data to make them available for statistical purposes. Where 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.
-  **2-8:** The Office of Management and Budget should promulgate a single, streamlined process for researchers external to the government to apply, become qualified, and gain approval to access government data that are not publicly available. Approval would remain subject to any restrictions appropriate to the data in question.

Data are an essential ingredient for the success of evidence-based policymaking, providing the foundation for basic descriptive statistics that describe the status quo as well as conclusions from the most sophisticated program evaluations. The routine administrative operations associated with many Federal activities result in the collection of data from program participants. These administrative data, along with the multitude of surveys and other data collections, have value for evidence building.

Data collection is resource intensive for the government and burdensome for the public. There is strong reason to carefully increase the use of existing data to the extent possible before imposing new burdens on individuals, businesses, and other organizations. Leveraging administrative data can improve statistical products, reduce the costs and administrative burden associated with creating statistics and conducting research and evaluation, and thereby greatly expand the evidence base for setting Federal policy (see the box “Reducing the Respondent Burden of Income Questions on Surveys”). The Commission believes the Federal government can more efficiently and effectively use the data it collects, while simultaneously enhancing privacy protection.

In a well-designed system, access to data and the protection of privacy can work hand in hand. The Commission’s conception of “secure data access” is prescribed narrowly to include access for exclusively statistical purposes—that is, the analysis of data to generate inferences about groups. A well-designed system for secure data access is also the means for achieving greater transparency with respect to how data are being used, an important element for increasing privacy and accountability.

Access to data held by the government should occur only in service to the public interest. Decisions about allowing data access must be calibrated according to a project’s potential public benefits, the sensitivity of a particular dataset, and any risk that allowing access could pose to confidentiality. Access can and should be restricted to eligible individuals who demonstrate an understanding of their obligations for data stewardship. Restricted access has been standard practice in many areas of government for decades. However, duplicative processes for accessing confidential data can distract from confidentiality and the promotion of transparency. The Commission believes

that access to confidential data for evidence building should be enhanced, but in a fashion that also improves privacy protections.

Findings

During the Commission’s fact-finding phase, numerous expert witnesses, commenters, and government agencies identified examples in which successful access to data generated information important for improving programs and policies. They often involved an individual or research team with sufficient persistence and creativity to successfully navigate—or even pioneer the use of—available processes for accessing data. For example, Stanford University Professor of Economics Raj Chetty and his collaborators produced path-breaking research on social mobility.¹ Such success stories, however, highlight common barriers to accessing critical data that, in too many cases, have prevented researchers from producing analyses of important policy questions. Many of these same challenges have been identified by past commissions and panels tasked with exploring the Federal Statistical System or making recommendations about policy in a particular area.² Accumulated experience offers valuable lessons about how to establish processes that can make ongoing analyses of important programs and policies a norm in government, rather than a rarity.

To access confidential data for the development of statistics, evaluation, and policy research, members of the evidence-building community today,

1. Raj Chetty, Stanford University, Commission Meeting, Washington, D.C., July 22, 2016.

2. National Academies of Sciences, Engineering, and Medicine, *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy* (Washington, D.C.: National Academies Press, 2017); *Federal Statistics: Report of the President’s Commission, Volume I-II* (Washington, D.C.: U.S. Government Printing Office, 1971); <https://catalog.hathitrust.org/Record/011325636> (accessed August 10, 2017); James T. Bonnen, Theodore G. Clemence, Ivan P. Fellegi, Thomas B. Jabine, Ronald E. Kutscher, Larry K. Roberson, and Charles A. Waite, “Improving the Federal Statistical System: Report of the President’s Reorganization Project for the Federal Statistical System,” *American Statistician* 35, no. 4 (November 1981, reprint): 184–196; National Commission on Hunger, *Freedom From Hunger: An Achievable Goal for the United States of America*, Recommendations of the National Commission on Hunger to Congress and the Secretary of the Department of Agriculture, 2015; <https://digital.library.unt.edu/ark:/67531/metadc799756> (accessed August 10, 2017); Commission to Eliminate Child Abuse and Neglect Fatalities, *Within Our Reach: A National Strategy To Eliminate Child Abuse and Neglect Fatalities* (Washington, D.C.: Government Printing Office, 2016).

both inside and outside the Federal government, must navigate a complex array of processes, protocols, and approaches. They must negotiate legal documents and bureaucratic processes that increase in volume and complexity when using data from multiple policy domains, jurisdictions, or agencies. Often, such processes consider the value proposition of data use only in the context of the mission of the originating agency, irrespective of its broader value. For example, generally Title 26 of the U.S. Code limits the use of tax data to those projects that would improve “tax administration.” The application of this narrow standard to research on human services or transportation, for example, may undervalue the available public good. These kinds of barriers limit the effective, efficient, and transparent use of existing data.

Federal departments are making greater use of their own administrative data for statistical activities today than in the past, and in some cases,

departments are also making such data available to external researchers. Still, significant barriers to data access remain. Starting a new statistical project using administrative data is complicated and time consuming. Even departments that regularly use administrative data for evidence building find it onerous to navigate each originating agencies’ existing processes for acquiring administrative data for statistical use. Agencies just starting to explore the power of administrative data for improving their programs and policies often waste months wading through various applicable laws with little result.

As the demand for statistics, evaluation, and policy research increase, so too must secure data access with strong privacy protections. The Commission identified four overarching challenges to realizing the vision of secure access to confidential data for evidence building that appropriately calibrates public benefits, privacy, and transparency.

Reducing the Respondent Burden of Income Questions on Surveys

Household survey data collection programs, including key U.S. Census Bureau programs, are finding it more difficult to obtain accurate income data from the survey population. Respondents have become less willing to participate in surveys and are increasingly reluctant to respond to questions about income.¹ When they do answer questions about income, they are providing less accurate responses.² The burden on respondents could be reduced and the accuracy of the data improved if statisti-

cal agencies were able to rely more on the income data the government already maintains to administer tax, income support, and social insurance programs. Recent work at the Census Bureau has begun to explore the potential of administrative sources to replace questions about selected income sources in household surveys such as the American Community Survey and the Current Population Survey. While this work is in the early stages and current data access is quite limited, the results are promising and suggest the administrative data can be a significant improvement.³

1. Christopher R. Bollinger, Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak, “Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch,” paper (June 2017); http://www2.gsu.edu/~ecobth/BHHZ_Trouble-in-the-Tails_6-8-2017.pdf (accessed August 10, 2017).

2. National Research Council, *Nonresponse in Social Science Surveys: A Research Agenda*, edited by Roger Tourangeau and Thomas J. Plewes, Panel on a Research Agenda for the Future of Social Science Data Collection, Committee on National Statistics, Division of Behavioral and Social Sciences and Education (Washington, D.C.: The National Academies Press, 2013); Bruce D. Meyer, Wallace K.C. Mok, and James X. Sullivan, “Household Surveys in Crisis,” *Journal of Economic Perspectives* 29, no. 4 (Fall 2015): 199–226.

3. U.S. Census Bureau, “American Community Survey Research and Evaluation Report,” Center for Administrative Records Research and Applications Memorandum Series 16–72015; Bruce D. Meyer, and Nikolas Mittag, “Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net,” National Bureau for Economic Research Working Paper 21676 (Washington, D.C.: 2015); C. Adam Bee and Joshua Mitchell, “The Hidden Resources of Women Working Longer: Evidence from Linked Survey-Administrative Data,” in *Women Working Longer: Increased Employment at Older Ages*, edited by Claudia Goldin and Lawrence F. Katz (Chicago: University of Chicago Press, 2016).

Challenge #1: The current legal framework can stand in the way of data stewardship and the sharing of data to support evidence building.

The current legal environment for data collection, protection, and sharing lacks consistency, leading to confusion and inefficiency among departments, external researchers, and other members of the evidence-building community. Laws governing the data lifecycle for any dataset include the statute that authorized the collection of the data, statutes generally applicable to data collection and management processes, and various legal provisions governing privacy and confidentiality protections. The Commission recognizes that some variation in the laws that govern the protection of data is sensible given the contextual nature of privacy. It is unclear, however, that all of the variation in the legal structure is intentional; rather, some of the variation may be the result of the complexity of government together with the various statutes having been developed independently at particular times and with particular goals in mind. The Congress and the President seemingly recognized as much when forming the Commission and charging it to review the net impact of accumulated laws and policies on evidence building.³

The authority to collect data typically is provided in a program or statistical agency's authorizing statute. In the case of program agencies, authorizing statutes may stipulate the nature and operations of the objectives and policies of the agency. Program statutes vary in their prescriptiveness with regard to confidentiality and the use of data for evidence building, with many statutes not addressing the issue at all.

The laws that authorize statistical agencies generally include restrictions on who can access data and for what purposes. While data collection was a primary intent of the laws that created these agencies, the existing structure has not produced the same level of coordination and integration that is typical in other countries because the Federal Statistical System is decentralized. One result is the inability of U.S. statistical agencies to share

data readily among themselves. In some cases, even when data sharing is allowable for agency staff, impediments remain to providing external researchers with appropriate access to combined data. In other countries with a more centralized statistical system, data sharing is both less needed and more straightforward.

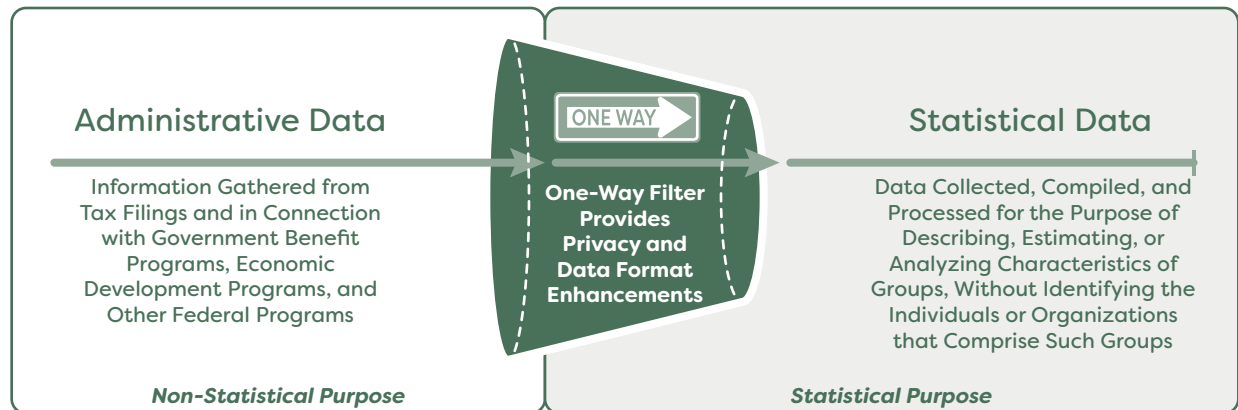
Both program and statistical agencies also must consider other applicable laws. These include the Paperwork Reduction Act (PRA), the Privacy Act, and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA). Generally the PRA does not provide the authority to collect data, but it does govern the process by which data collection occurs. The PRA specifically includes requirements for transparency with regard to the information that government collects and why, by extension implicating potential secondary uses of collected data. As the name implies, the PRA also attempts to reduce the burden of data collection on the public by encouraging data sharing between agencies through the establishment of government-wide statistical and information policies and coordination procedures.

Similarly, the Privacy Act does not provide the authority to collect data, but it does set requirements regarding how those data may be disclosed. Specifically, the Privacy Act requires public and individual notices about data held in government systems and limits disclosure of these data without consent. This includes restricting the secondary uses of data without consent unless one of a limited number of exceptions applies. For example, the Congress exempted from individual consent requirements data provided by another agency to the U.S. Census Bureau for the purposes of planning or carrying out a census or survey.⁴

CIPSEA was enacted to address some gaps and variations in existing statistical laws. (See the box, "The Confidential Information Protection and Statistical Efficiency Act of 2002.") While the authorizing statutes for some statistical agencies, such as the Census Bureau, provide collection authority together with strong confidentiality protections, those for some other agencies do not specifically address privacy and confidentiality. Many statutes also limit data sharing, even among Principal Statistical Agencies (PSAs), leading to inefficiencies such as duplicative data

3. Evidence-Based Policymaking Commission Act of 2016 (Public Law 114-140, Section 4(a), March 30, 2016).

4. Privacy Act, 5 USC § 552a(b)(4) (1974).

Figure 3. Functional Separation of Administrative and Statistical Data

The Confidential Information Protection and Statistical Efficiency Act of 2002

The Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) is a law with two components. The first provides a set of uniform confidentiality protections for data acquired under a pledge of confidentiality and for exclusively statistical purposes. The second provides a limited authority to share protected data. Among the motivations for CIPSEA was a need to ensure a consistent statutory authority for agencies handling statistical data to protect those data from non-statistical use.¹ The Federal government now has 15 years of demonstrated success with CIPSEA.²

The CIPSEA framework codifies the concept of “functional separation” (see Figure 3). Specifically, CIPSEA defines statistical activities, statistical purposes, non-statistical purposes, and qualified agents and agencies. Such definitions are essential to the application of CIPSEA, and

provide a prevailing standard for these concepts in law. Statistical purpose is defined under CIPSEA as “the description, estimation or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups” and “includes the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support” those purposes. Statistical activities are defined as “the collection, compilation, processing, or analysis of data for the purposes of describing or making estimates concerning the whole, or relevant groups or components within the economy, society, or the natural environment” and include “the development of methods or resources that support those activities.”

CIPSEA’s subtitle A fosters public trust by generally prohibiting disclosure of protected information in identifiable form, controlling access to and use of protected information, and ensuring that information is used exclusively for statistical purposes. While some statistical statutes explicitly provide for data protection, others, such as the authorizing statute for the U.S. Bureau of Labor Statistics, do not. CIPSEA Subtitle A provides a remedy for this.

1. U.S. House of Representatives, “Report to Accompany H.R. 2458,” House Report No. 107-787 (2002); <https://www.congress.gov/congressional-report/107th-congress/house-report/787> (accessed August 10, 2017).

2. Brian Harris-Kojetin, “CIPSEA at 15 Years: Benefits to Federal Statistics and Unmet Needs - Overview,” (paper presented at the 133rd Meeting of the Committee on National Statistics, Washington, D.C., May 12, 2017); http://sites.nationalacademies.org/dbasse/cnstat/dbasse_178400 (accessed August 10, 2017).

—continues

collections. CIPSEA was created as a partial answer. The law provides common statutory protections for data acquired for exclusively statistical purposes under a pledge of confidentiality and currently applies to all of the PSAs and to other recognized statistical units.

CIPSEA also established common penalties for any officers, employees, or agents of an agency who, knowing that disclosure is prohibited under CIPSEA, willfully disclose protected data collected

for exclusively statistical purposes to a person or agency not entitled to receive it, including a felony charge and imprisonment for up to 5 years and/or fines up to \$250,000. The CIPSEA penalties are equal to, or exceed, other such provisions in statutes, including provisions contained in Titles 13 and 26 of the U.S. Code.

For the Commission's purpose, the vital components of the CIPSEA legal framework include the ability to offer strong confidentiality protections,

The Confidential Information Protection and Statistical Efficiency Act of 2002—continued

Statistical Agencies or Units Recognized by OMB for CIPSEA Purposes⁴

Bureau of Economic Analysis	National Center for Science and Engineering Statistics
Bureau of Justice Statistics	Office of Research, Evaluation, and Statistics, Social Security Administration
Bureau of Labor Statistics	Statistics of Income Division, Internal Revenue Service
Bureau of Transportation Statistics	Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration*
Census Bureau	Microeconomic Surveys Statistical Unit, Federal Reserve*
Economic Research Service	National Animal Health Monitoring System, Animal and Plant Health Inspection Service*
Energy Information Administration	
National Agricultural Statistics Service	
National Center for Education Statistics	
National Center for Health Statistics	

* Recognized statistical units

CIPSEA grants the Director of the U.S. Office of Management and Budget (OMB) the authority to promulgate rules and provide implementing guidance, including the qualification and recognition of agencies or units that may exercise CIPSEA authority. In its CIPSEA implementation guidance, OMB recognized the 13 Principal Statistical Agencies (PSAs) as statistical agencies under CIPSEA, and also recognized three additional statistical units that had demonstrated functional separation in accordance with OMB's guidance.³

CIPSEA's subtitle B provides for the limited sharing of business data among three designated statistical agencies: the U.S. Bureau of Economic Analysis, the Bureau of Labor Statistics, and the U.S. Census Bureau. The authority is designed to reduce paperwork burden on businesses, improve comparability of economic statistics, and increase understanding of the economy.

3. "Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)," *Federal Register* 72, no. 115 (June 15, 2007): 33361; <https://www.gpo.gov/fdsys/granule/FR-2007-06-15/E7-11542/content-detail.html> (accessed August 10, 2017).

4. List of Principal Statistical Agencies and Recognized Units; https://obamawhitehouse.archives.gov/omb/inforeg_statpolicy/bb-principal-statistical-agencies-recognized-units (accessed August 10, 2017).

ensure exclusively statistical uses, and within that structure, enable limited and secure sharing of data.

CIPSEA permits the sharing of business data exclusively for statistical purposes among three PSAs to create statistical efficiency within the context of CIPSEA's legal framework.⁵ This specific provision enabled several new data sharing initiatives that have improved efficiency for all three agencies.⁶

Except for that provided under CIPSEA, the authority to share data for evidence building is rarely explicit. In cases where authorizing laws are ambiguous, agency interpretations ultimately govern access to and use of data. In some cases multiple agencies interpret the same law differently. This can cause confusion and limit the efficient use of existing data for evidence building.⁷ The complex web of statutes, regulations, and implementing guidance—or absence thereof—drives risk aversion in agencies, causes frustrations for the evidence-building community, and limits the value of data for statistical activities. In effect, the existing legal environment limits the government's ability to steward data responsibly as a valuable resource for the American people and for policymakers.

In the CEP Survey of Federal Offices, 52 percent of responding offices identified legal limitations as a major or moderate barrier to using data for statistics, evaluation, research, and policy analysis purposes. While 10 out of 13 PSAs and 8 out of 10 evaluation offices reported that legal limitations were a major or moderate barrier, less than half (41 percent) of other types of offices cited legal limitations as a major or moderate barrier.⁸ In other

words, even the parts of the Federal government that today are leaders in producing evidence report challenges in navigating the complex array of legal requirements to use data for evidence building.

Perhaps more than any other single data need, the Commission repeatedly heard calls for improving access to income, wage, and earnings data for evidence-building purposes. Effects on income are central to the evaluation of numerous Federal programs and activities, but existing laws and practices limit the ability of researchers both internal and external to the government to access even the income data the Federal government already collects (see the box "Income Data: Federal Tax and Unemployment Insurance Earnings Data"). Programs such as tax administration depend on compliance from the public, making it particularly important to ensure the privacy of records generated in the course of administering those programs. It is equally important, however, to calibrate the need for privacy with the public good that research findings based on such data can provide. Of the 22 offices responding to the CEP Survey of Federal Offices that it was hard to access Federal Tax Information for evidence building, half cited statutes prohibiting data sharing as the reason for the difficulty.

Another barrier identified in testimony to the Commission is the explicit prohibition in some laws on the collection or integration of data, which prevent the government from building evidence to improve Federal programs and policies. For decades, different iterations of evidence-prohibiting provisions have been included by either the Congress in statutes or the Executive Branch in directives and policies. Many of these prohibitions on collecting or accessing data to develop evidence exist in domains that have major implications for policy areas of national importance and they often involve programs that represent a substantial share of the Federal budget.

Through the Commission's fact-finding phase, stakeholders identified several examples of data collection, integration, and analysis bans. The Commission received substantial public comments regarding the student unit record ban, enacted in 2008.⁹ The student unit record ban effectively limits some aspects of evidence production related to pre-kindergarten through 12th grade and

5. Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) (Public Law 107-347, Title V, 2002).

6. For a summary of several of the successful data sharing initiatives that were enabled by CIPSEA Subtitle B, see U.S. Bureau of Labor Statistics (BLS), "Confidentiality of Data Collected by BLS for Statistical Purposes," (May 16, 2016); <https://www.bls.gov/bls/cipsea-report.htm> (accessed August 10, 2017).

7. U.S. Office of Management and Budget (OMB), "Barriers to Using Administrative Data for Evidence Building," white paper for the Commission on Evidence-Based Policymaking (Washington, D.C.: OMB, Executive Office of the President, 2016); https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/barriers_to_using_administrative_data_for_evidence_building.pdf (accessed August 10, 2017).

8. Based on 79 offices that reported they collect or use data for statistics, evaluation, research, or policy analysis or spend a portion of their budget for such purposes.

9. Database of Student Information Prohibited, 20 USC § 1015c.

postsecondary education by prohibiting the Federal government's development or maintenance of a new database with data on all students. The Commission heard many substantive comments about the student unit record ban, and received more feedback on the issue than on any other single topic within the Commission's scope. Nearly two-thirds of the comments received in response to the Commission's Request for Comments raised concerns about student records, with the majority of those comments in opposition to overturning the student unit record ban or otherwise enabling the Federal government to compile records about individual students.

Another ban brought to the attention of the Commission is embedded in the Workforce Innovation and Opportunities Act (WIOA), a program for workforce training and development.¹⁰ The ban is a prohibition on developing a national database of participants in WIOA training programs that includes personally identifiable information.¹¹ With these bans, and a similar ban covering data on Head Start participants, much of government's investment in human capital programs—from early childhood through adulthood—largely is not subject to the type of rigorous, national-scale evaluation that leveraging administrative data could make possible. Other examples of bans or restrictions on use of government data to conduct research populate the U.S. Code, although the Commission was not approached in public comments or by expert witnesses to address these other bans.

Prohibitions on data collection and use arguably conflict with the Commission's vision to improve government based on credible evidence. Within the broad array of factors that influence policymaking, a ban may be an appropriate tool for setting priorities or achieving certain other goals. In the context of evidence building, however, a ban on data collection or use cannot easily be reconciled with a goal of increased reliance on evidence to inform policymaking.

Challenge #2: Many high-value administrative data associated with Federally funded programs that could be useful to measur-

ing the outcomes of government programs and policies are collected at the state or local level and are inaccessible for evidence building.

Many Federal programs are administered by states and localities, most often in cooperation with the Federal government. Examples of programs that operate under this model include Medicaid, the Supplemental Nutrition Assistance Program (SNAP), Unemployment Insurance (UI) programs, Housing Assistance programs, and the Temporary Assistance for Needy Families (TANF) program. These programs have a broad variety of funding and administrative structures that include the Federal and state government as partners and often include cities, counties, and grantees as partners as well.¹² The administrative data generated by the operation of these programs are a valuable source of data for use in evidence building, but there are numerous, well-documented barriers to accessing and using these data.¹³

Barriers to accessing state-level data include: state laws or legal interpretations of Federal law that either expressly prohibit or tightly restrict data sharing for the purposes of evidence building; lack of capacity within states to transform the administrative data to make them suitable for evidence building; and administrative and/or procedural variations for accessing data across

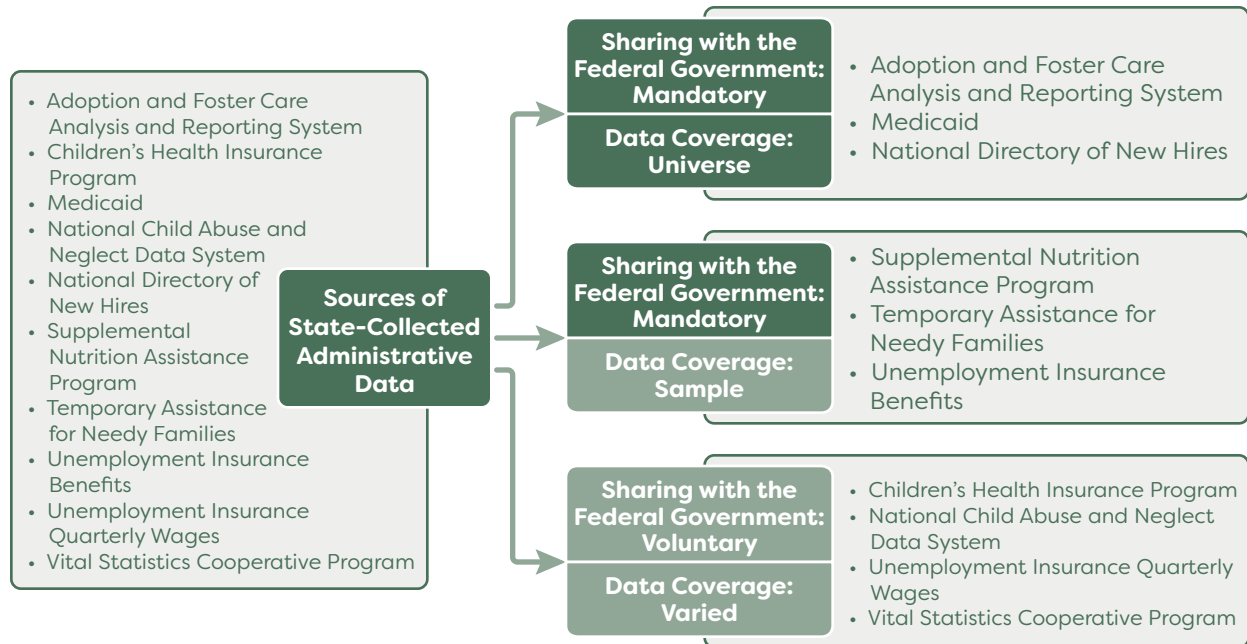
12. For example, the Federal government funds the full cost of SNAP benefits and reimburses the state for approximately 50 percent of its administrative costs. Medicaid is jointly funded by the Federal and state government, where the Federal government pays states for a specified percentage of program expenditures calculated as the Federal Medical Assistance Percentage. TANF operates as a block grant to states, and the program may be state or county administered, depending on the structure of the state.

13. OMB, "Barriers to Using Administrative Data for Evidence-Building;" Office of Planning, Research and Evaluation (OPRE), Administration for Children and Families, U.S. Department of Health and Human Services, "Using Administrative Data in Social Policy Research," OPRE Report 2016-62 (Washington, D.C.: OPRE, 2016); Northwestern University/University of Chicago Joint Center for Poverty Research, "Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development," V. Joseph Hotz, Robert Goerge, Julie Balzekas, and Francis Margolin, editors; Helen Lee, Anne Warren, and Lakhpreet Gill, "Cheaper, Faster, Better: Are State Administrative Data the Answer? The Mother and Infant Home Visiting Program Evaluation-Strong Start Second Annual Report," OPRE Report 2015-09 (Washington, D.C.: OPRE, 2015); and Kelly Maxwell, "Issues in Accessing and Using Administrative Data," OPRE Report 2017-24 (Washington, D.C.: OPRE, 2017).

10. Workforce Data Quality Campaign, submission to the Commission's Request for Comments.

11. Prohibition on development of national database, 29 USC § 3341(b).

Figure 4. Examples of State-Collected Administrative Data Productive for Evidence Building



For a list of additional datasets productive for evidence building, see Appendix D.

different state agencies or programs, leading to protracted and often duplicative administrative requirements. These barriers are compounded for projects in which multiple administrative datasets are to be combined, as each dataset has different access requirements. Similar barriers often exist to access data from local or other jurisdictions.

The Federal government invests significant funds to support the operation of many jointly administered Federal-state programs, providing a basis for some type of Federal reporting requirement. Currently, the volume and type of state-collected administrative data that are reported to Federal program agencies vary considerably across programs. Each program operates under its own set of statutory and regulatory requirements for submitting program data to their Federal sponsor, ranging from the Medicaid program—for which states are required to submit data on all beneficiaries to the Centers for Medicare and Medicaid Services (CMS)—to SNAP and TANF—under which states are required to submit only a sample of data to a Federal agency. Additionally, some of the data submitted to the Federal government have statutory use restrictions; for example, the individual-level

SNAP data submitted to the USDA may only be used for administrative or enforcement purposes, not for evidence building.¹⁴

Figure 4 describes the current landscape of data availability at the Federal level for a set of high-priority state-collected administrative datasets. The figure illustrates where the data sources diverge based on three important distinctions: (1) whether the provision of data from the state to the Federal government is mandatory or voluntary, (2) whether and to what extent the data being shared represent the universe or just a sample, and (3) whether the data contain sufficient personal identifiers to enable integration with other data sources.

Increasing access to administrative data documenting individual earnings is of particular importance, given the many Federal programs and policies that seek to have an impact on earned income. Accessing quarterly earnings data directly from all 50 states, the District of Columbia, and the territories can be challenging and time

14. See use restrictions in the Food, Conservation, and Energy Act of 2008 (Public Law 110-246).

Income Data: Federal Tax and Unemployment Insurance Earnings Data

Wage and income data exist in at least two forms, data from the Unemployment Insurance (UI) programs and Federal Tax Information (FTI). These data sources can be viewed as complementary, covering somewhat different populations and different types of income with differing periodicity. Specifically, the information from the UI programs consists of individual wage records and is available quarterly, whereas FTI offers a broader view of income earned from sources other than wages such as transfers, permits the analysis of household as well as individual income, and is available annually.

Federal law requires states to collect quarterly wage data as part of their administration of state UI programs operated as part of the Federal-state UI system.¹ The U.S. Department of Labor oversees and provides funds for the UI system, but each state administers its own UI program, which includes the collection of data on individual employees and their earnings in each quarter.

Researchers wishing to access UI data currently have three primary access points: originating state systems, the National Directory of New Hires maintained by the U.S. Department of Health and Human Services' Administration for Children and Families, and the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program. Yet, there are significant challenges to accessing and using each of these sources of UI data.

Accessing quarterly earnings data directly from multiple states can be challenging and time consuming, if not impossible. Both the National Directory of New Hires and the Longitudinal Employer-Household Dynamics Program have existing authority to allow researchers access to the wage record data for some statistical activities, but the authority is narrow, making secure access to the data excessively restrictive.²

1. The Federal Unemployment Tax Act (26 USC § 3301 et seq.) and titles III, IX, and XII of the Social Security Act (42 USC chapter 7) form the framework of the system.

2. The Wage Record Interchange System (WRIS), maintained by the Department of Labor (DOL), facilitates the exchange of UI wage records between states for performance accountability purposes, enabling improved reporting on the outcomes experi-

enced by participants in Federally funded state employment and training programs. Either a state or the Employment and Training Administration at DOL may propose research projects using the WRIS, but each proposed project must demonstrate a direct benefit to a qualifying program or activity and approval must be obtained individually from each state whose data will be used.

Researcher access points for FTI are the U.S. Treasury Department and the Census Bureau. Statutory restrictions on accessing data are the major barrier to the use of FTI. The Federal tax code (Title 26 of the U.S. Code) designates FTI as confidential—disclosure to any party is prohibited, except under explicit statutory exceptions. FTI can be used by Treasury employees and a limited set of approved researchers engaging in statistical projects in support of tax administration. The Census Bureau and the U.S. Bureau of Economic Analysis also are authorized limited access under Title 26 and associated regulations for a narrow set of uses related to improving their statistical products. The Title 26 limitations then transfer to any data products these agencies produce with comingled FTI and access to comingled microdata also requires Internal Revenue Service (IRS) approval.

The Treasury Department is the primary access point for FTI. There are several ways that Treasury provides access to tax data for research. One way is through the IRS Statistics of Income Division's Joint Statistical Research Program. An important characteristic of the Statistics of Income program is that it aims to limit, as much as possible, non-IRS researchers' direct access to directly identifiable taxpayer data by assigning Statistics of Income employees as co-researchers in all phases of the work, including assembling and cleaning the required data, performing analyses, and writing reports on the findings. For this reason, the program has very limited capacity. For example, in 2014, only 12 projects were selected from among 80 proposed. Numerous researchers have gained access to tax data through alternative means, including through the Offices of Tax Analysis and Tax Policy within the Treasury Department, which also engage in independent and occasionally joint research.

Data About Births and Deaths in the Population

Data on vital events (births and deaths) are of great value for statistics, evaluation, and policy research. For studies of health, housing, and other policies, there is perhaps no more important outcome variable than death. Vital records also are critical for public health programs and serve a variety of administrative purposes.

The U.S. National Center for Health Statistics (NCHS) is the Federal agency mandated to produce national health statistics based on a cooperative, decentralized system. Data from more than six million vital-event records are collected each year by vital registrars in all states and U.S. territories and transmitted to NCHS for processing and dissemination. NCHS has a statutory mandate (42 USC § 242k) to collect data annually from the records of births, deaths, marriages, and divorces in registration areas. Currently the data collection is limited to data from birth and death records (including fetal deaths), as NCHS discontinued the collection of individual-record marriage and divorce reports after 1995.

Data are collected through the Vital Statistics Cooperative Program. These data are provided through contracts between NCHS and vital registration systems operated in the 57 jurisdictions legally responsible for the registration of vital events, namely states. The contracts support the cost of training and technical assistance to help standardize data quality.

In 2007, NCHS released a new policy on the release of and access to vital statistics microdata to comply with state laws and policies.¹ The revised policy reflects the dual goals to make data available as widely as possible while being responsive to concerns about confidentiality.

The current agreement with the states on the re-release of restricted data containing geographic detail requires a review of all such data requests by the National Association for Public Health Statistics and Information Systems (NAPHSIS), which represents state vital registrars.² The review by NAPHSIS is conducted prior to the NCHS review, and applies to both Federal and non-Federal requests for restricted data files. Researchers in Federal agencies, as well as their on-site or off-site contractors, also can submit project proposals that request exact dates of vital events. If needed, the file with exact dates can include geographic detail. Non-Federal researchers (including Federal grantees) can gain controlled access to exact dates of vital event files only through the NCHS Research Data Center, with approval.

1. NCHS defines microdata as all raw data, including public use and restricted data files.

2. NCHS, National Vital Statistics System, “NCHS Data Release and Access Policy for Microdata and Compressed Vital Statistics Files;” https://www.cdc.gov/nchs/nvss/dvs_data_release.htm (accessed August 10, 2017).

consuming, and in many cases, not possible. Many residents of one state work in a different state; thus, evidence-building efforts frequently necessitate accessing quarterly earnings data from multiple states. While there are currently two possible avenues by which researchers can access data on quarterly earnings from multiple states through a single access point (see the box “Income Data: Federal Tax and Unemployment Insurance Earnings Data”), both avenues are relatively narrow and include numerous restrictions.

Federal resources and technical assistance to states for modernization of information systems

have enabled states to improve their capacity for evidence building. One example of this kind of investment has been the grants administered by the National Center for Education Statistics to support the development of Statewide Longitudinal Data Systems to assess K–12 education outcomes and investments, creating data systems at the state level similar to those prohibited by the student unit record ban at the Federal level. Such investments also have been valuable in enabling states to comply with reporting requirements. For example, while states are required to submit extensive Medicaid data to CMS, the Federal government spends

Longitudinal Employer-Household Dynamics Program

The Longitudinal Employer-Household Dynamics Program (LEHD) is an example of a partnership that incorporates voluntarily provided state-held administrative data with survey and other administrative data held by the U.S. Census Bureau to support statistical analysis and research. However, the current structure of the program limits broader use of the underlying state data for evidence building beyond the narrow purposes of LEHD.

Under the Local Employment Dynamics Partnership, which is a voluntary Federal-state partnership started in 1999, states agree to share historical and ongoing administrative records of Unemployment Insurance (UI) earnings data and the Quarterly Census of Employment and Wages data with the Census Bureau. In return, the Census Bureau produces a longitudinal data infrastructure from which new statistics about the dynamics of local employment and the locations of jobs and workers can be produced.

Researchers seeking access to LEHD data, including the UI wage record microdata, can apply to access the data through a Federal Statistical Research Data Center. The researcher must comply with the requirements for conducting research at a research data center. Among other things, research projects using the LEHD must advance the mission of the Census Bureau, a requirement that precludes many types of valuable research. It also must be consistent with applicable state law. The data use agreement that the Census Bureau enters into with each

state includes an option for the state to streamline the review and approval process by allowing the Census Bureau to make the determination as to whether a proposed research project is consistent with the uses authorized by the state, but currently only 12 states have selected this option.¹ All of the other states review and approve each project individually, meaning that data access is cumbersome and time-consuming, if it can be accomplished at all.

Some of the individuals who commented to the Commission noted that UI wage records remain underutilized because of limitations to access. Commenters expressed support for increasing access to the LEHD data. The data use agreements with each state typically need to be renewed every five years. While at present, 48 states, the District of Columbia, and two U.S. territories have agreements in place with the Census Bureau, their continued participation is not guaranteed. The UI program is governed and partially funded by the Department of Labor, but it does not have direct access to the microdata and is only able to access LEHD microdata for projects supporting the Census Bureau mission.

1. If the research project will include the release of state or substate-level specific results (as opposed to results from a group of states), the state will have the opportunity to review and approve the external research project. If the intent of the external researcher is to release pooled results only, states can choose to waive their opportunity to review each proposed project. If this option is not selected, even external projects that will release pooled results only would need to be reviewed and approved by the relevant states.

\$2–3 billion per year to support more efficient state information technology systems for tracking Medicaid eligibility, enrollment, and claims.¹⁵

Some Federal agencies also have developed productive partnerships with states and other jurisdictions to establish mutually advantageous systems for making data accessible for evidence building. For example, the Vital Statistics Cooper-

ative Program is based on the relationship between the U.S. National Center for Health Statistics and the 57 jurisdictions that collect vital records covering births and deaths in the United States (see the box “Data about Births and Deaths in the Population”). In exchange for collecting data in a uniform manner, states receive funding, training, and technical assistance.

Other examples of Federal-state partnerships are not necessarily financial in nature. For example, the Census Bureau offers data analysis services to state programs willing to make their data accessible

15. CEP staff correspondence with Department of Health and Human Services, Assistant Secretary for Planning and Evaluation staff in March 2017.

for Federal evidence building. One witness during a public Commission meeting—Erin Ulric from Colorado’s Women, Infant, and Children (WIC) Program—described how Colorado benefits from receiving statistics generated by the Census Bureau to monitor program performance and identify populations for targeted outreach. In this sense, secure access to the Colorado WIC data by eligible researchers enables the state to improve the operation of its program by receiving relevant aggregate statistics that inform day-to-day operations. The Longitudinal Employer-Household Dynamics Program (LEHD) also operates as a non-financial partnership (see the box “Longitudinal Employer-Household Dynamics Program”).

If the combination of laws that apply to Federal programs and policies can create confusion, the same can be especially true for state-held data. When individual jurisdictions interpret Federal laws, inconsistencies can arise in whether they permit the use of data on the very same program. Several Federal agencies have taken actions to address inconsistent interpretations across states. For example, to encourage greater use of data from TANF and SNAP, the U.S. Department of Health and Human Services’ Administration for Children and Families and the U.S. Department of Agriculture’s Food and Nutrition Service issued memoranda to state program directors clarifying what data uses are permissible under their respective program laws.¹⁶

Beyond accessibility, members of the evidence-building community noted two additional challenges related to using state-collected administrative data for evidence-building activities. First, Federal programs may not require states or local jurisdictions to collect the same kind of data in the same way. As a result, while data within a single state, program, or jurisdiction may be analyzed with relative ease, when data from multiple sites or states are brought together for analysis, substantial cleaning and consistency edits are

needed for comparability.¹⁷ Second, some states provide data on the full set of program participants or beneficiaries, while others provide only a representative sample. To be integrated with other data and used for evidence building, administrative datasets are more useful when they include all participants or beneficiaries. Supplying the universe file also can reduce the burden on states, as implementing robust sampling procedures requires technical expertise. Furthermore, when the datasets include personal identifiers so program participation can be statistically associated with outcomes of interest, such as college graduation or earnings, members of the evidence-building community are better able to produce valid and reliable estimates about program and policy outcomes.

Challenge #3: Cumbersome and idiosyncratic data access procedures create confusion, impose unnecessary costs, and are a barrier to evidence building, without always providing significant privacy benefits.

The Commission heard repeatedly about the difficulties that cumbersome and onerous procedures, often the result of idiosyncratic processes that vary across government, cause for members of the evidence-building community seeking to securely access confidential data. For researchers and evaluators external to government, no standard for applying for data access currently exists across government agencies, making it necessary to navigate different and varied processes for each agency.

The CEP Survey of Federal Offices found that all 13 PSAs, 5 out of 8 evaluation offices, and 29 percent of other offices (14 out of 49) provide external researchers access to the data they collect.¹⁸ Some agencies have formal programs for external researcher access, with well-established procedures for applying for access. Under these programs, researchers apply for access to data and the agency “qualifies” eligible researchers and approves their projects based on agency-defined

16. Susan Golonka, “Re: U.S. Census Bureau Request for TANF Data,” (memorandum, Washington, D.C.: U.S. Department of Health and Human Services, Administration for Children and Families, December 7, 2016); Thomas Louis and Richard Lucas, “Memorandum of Record Regarding the Sharing of State SNAP and WIC Recipient Data With the U.S. Census Bureau,” (memorandum, Washington, D.C.: U.S. Department of Agriculture and U.S. Department of Commerce, August 12, 2014).

17. Harry Hatry, “Using Agency Records,” in *Handbook of Practical Program Evaluation*, edited by Kathryn E. Newcomer, Harry Hatry, and Joe Wholey, 325–343 (Hoboken, N.J.: Wiley, 2015).

18. Two of the 10 responding evaluation offices and 18 of the 67 responding other offices reported they do not collect data.

processes. Commenters said these processes can often be slow and confusing. Other agencies do not have formal programs for external researcher access, so the application process can be much more ad hoc.

“Key issues encountered include: inconsistent, but often lengthy and cumbersome processes for accessing data. Each agency negotiates its own data sharing agreements, which can add up to significant time and cost for national studies that require data from multiple agencies in many states, particularly when there may be variations in exactly what the process requires and how long it takes.”

– Andrew Weigan, President and Chief Executive Officer at Social Policy Research Associates, Commission Public Hearing, San Francisco, February 9, 2017

These challenges are compounded when a researcher seeks to access multiple agencies’ data. In this situation, the researcher must apply to and be approved for access by all of the agencies. The result can create scenarios in which one agency provides an approval and another does not, necessitating ongoing negotiations between the agencies and the researcher.

Not only does this result in an inefficient process, these inefficiencies might not enhance confidentiality or privacy because the standards for access and use are not based on a common understanding of risk for linked data. Nor do these access procedures result in standard information about use and practice to provide transparency and an ability to be audited.

Moreover, inefficiencies in data access processes for evidence building create administrative expenses and researcher burdens that can impede Federally funded research. The costs associated with excessive administrative burdens were acknowledged by the Congress and the President in a somewhat different context with the enactment of the American Innovation and Competitiveness

Act in 2017, which found that “administrative costs faced by researchers may be reducing the return on investment of Federally funded research and development” and that “it is a matter of critical importance to United States competitiveness that administrative costs of Federally funded research be streamlined so that a higher portion of Federal funding is applied to direct research.”¹⁹

Some of the idiosyncrasies of data access procedures result from a lack of capacity within agencies to devote to this work. The Commission found that agencies vary significantly in their capacity to provide external researcher access to data. A lack of sufficient capacity restrains agencies’ ability to undertake key administrative functions necessary for providing access to confidential data, including the development and dissemination of metadata and even the government’s ability to explain processes and procedures to eligible researchers. A lack of capacity can also result in delays throughout a research project, from approving a researcher’s access request to reviewing the researcher’s output before it can be released.

The Federal Statistical Research Data Centers (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 (see the box “Federal Statistical Research Data Centers”). The Census Bureau and other Federal agencies have collaborated with host institutions to enable approved researchers to access confidential data from multiple agencies through a network of secure data enclaves. While effective for facilitating data access, the FSRDCs are still constrained by a lack of standard access procedures across the agencies making their data available through an FSRDC. In practice, this means that a researcher seeking data other than or in addition to Census Bureau data must still apply to the originating agency for access, which can often result in delays and administrative hurdles necessitating months or years of effort to gain access. Progress is being made on this front, with work on new data access agreements that will enable the FSRDC staff to approve projects on behalf of an originating agency. However, progress is only available to the extent applicable statutes provide agencies such discretion, as discussed earlier in this chapter.

19. American Innovation and Competitiveness Act (Public Law 114-329, Section 201, January 6, 2017).

Federal Statistical Research Data Centers

The U.S. Census Bureau, on behalf of the Federal statistical agencies, operates a network of 27 Federal Statistical Research Data Centers (FSRDCs) in 18 states and the District of Columbia to provide secure access to a range of Federal restricted-use microdata for statistical purposes.

The FSRDCs are partnerships between Federal statistical agencies and research institutions across the country. Because the current structure requires physical presence in order

to secure access to data, geographic proximity for researchers is key. The FSRDC network has grown organically with demand. The first remote FSRDC opened in Boston in 1994. The most recent FSRDCs opened at the University of Kentucky and the University of Colorado Boulder in August 2017. With these additions, the FSRDC network will have doubled in size since 2013. As of 2016, the FSRDC network supported about 900 researchers working on about 250 different projects.

Federal Statistical Research Data Center (FSRDC) Locations



Once access has been approved, researchers encounter other structural issues with FSRDCs such as the requirement to be physically present at an FSRDC site, the limited number of seats available at a given FSRDC, the limitations of current computing resources available, and the wait time to obtain disclosed results.²⁰ Further, through its

fact-finding phase, the Commission heard concerns that the FSRDC network, as it exists currently, would be overwhelmed by the additional demands of supporting an increase in data access for a higher volume of evidence building.

Members of the evidence-building community inside government face challenges similar to those confronted by external researchers. The non-standardized and lengthy processes by which access is negotiated and memorialized via an agreement

20. According to the Census Bureau, the average wait time in 2017 is approximately three weeks for obtaining disclosed results.

were among the most common data access barriers cited by stakeholders. Formal data access agreements (e.g., Memoranda of Understanding or MOUs) between two or more agencies can take years to develop. Delays associated with negotiating MOUs are compounded by the challenges described earlier in this chapter when legal authorities are inconsistent. A lack of clear legal authority can result in extended reviews and negotiations by lawyers within multiple offices or departments prior to granting access. An agency's ability to efficiently execute MOUs can also be impacted by capacity constraints.

“I think ideally you would have a much more streamlined [data access] procedure that was really oriented around statistical research.”

– Raj Chetty, Professor of Economics at Stanford University, Commission Meeting, Washington, D.C., July 22, 2016

In 2014, OMB developed an optional model agreement for agencies to consider, though uptake has been limited to date.²¹ Some states also have pursued similar agreements for use among internal state agencies. Michael Basil, General Counsel in the Illinois Department of Innovation and Technology, described the state's effort to develop an Enterprise Memorandum of Understanding (eMOU) in an effort to improve the use of data at the state level.²² In 2016, an expert panel sponsored by Actionable Intelligence for Social Policy developed model templates for Memoranda of Understanding to help states develop integrated data systems.²³ All of these efforts suggest opportunities for developing a common MOU template for use in the Federal government. In the CEP Survey of Federal Offices,

close to 30 percent (23 out of 79) of responding offices indicated that lack of staff, policies, and procedures to establish data sharing agreements in their own office made it hard for them to get access to the specific dataset they needed.²⁴

Challenge #4: Many Federal departments do not sufficiently or regularly assess their information resources to determine access controls and privacy protections appropriate to the sensitivity of the data.

As described earlier in this chapter, a variety of laws govern access to government data for evidence building. Federal agencies follow these laws, associated regulations, and agency legal interpretations when making decisions about data security protocols and access restrictions for the data they hold.

The Commission finds that many PSAs implement tiered access programs that set data access and security requirements based on an assessment of dataset sensitivity. Tiered access is an application of data minimization, a key privacy safeguard for evidence building as embodied in the Fair Information Practice Principles (described in Chapter 3). Data minimization means giving access to the least amount of data needed to complete an approved project. For example, an eligible researcher's project might earn approval for access to confidential information at a highly secure research data center that requires expert review of all output. Another researcher's project may need only access to a data query tool that runs an analysis, checks for disclosure risk without ever showing individual records, and provides group statistics (see the box “Data Query Tools”). If researchers find that a data tool meets their needs, they may never need to access the underlying confidential data. Similarly, researchers, journalists, and other members of the public may be able to answer their questions by using data that have been masked, de-identified, and approved for release as a public use file or by reference to published statistics. A well-designed and properly implemented data minimization strategy like tiered access can reduce the risk of unauthorized use and unintended harm to individuals.

21. Office of Management and Budget (OMB), “Guidance for Providing and Using Administrative Data for Statistical Purposes” (memorandum M-14-06, Washington, D.C.: OMB, Executive Office of the President, February 14, 2014).

22. Michael Basil, Illinois Department of Innovation and Technology, Commission Meeting, Washington, D.C., September 9, 2016.

23. John Pettila, Barbara Choh, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, and Natassia Rozario, “Legal Issues for IDS Use: Finding a Way Forward,” (Philadelphia: Actionable Intelligence for Social Policy, 2017); <https://www.aisp.upenn.edu/wp-content/uploads/2016/07/Legal-Issues.pdf> (accessed August 10, 2017).

24. Based on 79 offices that reported they collect or use data for statistics, evaluation, research, or policy analysis or spend a portion of their budget for such purposes.

Data Query Tools

Online query tools allow researchers to find answers to their questions without ever accessing individual records. Researchers submit requests for analysis through the tool. The query tool conducts the analysis behind a secure server wall and returns only the result to the researcher. This approach can be applied to support access to data in a way that limits disclosure risks.

The National Center for Education Statistics uses such technology in its DataLab application. The application allows analysts to

conduct research on restricted access data files without needing access to the underlying data. The simplest query tools typically include a table generator that allows the researcher to select variables for rows and columns; the query tools generate the results. More advanced query tools allow researchers to submit code for statistical analyses on the restricted use data. Query tools using underlying confidential data can automatically evaluate output for disclosure risk before returning the output to the researcher.

The National Center for Health Statistics uses a variation of tiered access based on data sensitivity. The agency produces a public use version of the National Health and Nutrition Examination Survey that has been masked to protect individual identities, but requires researchers needing access to human genome data to train and qualify as “sworn agents” of the National Center for Health Statistics before accessing the data in a secure facility. Today, Federal agencies can take advantage of new technologies that allow versions of confidential datasets that are safe for wide public use. Some of those technologies are described further in Chapter 3.

To determine appropriate access, Federal departments should consider the sensitivity of the data in context. A recent law review article on sensitive information indicated that “the conclusion that a particular type of information should be treated as sensitive gives rise to special rules of collection, use, and disclosure as a means to prevent security or privacy harm.”²⁵ Existing guidelines do not require data access decisions to be based on a formal analysis of the sensitivity of the data. The Commission did find, however, that most agencies conduct informal analyses of the sensitivity of data, and several agencies, including all PSAs, conduct formal analyses as part of their standard disclosure review processes.

While staff in Federal departments include experts with the ability to assess the sensitivity of the data, they can benefit from input from stake-

holders, including researchers and privacy advocates, to establish access restrictions based on law, context, and sensitivity. Such a review would ensure that departments give appropriate access to Federal data needed for evidence building while protecting sensitive personal information.

Even with a tiered access approach, the Federal government currently lacks a consistent and objective system for classifying statistical and administrative data based on sensitivity. The absence of such a classification system limits government’s ability to describe access restrictions and data security and privacy protocols appropriate for each level of sensitivity, in a way similar to the well-known classification and handling schedules for national security information.

Harvard University has developed a “data tags” system to code the sensitivity of various datasets that illustrates the potential of such a system. The assigned data tags govern the encryption protocols used for data storage and transmission as well as the requirements for data access.²⁶ The national archive of Denmark is considering using a similar system to assess the sensitivity of its datasets in preparation for the implementation in 2018 of the European Union’s General Data Protection Regulation.²⁷

25. Paul Ohm, “Sensitive Information,” *Southern California Law Review* 88 (2015): 7; <https://papers.ssrn.com/sol3/papers.cfm?abstract-id=2501002> (accessed August 10, 2017).

26. Latanya Sweeney, Mercè Crosas, Michael Bar-Sinai, “Sharing Sensitive Data with Confidence: The Datatags System,” *Technology Science* (October 16, 2015); <https://techscience.org/a/2015101601> (accessed August 10, 2017).

27. “EU General Data Protection Regulation (GDPR),” EU GDPR Portal; <http://www.eugdpr.org/> (accessed August 10, 2017).

Figure 5. Model of Sensitivity Levels for Federal Data

Level	Sensitivity	Description
5	Crimson	Maximally restricted. Highly sensitive. Identifiable records from data collected with a promise of confidentiality.
4	Red	Restricted. Sensitive. Identifiable records from data collected with a promise of confidentiality.
3	Yellow	Restricted. Crimson or Red datasets modified by technologies that mask individual records (e.g. data query tools, differential privacy).
2	Green	Minimally restricted. Not sensitive. Data files made available to the public but subject to procedures designed to raise accountability by users, such as registration before accessing.
1	Blue	Public data. Most safe. Open data.

Based on the Harvard model, Figure 5 provides an example of sensitivity levels that could be applied to Federal data.

Recommendations

The Commission concludes that a priority for evidence building now and in the future is to *equip the evidence-building community with secure, restricted access to data to facilitate the generation of high quality evidence about government programs and policies*. Recognizing the benefits of maximizing privacy protections and the utility of existing data, the Commission identified a series of recommendations to better enable secure, restricted access to increased amounts of the data already collected by government. Many of these data are highly relevant for informing and evaluating an array of Federal programs and policies. Based on the evidence examined by the Commission,

when taken together with the privacy enhancements proposed in Chapter 3, the following recommendations will safely harness the potential of data to produce more and better evidence in the future.



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 the implementation of innovative privacy-enhancing technologies.

Addressing the Federal government's capacity for evidence building should begin by enabling consistent, transparent, and accountable secure access to data. Recommendation 2-1 will establish NSDS to help address the existing gaps in the capacity to meet this need on a consistent, government-wide basis. Because the evidence-building community is highly decentralized, gaps exist in the consistent application of these functions and a service with the primary mission to address these gaps is needed (see Figure 6). Specifically, a service charged with facilitating access and ensuring protection of data for evidence building is a direct approach to ensure evidence is more routinely generated using privacy best practices.

The NSDS should support secure access to confidential datasets in a setting that meets stringent protection standards. The NSDS will apply strict data minimization techniques to ensure researchers accessing combined data will use datasets with as much information removed as is possible while still meeting the research need. When two or more datasets will be combined, only a narrow group of qualified and trained employees will have access to direct identifiers to conduct the linkage.

The NSDS also should be responsible for developing and implementing state-of-the-art methods to safely combine confidential administrative and survey data from two or more Federal departments for evidence building. These combined data will provide new opportunities for evidence building. The NSDS's existence will highlight how important it is for the Federal government to use the

data it already collects with an unparalleled level of transparency about who is using which data for what statistical purposes.

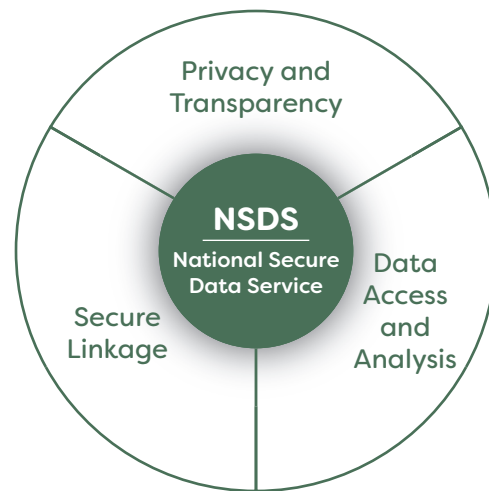
The NSDS should help develop and apply cutting-edge technologies to create highly privacy-protective versions of combined datasets. A well-implemented tiered-access system will limit the need for access to combined datasets containing identifiable records.

Like the existing PSAs, the NSDS must have an exclusively statistical mission. Distinct from PSAs, which typically have a specific topical focus, the NSDS will focus more functionally on data access and confidentiality for use in evidence building, especially across topical areas. The NSDS should be designated by the Congress and the President as a PSA under CIPSEA to afford the entity protections already in law and embodied in other recommendations throughout this report. The statute creating the NSDS should expressly prohibit its use for non-statistical activities.

The NSDS should be located in the U.S. Department of Commerce, building upon the Department's extensive portfolio of statistical and data-related bureaus and expertise existing within the Census Bureau, the U.S. Bureau of Economic Analysis, the U.S. National Institute of Standards and Technology, and the U.S. National Oceanic and Atmospheric Administration, among others. However, it should be situated in such a way as to provide independence sufficient to set strategic priorities distinct from any existing Commerce agency and to operate apart from policy and related offices. The NSDS also must be organized in such a way as to prioritize support for evidence building across government, rather than support specific to any one department, as is the case for existing PSAs.

The institutional placement and governance of the NSDS should be considered in terms of any implications for key objectives: (1) transparency and trust, (2) support for evidence building, (3) strategic coordination and cooperation, (4) confidentiality and security, (5) authority and flexibility, (6) scalable functionality, and (7) sustainability. An institutional placement within an existing Federal department has a practical benefit, particularly the ability to leverage shared services for administrative functions, as well as existing professional staff, established levels of public trust, and operational knowledge.

Figure 6. Key Functions of the National Secure Data Service (NSDS)



REC. 2-2: The NSDS should be a service, not a data clearinghouse or warehouse. The NSDS should facilitate temporary data linkages in support of distinct authorized projects.

The Evidence-Based Policymaking Commission Act charged the Commission with considering whether the country needed a data “clearinghouse.”²⁸ Consistent with this charge and after careful consideration of the issue, the Commission has concluded that a clearinghouse should not be created.

Specifically, the Commission interprets the word “clearinghouse” as connoting a data storage facility that permanently stores records from multiple databases from multiple agencies and, therefore, grows with each new data linkage. Previous panels and commissions either have come close to recommending this type of clearinghouse or did recommend one.²⁹ These previous efforts provoked well-founded concerns about the potential privacy harm such a clearinghouse could raise.

As further elaborated in Chapter 4, the Commission rejects the clearinghouse model in favor of the NSDS. The NSDS should be designed to link


28. Evidence-Based Policymaking Commission Act of 2016 (Public Law 114–140, Section 4(b), March 30, 2016).

29. As described in Chapter 4, a committee in the 1960s proposed creating a national data bank to address the decentralized nature of the Federal government’s data infrastructure.

data on an individual project basis only. In contrast to a clearinghouse, it should not lead to the establishment of a store of data that grows with every research project conducted. The data linked for a project through the NSDS should be kept structurally separate from other data linked through the NSDS for other projects. By strictly enforcing this design, the NSDS will further the goal of increased access to and use of data for specific research and evaluation efforts, without unduly increasing the potential for privacy harm.

The Commission envisions the NSDS working in concert with the existing network of PSAs to facilitate secure access to government data for evidence building conducted by those inside and outside of government. The PSAs currently house and manage much of the data critical to Federal evidence building in order to fulfill their statutory missions. They collectively have appropriate infrastructure, technical expertise, and legal authority to protect data. Therefore, the NSDS will both call upon PSAs to address requests by researchers and evaluators, and assist PSAs in furthering their own research to improve the quality and efficiency of existing statistical programs. Such a relationship will allow the NSDS to facilitate access and create approved linkages, while also ensuring the underlying data remain dispersed in their originating agencies.

The NSDS may also support evidence building in states and localities, especially as part of creating mutually beneficial partnerships around use of Federal, state, and locally held data. The NSDS could also offer analytical services to Federal, state, and local agencies that lack capacity to undertake the work on their own, providing both needed capacity and ensuring privacy best practices.

 **REC. 2-3: In establishing the 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 the 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.**


Weighing its charge and the “optimal arrangement” for data collected by government, the Commission concluded that more of the data the government holds should be accessible for statistical purposes to inform programs and policies. Such access optimally will be grounded in clearly articulated authority from the Congress for program agencies to make their data available for statistical purposes and for agencies envisioned as recipients to use the data for those purposes. Absent clear direction from the Congress, many Federal departments currently may choose not to use existing authorities or may interpret their statutes as lacking these authorities. To provide clarity about permissible statistical uses, the Commission recommends building on the existing CIPSEA framework.

The joint benefits to privacy and data access envisioned by the Commission can be realized simultaneously within an appropriate enabling legal framework. The existing CIPSEA framework, including the provision that authorizes the sharing of business data among three designated statistical agencies, permits effective statistical uses of data within a protected enclave. The Commission recommends that the CIPSEA framework be extended by providing NSDS with the powers and exemptions necessary to use, for statistical purposes, data collected by other agencies. To ensure that the acquisition and use of data by the NSDS occurs in a privacy-protecting manner that is at the frontier of what is feasible, the Commission’s recommendation also requires that NSDS adhere to specific new and stringent privacy qualifications outlined elsewhere in the Commission’s recommendations. The Commission does not intend for the new authority given to the NSDS to curtail existing authorities to acquire and link data currently being utilized by the PSAs, but does see value in encouraging all such activity to occur within the CIPSEA framework. In fact, the Commission believes that some PSAs will elect to use the NSDS to facilitate their own research with potential to address the quality, burden, or cost-efficiency of statistical production, and that the NSDS will call upon the PSAs to provide data to facilitate others’ research activities.

As a companion to the CIPSEA amendment, the Commission also recommends that the Congress and President amend the Privacy Act to extend to NSDS the existing exception to consent requirements that applies when agencies disclose records within a system of records to the Census Bureau. CIPSEA authorization alone may not be sufficient to

enable needed access to data, as agencies may also have to satisfy the Privacy Act's requirement to obtain consent before disclosing information to third parties. The recommendation acknowledges that the evidence-building landscape has evolved significantly since the enactment of the Privacy Act. With the Census Bureau exemption, the Congress demonstrated its belief that the provision of administrative data under strong confidentiality protections for exclusively statistical purposes was safe enough to exempt it from the Privacy Act's consent requirements. In light of the stringent privacy and transparency protections built into its procedures, the Commission believes there is justification to extend the same narrow exemption to the NSDS.


The Commission does not envision that the NSDS will be a locus of statistical production. That role will remain, as today, with the statistical agencies in which the Congress has vested the responsibility for such work. For the purpose of improving the efficiency and precision of key statistical indicators and datasets, these agencies should be able to engage with the NSDS to carry out research on and development of new statistical products using data linked under the authorities newly granted to the NSDS. But once research and development gives way to production, that activity—along with the linking of data necessary to support it—will need to revert to the relevant statistical agency or agencies. The Commission encourages the Congress to consider conforming amendments, including to CIPSEA and the Privacy Act, to allow these improvements in statistical production processes to occur.

 **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.**

The Commission identified several datasets that are highly valuable for evidence building (see Appendix D). Several of the statutes that govern use of these datasets contain explicit prohibitions or strong restrictions on use for the purposes described in this report. For example, the Higher Education Opportunity Act limits the use of data collected to operate Federal student financial assistance programs to “administration of financial aid

programs.”³⁰ Similarly, under Title 26 of the U.S. Code, the predominant purpose of research conducted using confidential Federal tax information must be “tax administration.”³¹ Likewise, the law generally applicable to the Census Bureau, Title 13 of the U.S. Code, permits access to data only for the narrow purpose of improving Census Bureau programs.

Where deemed appropriate, the Congress should act to amend relevant statutes to authorize access and use of data for statistical purposes. The recommendation encourages the Congress to review its prior determinations in the context of the desire for increased use of evidence and the modernized legal and security environment contemplated in this report.

 **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.**

Legislative bans on the collection and use of data create barriers to the use of information for evidence building. The Commission recognizes that the creation of bans may be a mechanism for setting priorities for the collection and use of data in a complex policy context. While it is clearly within the purview and responsibility of the Congress to establish such priorities, absent articulated criteria, bans on data collection and use create a serious impediment to evidence-based policymaking, and could make it difficult or impossible to hold government activity accountable. Provided that the collection and use of data occur consistent with the Commission's guiding principles, particularly with regard to privacy protection, such collection and use generally should be allowed.

The Commission concluded that any use of data may carry both value and the risk of harm. There is also, however, a risk of harm associated with the investment of billions of dollars in taxpayer funds without adequate monitoring and study of the effectiveness of the taxpayers' investment.

30. Higher Education Opportunity Act (Public Law 110-315, August 14, 2008); <https://www.gpo.gov/fdsys/pkg/PLAW-110publ315/pdf/PLAW-110publ315.pdf> (accessed August 10, 2017).

31. Internal Revenue Code, 26 USC § 6103(h).

In considering the application of a ban on data collection and use, the Congress should create a means to weigh the various, and sometimes competing, interests at stake. The Commission emphasizes that encouraging the review of bans does not mean privacy protections should be undermined or that data collection and use should be guaranteed. Rather, the Commission is encouraging the Congress to develop reasonable criteria to help ensure that legislators have the tools and opportunity to carefully weigh the implications of significant bans.

As noted previously, the Commission understands the complexity, motivations, and concerns involved in weighing privacy against other considerations. History is replete with examples of getting this balance wrong. Yet, with modern privacy-enhancing technologies, a well-developed legal framework, and a strong commitment to transparency, the Commission believes it is possible to use data to evaluate the effectiveness of policies and ensure accountability to stakeholders, while also protecting privacy. The Commission also believes it is incumbent on the Congress and the President to regularly and thoroughly consider the value of policies that preclude evidence building against the prospective public value of the evidence that could be produced.



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 solely for statistical purposes. The data should be available through a single Federal source for solely statistical purposes.

The Commission finds that increasing access to state-collected administrative data on quarterly earnings is of primary importance. Numerous Federal programs and policies seek to impact earned income. Data on quarterly earnings are already collected by states through their administration of unemployment insurance programs, but these data are not easily accessible by researchers for use in evidence-building activities. For a more complete picture of wage recipients, access to quarterly wage data from multiple states is critical.

The Commission identified several strategies for improving access to administrative data on quarterly earnings for statistical purposes, including expanding access to existing multi-state data sources such as the National Directory of New Hires or LEHD or developing a new national UI wage record system (see the box “Expanding

Expanding Access to Quarterly Earnings Data

The Commission identified several possible ways to achieve the goal of improving access to administrative data on quarterly earnings for statistical purposes. One approach could be to expand access for statistical purposes to wage records maintained by the National Directory of New Hires (NDNH) system or to wage records maintained by the U.S. Census Bureau as part of the Longitudinal Employer-Household Dynamics Program (LEHD).

Alternatively, the Congress and the President could create a national Unemployment Insurance (UI) quarterly earnings data system within the U.S. Department of Labor. A national UI quarterly earnings data system could build on the same UI wage record data that currently support both the NDNH system and the LEHD program. In fact, states consistently have expressed interest in being able to access inter-

state wage record data for statistical activities. Currently, they have uneven access to these data, despite the mandatory reporting requirement for NDNH and the existence of voluntary reporting for the LEHD program.

Creating a single system to which states report UI wage record data could reduce the burden on states, who currently report the same data to multiple entities, and could reduce the burden on Federal agencies that spend a significant amount of resources negotiating multiple agreements and memoranda of understanding with the states. A single system also could streamline access to this highly valuable data source for statistical activities, while enhancing the security and privacy of the data through development of standardized procedures for data submission and access.

Access to Quarterly Earnings Data”). Each of these approaches has strengths and weaknesses. As such, the Commission chose to focus its recommendation on the desired outcome—the availability of national administrative data on quarterly earnings, accessible from a single Federal source, for use in evidence building—rather than articulating a specific approach.



REC. 2-7: The President should direct Federal departments that acquire state-collected administrative data to make them available for statistical purposes. Where 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.

The Commission strongly believes that increasing access to state-collected administrative data is vital to efficiently increasing the volume and quality of evidence that can be produced and used for policymaking. Availability of state-collected administrative data at the Federal level is highly variable. In considering which administrative data to prioritize for enhanced accessibility, policymakers might consider the level of Federal investment in the program, the value of the data for evidence-based policymaking, and the level of difficulty that might be encountered in seeking to increase access to a particular data source. For example, SNAP, UI, and TANF represent three important programs in the current safety net in the United States, and each involves substantial Federal and state expenditures.³²

Currently there are two distinct challenges in using state-collected administrative data for Federal evidence-building purposes—the data that are available for evidence building are limited and what are available are not easily accessible. While the Federal government already has the statutory authority to collect a sample of linkable administrative

microdata from states on participants of all three programs, sample data have limited use in evidence building. Universe data, which includes all program participants or beneficiaries, are the data that have maximum value for evidence building. In addition, existing data are inaccessible for evidence building. Recent efforts by the Census Bureau to enter into data sharing agreements with states to access these data have been advancing, although slowly. As of July 2017, there were 15 states sharing SNAP data and 9 states sharing TANF data through this effort, with varying years of data available, limiting the ability to easily conduct multi-state analysis.³³

In cases where barriers to access are significant and deemed to be misaligned with both the amount of Federal investment in a program and evidence-building needs, the Congress should consider whether to enact statutory changes to require the submission by states of administrative data on all program participants that could be combined with other data sources for exclusively statistical activities. While significant, the challenges of expanding the existing statutory authority to require the submission of universe data in the case of SNAP, UI benefits, and TANF, for example, could be lessened considerably with the provision of supports to help states comply with a new requirement. The Commission expects that doing so for these three programs could lead to especially valuable information for policymakers and program managers.

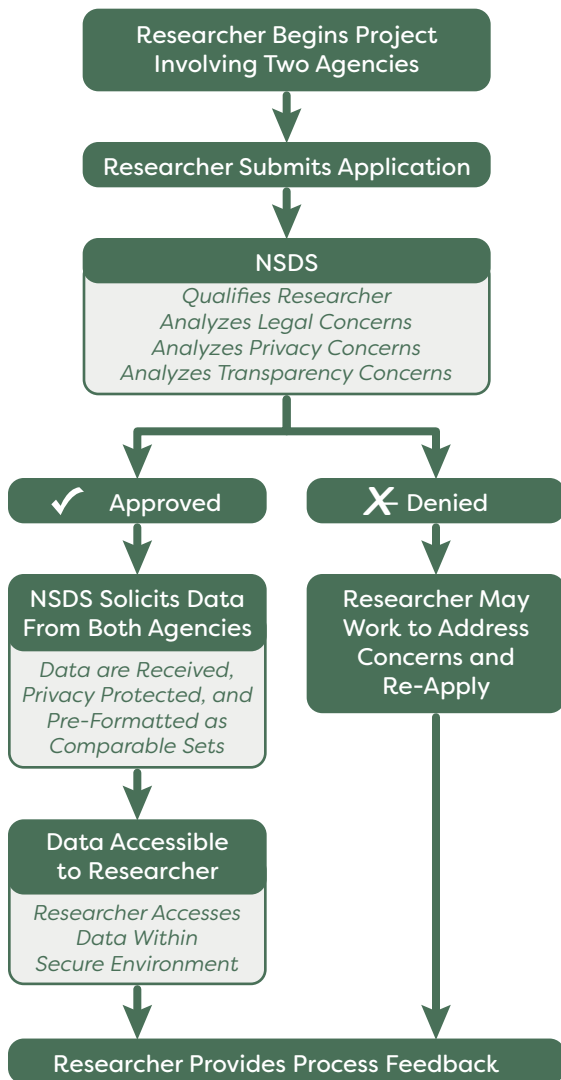
Recognizing the potential value of state-held datasets related to Federal programs, the Commission recommends that Federal departments conduct a thorough review of relevant statutes and regulations governing the accessibility of state-collected data as well as any statutory use restrictions related to evidence building to identify barriers that impede access to the data by qualified researchers seeking to use the data for statistical purposes. The Congress and the President should require Federal agencies to submit a report to the Congress summarizing the barriers they identify from a review of the relevant statutes and regulations governing these collections.

Across the existing jointly administered Federal-state programs in which the submission of universe administrative microdata is statutorily required, the Federal government provides a

32. Trudy Renwick and Liana Fox, *The Supplemental Poverty Measure 2015* (Washington, D.C.: U.S. Census Bureau, September 2016) table 5b; <https://www.census.gov/content/dam/Census/library/publications/2016/demo/p60-258.pdf> (accessed August 10, 2017).

33. U.S. Census Bureau, “Administrative Data Inventory,” <https://www2.census.gov/about/linkage/data-file-inventory.pdf> (accessed August 10, 2017).

Figure 7. Research Process After Instituting NSDS



significant amount of support to enable states to comply with reporting requirements. The provision of funding for information technology infrastructure and technical assistance can assist states in building the necessary data collection systems and reporting structures to be able to comply with a new requirement for increased data collection and transmission and to address issues of data integrity. Federal departments should explore and adopt incentives to support either compliance with a requirement for states to make data available to the Federal government or to incentivize the voluntary action of making administrative data available to the Federal government in cases where a requirement is not established.

REC. 2-8: The Office of Management and Budget should promulgate a single, streamlined process for researchers external to the government to apply, become qualified, and gain approval to access government data that are not publicly available. Approval would remain subject to any restrictions appropriate to the data in question.

When the regulatory framework permits data access for statistical purposes, researchers still need clear and well-documented protocols and procedures for data access. A centralized and known set of procedures for requesting access to data allows greater focus on protecting privacy and avoids inadvertently establishing onerous roadblocks to accessing data. A streamlined process will ensure that data access requests are all adequately scrutinized. In fact, the approach for improving access would acknowledge differences in the sensitivity levels of data and set appropriate access requirements and data security and privacy protocols in accordance with applicable laws or regulations.

The Commission recognizes OMB as the primary actor for Recommendation 2-8 in order to capitalize on OMB's existing legal authority and experience in setting information and statistical policy that applies government wide. OMB should include key stakeholders, both within and external to the government, throughout the process of establishing new procedures and standards related to access and sensitivity.

The NSDS will play a lead role in implementing these new standards (see Figure 7). NSDS will be responsible for consulting and collaborating with Federal departments to assign appropriate sensitivity levels to datasets they hold. Because sensitivity may change over time, it would be appropriate for sensitivity analyses to be reviewed whenever important new information about potential risks becomes available. Sensitivity levels could be based on a standard classification system for data sensitivity, as discussed earlier in the chapter.


Evaluators and researchers alike could similarly be more readily informed about what data are available at what levels of sensitivity with what access restrictions. As appropriate, the Federal government should provide detailed technical data documentation useful in planning projects and designing research studies, as discussed further in Chapter 4. ■


3


Enhancing Privacy Protections for Federal Evidence Building


Improve data quality, integrity, and security by updating laws to require risk assessments for data releases, embracing cutting-edge technologies, and prioritizing data stewardship and the public trust.

Recommendations

 **3-1:** The Congress and the President should amend the Privacy Act and the Confidential Information Protection and Statistical Efficiency Act (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 be made available only 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.

 **3-2:** The President should direct Federal departments, in coordination with the National Secure Data Service, to adopt state-of-the-art database, cryptography, privacy-preserving, and privacy-enhancing technologies for confidential data used for evidence building.

 **3-3:** The President should direct Federal departments to assign a senior official the responsibility for coordinating access to and stewardship of the department's data resources for evidence building in collaboration with senior department information technology, privacy, and other leaders. A Principal Statistical Agency head, or other appropriately qualified senior official, should serve this function.

 **3-4:** The Congress and the President should enact legislation to codify relevant portions of Office of Management and Budget Statistical Policy Directive #1 to protect public trust by ensuring that data acquired under a pledge of confidentiality are kept confidential and used exclusively for statistical purposes.

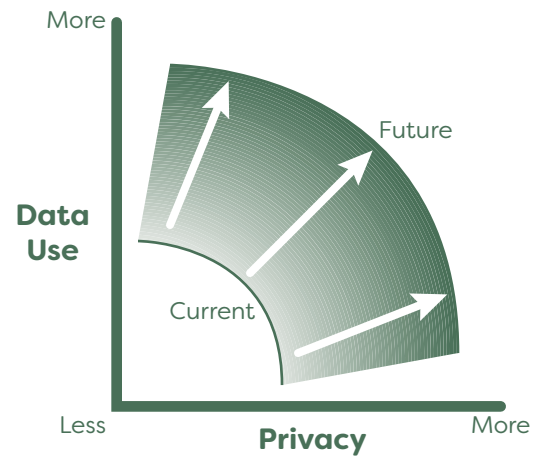
The Federal government has a legitimate need and responsibility to use data for evidence building. At the same time, the public has a legitimate interest in knowing that their government is protecting their privacy while using their data. With emerging technologies, the Commission believes that the country can have both (see Figure 8). The knowledge and technology now exist to achieve the Commission’s vision for improving secure access to confidential data while simultaneously enhancing privacy. Through law and policy, it is possible to ensure that Federal departments are consistently adopting the most up-to-date methods to protect data used for evidence building.

“Privacy, when properly implemented through good programs and early on, promotes innovation in every case and allows for and encourages the wide scale adoption of new technology. In fact, privacy, when implemented correctly, fosters more information sharing, not less.”

– Marc Groman, then-Senior Advisor for Privacy at OMB and Chair of the Federal Privacy Council, Commission Meeting, Washington, D.C., September 9, 2016

Throughout its deliberations, the Commission prioritized consideration of privacy on behalf of the American public. The importance the Commission attached to privacy is captured in the first of the guiding principles for evidence-based policymaking stated in Chapter 1: “Individual privacy and confidentiality must be respected in the generation and use of data and evidence.” The Evidence-Based Policymaking Commission Act called for the President and the Congress to appoint one-third of the Commissioners specifically for their expertise in “protecting personally identifiable information and data minimization.”¹ The Commission devoted two public meetings to testimony about privacy and two other meetings included substantial discussions of how state, Federal, and international governments

Figure 8. The Privacy/Data Use Frontier Curve



Privacy-Enhancing Technologies Allow Greater Protection of Privacy While Increasing Access to Data

provide secure access to confidential data. In addition, privacy considerations have underscored all of the Commission’s deliberations and priorities. Privacy, data security, confidentiality, and public transparency were frequent themes in written comments provided to the Commission and in each of the three public hearings. Finally, Commissioners and staff sought out additional experts on privacy-protective technologies and approaches to develop the recommendations in this chapter.

Just as the Fair Information Practice Principles guide the Federal government’s privacy policy, the Commission believes that these Principles should similarly guide the use of data for evidence-based policymaking in accordance with the Commission’s guiding principle on privacy (see the box “Fair Information Practice Principles”). By considering each of the Fair Information Practice Principles in the context of evidence building, the Commission has taken an ethical approach to data use that addresses both privacy and the need to generate accurate and reliable evidence. Many of the Commission’s recommendations in other chapters address transparency, individual participation, purpose specification, data minimization, use limitation, and accountability and auditing. The recommendations in this chapter specifically address the principle of data quality and integrity and the principle of security, and also relate to other principles, especially transparency.

1. Evidence-Based Policymaking Commission Act of 2016 (Public Law 114–140, Section 3(a), March 30, 2016).

Fair Information Practice Principles¹

- **Transparency:** notify individuals regarding collection, use, dissemination, and maintenance of personally identifiable information (PII)
- **Individual Participation:** involve the individual in the process of using PII and, to the extent practicable, seek individual consent for the collection, use, dissemination, and maintenance of PII
- **Purpose Specification:** articulate the authority that permits the collection of PII and specifically articulate the purpose(s) for which the PII is intended to be used
- **Data Minimization:** collect PII that is directly relevant and necessary to accomplish the specified purpose(s) and only retain PII for as long as is necessary to fulfill the specified purpose(s)
- **Use Limitation:** use PII solely for the purpose(s) specified
- **Data Quality and Integrity:** ensure that PII is accurate, relevant, timely, and complete
- **Security:** protect PII through appropriate security safeguards
- **Accountability and Auditing:** be accountable for complying with these principles and audit the actual use of PII to demonstrate compliance

1. U.S. Department of Health, Education, and Welfare, *Records, Computers, and the Rights of Citizens*, Report of the Secretary's Advisory Committee on Automated Personal Data Systems (Washington, D.C., July 1973); <https://www.justice.gov/opcl/docs/rec-com-rights.pdf> (accessed August 10, 2017); The White House, "Appendix A. Fair Information Practice Principles," *National Strategy for Trusted Entities in Cyberspace: Enhancing Online Choice Efficiency, Security and Privacy* (2011); https://obamawhitehouse.archives.gov/sites/default/files/rss_viewer/NSTICstrategy_041511.pdf (accessed August 10, 2017).

While privacy harms can come from many sources, the Commission's recommendations focus on establishing laws, policies, and procedures to minimize the risk that confidential data used for evidence building may be misused to cause privacy harm. Privacy scholar Ryan Calo notes that privacy harms can be both subjective and objective: "The subjective category of privacy harm is the perception of unwanted observation. This category describes unwelcome mental states—anxiety, embarrassment, fear—that stem from the belief that one is being watched or monitored... The objective category of privacy harm is the unanticipated or coerced use of information concerning a person against that person. These are negative, external actions justified by reference to personal information."² Privacy harm occurs when an individual is adversely affected because of misuse of data; aspects of this harm can also affect businesses or other entities if, for example, their confidential data are inappropriately accessed and used. Respecting privacy in the generation and use of

evidence means minimizing the risk of such privacy harms.

Risk is "a measure of the extent to which an entity is threatened by a potential circumstance or event, and typically is a function of: (i) the adverse impact, or magnitude of harm, that would arise if the circumstance or event occurs; and (ii) the likelihood of occurrence."³ Risk can come from bad actors who actively attempt to break into confidential data that they know they are not authorized to use and from well-intentioned people who inadvertently violate the procedures established to protect privacy. Risk also can come from outdated data encryption protocols and uneven capacity to apply up-to-date data protection techniques. An increasingly important source of risk is the growing number of publicly available data sources containing personal information that can be combined with publicly available datasets designed to be de-identified, thus permitting the

2. Ryan Calo, "The Boundaries of Privacy Harm," *Indiana Law Journal* 86, no. 3 (2011); <https://ssrn.com/abstract=1641487> (accessed August 10, 2017).

3. Office of Management and Budget, "Managing Information as a Strategic Resource," Circular A-130 (updated July 27, 2016); <https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/OMB/circulars/a130/a130revised.pdf> (accessed August 10, 2017).

connection of identities with previously private, sensitive information. This risk must be managed.

This chapter describes a need to better protect individuals from privacy harms caused by unauthorized use or disclosure of Federal confidential data. The Commission developed specific recommendations about the quality, integrity, and security of confidential data to prevent individuals and organizations from being identified without authorization. Private sector research has contributed to rapid advances in computer science that make data safer and easier to access. Adoption of these emerging technologies can accelerate secure access to the government's confidential data resources for evidence building. Federal departments need leaders with the skills and vision to fully harness the promise of privacy-protective data stewardship.

“What people care most about is not simply restricting the flow of information but ensuring that it flows appropriately.”⁴

– Helen Nissenbaum, Professor of Information Science at Cornell Tech

Findings

Protecting privacy has been a priority for Federal evidence building for many decades. Many Federal departments use modern methods to keep data secure and confidential, but government practices must adapt to new threats and take advantage of technologies that better protect data.

For decades, Principal Statistical Agencies (PSAs) such as the U.S. Census Bureau have demonstrated responsible stewardship of data collected through censuses and surveys. These data are used to create important national statistical indicators, such as the unemployment rate, gross domestic product, health outcome statistics, and crime statistics. Public trust in the Federal Statistical System historically has been high, with a majority of Americans concluding that government can be trusted to produce statistics that

are respectful of privacy.⁵ Not surprisingly, public trust in the accuracy and validity of statistical data reflect the public's trust in the statistical agencies that produced them.⁶

The laws governing PSAs give them the authority to protect confidential data by allowing use of the data for exclusively statistical purposes. Statistical purposes include analyzing individual records along with many other individual records to produce descriptions of groups of people or creating averages about groups within society, the economy, or the environment. PSAs cannot and do not use data records about individual people, businesses, or other entities for non-statistical purposes, such as administering a program, determining benefits, or enforcing laws.

PSAs pioneered and continue to advance “statistical disclosure limitation” techniques, discussed below, with the purpose “to ensure that the risk of disclosing confidential information about identifiable persons, businesses or other units will be very small.”⁷ PSAs have a long history of providing secure access to the least amount of data needed for approved statistical purposes and of ensuring that researchers abide by strict rules to protect confidentiality.

Reducing barriers to accessing and using administrative data for evidence building requires the same kind of attention to confidentiality that is embedded in the culture and history of the PSAs. Through the establishment of disclosure review boards, PSAs apply best practice statistical disclosure limitation techniques and assess the risk of public data releases. Disclosure review boards assess proposed public data releases to ensure that enough identifying information has been removed or masked to protect confidentiality. All Federal

5. Jennifer Hunter Childs, “Understanding Trust in Official Statistics in the U.S. and Implications For Administrative Record Use” (prepared for the Commission, Washington, D.C., December 2016).

6. Melissa Mitchell, Jennifer Hunter Childs, and Morgan Earp, “Monitoring and Detecting Shocks That Influence Public Trust Towards the Federal Statistical System” (paper presented at the 68th Annual Conference of the American Association of Public Opinion Research, Boston, MA, May 2013); http://www.aapor.org/AAPOR_Main/media/AnnualMeetingProceedings/2013/Session_H-5-2-Mitchell.pdf (accessed August 10, 2017).

7. Federal Committee on Statistical Methodology (FCSM), “Report on Statistical Disclosure Limitation Methodology,” Statistical Policy Working Paper 22 (Washington, D.C.: FCSM, December 2005); <https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/spwp22.pdf> (accessed August 10, 2017).

4. Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford, CA: Stanford University Press, 2010).

departments should do the same. When government pledges to keep data confidential, the data should have strong protections, and data use should generally be made known to the American public.

The Federal government's Open Data initiative has made more government data publicly available than ever before. Even so, "agencies must review the information collected or created for valid restrictions to release to determine whether it can be made publicly available, consistent with the *Open Government Directive's* presumption in favor of openness, and to the extent permitted by law and subject to privacy, confidentiality pledge, security, trade secret, contractual, or other valid restrictions to release."⁸

Throughout its fact-finding phase, the Commission learned about potential threats to the confidentiality of Federal data used for evidence building and heard from experts about new and emerging methods to address such threats. The findings presented in this chapter highlight three challenges that keep the Federal government's confidentiality protections from being as strong as possible. First, laws governing Federal data sometimes do not require a risk assessment or disclosure review prior to the public release of de-identified confidential data. Second, Federal departments must adapt to new threats to information security and privacy and take advantage of emerging technologies that better protect data. Third, many Federal departments lack senior leadership focused on or experienced in data stewardship for evidence building, and the PSAs, while often best positioned to provide such leadership, have uneven independence and authority.

Challenge #1: Key laws governing Federal data do not require the application of best practice statistical disclosure limitation techniques or a risk assessment prior to the public release of de-identified confidential data.

As discussed in Chapters 1 and 2, two laws that govern much of the confidential Federal data used for evidence building are the Privacy Act of 1974⁹

and Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA).¹⁰ These laws seek to protect confidentiality by requiring that released records be in a form that is "not individually identifiable"¹¹ (Privacy Act) or "prevents the identity of the respondent...[from being] reasonably inferred by either direct or indirect means"¹² (CIPSEA). The Privacy Act and CIPSEA, as well as other Federal laws governing confidential data, share a basic goal: to ensure that information given to the Federal government under a promise of confidentiality is not released to anyone in identifiable form, except as allowed by law.

As described in Chapter 2, CIPSEA authorizes access to identifiable confidential data for exclusively statistical purposes by (1) trained staff in PSAs and recognized statistical units and (2) by external researchers who are trained and qualified by PSAs. Even then, direct identifiers are removed from analysis files and their access protected and limited to a few technical staff. Several Federal agencies that collect confidential administrative data under the Privacy Act also provide secure access to identifiable information for research purposes.

PSAs release "de-identified" confidential data collected under CIPSEA as part of their mission and legal authority to disseminate data in their topical domains. Many program agencies that collect confidential information under the Privacy Act also release "de-identified" data under their own statutory authorities. These public-use files are easily accessible and widely available; many public-use files can be downloaded from department websites.

In the Federal government today, however, there is great variation in the meaning of "de-identified" data. Following CIPSEA implementation guidance from OMB, statistical agencies routinely go beyond removing direct identifiers to apply appropriate statistical disclosure limitation techniques before seeking disclosure review board approval to release de-identified datasets.¹³ In contrast, program

8. Office of Management and Budget (OMB), "Open Data Policy: Managing Information as an Asset" (memorandum M-13-13, Washington, D.C.: OMB, Executive Office of the President, November 8, 2016); <https://project-open-data.cio.gov/policy-memo> (accessed August 10, 2017).

9. Privacy Act, 5 USC § 552a et seq.(1974).

10. Confidential Information Protection and Statistical Efficiency Act of 2002," (CIPSEA), (Public Law 107-347, Title V, 2002).

11. Privacy Act, 5 USC § 552a (b)(5) (1974).

12. Confidential Information Protection and Statistical Efficiency Act of 2002," (CIPSEA), (Public Law 107-347, Section 502, 2002).

13. "Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)," *Federal Register* 72, no. 115 (June 15, 2007): 33361; <https://www.gpo.gov/fdsys/granule/FR-2007-06-15/E7-11542/content-detail.html> (accessed August 10, 2017).

agencies routinely remove direct identifiers following legal guidance from general counsels about implementing the Privacy Act, but only some apply rigorous statistical disclosure limitation techniques before releasing de-identified data. According to the Federal Committee on Statistical Methodology's Working Paper 22, "agencies that need to protect data should move as far as possible toward the use of a small number of standardized disclosure limitation methods whose effectiveness has been demonstrated."¹⁴

Federal departments run the continuum of capacity and expertise to apply state-of-the-art methods for creating de-identified data that minimizes the risk of re-identification—that is, the chance that any individual in a publicly released dataset could be uniquely identified. Minimizing the risk of re-identification is important because of the potential that a bad actor could inappropriately figure out which record belongs to an individual and inflict harm as a result.

Statisticians, computer scientists, and the

Federal departments that collect and store confidential data are increasingly concerned about the risk of privacy harm that comes from data security breaches and unauthorized re-identification of individual identities in confidential data. The Federal government has an obligation to develop strategies to assess and minimize the impact of these risks when it uses government data for evidence building. The Commission's witnesses indicated that new technologies can be part of the solution.

Preventing Unauthorized Re-Identification

Early laws on privacy in the United States generally described a relatively short and finite list of direct identifiers that could easily be removed to create de-identified data for broad distribution. Accepted practice for many years meant simply removing direct identifiers such as names, addresses, and dates of birth. By the late 1970s, however, the Federal Committee on Statistical Methodology recognized that de-identification of data required far more than simply removing elements from a dataset and advocated the adoption of more formal

14. FCSM, *Report on Statistical Disclosure Limitation Methodology*: 6.

Statistical Disclosure Limitation Techniques from Working Paper 22

"To reduce the potential for disclosure... public use data files:

1. Include data from only a sample of the population,
2. Do not include obvious identifiers,
3. Limit geographic detail, and
4. Limit the number and detailed breakdown of categories within variables on the file.

Additional methods typically used to disguise high risk variables include:

1. Truncation of extreme codes for certain variables (top or bottom-coding),
2. Recoding into intervals or rounding,

3. Adding or multiplying by random numbers (noise),
4. Swapping or rank swapping (also called switching),
5. Selecting records at random, blanking out selected variables and imputing for them (also called blank and impute), and
6. Aggregating across small groups of respondents and replacing one individual's reported value with the average (also called blurring)."¹⁵

15. FCSM, *Report on Statistical Disclosure Limitation Methodology*: 24–25.

statistical disclosure limitation methods to protect individuals' privacy in public data releases.¹⁵

The potential danger of re-identification from indirect identifiers and contextual information in datasets stripped of direct identifiers became more widely recognized in the 1990s. As a result, PSAs increasingly began to rely on statistical disclosure limitation techniques to mask the data. The Federal Committee on Statistical Methodology's Working Paper 22, issued in 2005, stated: "Every agency or unit within an agency that releases statistical data should be capable of selecting and applying suitable disclosure limitation procedures to all the data it releases."¹⁶ By the early 2000s, the prevailing way to implement strong de-identification protocols was to use statistical disclosure limitation techniques on a release-by-release basis to reduce the risk of re-identification that could potentially cause privacy harm.

Today, a variety of statistical disclosure limitation techniques, methods, and technologies are used to reduce the identifiability of a dataset. One relatively new technique, *k*-anonymity, protects confidentiality by adjusting the data to make sure that all combinations of potentially identifiable information reflect a minimum number of individuals (*k*).¹⁷ No existing statistical disclosure limitation method, however, is able to completely eliminate the risk of re-identification. As Simson Garfinkel, formerly of the National Institute of Standards and Technology, the Federal government's standard-setting body for information technology, explains:

"...all data exist on an identifiability spectrum. At one end...are data that are not related to individuals (for example, historical weather records) and therefore pose no privacy risk. At the other end...are data that are linked directly to specific individuals. Between these two endpoints are data that can be linked with

effort... In general, de-identification approaches are designed to push data to the left [towards not being related to individuals] while retaining some desired utility, lowering the risk of distributing de-identified data to a broader population or the general public."¹⁸

Current best practice statistical disclosure limitation methods may become less effective in the future because of the growing availability of public and private information about individuals and computing techniques to exploit them. Adding to currently available information, Federal departments are now directed to release as much data as possible to the public under the Open Data initiative. However, advances in computer science and technology have produced promising new enhanced statistical disclosure limitation techniques that have the potential to provide stronger confidentiality protections by minimizing the risk of re-identification.

Assessing the Risk of Re-Identification

The Privacy Act and CIPSEA do not currently *require* agencies to assess the risk of unauthorized re-identification for de-identified confidential data; however, the CIPSEA implementation guidance includes the requirement that "for CIPSEA protected information, the agency as well as any agent accessing the information shall ensure that any dissemination of information based on confidential information is done in a manner that preserves the confidentiality of the information."¹⁹ Today, each agency that releases de-identified confidential data has to decide for itself how it will assess the risk of re-identification and the resulting potential for harm to individuals from misuse of individual information.

One demonstrated approach for carrying out the responsibility to assess risk is to establish a disclosure review board. Most of the 13 PSAs and a few other Federal agencies have established formal disclosure

15. FCSM, "Report on Statistical Disclosure and Disclosure Avoidance Techniques," Statistical Policy Working Paper 2 (Washington, D.C.: FCSM, 1978); <https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/spwp2.pdf> (accessed August 10, 2017).

16. FCSM, *Report on Statistical Disclosure Limitation Methodology*: 2.

17. Latanya Sweeney, "K-Anonymity: A Model for Protecting Privacy," *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems* 10, no. 5 (2002): 557-570.

18. Simson L. Garfinkel, "De-Identification of Personal Information," National Institute of Standards and Technology (NIST) Internal Report 8053 (Gaithersburg, MD: NIST, October 2015): 5; <https://dx.doi.org/10.6028/NIST.IR.8053> (accessed August 10, 2017).

19. "Implementation Guidance for Title V of the E-Government Act, CIPSEA."

review boards that assess the risk of re-identification in the agencies' public data releases.²⁰ Having a disclosure review board allows a department to assess the underlying risk of re-identification prior to public data release and assess whether the statistical disclosure limitation methods applied to the data have effectively mitigated the risk. All PSAs conduct disclosure review and typically document the results in "safe to release" memos or disclosure review reports for public data releases.

The establishment of a centralized disclosure review board within a department has added benefits, in that each data release can be considered within the context of all other planned data releases across the department (see the box "Establishing a Centralized Disclosure Review Board"). In addition, a centralized disclosure review board can leverage expertise from across the department, bringing together staff that have technical skills, such as mastery of statistical disclosure limitation methods, and specialized program knowledge about the contents of a particular dataset.

20. FCSM, *Report on Statistical Disclosure Limitation Methodology*.

Data collected and released by state and local governments, private companies, non-profit organizations, and researchers contribute to the cumulative amount of information available about individuals and businesses that could be used for re-identification. Researchers recently demonstrated how data on air and dust samples from 50 homes in two communities in California could be combined with data released under the Safe Harbor provisions of the Health Insurance Portability and Accountability Act (HIPAA) to "uniquely and correctly identify [in one community] 8 of 32 (25 percent) by name and 9 of 32 (28 percent) by address."²¹

Many programs have released de-identified public-use data files for decades without formally assessing risk. However, the broader context within which these releases are occurring has changed as the amount of information about individuals

21. Latanya Sweeney, Ji Su Yon, Laura Perovich, Katherine E Boronow, Phil Brown, and Julia Green Brody, "Re-identification Risks in HIPAA Safe Harbor Data: A Study of Data From One Environmental Health Study," *Technology Science* (August 28, 2017).

Establishing a Centralized Disclosure Review Board

In 2013, the U.S. Department of Education established a centralized disclosure review board (the ED-DRB) to review and approve privacy protections for public releases of the department's administrative data. Prior to the ED-DRB's creation, each of the department's principal offices was responsible for selecting and applying privacy-protection methods for their own data releases. This decentralized approach was problematic, as data releases by one office could affect the re-identification risk of related data released by other offices. Senior department leaders recognized that decentralization of risk assessments resulted in substantial variation across offices about the acceptable level of re-identification risk in public data releases.

The Education Department chartered the ED-DRB to be the central point for review and approval of privacy protections for the majority of the department's administrative data releases. The board has the technical expertise nec-

essary to ensure that the department releases as much usable data as possible while protecting privacy. The ED-DRB consists of staff with expertise in privacy law, policy priorities, disclosure risk assessment, and the application of statistical disclosure limitation techniques, as well as subject matter experts from data-releasing offices in the department. Education Department staff participate in the ED-DRB as one part of their regular duties. Since its creation, the ED-DRB has been a model for other agencies seeking to improve the consistency and efficacy of the disclosure review process for administrative data releases.¹

1. Federal CIO Council Innovation Committee, *The Data Disclosure Decision: Use Case Study* (Washington, D.C.: Government Printing Office, 2015); https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/1151/downloads/2015/03/The-Data-Disclosure-Decision-Department-of-Education-Case-Study_Mar-2015.pdf (accessed August 10, 2017).

that is publicly available has grown and the technology that can permit unauthorized re-identification has improved. Within the Federal government alone, the Open Data initiative made over 150,000 datasets accessible through a single website, including many administrative datasets never before released to the public.²² While releasing these data can generate tremendous value, enabling entrepreneurs to produce better products and departments to understand their work better, it is important that the risks that may accompany data releases also be assessed.²³ The Commission learned that a robust approach to risk assessment, one that considers the availability of external data in deciding what is safe to release, is paramount to the success of all future evidence building.

Agencies that link electronic records for internal government administrative purposes related to financial benefits are required today to assess the risk of the linkage and develop procedures for safeguarding the data.²⁴ The system used for the assessment is cumbersome and could be simplified, and statistical activities are exempt. However, the transparent accountability portion of the process is instructive for evidence building. Agencies must issue a *Federal Register* notice and post the written agreement between the relevant agencies on each agency's website. For example, the U.S. Department of Education and the U.S. Department of Veteran's Affairs have a publicly available agreement to link administrative data to determine eligibility for veteran's educational benefits.

Challenge #2: Federal departments must adapt to new threats to information security and privacy and take advantage of emerging technologies that better protect data.

The risk of re-identification is not the only source of potential privacy harm from personal information collected by the Federal government. Staff with otherwise legitimate reasons to access personal information may misuse such access, for example, by browsing a celebrity's or acquaintance's

record. In the absence of strict protocols for using confidential data, an inexperienced researcher may attempt to take a screen shot of findings that have not been through the required disclosure review or use other means to remove information not approved for release. Federal departments are required to have internal controls to prevent and address poor data handling practices and to impose penalties for any procedural violations. Federal guidance, however, "does not prescribe how management designs, implements, and operates an internal control system."²⁵

To keep data safe from misuse, the Federal government must keep data secure and ensure appropriate training, controls, and accountability. Data security is the development and implementation of "management, operational, and technical aspects of protecting the confidentiality, integrity, and availability of federal information and information systems."²⁶ Many Federal agencies collect identifiable information in accordance with their statutory missions and store those data on government servers. A breach of a data system, or a misuse of data, resulting in the release of identifiable information could lead to an increased risk of individuals being re-identified in a dataset released to the public under CIPSEA or the Privacy Act.

Data security in Federal departments today is generally the responsibility of Chief Information Officers who implement standards promulgated by OMB in consultation with the National Institute of Standards and Technology. A recent report from the Commission on Enhancing National Cybersecurity sought to develop "actionable recommendations for securing and growing the digital economy by strengthening cybersecurity in the public and private sectors."²⁷ Cyberattacks on confidential data increase the risk of privacy harm and may erode public trust.

22. DATA.gov, "Federal Agency Participation;" <https://www.data.gov/metrics> (accessed August 10, 2017).

23. DATA.gov, "Federal Agency Participation."

24. Computer Matching and Privacy Protection Act of 1988 (Public Law 100-503); <https://www.gpo.gov/fdsys/pkg/STATUTE-102/pdf/STATUTE-102-Pg2507.pdf> (accessed August 10, 2017).

25. Government Accountability Office (GAO), *Standards for Internal Control in the Federal Government* (Washington, D.C.: GAO, September 2014): 8; <http://www.gao.gov/assets/670/665712.pdf>.

26. Marianne Swanson, Joan Hash, and Pauline Bowen, *Guide for Developing Security Plans for Federal Information Systems*, National Institute of Standards and Technology (NIST) Special Publication 800-18 (Gaithersburg, MD: NIST, February 2006): 13; <http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-18r1.pdf> (accessed August 10, 2017).

27. Commission on Enhancing National Cybersecurity, "Report on Securing and Growing the Digital Economy" (December 1, 2016): i; <https://www.nist.gov/sites/default/files/documents/2016/12/02/cybersecurity-commission-report-final-post.pdf> (accessed August 10, 2017).

Well-publicized cybersecurity incidents in the past have affected Federal systems and resulted in the unintentional release of millions of confidential records, raising concerns about vulnerabilities in the level of data security of Federal systems. Nearly all Federal agencies must assess and secure information systems under the Federal Information Security Management Act (FISMA).²⁸ PSAs and the three other statistical units authorized to protect data under CIPSEA have an explicit requirement to “ensure the physical security and information systems security where data protected under CIPSEA are accessed and stored.”²⁹ The National Institute of Standards and Technology is currently collaborating with Federal departments and the private sector to establish new engineering standards and privacy risk assessments intended to minimize security risk in Federal information technology systems.³⁰

PSAs respond to data security mandates under CIPSEA and individual authorizing statutes by requiring computer systems that store confidential data be kept completely separate from other systems. In practice, this means that PSAs store confidential data used for evidence building on servers that are not connected to other data systems within their departments. Data can be extracted from these servers only by authorized individuals for exclusively statistical purposes such as to create the PSA’s own production statistics or to combine data across authorized agencies for approved purposes.

The Federal evidence-building community must keep pace with evolving technology to maintain the security of data and the trust of the American public. While there are multiple approaches to protecting data against misuse, the Commission believes that technology offers part of the solution for enhancing privacy while allowing more use of data to satisfy the growing demand for evidence to support policymaking.

New and evolving technologies have the potential to improve data security, minimize risk

of re-identification, guard against other forms of privacy harm, and allow for increased use of data to support evidence-based policymaking. In particular, privacy-protective technologies and protocols can expand access to data while potentially improving, rather than adversely affecting, privacy. Because technology is always evolving, the Commission does not endorse the adoption of any one specific approach. Rather, the Commission acknowledges the critical need for leadership and partnerships that will advance and accelerate the adoption of technologies that improve authorized access and protect privacy.

As the Commission learned, new approaches hold the potential for enhancing privacy protection. Examples include differential privacy, Secure Multiparty Computation, and synthetic data (see the box “Emerging Approaches That Enhance Privacy Protections”).

“Differential privacy is a mathematical guarantee that an individual data contributor will not be affected, adversely or otherwise, by allowing her data to be used in any study or analysis, no matter what other studies, datasets, or information sources, are—or will become—available.”

– Cynthia Dwork, Microsoft Research (former), Commission Meeting, Washington, D.C., September 9, 2016

The Census Bureau has led efforts among PSAs to use enhanced statistical disclosure limitation tools now made possible by cutting-edge technologies. For example, the Census Bureau is beginning to use synthetic data to create data tools that answer interesting policy questions while using differential privacy to measure the risk of re-identification and keep it below a defined threshold (see the box “OnTheMap: Differential Privacy in Practice at the Census Bureau”).

Adopting new privacy-protective technologies will pose implementation challenges even once the methods have been sufficiently demonstrated in practice. One substantial challenge, for example, is that many departments currently use

28. E-Government Act of 2002 (Public Law 107-347, Title III).

29. “Implementation Guidance for Title V of the E-Government Act, CIPSEA.”

30. Sean Brooks, Michael Garcia, Naomi Lefkowitz, Suzanne Lightman, and Ellen Nadeau, *An Introduction to Privacy Engineering and Risk Management in Federal Systems*, NIST Internal Report 8062 (Gaithersburg, MD: NIST, January 2017); <http://nvlpubs.nist.gov/nistpubs/ir/2017/NIST.IR.8062.pdf> (accessed August 10, 2017).

Emerging Approaches That Enhance Privacy Protections

Differential privacy is a formal privacy concept that applies to analyses performed over collections of sensitive personal information.¹ It provides a quantifiable measure of the excess privacy risk any individual may incur due to their data being included in an analysis, as compared with their data not being included. Because the application of differential privacy includes the mathematical computation of the risk to privacy loss in any given analysis, parameters are established that set bounds on what amount of privacy loss is acceptable. These parameters are then applied to control the tradeoff between protection of individual privacy on the one hand and accuracy of the performed analyses on the other.

Research over the last decade has provided solid theoretical grounds for differential privacy. Algorithms realizing optimal privacy-accuracy tradeoffs are emerging for a variety of statistical and machine learning tasks. In particular, these algorithms can enable privacy-preserving computations using data that more traditional techniques would need to redact. Notably, differential privacy is currently the only framework with mathematically proven bounds on the accumulated privacy risk resulting from the composition of several analyses. The first real-world uses now exist (by the U.S. Census Bureau and Internet companies such as Apple, Google, and Uber).

Secure Multiparty Computation, or SMC, is a method for conducting privacy-preserving data analysis.² The concept of SMC has been in existence for over 40 years. However, recent advances in algorithms and computational power have made it feasible to achieve

SMC in the real world.³ The technique allows one party to jointly compute a function with another party, while never revealing one another's private data. SMC and its variations rely on mathematical principles based on cryptography to compute answers over databases from different locations and controlled by different organizations without either side needing to see individual records of the other. In short, SMC enables distributive computing that allows data owners to maintain their data "in their own silos" and yet compute results on the combined data.

Synthetic Data: Verification Model: A synthetic dataset mimics an original dataset, by replacing actual values from the original data with altered values that still retain many (but not necessarily all) important statistical properties of the real dataset.⁴ Synthetic datasets include more detailed data with lower disclosure risk because the original data are changed. But for the same reason, analysts may be uncertain about the precision of the results. Using a verification server is one way to address this uncertainty. The verification server runs the analysis on the original data and then tells the analyst how different the results are between the two datasets. If the analyst feels that the result using synthetic data is "good enough," then they may choose to work solely with the synthetic data; but if the analyst feels that the result using synthetic data is not "good enough," they may choose to pursue access to the original dataset.

1. National Academies of Sciences, Engineering, and Medicine. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy* (Washington, D.C.: National Academies Press, 2017).

2. Yehuda Lindell and Benny Pinkas, "Secure Multiparty Computation for Privacy-Preserving Data Mining," in *Encyclopedia of Data Warehousing and Mining*.

3. Dan Bogdanov, Liina Kamm, Baldur Kubo, Reimo Rebane, Ville Sokk, and Riivo Talviste, "Students and Taxes: a Privacy-Preserving Study Using Secure Computation," *Proceedings on Privacy Enhancing Technologies* no. 3 (July 2016); <https://doi:10.1515/popets-2016-0019> (accessed August 10, 2017).

4. National Institutes of Health Office of Behavioral and Social Science Research, "Synthetic Data: Protecting Data Privacy in an Era of Big Data" (October 15, 2015); <https://obssr.od.nih.gov/synthetic-data-protecting-data-privacy-in-an-era-of-big-data> (accessed August 10, 2017).

techniques that assess and seek to minimize the risk of re-identification of a particular data release as a stand-alone matter. In contrast, full implementation of differential privacy requires determining the cumulative re-identification risk of data releases over the entire life-span of the data. The widespread adoption of techniques such as differential privacy would require departments to consider and manage the myriad planned and unplanned future uses of data.

Even with these limitations, the promise of “provable privacy” methods such as differential privacy is that the level of risk of re-identification can be quantified, thus allowing any dataset to be used for evidence building while minimizing privacy risk.

Another implementation challenge is ensuring that enhanced statistical disclosure methods do not change the data in ways that increase the difficulty of reproducing research results. The Commission’s guiding principle on humility points to the need for multiple researchers to replicate results to verify the credibility of evidence. Further, professional journals sometimes require researchers to demonstrate that analyses can be reproduced prior to publication.³¹ In adopting new

technologies and techniques to provide secure access to data, the government must consider the need for reproducibility.

Challenge #3: Many Federal departments lack senior leadership focused on data stewardship, and PSAs have uneven independence and authority to protect confidential data used for evidence building.

The Commission found that evidence-based policymaking requires leaders who recognize the value of generating more and better evidence as a mechanism for improving government services and holding government accountable. Federal leadership that acknowledges and values the importance of evidence building must support the use of Federal data assets for statistical purposes. Without sustained leadership and support within Federal departments for data stewardship, opportunities to improve and appropriately leverage data resources for evidence building will be missed.

Meaningful collaboration within departments is essential if administrative data are to be harnessed as an integral part of evidence building. For example, the Commission to Eliminate Child Abuse and Neglect Fatalities concluded in 2016 that the inability to coordinate data collection and analysis activities across child protection programs posed a real risk to children’s lives. One of the

31. For example, the *American Journal of Political Science’s* “Replication and Verification Policy,” <https://ajps.org/ajps-replication-policy> (accessed August 10, 2017). Additionally, the *American Economic Review* only publishes papers if the data used in the analysis are documented and available for replication; <https://www.aeaweb.org/journals/policies/data-availability-policy> (accessed August 10, 2017).

OnTheMap: Differential Privacy in Practice at the Census Bureau

OnTheMap is an online mapping and reporting application that shows areas where people work and where workers live. OnTheMap was developed through a partnership between the Census Bureau and its Local Employment Dynamics partner states. OnTheMap is an example of differential privacy in practice. In “Privacy: Theory meets Practice on the Map,” researchers describe how technology was used to create OnTheMap. “The algorithm used to anonymize the data for the... mapping application is known as the synthetic data generation,

which is becoming popular in the statistical disclosure limitation community.”¹ Without the use of privacy-preserving technologies, OnTheMap would not be able to provide the detail it does.

1. Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber, “Privacy: Theory Meets Practice on the Map” (paper presented at the IEEE 24th International Conference on Data Engineering, Cancun, Mexico, April 2008); <http://www.cse.psu.edu/~duk17/papers/PrivacyOnTheMap.pdf> (accessed August 10, 2017).

Commission's recommendations was to "improve and support data collection about child abuse and neglect fatalities of...children, and integrate the data into national databases for analysis, research, and the development of effective prevention strategies."³² Coordinating evidence-building activities and supporting use of administrative data across departments leads to useable inventories, data resources, and analytic capabilities for responding to specific policy and research questions.

All Federal departments have a Chief Information Officer (CIO) and a Senior Agency Official for Privacy (sometimes called a Chief Privacy Officer). According to OMB Circular A-130, CIOs play an important role in the management of Federal information resources. In addition to advising department heads on "the design, development, and implementation of information resources," CIOs help prioritize department expenditures on information resources.³³ The Senior Agency Official for Privacy has responsibility for "implementation of privacy protections; compliance with Federal laws, regulations, and policies relating to privacy; [and] management of privacy risks at the agency."³⁴ Other senior leaders with influence on department information resources include Chief Financial Officers and Chief Performance Officers. The Commission also recommends that departments identify a senior official to serve as Chief Evaluation Officer (Recommendation 5-1).

While each of these officials has a role to play in data stewardship, each also has an expansive portfolio of responsibilities that prevents their prioritizing data management and secure access to data resources for evidence-building activities. The Commission learned that none of these senior leaders is explicitly charged with managing the portfolio of department data resources in support of Federal evidence building. Because senior leaders with a focus on data stewardship for evidence building are not at the table when priorities are discussed and resources allocated, departments cannot take full advantage of Federal data to build

evidence about programs and policies.

The Commission found that PSAs are well positioned to become department leaders to facilitate the transformation and analysis of administrative data for evidence building. Currently, however, PSAs vary in readiness to take on greater responsibility for supporting the larger evidence-building community. Some PSAs have strong statutes that ensure their independence, place them high in department hierarchies, and authorize them to protect their confidential data. Other PSAs assert their need for independence and authority through policy and practice but may lack strong legal protections.

Lack of independence can be a challenge for some PSAs. The Federal Information Technology Acquisition Act (FITARA) of 2014 led department information technology leadership to consolidate Federal data centers and coordinate information technology acquisitions to promote consistency and efficiency.³⁵ These well-intentioned efforts could hamper the ability of the evidence-building community to limit access to confidential data. As Katherine (Kitty) Smith of the Council of Professional Associations on Federal Statistics (COPAFS) noted in comments to OMB about the roles and responsibilities of Federal statistical agencies, "information technology systems that are out of the direct control of the statistical agency can: result in delays in the retrieval of and dissemination of statistical data; impose restrictions on the accompaniment of transparent explanations of methodology with the data; violate the integrity of statistical information; and, very importantly, endanger the statistical agencies' ability to follow through on their pledges of confidentiality and non-disclosure."³⁶

In public comment, witnesses expressed concern about independence for PSAs and recommended that PSAs should have sufficient legal authority implemented or restored. Margaret Levenstein of ICPSR, a consortium of more than 750 academic institutions and research organizations, said that "undermining [the public's trust in Federal statistical agencies] undermines statistical measurement

32. Commission to Eliminate Child Abuse and Neglect Fatalities, *Within Our Reach: A National Strategy To Eliminate Child Abuse and Neglect Fatalities* (Washington, D.C.: Government Printing Office, 2016): 14.

33. U.S. Office of Management and Budget (OMB), "Managing Information as a Strategic Resource." 14.

34. OMB, "Managing Information as a Strategic Resource." 36.

35. Federal Information Technology Acquisition Reform Act, 40 USC § 11319(b)(1)(A).

36. Katherine R. Smith (COPAFS) to Katherine K. Wallman, Comment on OMB Statistical Policy Directive, June 18, 2014; <http://www.copafs.org/UserFiles/file/handouts/CommentonOMBStatisticalDirective%20.pdf> (accessed August 10, 2017).

as well as the effectiveness of the programs upon which statistics are based.”³⁷ Clyde Tucker of the American Statistical Association cited a 2015 letter from 20 former PSA heads that said “the independence of a federal statistical agency is a critical element in an agency producing objective and credible statistical data....such autonomy should include control over an agency’s planning, budget, press releases, and information technology.”³⁸

Several commenters asked the Commission to reinforce the importance of PSAs and the laws and policies that undergird them. The public’s willingness to trust PSAs to protect confidential information must be maintained and strengthened.

OMB Statistical Policy Directive #1 encourages Federal departments to recognize the need for integrity and objectivity in Federal statistics (see box, “Four Responsibilities of Statistical Agencies from OMB Statistical Policy Directive #1”). Policy Directives do not have the same force as statutes, however, so legislation to codify certain key

37. Margaret Levenstein, ICPSR, submission to the Commission’s Request for Comments.

38. Clyde Tucker, American Statistical Association, Commission Public Hearing, Washington, D.C., October 21, 2016.

elements of Statistical Policy Directive #1 would help PSAs further protect confidential data used in evidence building.

Recommendations

Achieving the Commission’s vision for enhancing privacy while also making better use of data for evidence building will require some significant changes. Departments will need to conduct risk assessments for public releases of de-identified confidential data and to adopt cutting-edge technology for data security, integrity, and confidentiality. This vision also calls for data stewardship coordinated among senior leaders within Federal departments and the independence of PSAs to protect confidential data. The Commission concluded that the Federal government must *improve data quality, integrity, and security by updating laws to require risk assessments for data releases, embracing cutting-edge technologies, and prioritizing data stewardship and the public trust.*

The recommendations below support the Commission’s vision by allowing secure access to confidential data in a privacy-protective manner, enabling the accountable production of evidence.

Four Responsibilities of Statistical Agencies from OMB Statistical Policy Directive #1¹

- 1. Relevance** requires “communication across and within departments with planning information collection and dissemination activities” and that statistical agencies collect program and policy-relevant information from “administrative records collected and maintained by the agency, or other government agencies.”
- 2. Credibility** requires that agencies “apply sound statistical methods to ensure statistical products are accurate.”
- 3. Objectivity** requires that agencies “produce data that are impartial, clear, and complete and are readily perceived as such by the public” and further requires that “statistical units must function in an environment that is clearly separate and autonomous from the other administrative, regulatory, law enforcement, or policymaking activities within their respective Departments.”
- 4. The trust** of information providers requires agencies to ensure “confidentiality and exclusive statistical use” of their data.

1. OMB, “Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units,” *Federal Register* 79 (December 2, 2014): 71610; <https://www.gpo.gov/fdsys/pkg/FR-2014-12-02/pdf/2014-28326.pdf> (accessed August 10, 2017).

When taken together with the privacy-protective data access and use recommendations proposed in Chapters 2 and 4, these recommendations seek to safely harness the potential of data already collected by government to produce more and better evidence in the future.

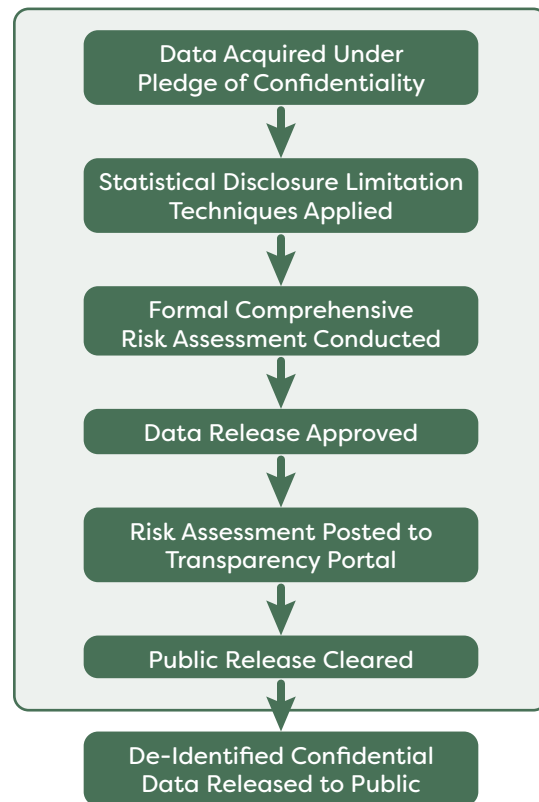
REC. 3-1: The Congress and the President should amend the Privacy Act and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) to require Federal departments to conduct a comprehensive risk assessment on de-identified confidential data intended for public release. De-identified confidential data subject to the Privacy Act and CIPSEA should be made available only 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.

The Commission concludes that existing laws such as the Privacy Act and CIPSEA should require a comprehensive risk assessment and disclosure review board approval before public release of de-identified confidential data. With a small change in how government approaches the management of confidential data, public data releases can be made much more secure. Many PSAs have the expertise to do this now and are a valuable resource for program agencies that want to make their data releases more secure.

Establishing a requirement in law for risk assessment provides a basis for the entire Executive Branch to more seriously and deliberately approach risk management for publicly releasing data. OMB should collaborate with Federal departments and other stakeholders to establish implementation guidance that requires a risk assessment and approval by a disclosure review board of any de-identified data released under the Privacy Act or CIPSEA. The Congress should consider whether other statutes governing privacy for Federal confidential data should be subject to a similar requirement.

Under the approach envisioned by the Commission, Federal departments will conduct comprehensive risk assessments that include an analysis

Figure 9. Improving Confidentiality Protections During Public Release of De-Identified Data




of all known sources of risk (see Figure 9). Departments should consider the risk of unauthorized re-identification in the context of other publicly available data. As a starting point, departments should prioritize risk assessments for de-identified confidential datasets planned for public release and expand risk assessment requirements over time to include all public data releases.

Departments may choose to go through existing disclosure review boards, such as those in the PSAs, for approval to release de-identified confidential data under the Privacy Act or CIPSEA. Other departments may choose to establish a separate disclosure review board for this purpose or establish an entirely new disclosure review board if no such centralized function currently exists. In addition, the National Secure Data Service (NSDS) will establish a disclosure review board to provide review services, extending the capacity of Federal departments.

Departments should be required to publicly post risk assessments as well as the general steps taken to mitigate risk from public releases of de-identified confidential data covered by the Privacy Act and CIPSEA. Risk assessments should be posted to the NSDS transparency portal and include a mechanism for the public to report information about additional sources of risk (see Recommendation 4-3). Either OMB or a department could request the initiation of a new risk assessment when new information comes to light.

Transparency about risk assessments means that the public will know what the government considered in its assessment of risk when it assesses risk and why the government now believes the data can be released as de-identified. Public risk assessments also make it easy for agencies to learn from one another about the implementation of enhanced statistical disclosure limitation techniques.


 **REC. 3-2: The President should direct Federal departments, in coordination with the National Secure Data Service, to adopt state-of-the-art database, cryptography, privacy-preserving, and privacy-enhancing technologies for confidential data used for evidence building.**

To implement this recommendation, the Commission proposes that the NSDS should have the initial and generative role in this enhancement of privacy protections. The NSDS should lead the implementation and application of new privacy-related technologies in the Federal evidence-building community. OMB should consult with the National Institute of Standards and Technology on the most appropriate privacy-protective data exchange standard for use across Federal departments and Federally funded projects in support of evidence-building, consistent with or useful for operational or administrative data exchange standards. Finally, this recommendation requires the active participation of all Federal departments, as they will need to periodically review their compliance with the National Institute of Standards and Technology standards for the security of confidential data in transmission and storage.

The proposal to establish the NSDS in statute should include authority for the NSDS to be the

Federal government's lead on the implementation and application of cutting-edge data management technologies. OMB should review publicly available risk assessments to determine if Federal departments are adopting the most advanced disclosure limitation methods, techniques, and technologies. OMB should set standards that require all agencies to adopt new methods as they are shown to be feasible in the Federal context. The National Institute of Standards and Technology should continue its efforts to ensure security protection for data exchange, transmission, and storage.

The evidence-building community should be at the forefront of applying cutting-edge technologies to improve data security and privacy. A National Academies of Sciences, Engineering, and Medicine report recently recommended that "Federal statistical agencies should adopt modern database, cryptography, privacy-preserving, and privacy-enhancing technologies."³⁹ The Commission's Recommendation 3-2 echoes and extends the report's recommendation. Standards governing data security for all Federal agencies should be adapted and strengthened as new technologies emerge that provide greater privacy protections. Leaders within Federal departments should acknowledge and adhere to special data stewardship requirements for confidential data used for evidence building.

 **REC. 3-3: The President should direct Federal departments to assign a senior official the responsibility for coordinating access to and stewardship of the department's data resources for evidence building in collaboration with senior department information technology, privacy, and other leaders. A Principal Statistical Agency head, or other appropriately qualified senior official, should serve this function.**

Each department should assign to a senior official the duties of Senior Agency Official for Data Policy and the responsibility for coordinating the

39. National Academies of Sciences, Engineering, and Medicine. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy* (Washington, D.C.: National Academies Press, 2017): 93.

department's statistical needs and information policy in support of evidence building.

The person assigned this responsibility would need to have the technical and substantive expertise to be effective in promoting secure evidence building within the department. Because of the imperative to protect confidential data, the designated senior leader would need extensive knowledge of data confidentiality laws, policies, and practices including experience with applying statistical disclosure limitation techniques in public releases of data. Other necessary technical skills would include knowledge of strategies to collect and manage data used for evidence building, data analysis methods appropriate to data used for evidence building, and emerging strategies to provide secure access to confidential data in a manner useful for analysis. In addition, this senior official would need a connection to and understanding of the broader evidence-building community, including a deep appreciation for the importance of keeping Federal data authorized for evidence building on the exclusively statistical purposes side of the functional wall of separation described in Chapter 1.

The Commission believes that this function is essential as the Federal government increasingly uses its existing data resources to generate valid evidence. Today, leadership for evidence building is fragmented in or absent from many departments. The senior data policy official would be charged to anticipate that all of the department's data may be assets for evidence building. They also would be charged to ensure that legal, policy, and technical data stewardship requirements are met when using data for evidence building.

Senior Agency Officials for Data Policy would have both inward- and outward-facing roles. Within their own departments, the senior data policy officials would: (1) collaborate with other senior department leaders in information technology, finance, privacy, and evaluation to prioritize the use of data for evidence building; (2) ensure that program and statistical agencies determine the sensitivity of data resources and establish appropriate access controls (see Recommendation 2-8); and (3) promote the use of enhanced statistical disclosure limitations techniques to protect confidentiality and expand access for evidence building. In their external coordination role, senior data policy officials would

serve as the department's liaison to the NSDS and advise the NSDS, other Federal departments, and OMB about the best way to use existing department data beneficially for exclusively statistical purposes.

In many cases, departments would not have to create a new position or hire a new person to fill the role. Because of the expertise required, in most cases the head of a PSA should fill the position. In other cases, or in departments that do not have a PSA, a senior official with expertise in evaluation or policy-relevant research may be qualified to provide such leadership. Regardless, collaborative leadership around the coordination and management of data resources within a department is important for the efficient production of evidence.



REC. 3-4: The Congress and the President should enact legislation to codify relevant portions of Office of Management and Budget Statistical Policy Directive #1 to protect public trust by ensuring that data acquired under a pledge of confidentiality are kept confidential and used exclusively for statistical purposes.

The Congress and the President should enact legislation to codify those elements of Statistical Policy Directive #1 related to supporting “the quality and objectivity of Federal statistical information.”⁴⁰ OMB issued the directive in 2014 to affirm the long-acknowledged, fundamental responsibilities of Federal statistical agencies and recognized statistical units in the design, collection, processing, editing, compilation, storage, analysis, release, and dissemination of statistical data and information. The Commission recommends that the requirements embodied in the directive be embedded in law. The Congress and the President should provide each PSA with the authority and institutional support necessary for ensuring confidentiality and maintaining the integrity and objectivity of Federal statistics.

Recommendation 3-4 further strengthens statistical institutions by placing in law the expectations of independence outlined in OMB's

40. OMB, “Statistical Policy Directive No. 1.”

Statistical Policy Directive #1. The placement of these responsibilities explicitly in statute strengthens the government's commitment to the policies, makes the principles more difficult to change, and increases the likelihood that violations of the responsibilities would be reported under whistleblower protections for Federal

employees established in Title 5 of the U.S. Code.⁴¹ Placing the directive in law also provides PSAs a stronger basis from which to defend their need for independent information technology resources and tailored procedures to secure the confidentiality of Federal data used for evidence building. ■






41. Prohibited Personnel Practices, 5 USC § 2302.

4

Modernizing America's Data Infrastructure for Accountability and Privacy

Empower government to develop state-of-the-art capacity to securely combine existing data and provide secure data access for exclusively statistical purposes in a privacy-protective and transparent way.

Recommendations

-  **4-1:** The National Secure Data Service (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.
-  **4-2:** The NSDS should establish a Steering Committee that includes representatives of the public, Federal departments, state agencies, and academia.
-  **4-3:** To ensure exemplary transparency and accountability for the Federal government's use of data for evidence building, the 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.
-  **4-4:** The NSDS should have specific administrative and implementation flexibilities including the ability to leverage public-private partnerships and to collect and retain user fees.
-  **4-5:** The Office of Management and Budget should increase efforts to make information available on existing Federal datasets including data inventories, metadata, and data documentation in a searchable format.

The Federal evidence-building community needs a modern institutional infrastructure that facilitates evidence building and incorporates strong privacy protections. Today, some elements of that infrastructure exist within individual silos, focused on particular agency topics or tasks. Achieving the Commission's vision of routine, secure evidence production for the public good requires a new infrastructure aiding Federal departments and outside researchers by improving the amount and quality of evidence while enhancing privacy.

While the risk assessment described in Chapter 3 (Recommendation 3-1) was designed for public releases of de-identified confidential data, the recommendations in this chapter seek to promote evidence generation possible only by securely harnessing the value of detailed data that cannot be made publicly available, and only for exclusively statistical purposes.

This chapter expands on Recommendation 2-1 to establish the National Secure Data Service (NSDS) by describing several of the implementation steps necessary to make the envisioned service possible. The NSDS described in Chapter 2 is intended to be:

- A service for qualified researchers seeking approval to securely analyze confidential government data for exclusively statistical purposes. The NSDS will apply data minimization techniques to protect combined data, require researchers to complete training, and subject researchers to stringent monitoring for adherence to privacy-protective protocols.
- A center for unparalleled transparency about government uses of confidential data for exclusively statistical purposes. The NSDS will make information about data sensitivity, risk assessments for public release of de-identified confidential data, and individual evidence-building projects available in ways that exceed previous transparency and accountability for evidence building in the United States.
- A developer and implementer of state-of-the-art methods to safely combine confidential data. The NSDS will temporarily and securely combine data and create analysis files with the minimum amount of data needed for approved projects with exclusively statistical purposes.

- A developer and implementer of new and secure approaches to data access and analysis. The NSDS will be designated as a Principal Statistical Agency (PSA) with an exclusively statistical mission, with the authority to protect data under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) and the resources and capacity to implement emerging privacy-protective technologies.

The Congress and the President specifically charged the Commission to consider whether the country needed a data “clearinghouse.”¹ While the term was not defined in the Commission's statute, “clearinghouse” evokes images of past proposals to consolidate large amounts of confidential data into a central location. For example, in the 1960s, the Kaysen Committee proposed creating a national Data Bank to address the decentralized nature of the Federal Statistical System and the Federal government's data infrastructure. The Kaysen Committee determined that decentralization prohibited effective use of socioeconomic data by researchers both inside and outside government, and recommended the creation of a large warehouse of data about the American public.² While the Commission on Evidence-Based Policymaking came to a similar conclusion about the value of data, it concluded that setting up a data warehouse of the sort envisioned by the Kaysen Committee would create an attractive target for misuse of private data. The Commission on Evidence-Based Policymaking opposes the creation of any such “clearinghouse” and proposes an entirely different solution for addressing the challenges of data access. (See the box “How is the National Secure Data Service Different from the Data Bank?”.)

Technology and privacy-protective approaches have advanced greatly over the past 50 years since the Data Bank proposal was rejected, and they will continue to evolve. New approaches to secure access to data for evidence building feature stronger privacy protections than at any point in history. The Commission's approach to implementing

1. Evidence-Based Policymaking Commission Act of 2016 (Public Law 114–140, Section 4(b), March 30, 2016).

2. Rebecca S. Kraus, “Statistical Déjà Vu: The National Data Center Proposal of 1965 and Its Descendants” (paper presented at the Joint Statistical Meetings, Miami Beach, FL, August 1, 2011); <https://www.census.gov/history/pdf/kraus-natdatacenter.pdf> (accessed August 10, 2017).

How is the National Secure Data Service Different from the Data Bank?

While both the 1960s Data Bank and the National Secure Data Service (NSDS) proposals seek better use of government data for statistical activities, that is where their similarity ends. The Data Bank was designed as a massive warehouse of constantly cumulating government-held data about the American public that could have been used for diverse purposes. The NSDS instead is a service that brings together as little data as possible for as little time as possible for exclusively statistical purposes. The NSDS is not intended to store data and users of the service would not be permitted to conduct analyses for non-statistical purposes under the Commission's proposed legal and operational framework.

While the Data Bank proposal would have stored individual-level data, the proposed NSDS is designed to temporarily link data and

remove personal identifiers before the data can be analyzed for exclusively statistical purposes. The NSDS takes privacy to a new level by applying state-of-the-art protections and ensuring that, at every step, the American public is notified about the uses of data to hold government accountable on a project-by-project basis. The NSDS proposal requires a privacy audit for public transparency and accountability.

In short, the Commission does not recommend a single "clearinghouse" that collects massive stores of data to be warehoused indefinitely, as was envisioned in the 1960s Data Bank proposal. In fact, the Commission concluded that a single national "clearinghouse" with all types of data is simply not necessary, nor does it represent best practices in privacy protection.

Key Similarities and Differences Between the Data Bank and the NSDS

Proposal Feature	1960s Data Bank	2017 NSDS
Limits data uses to exclusively statistical purposes		✓
Allows public participation in setting policies and priorities		✓
Intended to improve researcher access to confidential data	✓	✓
Restricts access to data with personally identifiable information (PII) and keeps access temporary		✓
Approves projects individually		✓
Includes an appeals process for project approval		✓
Includes a publicly available inventory of approved projects		✓
Implements state-of-the-art privacy protection techniques		✓
Makes transparent all data access for statistical purposes		✓
Includes an auditing function		✓

Recommendation 2-1, elaborated upon in this chapter, aims to take full advantage of the opportunity created by these new approaches.

However, the Federal government currently lacks the complete infrastructure needed to routinely, efficiently, and securely generate evidence about government policies and programs using confidential data. The infrastructure that does exist is not designed to meet government-wide needs; does not have the capacity to meet expanded demand from inside and outside government; is uneven in its application of best practice technologies and techniques for secure data access, linkage, and analysis; and does not always prioritize public transparency and accountability.

This chapter includes recommendations that elaborate on aspects of NSDS's implementation and build on a vision of the NSDS as a source of important shared services to improve government-wide data infrastructure for evidence building. The Commission envisions that NSDS will help government increase the appropriate use of data for exclusively statistical purposes, while keeping confidential data safe and secure. NSDS will have the capacity and capability to support secure analysis of single or combined confidential, restricted-use datasets while minimizing the risks of privacy harm from evidence-building activities. Given the breadth of the existing gaps in infrastructure, the Commission recommends the establishment of NSDS to address multiple capacity needs simultaneously to usher in a new era of evidence building in the United States.

Findings

The Commission learned from witnesses and submitted statements about gaps and unevenness in infrastructure supporting evidence building. The findings presented in this chapter highlight four challenges to routine and secure evidence generation to improve government policies and programs. These findings relate specifically to the use of confidential government data for exclusively statistical purposes. First, the Federal government currently has pockets of expertise in privacy-protective data linkage, access, and analysis, but this capacity does not meet government-wide needs for evidence. Second, the identification and adoption of new technologies that can increase data access while enhancing privacy is slow and uneven across Federal agencies. Third, Federal agen-

cies abide by legal requirements for transparency about data systems, but have made only limited attempts at public transparency about individual projects using confidential data for exclusively statistical purposes. Finally, the lack of a comprehensive inventory of data useful for evidence building, lack of complete and accurate technical documentation for many government datasets, and inconsistent definitions of key concepts across datasets inhibits routine evidence building.

Challenge #1: Existing Federal government privacy-protective data linkage, access, and analysis capabilities do not meet government-wide needs for evidence.

The Commission received testimony from Federal agencies using state-of-the-art methods to provide privacy-protective access to securely linked confidential data for exclusively statistical purposes. Within the Federal government, pockets of excellence can be found for most of the statistical activities needed to routinely generate evidence.

Some Federal departments are using state-of-the-art methods for secure data linkage, access, and analysis in response to their specific statutory missions. These departments have expert staff with experience implementing privacy-protective techniques in a Federal context. This expertise constitutes a valuable resource for the Federal evidence-building community, but it is currently concentrated in different agency silos and focused on specific agency purposes. There are pockets of excellence; there is no center of excellence capable of serving government-wide needs for privacy-protective data linkage, access, and analysis for evidence building.

A great deal of expertise relevant to Federal evidence building resides in the U.S. Department of Commerce. The U.S. Census Bureau has built the capacity to securely link demographic data for exclusively statistical purposes in support of Census Bureau operations and demographic and socioeconomic research (see the box "Center for Administrative Records Research and Applications"). The Census Bureau's Center for Economic Studies (CES) links and archives business and economic data used for research and analysis. The Census Bureau also coordinates a network of Federal Statistical Research Data Centers (FSRDCs) to provide secure access to confidential data (see the

Center for Administrative Records Research and Applications

The Center for Administrative Records Research and Applications (CARRA) at the U.S. Census Bureau strategically uses administrative data from Federal, state, and third party providers for exclusively statistical activities in support of the Census Bureau's operations and demographic and socioeconomic research. Linked data are also used by researchers at the Census Bureau's Center for Economic Studies and through the Federal Statistical Research Data Centers.

As of 2017, CARRA's portfolio includes 12 pilot research and evaluation projects spanning housing, health, welfare, education, and labor. The initiative also includes six pilot projects in partnership with Chapin Hall at the University of Chicago that combine state and local administrative data with information accessible to the Census Bureau. Existing CARRA partnerships with experts in academia support CARRA's ongoing efforts to improve capabilities for securely combining confidential data.

CARRA benefits from a number of legal authorities applicable to the Census Bureau. Foremost, the Census Bureau has broad authority under Title 13 (Census statute) to acquire administrative records. Title 13 authorizes the Census Bureau to use records previously collected by other Federal agencies and state, tribal, or local governments, as well as private

organizations; and directs Census to seek out this information instead of conducting direct inquiries. Further, the Census Bureau has a specific exemption under the Privacy Act that allows Federal agencies to disclose records to the Census Bureau without prior written consent of the individual if it is for the purposes of planning or carrying out a census or survey or related activity.

Census is authorized to access Federal Tax Information for individuals and businesses under Title 26 (IRS statute). CIPSEA authorizes the Census Bureau, the U.S. Bureau of Labor Statistics, and the U.S. Bureau of Economic Analysis to share business data for exclusively statistical purposes. While these authorities permit the Census Bureau to access many different datasets, in practice data sharing agreements are negotiated separately and governed by applicable laws and policies.

Under Title 13, CARRA can establish agreements to access data held by Federal agencies, states, and some other jurisdictions for exclusively statistical purposes. The list below highlights a portion of the data CARRA currently acquires to support the Census Bureau's mission.¹ State administrative records not only vary on the number of states participating, but also in the years of data available.

Federal administrative data

- Medicaid Statistical Information System
- Office of Personnel Management Personnel System
- Public and Indian Housing Information Center
- Selective Service System
- Small Business Administration Loan Guarantee Program
- Social Security Old Age, Survivors, and Disability Insurance
- Tax Information
- United States Postal Service Change of Address

State administrative data

- Child Care Development Fund (1 state)
- Low Income Home Energy Assistance Program (1 state)
- Special Supplemental Nutrition Program for Women, Infants, and Children (6 states)
- Supplemental Nutrition Assistance Program (15 states)
- Temporary Assistance for Needy Families (9 states)
- Unemployment Insurance quarterly wages (48 states)

1. U.S. Census Bureau, "Administrative Data Inventory" as of July 18, 2017; <https://www2.census.gov/about/linkage/data-file-inventory.pdf> (accessed August 10, 2017).

box “Federal Examples of Secure Access to Confidential Data”). In addition, the U.S. Bureau of Economic Analysis, a PSA that produces a broad range of economic data; the U.S. National Oceanic and Atmospheric Administration, an agency with many scientific and technical data resources; and the National Institute of Standards and Technology are all part of Commerce.

PSAs, and some program agencies, provide secure access to their confidential data for statistical purposes authorized by law (see examples in the box “Federal Examples of Secure Access to Confidential Data”). The CEP Survey of Federal Offices found that 10 out of 13 PSAs, 5 out of 8 responding evaluation offices, and 39 percent (19 out of

49) of other responding offices have provided other Federal agencies access to data they collect.³ All PSAs have external researchers who access data, while the same is true of 75 percent of responding evaluation offices and 31 percent of other responding offices.

The Commission considered whether this existing infrastructure could, with little new additional authority or resources, meet government-wide needs for evidence building. Much of the government’s staff expertise on securely linking confidential data resides at the Census Bureau,

3. Two of the 10 responding evaluation offices and 18 of the 67 responding other offices reported that they do not collect data.

Federal Examples of Secure Access to Confidential Data

Administrative Data Research Facility: The Administrative Data Research Facility is a pilot project that enables secure access to analytical tools, data storage and discovery services, and general computing resources for users, including Federal, state, and local government analysts and academic researchers. The Census Bureau and academic partners developed the project as part of the collaborative Training Program in Applied Data Analytics sponsored by the University of Chicago, New York University, and the University of Maryland.¹ It is currently operating as a pilot with users accessing the Facility as part of the training program. The Facility operates as a cloud-based computing environment, with Federal security approvals, which currently hosts selected confidential data from the U.S. Department of Housing and Urban Development and the Census Bureau, as well as state, city, and county agencies, and an array of public use data.

Data Licensing System: The National Center for Education Statistics (NCES), one of the Principal Statistical Agencies (PSAs), is part of the Institute of Education Sciences (IES) in the U.S. Department of Education. It is the primary Federal entity for collecting and analyzing

data related to education in the United States and other nations. NCES makes detailed data available to external researchers through the IES restricted-use data licensing system. NCES loans researchers restricted-use data through a license (contract) among IES, the user, and the user’s organization (e.g., university, research institution, or company). Licensed users agree to the terms of the IES confidentiality law; IES performs regular audits to confirm compliance with required procedures. Because of this audit requirement, IES will only loan restricted-use data to qualified organizations in the 50 States and Washington, D.C. Researchers seeking a data license must complete an online formal request and research proposal.²

Federal Statistical Research Data Centers: The Federal Statistical Research Data Centers (FSRDCs) are partnerships between Federal statistical agencies and leading research institutions. They are secure facilities housed in partner institutions and managed by the U.S. Census Bureau to provide secure access to a range of Federal restricted-use microdata for statistical purposes only. All FSRDC

—continues

1. “Training Program in Applied Data Analytics;” <http://www.applieddataanalytics.org> (accessed August 10, 2017).

2. U.S. Department of Education, National Center for Education Statistics, “Learning Procedures: Frequently Asked Questions;” http://nces.ed.gov/statprog/instruct_licensing_faq.asp (accessed August 10, 2017).

and particularly in the Center for Administrative Records Research and Applications (CARRA). As currently constituted, however, CARRA cannot meet all of the project linkage needs of other Federal departments and qualified external researchers. It has neither the authority in law nor the capacity to do so. The FSRDCs and PSAs provide secure access to confidential data; none of them has the authority or capacity to provide expanded access for increased evidence generation government wide. Some individual Federal projects at the Census Bureau and elsewhere are piloting new technologies for secure data analysis (see the box “OnTheMap: Differential Privacy in Practice

at the Census Bureau” in Chapter 3), but no one agency today serves as a model of best-practice application of cutting edge privacy-protecting technologies.

The Commission’s public meetings, hearings, Request for Comments, and Survey of Federal Offices yielded numerous examples of unintentional barriers to evidence building erected due to unclear legal frameworks as described in Chapter 2. Other barriers arise because of uneven capacity across Federal departments. Some Federal departments lack PSAs; some evaluation and policy research offices lack technical capacity; and some program agencies fail to prioritize evidence building.

Federal Examples of Secure Access to Confidential Data— continued

researchers must obtain Census Bureau Special Sworn Status. FSRDC data enclaves meet the physical and environmental protections required by Federal statistical agencies for use of restricted-access microdata, including Federal Tax Information. FSRDC researchers can collaborate with other FSRDC researchers across the United States through the secure computing environment.

Joint Statistical Research Program: The Statistics of Income Division at the Internal Revenue Service collects and processes tax data and provides access to tax microdata to those users authorized under statute—including certain government agencies—for tax policy analysis, research, and statistical purposes. In addition, the Statistics of Income Division makes microdata available to researchers through its Joint Statistical Research Program. The goals of the program are to (1) provide new insights and advance the understanding of the ways existing tax policies affect individuals, businesses, and the economy, (2) suggest tax policy solutions to advance the common good, and (3) provide new understanding of taxpayer behavior that could impact the administration of the U.S. tax system. Applications for the program are accepted approximately every two

years, resources permitting.³

Virtual Research Data Center: The Centers for Medicare and Medicaid Services (CMS) is responsible for administering the Medicare, Medicaid, and Children’s Health Insurance programs. To support these functions, CMS collects, generates, and stores financial, health care, and other sensitive information. With appropriate safeguards, CMS provides access to microdata for statistical and research uses to Federal agencies, Federal grantees, and other approved researchers on encrypted data files provided on external media or through a Virtual Research Data Center. Through the virtual data center, approved researchers can access and perform their own analysis and manipulation of CMS data virtually from their own workstations. The virtual data center helps CMS meet additional demand from data users while also ensuring data privacy and security. It also provides researchers with more timely access to data in a more cost-effective manner.⁴

3. Internal Revenue Service, Statistics of Income, “Joint Statistical Research Program Call for Proposals,” <https://www.irs.gov/pub/irs-soi/16jsrpapplication.pdf> (accessed August 10, 2017).

4. Centers for Medicare and Medicaid Services (CMS), “Virtual Research Data Center (VRDC),” <http://www.resdac.org/cms-data/request/cms-virtual-research-data-center> (accessed August 10, 2017).

Some states and local jurisdictions told the Commission that basic technical services could substantially improve their capacity to support evidence-based policymaking about their own programs and policies as well as those implemented in partnership with the Federal government.⁴ At a Commission public hearing in February of 2017, Joy Bonaguro, Chief Data Officer for the City and County of San Francisco, testified that “matching and linking data across systems can be a challenge

and burden for local entities due to the expertise required and may [disincentivize] research partnerships. A central service could lend consistent technical expertise to the task of data matching.”⁵ Such technical services could be useful to Federal departments facing capacity challenges as well.

CARRA’s approach to using statistical and administrative data for statistical activities over the past several years demonstrates the possibility and potential of making better use of these data for

4. Communication with National Association of State Workforce Agencies Labor Market Information Committee, Chicago, IL, April 20, 2017.

5. Joy Bonaguro, City and County of San Francisco, Commission Public Hearing, Washington, D.C., February 9, 2017.

Examples of Secure Access to Confidential Data in Other Countries

Administrative Data Research Network (United Kingdom): The Administrative Data Research Network is a national partnership in the United Kingdom that provides trained social and economic researchers access to de-identified, linked administrative data in a secure environment. The Network was founded in 2013 by the Economic and Social Research Council—an independent, publicly funded entity—to allow more use of government data in social and economic research, with safety measures in place to protect sensitive data. A goal of the Network’s work is to develop high standards for sharing, linking, and matching records securely and consistently. The Network continues to research data linkage methodologies and evaluate new technologies. It also recognizes that public trust through public engagement is the most important element for success. The Network’s processes were designed with a great deal of public consultation. The Network’s public engagement strategy is designed to promote sustainability through transparency and public understanding.¹

1. United Kingdom, Administrative Data Research Network, “Communication and Public Engagement,” <https://adrn.ac.uk/about/network/ads/communication-and-public-engagement> (accessed August 10, 2017).

Research Data and Service Centre (Germany): The German Bundesbank compiles data for monetary, financial, and external statistics and produces a comprehensive set of indicators under its legal mandate to develop a broad spectrum of user-oriented economic data. The Bundesbank established the Research Data and Service Centre to facilitate data linkage and researcher access. The Bundesbank’s program provides visiting researchers with on-site access to microdata on banks, securities, investment funds, enterprises, and households. Researchers seeking to access Bundesbank data must submit an application that includes a description of the project, its hypotheses and methods, and a justification for the datasets required. Staff at the Bundesbank’s program work with researchers in advance of their application to help them refine their research questions and identify the data sources they will need.²

Research Services Division (Denmark): Statistics Denmark, under the Ministry for Economic and Interior Affairs, is responsible for creating official statistics on Danish society. The

—continues

2. Germany, Bundesbank, “Research Data and Service Centre,” <http://www.bundesbank.de/Navigation/EN/Bundesbank/Research/RDSC/rdsc.html> (accessed August 10, 2017).

government-wide evidence building. That CARRA safely and confidentially links data for exclusively statistical purposes under the law without privacy breaches demonstrates this can be done well when appropriate authorities and privacy protections are in place.

Challenge #2: Identification and adoption of emerging technologies and techniques for privacy-protective data access, linkage, and analysis are too slow and uneven in government to effectively support evidence building.

As described in Challenge #1, few Federal agencies are providing privacy-protective access to confidential data for exclusively statistical purposes, and even fewer are securely linking confidential data for evidence building. All PSAs and a few other Federal agencies use rigorous statistical disclo-

sure limitation techniques to protect confidential data and provide highly restricted access to confidential data for exclusively statistical purposes (see Chapter 3).

Some of today's most innovative and informative evidence-building projects rely on highly restricted access for researchers to confidential data. Examples include (1) research on intergenerational mobility by race and ethnicity using tax and survey data,⁶ (2) assessments of the long-term effects of antipoverty policies that connect welfare data with earnings records,⁷ and (3) development

6. Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," National Bureau of Economic Research (NBER) Working Paper 19843 (January 2014); <http://www.nber.org/papers/w19843> (accessed August 10, 2017).

7. For example, "Experimental Estimates of the Long-Run Impacts of Welfare Reform on Participants and Their Children" and ongoing work by Hilary Hoynes, Jordan Matsudaira, Pauline Leung, and Zhuan Pei.

Examples of Secure Access to Confidential Data in Other Countries—continued

Research Services Division of Statistics Denmark provides researcher access to linked data with identifiers removed for statistical purposes only. Since 1968, every person in Denmark has received a Personal Identification Number. Each person is also associated with a dwelling identification and possibly an enterprise or business identification. These three keys can be used to link multiple sources of administrative and survey data for individuals and family units over time. Researchers affiliated with an approved Danish research environment may submit an application to the Research Services Division to access Statistics Denmark data. Data may be accessed only through authorized research environments, which include universities, sector research institutes, ministries, and non-profit foundations.³

Secure Data Access Centre (France): France's Secure Data Access Centre (the Centre

d'Accès Sécurisé aux Données, or CASD) was founded in 2009. One distinct feature of this system is a secure device, the SD-Box, that allows researchers to remotely access and analyze confidential data from CASD's servers. The SD-Box is installed at approved researchers' own institutions. With their access card and fingerprint confirmation, researchers access everything they need at their local workstation—including data, statistical software, storage, and computing power—when, in fact, all of these are on the secure servers inside CASD. CASD remotely authenticates, monitors, and configures the SD-Box and all communications between it and CASD servers take place through encrypted tunnels. Because there is centralized security management, CASD can push upgrades and patches to each device. SD-Boxes can also be remotely disabled. All SD-Boxes are identical and easily replaced in the event of an equipment failure.⁴

3. Denmark, Statistics Denmark Research Services Division, "Data for Research;" <http://www.dst.dk/en/TilSalg/Forskningsservice> (accessed August 10, 2017).

4. France, "Secure Data Access Centre;" <https://casd.eu/en> (accessed August 10, 2017).

of improved measures of poverty and the income distribution that connect survey, tax, and program data.⁸ Such projects can only be conducted if linkages among datasets can be made in an environment in which privacy and confidentiality are protected.

*“The collection of accurate and complete information... is predicated on trust... Without that trust, data collection will be undermined before it even begins...”*⁹

– Thomas L. Mesenbourg, Former U.S. Census Bureau Deputy Director

Today, PSAs and a few other Federal agencies that provide secure access to confidential data for statistical purposes have a number of protections in place as required by statute and policy. Federal staff and outside researchers accessing confidential data must complete annual training in information technology security and data stewardship. Agencies restrict access to person identifiers such as name and date of birth to a very few technical staff under tightly controlled conditions. Other staff and outside researchers almost never receive approval to access personal identifiers. Any statistical research output that uses confidential data undergoes statistical disclosure review before public release to ensure confidentiality is not compromised. During the Commission’s December meeting, government officials, including from the Census Bureau and the Statistics of Income Division in the Internal Revenue Service, indicated that to their knowledge no breach of their confidential data used for evidence building had ever occurred.¹⁰

8. Bruce D. Meyer and Nikolas Mittag, “Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net,” NBER Working Paper 21676 (October 2015); <http://www.nber.org/papers/w21676> (accessed August 10, 2017).

9. Thomas L. Mesenbourg, “How Statutory and Regulatory Changes May Create New and Unintended Challenges for Statistical Agencies” (paper presented at the International Society of Scientometrics and Informetrics Conference, Durban, South Africa, July 2011).

10. Barry Johnson, Internal Revenue Service, and Ron Jarmin, Census Bureau, Commission Meeting, Washington, D.C., December 12, 2016.

However, the absence of past breaches does not mean future breaches are impossible, particularly as technologies evolve or bad actors focus on different types of data.

As described in Chapters 2 and 3, a legal infrastructure must exist to enable secure data access for evidence building and to take advantage of new privacy-protective technologies. Among those agencies providing secure access to confidential data, some have the expertise and resources to identify and implement feasible new technologies that hold the promise of allowing more evidence generation in more secure ways (see the box “Emerging Approaches That Enhance Privacy Protections” in Chapter 3). While these privacy-protective approaches are in use today, they are not available at the scale necessary to enhance evidence production across government. The Census Bureau’s capacity, for example, is limited to a handful of projects, with new cutting-edge technologies limited to even fewer applications.

The Commission heard from several other countries about models for using new technologies for secure data access (see the box “Examples of Secure Access to Confidential Data in Other Countries”).

Today, the Federal evidence-building community lacks a coordinated focus on harnessing new technologies to provide more secure data access for exclusively statistical purposes. The government’s capacity to take advantage of emerging privacy-protective technologies is spread across a number of different agencies. The focus of individual agencies is rightfully on their own missions and not oriented towards government-wide evidence building. No one agency has sufficient resources and staff expertise to develop and implement state-of-the-art technologies or to push the evidence-building community to continually innovate to address potential threats and risks to the American public’s confidential data. Achieving the Commission’s vision means addressing these challenges.

Challenge #3: The existing infrastructure for accessing, linking, and analyzing confidential data for evidence building does not always prioritize state-of-the-art transparency and oversight.

From collecting data to generating evidence about government programs and policies, members of the Federal evidence-building community should

tell the American people how their information is being used and for what benefit to society. Transparency means giving the public information about how the government is using their data to improve its effectiveness and efficiency. The Federal evidence-building community also should establish oversight and accountability mechanisms to verify that it is using best practice procedures to protect personal information.

For data to be useful for evidence building, they must be complete and accurate. The cooperation of the American people is central to the quality and accuracy of survey and administrative data used for evidence building. Telling the American people about how the government is using their data to improve accountability and serve the public is ethical and conforms to the Fair Information Practice Principles (FIPPs) on transparency and accountability and auditing described in Chapter 3. Only with such transparency and accountability can the public trust needed to collect and use personal data for evidence building be preserved.

Further, for evidence to make a difference in government effectiveness and efficiency, policymakers, program administrators, and the public must trust the results. Without trust, the results will lack credibility. Without credibility, the results are more likely to be ignored, wasting the time and money used to generate them. Even worse, lack of credibility could jeopardize the goal of generating more and better evidence to improve government and monitor the state of the nation. Trust is critical to evidence-based policymaking.

The tradition and culture of data stewardship in some parts of the Federal evidence-building community today should inform the efforts of all Federal departments that provide secure access to confidential data. The PSAs responsible for much of the current data infrastructure operate within a framework of laws, regulations, policies, and practices that emphasize the importance of respecting individuals' privacy and protecting the confidentiality of their information. These historical protections must be augmented with additional measures to increase transparency and accountability about how data are accessed and used for exclusively statistical purposes.

Transparency

The Congress and the Executive Branch require public notice when the government collects con-

fidential information from the American people. To maintain the public's trust, the evidence-building community must take reasonable steps to ensure the public is aware of how data are collected, safeguarded, shared, and used, and what benefits those uses may provide for society. These are long recognized aspects of transparency for evidence building. However, the Commission finds that the Federal evidence-building community today does not consistently nor effectively use best practice transparency processes for public accountability.

During the Commission's September 2016 meeting, then-Senior Advisor for Privacy at the U.S. Office of Management and Budget (OMB) and Chair of the Federal Privacy Council, Marc Groman, emphasized the importance of transparency:

[As] the United States Government we have an obligation to be transparent to our citizens about what we collect and how we use it. We make representations about use and we have to honor those representations...[by explaining] how the data collect[ed] will be used. And that is a core principle of the Privacy Act or other laws and of privacy generally.¹¹

The Privacy Act requires a System of Records Notice to promote public transparency. Government agencies issue public notices when a "system of records" containing personally identifiable information is being created or expanded. These public notices disclose what information is collected and for what purpose. However, the notices are made available in a way that can be difficult for the public to access; members of the public may not know where to find them or understand the technical information they contain. While Groman discounted the public's interest in the notices given this challenge, he added: "I can tell you that advocacy groups do read it; [inspectors general] do read it. And it provides actually a benefit, because it's about transparency."

An important limitation is that these notices are for systems of records, not for individual projects. Today, no notice is required for individual statistical or evidence-building projects. Some other countries provide public notice about individual

11. Marc Groman, U.S. Office of Management and Budget, Commission Meeting, Washington, D.C., September 9, 2016.

projects that link confidential data (see the box “Case Study: Transparency at Statistics Canada”).

Federal departments also are required to gain approval for Information Collection Requests under the Paperwork Reduction Act for new data collections from 10 or more individuals. These requests must go through two public comment periods before final approval. Some statistical agencies go even further when conducting new primary data collections and engage stakeholder groups prior to data collection to incorporate feedback

and improve the quality of the data collection.

In sum, members of the public currently have multiple opportunities to learn about and provide feedback on government’s information collection activities before they are initiated. Such transparency is valuable but applies to systems of records, rather than to individual linkage and analysis projects, and public notices are sometimes difficult to find, access, and understand.

When collecting administrative information from program participants or beneficiaries,

Case Study: Transparency at Statistics Canada

Statistics Canada’s policies on using administrative data address how that agency protects privacy and provides transparency on the use and linkage of its administrative data about Canadians. The policy’s objective is to “maximize the efficiency and effectiveness of the use of administrative data in Statistics Canada’s statistical programs. Statistics Canada achieves this objective through a corporate strategy intended to influence, access, use, and manage the administrative data supplied to Statistics Canada, and to maintain public trust while doing so.”¹ Statistics Canada intends the policy to communicate its statistical use of administrative data to stakeholders and the public in a proactive, coherent, and transparent manner.

Under Statistics Canada’s *Principles for the Statistical Uses of Administrative Data*, statistical uses are kept functionally separate from administrative uses through strong legal, policy and organizational safeguards. In particular, the confidentiality of all identifiable administrative data obtained by Statistics Canada for statistical purposes is protected by law, and the data are only accessed within Statistics Canada based on demonstrated needs. Statistics Canada carefully considers the impact on privacy in using identifiable administrative data for a purpose that was not envisaged at the time of original data collection, particularly when

administrative data are combined with other data. It is transparent in its use of administrative data, and communicates benefits of using existing data as well as measures taken to protect confidentiality.

Statistics Canada employs individuals with dedicated roles to promote transparency. The Director General of the Communication Division is responsible for developing and implementing a plan that shows Statistics Canada’s commitment to the transparent use of administrative data. Statistical and legal analysts in the Information Management Division are responsible for maintaining and managing an inventory of all approved data linkages. Dedicated websites provide another layer of transparency.²

As part of its governance over microdata linkages, Statistics Canada has pre-approved specific types of linkages. The linkages involved are those in which privacy risks and situations of potential conflict of interest are low and where procedures to mitigate risk to confidentiality and privacy are in place. All other microdata linkages must undergo a prescribed review and approval process, which involves the submission of documented proposals to senior management. When such linkages include personal information, a summary of the linkage is posted on Statistics Canada’s website.

1. Statistics Canada, “Policy on the Use of Administrative Data Obtained under the Statistics Act,” http://www.statcan.gc.ca/eng/about/policy/admin_data (accessed August 10, 2017).

2. Statistics Canada, “Linkage depository,” <http://www.statcan.gc.ca/eng/sdle/status>. See also “Studies, research papers and technical papers,” <http://www5.statcan.gc.ca/researchers-chercheurs/> (both accessed August 10, 2017).

government agencies sometimes indicate that data will only be used for specific purposes, such as “only for processing a benefit application.” Such promises generally mean the Federal government cannot use those data for any other purpose, including statistical analysis for evidence building, without re-contacting the individual. However, when an agency uses its own administrative data for exclusively statistical purposes, it generally does not need to re-contact individuals. An exception under the Privacy Act allows agencies, in some cases, “to engage in an intra-agency disclosure of data for statistical purposes...”¹²

Federal agencies today operate with limited transparency about individual projects seeking to link and analyze confidential data for exclusively statistical purposes. Even agencies at the forefront of Federal evidence building, like CARRA and CES, do not currently have easily accessible information about approved projects, qualified researchers, and datasets linked for evidence building. Further, because CARRA and CES are located within the Census Bureau, they do not have their own independent advisory boards. A recent Gallup survey reported that nearly two-thirds of respondents were more likely to support combining

data when an independent group helped to ensure the presence of privacy protections.¹³

CARRA and CES operate with only the public transparency and oversight that are currently standard practices for Federal agencies, such as through the annual appropriations process when the Congress reviews programs and activities. The public does not have meaningful opportunities to provide guidance or feedback about the individual projects that are underway. Some states have begun to develop procedures for public transparency about evidence building that could be a model (see box “Case Study: Mississippi’s LifeTracks System”).

Privacy Auditing

Transparency in Federal evidence building requires verifying that the Federal government implements clear, strict rules about data access and follows through on its promises to the American people to keep data for evidence building safe and secure.

One way to promote public trust through accountability is to require privacy audits, which assess available protections within an organization against legal requirements and privacy best practices. Privacy audits document the status of an organization’s risk associated with information misuse and make recommendations to limit this risk. A privacy audit also reviews an organization’s

12. U.S. Office of Management and Budget (OMB), Guidance for Providing and Using Administrative Data for Statistical Purposes” (memorandum M-14-06, Washington, D.C.: OMB, Executive Office of the President, February 14, 2014); <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf> (accessed August 10, 2017).

13. CEP staff correspondence with Census Bureau staff in April 2017.

Case Study: Mississippi’s LifeTracks System

Mississippi LifeTracks is a state longitudinal data system that, in addition to programmatic uses, allows for the analysis of administrative data from multiple state agencies to assess education and workforce outcomes in the state.¹ LifeTracks is funded through a combination of U.S. National Center for Education Statistics grants and annual state appropriations. One challenge Mississippi faced in creating the LifeTracks system was conveying the value of the system. To accomplish this, Mississippi

devoted a portion of their website to public accountability. The website lists approved projects and completed projects and cites state-level statistics based on their results. This demonstration of the value of the program and the useful information it provides has been suggested as the key to the system’s sustainability. LifeTracks aims to hold state government accountable for implementing effective programs and policies. At the same time, the public can hold government accountable for the use of data in the LifeTracks System because of the program’s transparency efforts.

1. Domenico Parisi, Mississippi State University, Commission Meeting, Washington, D.C., January 13, 2017.

provided and used for statistical activities, information about their meaning and intent can be lost. Making appropriate use of administrative data for evidence building requires the development of data documentation, which can be time intensive and require a deep knowledge of the administrative data collected.¹⁸ OMB's report on a comprehensive government inventory described the minimal amount of metadata required under the Open Data initiative; however, not all agencies have met this minimum requirement.¹⁹ Commission witnesses requested much more robust metadata for evidence building.

Complete and accurate metadata can help researchers determine whether access to certain data even makes sense for their specific project and allow them to assess data quality prior to seeking access to microdata. In fact, incomplete or inaccurate metadata may result in researchers seeking access to confidential data that will not answer their research question. Having complete and accurate metadata is vital to improving secure data linkage and access. In fact, standardized metadata is an essential element of employing privacy-preserving technologies such as Secure Multiparty Computation (see Chapter 3).

In addition to emphasizing the importance of the Federal government maintaining a list of government-held datasets and providing basic information about mechanisms for accessing the datasets, many stakeholders called for a list of variables in the datasets as the most basic form of metadata. Stakeholders described metadata as a critical component of assessing data quality and fitness for use; the Commission heard numerous recommendations to create robust administrative dataset documentation by standardizing metadata.²⁰

The CEP Survey of Federal Offices found that all 12 PSAs that answered the survey item on documentation already create metadata that in-

cludes variable definitions, units of measurement, and response ranges. Also, responses to the CEP survey suggested that overall, offices that used a higher percentage of their budget for statistical, evaluation, research, and analysis purposes more frequently created robust metadata. But among Federal departments administering programs and collecting data that can be useful for evidence building, the completeness and accuracy of metadata vary widely.

Harmonization of Definitions

Different Federal agencies have different definitions of common terms and key concepts in their data systems. Sometimes, the same agency has multiple different definitions of the same concept. These differences evolved because of how the terms are defined in law or because of specific program needs.

This variation in definitions for common terms and key concepts in government presents challenges in analyzing data across agencies and sometimes across jurisdictions. At times, these variations are acknowledged and planned for or addressed by programs. In 2008, the Special Supplemental Nutrition Program for Women, Infants, and Children at the U.S. Department of Agriculture updated guidance to states about what data elements should be included in data systems, and clarified definitions needed for managing case-loads.²¹ For some concepts with government-wide applicability, OMB issues standards to create common definitions, such as the Standard Occupational Classification or a recently revisited standard on race and ethnicity.²²

Multiple witnesses in the Commission's public hearings and other commenters highlighted the need to harmonize definitions and the meaning

18. Aileen Rothbard, "Quality Issues in the Use of Administrative Data Records," in *Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government*, edited by John Fantuzzo and Dennis P. Culhane (New York, NY: Palgrave MacMillan, 2016): 72–103.

19. U.S. Office of Management and Budget, "Comprehensive Data Inventory."

20. Data quality pertains to data completeness, data accuracy, and consistent definitions of data elements. The fitness for use is determined based on the intended use, and may be limited by data quality or other challenges associated with overly aggregated data forms, use restrictions, or ethical considerations.

21. U.S. Department of Agriculture (USDA), Food and Nutrition Service, *Functional Requirements Document for a Model WIC Information System* (Washington, D.C.: USDA, 2008); <https://www.fns.usda.gov/sites/default/files/apd/FReD%20v2.0%20Final.pdf> (accessed August 10, 2017).

22. U.S. Office of Management and Budget, "Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity," *Federal Register* 81 (September 30, 2016); <https://www.federalregister.gov/documents/2016/09/30/2016-23672/standards-for-maintaining-collecting-and-presenting-federal-data-on-race-and-ethnicity>; OMB, "2010 Standard Occupational Classification (SOC)—OMB's Final Decisions," *Federal Register* 74 (January 21, 2009); <https://www.bls.gov/soc/soc2010final.pdf> (both accessed August 10, 2017).

Interagency Employer Data Matching Workgroup

An interagency workgroup developed strategies for matching and reusing data on U.S. employers across Federal datasets for statistical purposes. In 2016, the group published a white paper that summarized key issues for matching employer data and identified best practices that could be implemented across agencies.¹

Matching employer data can be complex because of the different levels of data (establishment versus firm or enterprise), the par-

ent-child relationships within firms, and the dynamic nature of such relationships. The workgroup found two primary challenges in matching employer data collected by the Federal government: (1) the lack of a common universal identifier and (2) poor quality of the underlying identifying data.

The interagency group identified the most important data elements for matching employer data across Federal agencies. The most common elements that are most essential to matching include Employer Identification Number, Legal Entity Identifier, legal business name, trade name, physical and mailing address, and country code.

1. Employer Data Matching Workgroup. “Employer Data Matching Workgroup White Paper.” (white paper for the Commission on Evidence-Based Policymaking, Washington, D.C. 2017).

of outcome metrics across Federal laws. Several witnesses expressed a desire for the Federal government to establish common definitions for performance metrics across the Higher Education Opportunity Act, the Workforce Innovation and Opportunity Act, and the Perkins Career and Technical Education Act to reduce burden and improve opportunities for integrating data across programs with similar goals and target populations. Another witness identified inconsistencies in the application of geographic definitions across the Federal government and suggested they be further harmonized.²³ As described in the box “Interagency Employer Data Matching Workgroup,” some Federal agency personnel recognize the importance of consistent definitions for key matching variables to combine economic data on firms and businesses for statistical analysis.

Recommendations


The Commission’s strategy seeks to extract the maximum value of data for public good while min-

imizing the possible risk of harm to individuals or organizations. The Commission concluded that the Congress and the President must *empower government to develop state-of-the-art capacity to securely combine existing data and provide secure data access for exclusively statistical purposes in a privacy-protective and transparent way.*

In the Commission’s vision, linkages of confidential data and access to those data for solely statistical purposes will occur in a highly protective privacy, legal, and technological environment, described in Chapters 2 and 3. Project-by-project reviews will help the evidence-building community work toward the ideal of permitting access to data for projects with great promise for improving the public good, while limiting the risk of re-identification and other privacy harms. At the same time, a more transparent process for statistical analysis of confidential government data by researchers and government agencies will reduce unintended inefficiencies that can inadvertently hamper the generation of evidence.

The Commission offers five recommendations regarding the implementation of the NSDS as a service to support privacy-protective, transparent, and accountable evidence generation.

23. Timothy Slaper, Indiana Business Research Center, Indiana University, Commission Public Hearing, Chicago, January 5, 2017.

 **REC. 4-1: The National Secure Data Service (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.**

The Census Bureau already has established, in pilot form, some of the capacities that are central to the Commission's vision for the NSDS. CARRA and the CES at the Census Bureau each work to combine data that can be made available for researchers, primarily through the FSRDCs. The Commission concluded that the Census Bureau's current operations, and other capacity throughout the Federal government, form a starting point for implementing Recommendations 2-1 and 2-2. Building on the Census Bureau's history and culture of data stewardship and public trust, NSDS can be a cornerstone of future Federal evidence building with a focus on secure access to Federal data, public transparency, and a consistent set of legal and technological privacy protections.

The existing Census Bureau infrastructure in CARRA and CES supports, on a small scale, internal research and operational uses of administrative data and, in partnership with the FSRDC network, some externally motivated research projects using administrative data. The current infrastructure should be expanded to support increased evidence building. Rather than build an entirely new and potentially duplicative set of functions, a cost-effective approach to implementing the NSDS is to adapt and scale the existing elements of CARRA and CES and build upon additional expertise throughout the Federal government and its partners in the evidence-building community.

To maximize efficiency, the President and the Congress should reallocate designated staff and resources from the Census Bureau to form the core infrastructure for NSDS, supplementing those resources as needed with expertise from other large data collecting and disseminating components of the government. To achieve an "optimal arrangement" for expanding evidence building, the Com-

mission also recommends that NSDS should have (1) an institutional placement that gives it necessary independence, (2) designation as a PSA with authority to protect data under CIPSEA, (3) separate funding, and (4) an improved governance structure offering enhanced transparency.

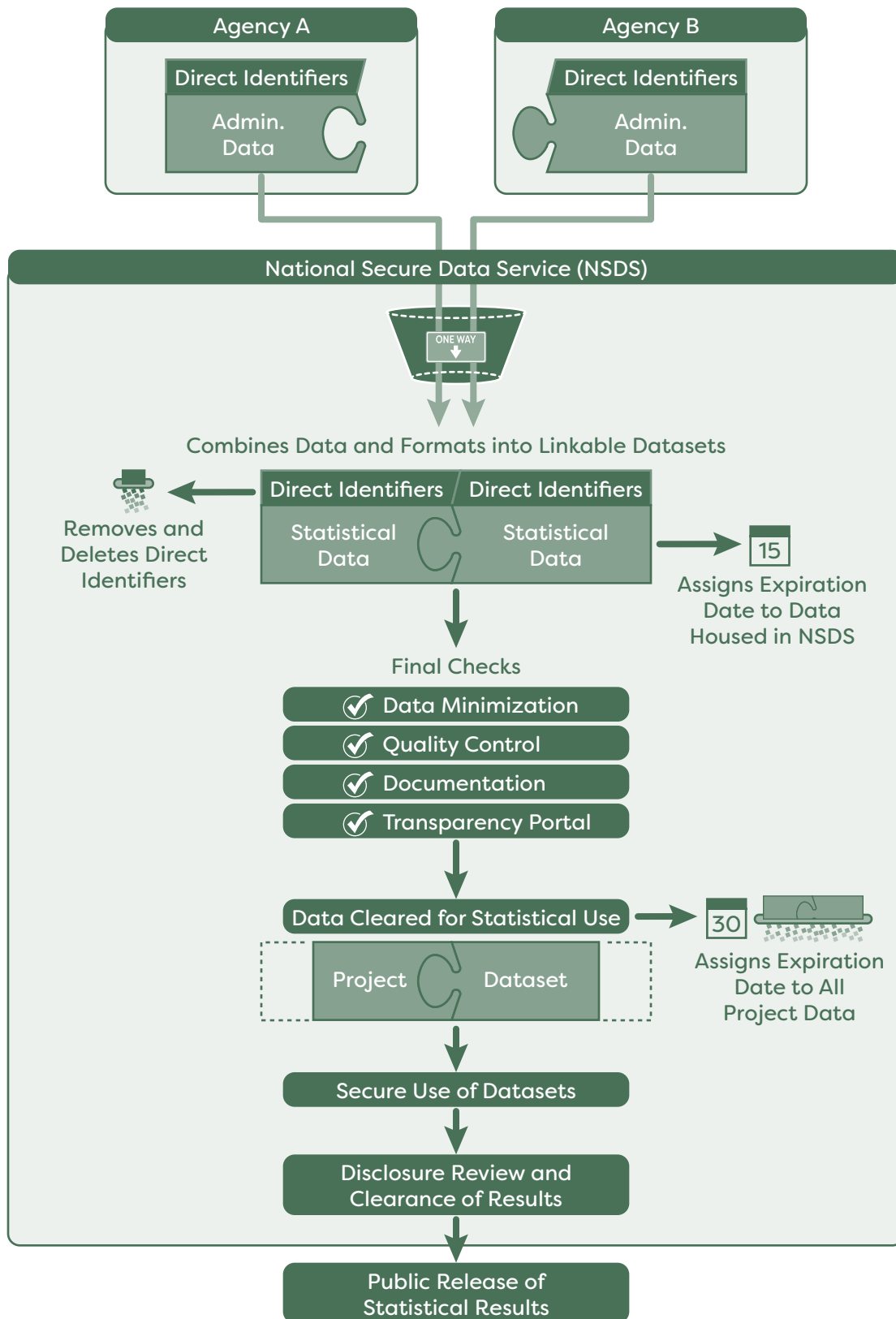
Using the expertise that already exists, NSDS will be well placed to quickly begin implementing activities, such as developing agreements with data providers, developing model approaches for securely sharing data across agencies, and facilitating secure access to data for researchers external to government with approved projects that serve the public good (see Figure 10). The Commission envisions that when the relevant parts of CARRA and CES become part of the NSDS they would continue to manage the FSRDC network and that researchers external to government will continue to access confidential data through the FSRDC network.

The Commission's vision requires that the Federal government stay on the cutting edge of technological approaches to protecting data privacy and security. The technical approaches for combining data require a high degree of expertise and continuous training to keep abreast of rapidly evolving methods.

Concentrating this expertise in a Federal center for excellence means that the government can develop best practices for data protection and secure access. Substantially restricting the individuals who have access to identifiable records on individuals or businesses will provide a higher degree of safety and security for confidential data.

Bringing expertise together within government will help identify and implement feasible new techniques and technologies. Specifically, the NSDS should be a center for excellence in developing, using, and sharing best practices for transparently and securely using confidential data for evidence building. New privacy-protective techniques such as Secure Multiparty Computation (described in Chapter 3) may allow individuals in the Federal evidence-building community to combine data and conduct analyses without directly accessing or storing information. The NSDS should be a leader within government for advancing that goal. The NSDS could also implement provable privacy methods such as differential privacy (described in Chapter 3) to create datasets with minimal and quantifiable risk of re-identification. Furthermore,

Figure 10. The NSDS in Action



best practices should be shared throughout the government and lead to more secure data access and better privacy protections across the Federal evidence-building community.

Because of the existing pilot activities at Census, the implementation of NSDS can begin immediately once the Congress and the President authorize its establishment. Work may then continue in phases, with the core activities that support NSDS fully devolved from the Census Bureau within two to three years.

A phased approach to establishing and implementing NSDS will allow it to conduct its own pilot projects, develop data documentation and integration procedures, and develop and enhance relationships with data providers that allow opportunities for public input (see the box “Potential Phases for NSDS Implementation”).

The NSDS should also have the capacity to provide technical services such as secure data linkage and analytical services on a fee-for-service basis for states, local governments, and other jurisdictions. While technical and analytic services should not be an initial priority for the NSDS, a basic level of such services may be both necessary and desirable in the long term to help improve the capacity for evidence building throughout the nation and to encourage collaboration across levels of government. Researchers may also wish to avail themselves of technical and analytic capacity as it

develops and expands at NSDS.

Building on the existing capacity in the Federal government, and particularly at the Census Bureau, and supplementing it with resources drawn from across government will ensure that the expertise for evidence building can be available not just for projects to advance the Census Bureau’s mission, but government wide.



REC. 4-2: The NSDS should establish a Steering Committee that includes representatives of the public, Federal departments, state agencies, and academia.

The public must have a voice in how the Federal government uses data for evidence building. Public participation ensures the NSDS’s policies and practices remain focused on exclusively statistical purposes for the public good. The Commission concluded that a service like NSDS requires input about how data are used from representatives of the public, privacy and technology experts, Federal departments, state agencies, research institutions, and other stakeholders.

A representative steering committee for NSDS will allow individuals to advise the service on strategic policies and provides one avenue for organizational accountability. The Commission specifically recommends that such an advisory

Potential Phases for NSDS Implementation

Phase I – Legal Authority. New legal authority is enacted to establish the National Secure Data Service (NSDS) in its new location, with a stated transition period of 2-3 years.

Phase II – Relocation and Staffing. The Secretary of Commerce hires an administrator. NSDS appoints the steering committee. The relevant assets, personnel, and resources from the Census Bureau are transferred to NSDS and the administrator brings in staff from other government agencies and outside government as needed.

Phase III – Pilot Projects and Data Documentation. NSDS works with researchers and other stakeholders to build out new privacy-protecting capacities and conduct a series of pilot projects for scaling implementation. The NSDS launches a transparency portal.

Phase IV – New Development and Implementation. NSDS develops new relationships with data providers to meet user needs and continues to scale activities as necessary based on demand. NSDS expands efforts to implement state-of-the-art technologies and techniques for secure data linkage and access and assesses the quality of the approaches.

committee be charged with the following responsibilities:

- Consult with the Secretary of Commerce about budget and resource allocation decisions related to NSDS operations and privacy protections.
- Conduct strategic planning for NSDS.
- Develop and ensure appeals processes are implemented for applications that are denied, for privacy concerns raised after a project is approved, and for administrative sanctions for non-compliance with applicable policies and procedures.
- Oversee and approve an annual report to the Congress and the President providing appropriate performance indicators, evaluating progress toward fulfilling NSDS's purpose, and describing implementation challenges including those that may require new statutory authority.
- Receive and review audit reports on privacy-protective protocols and practices; identify any ameliorating steps to be taken to reduce privacy risk within the mission of NSDS.

While a representative steering committee can provide high-level direction and support for NSDS, it would not be feasible or desirable for such a committee to manage the day-to-day operations and provide leadership for the NSDS. The Commission believes there should be a senior administrator empowered to administer, oversee, and coordinate the activities of the NSDS, such as implementing regulations and developing procedures to acquire access to data from Federal departments and state agencies.

As part of the public participation process, the administrator should ensure that individual projects are approved in consultation with data providers. Information about those projects will then be made available as described in Recommendation 4-3, and will include who the eligible researcher is and what data they are using for what purpose.

Because the American public expects the Federal government to manage the risks of using data for evidence building, the administrator must also prescribe data access control methods consistent with OMB policies and standards, including establishing data use agreements, training for staff and researchers, and determining penalties for inap-

propriate use or failure to follow procedures. Further, the administrator must ensure compliance with applicable Federal laws and policies, especially regarding data confidentiality and security.

The administrator should provide leadership across the Federal evidence-building community to use state-of-the-art data protection and access methods and technologies. However, the administrator should not be expected to also fill the role of privacy officer for the NSDS. A separate, designated NSDS official should be appointed to enforce the full breadth of privacy-protective approaches necessary and expected for the level of data use in the NSDS anticipated by the Commission, the Congress, and the President. The designated privacy official should be given a role in scrutinizing all data linkage activities, data access activities, and other data uses for evidence building.

A designated privacy official at the NSDS could: (1) support OMB in privacy standard setting activities; (2) establish and promulgate NSDS privacy policies, including for researcher qualification, project approval, and data stewardship training; (3) ensure NSDS compliance with Federal privacy laws, regulations, and policies; (4) develop mandatory privacy and data stewardship training for NSDS staff and researchers; (5) review risk assessments when confidential data are combined across departments; and (6) communicate and coordinate with privacy officers in statistical agencies and departments.



REC. 4-3: To ensure exemplary transparency and accountability for the Federal government's use of data for evidence building, the 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.

The Commission believes that advancing beyond the status quo and achieving unparalleled transparency means (1) telling the public about how government data are used for evidence building and (2) regularly auditing adherence to privacy-protecting policies to assure the public that the Federal evidence-building community only uses data for approved exclusively statistical purposes.

The President, through collaboration between the NSDS and OMB, in its government-wide

standard setting role, should use existing statutory authority to establish a clear example of transparency in Federal evidence building. The President should authorize a single internet-based portal that gives details about each project using integrated confidential data for evidence building or each external researcher accessing confidential data for evidence building. The NSDS is the logical place to develop, host, and maintain the transparency portal because of its role in combining data across the topical domains of Federal departments.

Approved projects facilitated by NSDS will be an important input to the portal. In addition, other PSAs should provide NSDS details on projects they undertake with other agencies to link confidential data as well as approved projects for external researcher access to confidential data. The information in the portal should minimally include for each approved project: the project name (which some Federal agencies already publish), a project abstract, the datasets involved, the products of the analyses, and the name(s) and contact information of the approved researcher(s). The transparency portal should also feature the transparency strategies discussed as part of Recommendation 3-1 and could include those discussed under Recommendation 2-8. These strategies include a public inventory of data available for evidence, including an analysis of the sensitivity of the data and publicly available risk assessments for the public release of de-identified confidential data under the Privacy Act and CIPSEA. The portal also should include a mechanism for public feedback.

Once established, the NSDS transparency portal will provide a valuable resource for the public to understand how data are being used and documentation to establish an audit trail. The Congress and the President should choose to assign responsibility for auditing to the Government Accountability Office or to a designated inspector general. Auditors should ensure compliance with privacy protocols and ensure that enforcement mechanisms are appropriately applied. Further, auditors should be responsible for reviewing the data use and handling procedures and processes in place for confidential data. The NSDS should establish and auditors should monitor a “near miss” reporting system for procedural or process violations to create a feedback and learning cycle about risks. Such a system could include descriptions of any violations and re-identifications that may occur, determinations of cause(s), actions taken to

mitigate those cause(s), and a process for revising procedures to reduce the risk of similar future incidents.

Finally, NSDS should report annually to the Congress, the President, and the public about uses of confidential data for evidence building and identify and discuss any problems that arise. Other accountability mechanisms are established through other Commission recommendations; for example, Recommendation 2-8 calls for a single, streamlined process for external researcher access that is transparent about data sensitivity and sets data access restrictions appropriately, and Recommendation 3-1 provides for the public posting and use of risk assessments for public release of de-identified confidential data.



REC. 4-4: The NSDS should have specific administrative and implementation flexibilities including the ability to leverage public-private partnerships and to collect and retain user fees.

In order to implement the NSDS in a manner that allows it to serve the government as a whole and continually adapt to emerging needs and changing technology, some administrative flexibility beyond the usual government practices will be required. Specifically, success for the NSDS will require legal authority for grant-making, cooperative agreements, workforce development, and other activities that facilitate engagement with partner organizations to accelerate the development of new methods and technologies. Such authority could include the ability to sponsor a Federally Funded Research and Development Center to support research and development related to innovative privacy-protective approaches. By encouraging research and development in privacy-enhancing and secure access technologies, the NSDS can both capitalize on and encourage innovation in the private sector with joint benefits to government. Development of new technologies related to providing data access, such as virtual data access technologies, will help address increased demand.

The NSDS should have the authority to collect and spend user fees, with sufficient flexibility to adjust rates based on changes in demand or other factors. Explicit authority should be provided in appropriations bills for other Federal departments

to transfer funds to the NSDS to support its centralized activities, as noted in Chapter 5.

The NSDS can provide a clear public benefit by facilitating generation of evidence about government programs and policies. It will likely not be feasible, however, for the NSDS to cover its expenses entirely with user fees. Therefore, the Commission believes that while user fees for some costs associated with the NSDS can support “self-funding,” other funding mechanisms will be necessary to achieve the Commission’s vision. Such mechanisms could include a direct appropriation or reimbursable funding from other departments. The NSDS will need some level of flexible resources to ensure it can implement cutting-edge technologies that enhance privacy.



REC. 4-5: The Office of Management and Budget should increase efforts to make information available on existing Federal datasets including data inventories, metadata, and data documentation in a searchable format.

Administrative datasets often contain extensive records about individuals and businesses receiving government benefits and services, sometimes including multiple entries on transactions. Given their size, frequent updates and revisions, and data structures, use of administrative data for evidence building requires knowing which data are available and what they mean.

The Commission intends that while data will be accessible through the NSDS, most of those data will remain in their original locations in the future. The NSDS will only maintain a minimal core set of data for linking confidential data within its

secure environment. The technical documentation requirements of new technologies for securely accessing data such as Secure Multiparty Computation further guide the Commission’s conclusion that improvements to the government’s data inventory, metadata, and definitions are necessary precursors to increased evidence generation.

OMB’s ongoing effort beginning in 2013 to develop a data inventory is a productive start, but much more needs to be done to make this information complete and useful for evidence building. The establishment of a comprehensive searchable inventory, through which the public can learn about the data that government collects, would improve transparency. With robust metadata, researchers inside and outside government will be better able to identify which data are needed and useful for answering policy questions, conducting program evaluations, and reducing inefficient and unnecessary data requests. These metadata should be sufficient to allow users to assess quality and implement privacy-enhancing data matching technologies, such as Secure Multiparty Computation. The metadata must also allow the NSDS and OMB to assess the sensitivity of data in specific contexts and to determine access restrictions and privacy protection protocols as described in Recommendation 2-8.






The Commission recognizes that, despite their potential for adding value, developing a data inventory, technical documentation, and consistent definitions for key concepts across datasets is a burden and challenge. In the short term, the Commission recommends prioritizing metadata and other technical documentation related to the datasets researchers consider important for evidence building, including those described in Appendix D. ■

5

Strengthening the Evidence-Building Capacity within the Federal Government

Build and maintain a strong Federal infrastructure for the sustained production and use of evidence.

Recommendations

-  **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.
-  **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.
-  **5-3:** The Congress and the President should direct the Office of Management and Budget (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.
-  **5-4:** The Congress and the President should align administrative processes to support evidence building, in particular by streamlining the approval processes for new data collections and using existing flexibilities in procurement policy.
-  **5-5:** The Congress and the President should ensure sufficient resources to support evidence-building activities about Federal government programs and policies.

Implementing the features of the Commission's vision to improve the use of data for evidence building requires more than just improving access to data and privacy protections. Government must have the capacity to analyze data, and then apply insights to inform policymaking. This chapter addresses the alignment and empowerment of the Federal evidence-building community—including those who generate, manage, and analyze data, those who transform information into evidence, and those who support those functions through the routine processes of government.

“Data are necessary, but not sufficient, to create evidence.”

– Naomi Goldstein, Deputy Assistant Secretary for Planning, Research, and Evaluation in the Administration for Children and Families, Commission Meeting, November 4, 2016

Today, evidence building takes place unevenly across the government, both within the internal Federal evidence-building community and across partnerships with external members of the evidence building community. Within departments inside the Federal government, key actors include Principal Statistical Agencies (PSAs) and other statistical programs, evaluation and policy research offices, program administrators, performance management offices, policy analysis offices, and privacy offices. Often the work of these offices occurs in silos, leading to duplication of effort or missed opportunities for collaboration, and efforts to coordinate evidence building across the Federal government have been challenging. In addition, administrative functions that support Federal government operations across departments are poorly aligned to support and prioritize evidence building. Thus, major gaps remain in the Federal evidence-building community's capacity, which hamper the ability to build on the community's many strengths to increase evidence production.

Federal departments must enhance their capacity for evidence building to support the growth of evidence-based policymaking. Those generating, supporting, and using evidence within and across departments throughout the Federal government must be empowered and organized to work together and accomplish shared goals. To

maximize evidence building, Federal departments must have the capacity to support the full range of analytic approaches required for evidence building, including the development of statistics, evaluation, and policy research. These functions must be operational, appropriately resourced, and well-coordinated both within and across departments. Strong leadership within government that prioritizes evidence building and creates the demand for evidence is vital for institutionalizing these functions. Without a strong institutional foundation, other recommendations related to improving data access, establishing the National Secure Data Service (NSDS), and implementing enhanced privacy protections will not have the comprehensive impact that is needed.

Findings

The Commission received input from numerous stakeholders, both internal and external to the Federal government, identifying challenges to increasing the capacity of the Federal evidence-building community and numerous ways in which the practices of the Federal government could be streamlined to increase the volume, quality, and timeliness of evidence production. The collective input stressed the importance of addressing these challenges directly. Based on this feedback, the Commission identified five primary challenges that currently restrict the capacity of the Federal government's evidence-building community.

Challenge #1: The capacity of Federal departments to support the full range of evidence-building functions is uneven, and where capacity for evidence building does exist, it is often poorly coordinated within departments.

Currently, some departments operate all, and other departments operate only some, of the evidence-building functions necessary to support evidence-based policymaking. Where these activities are taking place in different units across a department, these efforts must also be coordinated to maximize the department's capacity to fully address a specific research or policy question. In pursuit of more and better evidence, departments may choose to pursue different avenues for

increasing their capacity for evidence building. In some cases, departments will opt to strengthen their human capital by hiring staff or establishing units with particular expertise. In other cases, departments may decide that leveraging expertise outside of government is a more efficient approach to expanding capacity.

Regardless of how a department may pursue expanded capacity, the coordination of evidence-building functions is vital for achieving maximum efficiency. Currently, some departments prioritize coordination across units while other departments lack a coordinating infrastructure. The following section (1) describes each of the evidence-building functions needed within Federal departments, (2) offers existing approaches for expanding the evidence-building capacity of departments through human capital strategies and external partnership arrangements, and (3) addresses the importance of coordinating evidence-building efforts within a Federal department.

Evidence-Building Functions within Federal Departments

Key actors within the Federal evidence-building community include: the PSAs and other statistical programs, evaluation and policy research offices, program administrators, performance management offices, and policy analysis offices. These are central entities that support an organizational culture that enables evidence building.

Thirteen PSAs form the centerpiece of the larger Federal Statistical System, which also includes dozens of statistical programs spread across Federal departments, embedded within programmatic agencies. For many decades, the PSAs have collected most of their data by conducting surveys of individuals, businesses, and other organizations. PSAs now increasingly supplement these surveys with administrative data, creating new opportunities, and also new relationships with program agencies. Because of their existing legal confidentiality protections, culture of data protection, and substantive and technical expertise, PSAs can serve as building blocks for the modern system of evidence building in the United States. The growing capabilities of many PSAs to transform and analyze administrative data should be expanded to meet the needs for government evidence building and the public good.

“Organizational culture is an ongoing dialogue between leaders and others in their organization about ‘how we do things around here.’ Data and evidence are necessary but not sufficient. Continuing dialogue based on the data and evidence is critical.”

– Seth Harris, Former U.S. Department of Labor Acting Secretary, Commission Meeting, March 13, 2017

The evaluation and policy research functions are newer additions to the evidence-building community in many departments. The American Evaluation Association defines evaluation as “assessing the strengths and weaknesses of individual programs, policies, and organizations to improve their effectiveness, efficiency, and worth... It uses systematic data collection and analysis to address questions about how well government programs and policies are working, whether they are achieving their objectives, and, no less importantly, why they are or are not effective.”¹ Policy research is related to evaluation and often precedes formal evaluation, but plays a fundamentally different role in evidence building.² Because these functions are less institutionalized in many Federal departments, they require different solutions to bolster their development compared with longer standing functions such as the development of statistics. While some excellent examples of evaluation and policy research offices exist in the Federal government today, many departments will likely need additional capacity to meet the Commission’s vision of cultivating these two especially important types of evidence building.

1. American Evaluation Association, “An Evaluation Roadmap for a More Effective Government” (Washington, D.C.: AEA, October 2013); <http://www.eval.org/d/do/472> (accessed August 10, 2017).

2. Research includes basic and applied activities conducted in the pursuit of new knowledge, the former directed at general applications or theory and the latter focused on practical objectives. “Policy research” is used synonymously with “policy-relevant research” both of which are considered gaining knowledge about programs and policies. See OMB, Section 84: Character Classification (Schedule C) in “Preparation, Submission, and Execution of the Budget” (Circular A-11, Washington, D.C.: OMB, July 2016).

Program administrators play an increasingly important role in evidence-building activities, as more administrative data collected for program administration purposes are being transformed for use in evidence building. In addition, program administrators are one audience for the evidence being generated, as they have the capacity to administratively implement program and policy adjustments based on the results of research and evaluation.

Performance management offices are responsible for advising and assisting departmental leadership to ensure that the mission and goals are achieved cost effectively through strategic and performance planning, measurement, analysis, regular assessment of progress, and the use of high-quality performance information and other evidence to improve results.³

Centralized policy analysis offices support many departments today and typically perform an important translation function. Policy analysis translates knowledge gained from evaluation and policy research both inside and outside government, making the information accessible to program administrators and operators, policymakers, including the Congress, and the public. The dissemination of information helps inform and spur action. These “knowledge brokers” serve as intermediaries between evidence users and producers, and play an important role within the evidence-building community.⁴ Increasingly, as a greater volume of evidence is produced in the future, knowledge brokering efforts also will need to expand. Policy analysis offices sometimes sponsor or rely on formal dissemination tools for evidence, such as the What Works Clearinghouse operated by the U.S. Department of Education, and the Clearinghouse for Labor Evaluation and Research, operated by the U.S. Department of Labor, which offer compilations of research and evaluation to provide grantees with a valuable starting point for determining which interventions may be most appropriate to their popula-

tions, geographies, or contexts.⁵

To the extent that capacity is uneven in government, important evidence-building activities may be crowded out by other priorities within departments. Departments face numerous pressures to respond to immediate and important requests that limit the ability to deploy staff or funding for evidence building. For example, prioritizing funding for new programs and initiatives may mean longstanding programs receive less attention for measuring or assessing program outcomes. Program administrators whose performance is assessed based on their meeting annual targets may be reluctant to devote program resources to longer-term evidence-building activities. In most departments, the development of annual performance measures aimed at providing timely information for decisions in the budget process itself consumes an array of resources. While performance measurement (or monitoring) is an important component of the broader field of evaluation,⁶ a focus on short-term activities may jeopardize the ability to prioritize long-term projects that do not align with the time horizons of policymakers and program administrators. The capacity must exist for both short- and long-term activities.

Mechanisms for Expanding Capacity for Evidence Building in Federal Departments

Strengthening the program evaluation function within the Federal government is an important first step in expanding evidence building. Several strong evaluation units currently exist across government and work with contractors, other government staff, and academic institutions to implement government’s evaluation function. Such offices include the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families (ACF) of the U.S. Department of Health and Human Services (HHS) and the Chief Evaluation Office within the

3. OMB, “Overview of Federal Evidence-Building Efforts” (white paper for the Commission on Evidence-Based Policymaking, Washington, D.C.: OMB, 2016); https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/overview_of_federal_evidence_building_efforts.pdf (accessed August 10, 2017).

4. Karol Olejniczak, Estelle Raimondo, and Tomasz Kupiec, “Evaluation Units as Knowledge Brokers: Testing and Calibrating an Innovative Framework,” *Evaluation* 22, no. 2 (April 2016): 168–189.

5. What Works Clearinghouse; <https://ies.ed.gov/ncee/wwc>; Clearinghouse for Labor Evaluation and Research; <https://clear.dol.gov> (both accessed August 10, 2017).

6. Kathryn Newcomer and Clinton T. Brass, “Forging a Strategic and Comprehensive Approach to Evaluation Within Public and Nonprofit Organizations,” *American Journal of Evaluation* 37, no. 1 (March 2015): 80–99.

Department of Labor (DOL). (See the box “U.S. Department of Labor—Chief Evaluation Officer”). OMB supports evaluation activities across government and encouraged the administrative creation of a voluntary coordinating unit called the Interagency Council on Evaluation Policy.

As noted in the Analytical Perspectives section of the *President’s Fiscal Year 2018 Budget*, “Centralized or chief evaluation offices play an important role in an evidence infrastructure that can develop and sustain agency capacity to build and use evidence.”⁷ In this vein, the Commission received testimony suggesting that a Chief Evaluation Officer could assume responsibility for (1) establishing department-wide evaluation and research policies that encourage rigor, credibility, independence, and transparency; (2) coordinating and supporting technical expertise for evaluation and research within the department; and (3) identifying and setting priorities for departmental program evaluation and policy research, with appropriate attention to the mission and context of each department. Particularly in very large or decentralized organizations, departments may have multiple centers for evaluation, which could be coordinated by a Chief Evaluation Officer.

7. OMB, “Building and Using Evidence To Improve Government Effectiveness,” in *Analytical Perspectives, Budget of the United States Fiscal Year 2018* (Washington, D.C.: OMB, May 2017); https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/budget/fy2018/ap_6_evidence.pdf (accessed August 10, 2017).

To both support and benefit from routine evidence building, Federal departments require a range of expertise including in-house expertise to design and conduct evaluations or surveys, advanced technical knowledge required for combining or analyzing data, and expertise in writing and managing highly technical evidence-building contracts. In seeking to increase capacity for evidence building, some Federal departments may choose to establish new units or expand the functions of an existing unit. The CEP Survey of Federal Offices showed that over half (54 percent) of responding offices reported difficulty hiring staff with the appropriate skill set to match the work requirements for evidence-building activities.⁸ Functions that are newer in government, such as program evaluation, have particular challenges in attracting and retaining qualified staff and, therefore, may require some adjustments in existing Federal human resource policies. Today, while the U.S. Office of Personnel Management (OPM) recognizes statisticians, economists, information specialists, and policy analysts as official and distinct occupations within government, no occupational series exists specifically for the field of program evaluation.

External partnerships also can be an effective way to expand government capacity for production

8. Based on 79 offices that reported that they collect or use data for statistics, evaluation, research, or policy analysis or spend a portion of their budget for such purposes.

U.S. Department of Labor—Chief Evaluation Officer

The U.S. Department of Labor (DOL) is currently the only Cabinet-level department to have a Chief Evaluation Officer. Established in 2010, the Chief Evaluation Office, led by the Chief Evaluation Officer, coordinates, manages, and implements DOL’s evaluation program. The Chief Evaluation Office is an independent evaluation office, located organizationally in the Office of the Assistant Secretary for Policy. At the department level, the Chief Evaluation Officer:

1. Designs, funds, and oversees the implementation of a portfolio of evaluation proj-

ects designed to build evidence about DOL programs and policies.

2. Complements, not duplicates, agency evaluation functions.
3. Raises the quality of evaluations and awareness and knowledge of evaluation methodology and standards.
4. Improves use and dissemination of evaluation results.
5. Improves access to, quality of, and use of data.

of evidence by leveraging the expertise of the non-Federal evidence-building community. The Robert Wood Johnson Foundation encouraged the Commission to consider how “partnerships between researchers, service providers, and large institutions—both governmental and corporate—can improve the capacity to develop evidence.”⁹ Engaging successfully with external partners relies on having staff and structures in place to establish and support collaboration with academics, non-profits, foundations, and businesses. While tools like grant-making or cooperative agreements have been used to good effect at some agencies to build external partnerships, not all agencies have the necessary authority to use these tools for partnerships.

Partnerships can be especially useful for smaller agencies with limited capacity. Agencies must continually develop their workforce and increasingly need staff with specialized expertise in privacy protections, data security, computer science, and statistical methods. Partnering with non-governmental entities can be an efficient way to

expand staff capacity and build staff expertise. Federal departments have mechanisms for hiring staff from colleges and universities and other institutions on a temporary basis, such as the Intergovernment Personnel Act (IPA) Mobility Program. Assignments under the IPA program are intended to “facilitate cooperation between the Federal government and the non-Federal entity” and to serve a “sound public purpose.”¹⁰

Coordination of Evidence-Building Functions within Federal Departments

Currently, the coordination of evidence-building activities within departments varies widely. In some departments, very little coordination exists, whereas other departments place a high priority on coordination. For example, HHS supports an entire unit, the Office of the Assistant Secretary for Planning and Evaluation (ASPE), which is responsible for major activities in policy coordination, legislation development, strategic planning,

9. Robert Wood Johnson Foundation, submission to the Commission’s Request for Comments.

10. Office of Personnel Management, “Hiring Information: Intergovernment Personnel Act,” <https://www.opm.gov/policy-data-oversight/hiring-information/intergovernment-personnel-act> (accessed August 10, 2017).

Case Study: HHS Data Council

The Data Council within the U.S. Department of Health and Human Services (HHS) is an example of successful coordination of data resources within a Federal department.¹ The HHS Data Council is the principal internal advisory body to the Secretary for health and human services data and statistical policy, and facilitates collaboration and coordination among various offices involved in evidence building, meeting monthly to discuss department-wide issues related to data and statistical policy. The HHS Data Council plays an essential role leading the development and implementation of an integrated data collection strategy, coordination of analysis activities, development and implementation of

data standards and statistical policies, and privacy protection policies. An important activity of the Data Council includes leveraging HHS statistical and administrative data collections to support evidence building. The HHS Data Council develops recommendations regarding the collection, analysis, and dissemination of data to guide future decisions and enhance the health and well-being of Americans.

The Council is co-chaired by the Assistant Secretary for Planning and Evaluation (ASPE) or their designee, the head of an operating division or their designee (currently the Director of the Agency for Healthcare Research and Quality), and the Director of the National Center for Health Statistics, who also serves as the senior advisor to the Secretary on health statistics. The office of the ASPE serves as Executive Director for the Council.

1. U.S. Department of Health and Human Services, “HHS Data Council: An Introduction,” <https://aspe.hhs.gov/hhs-data-council-introduction> (accessed August 10, 2017).

policy research, evaluation, and economic analysis across the entire department. (See the box “Case Study: HHS Data Council”).

Harnessing administrative data as a more integral part of evidence building heightens the need for meaningful collaboration within departments. Coordination within a department is vital for the efficient production of evidence.

Challenge #2: Federal departments frequently do not have an integrated approach or a long-range plan for evidence building.

To make evidence-based policymaking routine in the Federal government, a broad base of evidence and information must be available in order to inform decision-making. Departments can benefit from an organized approach by which they regularly identify short- and long-term priority research and policy questions relevant to the department’s mission and legal responsibilities. The result of this exercise is a learning agenda, which can be used by leadership to prioritize the set of research and policy questions to be pursued by the department over a given period. The evidence-building community within a department can in turn use the learning agenda as a coordination tool.

The effective implementation of learning agendas requires sustained leadership and support. The various units within the evidence-building community in a department must work together to determine the best approach to answering a priority research or policy question, and to allocate the work appropriately across the different evidence-building functions. A well-coordinated evidence-building community within a Federal department will be able to leverage the various strengths of each function, resulting in a more robust project.

Learning Agendas

The development of a learning agenda is a concrete step that some Federal departments have taken towards prioritizing evidence building. Learning agendas are defined as “a set of broad questions directly related to the work an agency conducts that, when answered, enable the agency to work more effectively and efficiently, particularly pertaining to evaluation, evidence, and de-

cision-making.”¹¹ Such learning agendas typically identify the most pressing research and policy questions facing a department at a particular time, and provide departmental leaders with a mechanism to prioritize research questions within budget and policy timeframes.

The Office of Policy Development and Research within the U.S. Department of Housing and Urban Development (HUD) developed a learning agenda titled the “Research Roadmap.”¹² The Research Roadmap outlines the key research questions relevant to HUD’s mission, and enables HUD leadership to be responsive to OMB’s direction to strengthen the use of evidence and technology to improve government performance. HUD refreshes the Research Roadmap every five years, engaging a broad stakeholder group both inside and outside of the department, to ensure that the priorities captured reflect the needs of the department and the broader field. Learning agendas also can be used to communicate research priorities to external partners to help catalyze targeted evidence-building activities outside the Federal government.

A Portfolio of Evidence

Evidence building and use must become a regular feature of program oversight and operations to promote continual learning, program refinement, and accountability. Today, nearly every Federal government program collects information to support program operations and monitor performance. The data gathered through these activities provide valuable basic information about the scope and reach of programs, such as identifying the number of participants in a program. Departments generally need more in-depth information to understand how a program is implemented, the costs of administering the program, and the extent to which policies yield anticipated results. Policy decisions should be reviewed using a broad set of methodologies, including descriptive statistics, process studies, implementation evaluations,

11. U.S. Agency for International Development, “Implementing a Learning Agenda Approach,” https://usaidlearninglab.org/sites/default/files/resource/files/defining_a_learning_agenda.pdf (accessed August 10, 2017).

12. U.S. Department of Housing and Urban Development (HUD), *HUD Research Roadmap: 2017 Update* (Washington, D.C., January 2017); <https://www.huduser.gov/portal/pdf/ResearchRoadmap-2017Update.pdf> (accessed August 10, 2017).

impact evaluations, using randomized control trials where appropriate, and meta-analysis. The results of all relevant evidence-building activities is referred to as a portfolio of evidence.

The American Evaluation Association suggests that in selecting a methodological approach, the first step is to “identify the important evaluation questions [needing to be] answered to effectively direct the future of Federal programs” allowing evaluators to then “identify which scientific methods are best suited to answer those questions.”¹³ The Commission finds that the need for evidence about programs is context specific and should be tailored to produce the most relevant, valid, and reliable information possible for evaluating individual programs and policies.

Two of the Commission’s guiding principles for evidence-based policymaking outlined in Chapter 1 are particularly relevant when considering both the generation of evidence and the application of evidence by policymakers. The principle of rigor states “evidence should be developed using well-designed and well-implemented methods tailored to the question being asked.” For example, when seeking to determine program impacts, random assignment generally is preferable for developing causal statements. The Institute of Education Sciences and the Directorate for Education and Human Resources of the National Science Foundation established guidelines for evaluation that conclude “Efficacy, effectiveness, and scale-up research should use study designs that will yield impact estimates with strong causal validity . . . generally and when feasible, they should use designs in which the treatment and comparison groups are randomly assigned.”¹⁴

In another principle adopted by the Commission, policymakers are urged to approach evidence with humility. The Commission’s humility principle says “Care should be taken not to over-generalize from findings that may be specific to a particular study or context.” Even when evaluated using methods appropriate to stated questions, contexts may differ, circumstances can change, and fidelity to program design may diminish over time while conducting an evaluation.

The Commission identified numerous examples of Federal programs developing increasingly rigorous portfolios of evidence, but these programs are still the exception, rather than the rule. One example is the Teen Pregnancy Prevention Program administered by HHS, which was designed to address high teen pregnancy rates in the United States by replicating evidence-based models and testing innovative strategies. Evidence building was woven into the program from the start, including a full range of studies from implementation assessments to impact evaluations, using random assignment when appropriate. The program is administered by the Office of Adolescent Health, which used a multipronged approach to evaluation, including the establishment of uniform performance measures, a series of grantee-led evaluations, and multiple cross-site evaluations. The purpose of this framework was to address the question of whether the replicated evidence-based teen pregnancy prevention programs and the new, innovative strategies for preventing teen pregnancy were effective.

Evelyn Kappeler, former Director of the Office of Adolescent Health at HHS, described the program as follows:

Our experience with the first cohort of [Teen Pregnancy Prevention grantees] brings to light several important issues regarding the capacity of the organizations to conduct rigorous evaluations. Many of these lessons have applications for government entities, which are conducting evaluations in community settings. The challenges and lessons learned include the importance of providing for a planning and piloting period early in the grant cycle, measuring and monitoring fidelity and adaptations, ensuring high quality program implementation, incorporating evaluation effectively once program implementation has already begun, ensuring strong contrast between treatment and control, providing intensive evaluation technical assistance, and disseminating all evaluation results transparently.¹⁵

13. George Grob, American Evaluation Association, Commission Public Hearing, Washington, D.C., October 21, 2016.

14. Institute of Education Sciences, U.S. Department of Education, and the National Science Foundation, *Common Guidelines for Education Research and Development* (August 2013); <https://ies.ed.gov/pdf/CommonGuidelines.pdf> (accessed August 10, 2017).

15. Evelyn Kappeler, Office of Adolescent Health, Commission Meeting, Washington D.C., November 4, 2016.

The challenges cited with regard to the Teen Pregnancy Prevention program demonstrate the complexity of program evaluation. These challenges can make it difficult to keep the attention of leadership on evidence building over time. Even with an appropriate organizational structure and sufficient staff capacity, Federal departments still may not prioritize evidence building among the competing demands on their resources without the right set of incentives and expectations. As the quote from Ms. Kappeler suggests, developing and sustaining Federal leadership in support of evidence building is a formidable undertaking. Senior department leaders play an important role in making the generation and use of evidence routine in Federal policymaking.

Today, Federal departments lack individual incentives for senior leaders to produce and use evaluation and policy-relevant research. While the Commission does not offer recommendations to address individual motivation, it learned that departmental approaches could include incorporating measures into the performance reviews of career senior leaders that reflect their effectiveness in producing and using evidence. Similarly, the Senate could use the confirmation process to seek affirmative responses regarding a political appointee's support for producing and using evidence to inform decision-making.

The political process can itself introduce uncertainties about the value of evidence and its use in policymaking. A report from the Bipartisan Policy Center observes that “the fragmented nature of the legislative process, the clash of partisan interests, and the labyrinth of legislative procedures and practices make it more difficult to establish a coordinated structure for using research and evaluations in a more systematic fashion in the legislative process.”¹⁶ Establishing clear expectations through reauthorizations of programs, development of regulations, the budgeting process, and other routine interactions between the President, the Congress, and Federal departments will reinforce the demand for more and better evidence.

Program Design

The statute that established the Commission's charge includes multiple references to embedding

evaluation and policy research within program design. One challenge is that programs, policies, and regulations are generally designed without involving the evidence-building community. Engaging the evidence-building community as partners during the program design process will help ensure that the program can be rigorously evaluated and that research questions will meet the needs of program managers and policymakers. Barriers to early collaboration between evaluators and program designers include limited evaluation capacity and expertise, program offices' concerns about how evaluation results will be used, and lack of access to quality data. Several public comments submitted to the Commission highlighted core elements that must be incorporated into program design, including the following:¹⁷

- Clearly stated goals, objectives, logic, outputs, and desired outcomes.
- Recognition of roles for different types of analyses based on program, policy, or regulation stage and development (e.g., prospective, pilots, implementation, retrospective, systematic reviews/meta-analysis).
- Incentives for evidence production.
- Sufficient legislative authority or flexibility to enable evaluation, including mechanisms to encourage innovation and to conduct impact evaluations, using random assignment when appropriate.
- Incorporation of data collection needs and requirements; a focus on data quality to ensure meaningful and useful information is gathered in order to assess program costs and outcomes.

Nearly all programs collect some form of administrative data about program beneficiaries or the services offered. How programs use that information, however, varies greatly. As the evidence-building community increasingly relies on those data, programs and evidence building

16. Bipartisan Policy Center, *Congress and Evidence-Based Policymaking: Creating a 21st Century Legislature* (Washington, D.C.: Bipartisan Policy Center, 2017): 13.

17. Based on comments submitted to the Commission by J-PAL North America, Laura and John Arnold Foundation, and American Evaluation Association; <https://www.cep.gov/library/testimony.html> (accessed August 10, 2017).

become increasingly interdependent. The information collected by programs must be analyzed or made available to those who can interpret the data in a manner that does not compete with, but rather enhances, programs' ability to implement day-to-day operations and achieve their intended goals in the most cost effective manner. Leadership that recognizes the value of evaluation and other evidence building, and the role of administrative data in those activities, can ensure that this approach becomes the norm. The Commission finds that expectations for evaluation established by the Congress and the President can promote continuous production of valid and reliable evidence about programs and policies and their costs.

Challenge #3: The current organizational structure of OMB does not optimize the agency's ability to coordinate evidence building across the Federal government.

Federal departments often struggle to coordinate evidence-building activities within their own departments, and coordinating across departments can be even more challenging. Building evidence within topical silos like labor, transportation, and education alone misses opportunities to address cross-cutting research and policy questions. For example, questions about the impact of housing assistance on health outcomes or the impact of food security on child development would necessarily rely on the integration of data resources managed by two separate Federal departments. A lack of coordination across government for evidence building leads to unnecessary burden and cost from duplicative data collection, missed opportunities for programmatic collaboration, and a less robust response to a crosscutting policy or programmatic question.

The Critical Role of OMB

OMB serves as the central coordinator for many government-wide processes, but currently, responsibilities related to evidence building are dispersed across the organization without consistent, sustained internal coordination. By law, the Office of Information and Regulatory Affairs (OIRA) is the unit within OMB responsible for developing government-wide information, privacy, confidentiality, and statistical policy. It implements such

policy through required transactional reviews of agency information collection requests, systems of record notices, and other related reporting and by chairing interagency councils, such as the Privacy Council and the Interagency Council on Statistical Policy. The law also requires OIRA to ensure the sufficiency of budget proposals for statistical activities. These responsibilities coexist with OIRA's regulatory review and coordination function, which accounts for the majority of OIRA's resources. In addition to OIRA, OMB also houses several other statutorily created offices that play a role in evidence building. These other offices include the CIO, who oversees information technology policies that have direct bearing on information privacy, confidentiality, metadata and Open Data; the Chief Performance Officer, charged in law with overseeing performance management activities within the Federal government; and the Office of Federal Procurement Policy, which helps set Federal contracting policies. OMB also houses a small team focused primarily on Federal program evaluation. These offices work together with OMB's Resource Management Offices to prioritize policy and resource allocations across the Executive Branch. Each office reports to a different political appointee, and each can issue guidance or directives to agencies that enable or discourage activities related to evidence building.

The capacity of OMB to effectively coordinate the Federal evidence-building community has been complicated by the ways in which the roles of these offices have evolved over time, resulting in confusion or inconsistent guidance for agencies. For example, the PRA expressly directs that OIRA facilitate government-wide "information management" policies.¹⁸ The E-government Act of 2002 fractured implementation responsibilities within OMB by establishing a separate office charged with responsibility for all information technology policy, overlapping with the information policy function in OIRA. Perhaps more problematic, many agency organizational structures developed to mirror OMB's organizational arrangement. As the demand for evidence to support the policymaking process continues to grow, the operational silos within OMB will likely only become more constraining for the timely production of evidence across government.

18. Paperwork Reduction Act, 44 USC § 3503 (1995).

Due to OMB's current organizational structure, departments typically need to coordinate with multiple offices within OMB before launching new evidence-building activities. This structure places the burden of coordination on individual departments, rather than OMB, and further challenges the ability to coordinate throughout the evidence-building community. Because OMB is the hub of evidence and information policy in the Federal government, the Commission finds that fragmentation of its evidence-building functions hampers its ability to sufficiently prioritize and coordinate evidence building. The Commission finds further that ensuring OMB is structured to effectively coordinate the evidence-building community is crucial for realizing the full benefits of the Commission's recommendations.

Coordinating Evidence Building Across Federal Departments

In addition to improving the organizational structure of OMB, the Commission finds that interagency coordination is vital for the efficient operation of the Federal evidence-building community (see the box "Why Is Coordinating Evidence Building across Government Important?"). OMB already convenes and participates in several interagency coordination groups related to evidence building. One key group, the Interagency Council on Statistical Policy,¹⁹ has a statutory role in advising OMB on the needs and policies of the statistical system.

19. Paperwork Reduction Act, 44 USC § 3504(e)(8) (1995).

Why is Coordinating Evidence Building Across Government Important?

Coordination of evidence-building activities across government is particularly important when a research or policy question spans multiple topical domains. For example, to understand the impact of federal student aid and the GI Bill on veteran educational outcomes, one approach would be to combine data from both the U.S. Department of Education (ED) and the U.S. Department of Veterans Affairs (VA). However, both agencies track veterans who are students differently. Specifically, ED tracks student outcomes, but cannot identify which students in its database are veterans, in part because of a skip pattern introduced on ED's Free Application for Federal Student Aid around the same time that the Post 9/11 GI Bill was enacted, causing most students not to be asked if they are veterans. The exact opposite is true at the VA. The VA tracks which veterans are using the GI Bill, but does not track student outcomes, at least not robustly.

The problem becomes more prominent when one considers the differing methods by which ED and VA track colleges and the stu-

dents attending them. VA tracks colleges using VA "Facility Codes," which have no relation to either of the two different methods of tracking colleges used by ED. ED officials have spent years trying to build a crosswalk between the two separate systems within the department for tracking colleges, and currently, both appear in ED's Integrated Postsecondary Education Data System database. Crosswalking the data collected by ED and VA is even more complicated, and is especially difficult in the case of for-profit education companies with numerous campus and online locations. Harmonizing the VA and ED data collections to permit the tracking of colleges and the students who attend them is important to understanding the success of hundreds of billions of dollars of taxpayer investment in Federal student aid and the GI Bill.¹

1. Carrie Wofford, Veterans Education Success; Mark Schneider, American Institutes for Research, Commission Public Hearing, Washington D.C., October 21, 2016.

Additional coordinating committees, including the Privacy Council, Performance Improvement Council, and Interagency Council on Evaluation Policy, emerged more recently.²⁰ These groups provide valuable communities of practice within the Federal government to support knowledge generation, problem solving, best practice dissemination, and shared services. The Commission encourages the continuation and prioritization of the leadership of these Councils in order to bring maximum value to the Federal evidence-building community.

Challenge #4: Administrative processes are not tailored or aligned to support evidence-building activities.

During the Commission's fact-finding phase, both governmental and non-governmental witnesses identified numerous administrative barriers within the Federal government that hamper the efficient production of evidence. In a follow-up to the Commission public meeting held in November 2016, the Interagency Council on Evaluation Policy submitted a list of the top barriers to increasing Federal evidence-building capacity, highlighting bureaucratic barriers that "discourage evaluation, create inefficiencies and impose additional costs when conducting a Federal evaluation, particularly issues related to PRA, interagency agreements, and procurement."²¹ Evidence building requires support from administrative functions, but these functions are not always aligned or tailored to fulfill the particular needs of the evidence-building community.

Procurement

Procurement is the process by which government acquires supplies and services using funding appropriated by the Congress, including through grants and contracts.²² In fiscal year 2016, the Fed-

eral government awarded just under a half trillion dollars in contracts to entities to carry out work on behalf of the Federal government.²³ Contracts are an important mechanism used by Federal departments to carry out research and evaluation activities. Across government, numerous offices often identify contract procurement as a barrier to the efficient production of evidence. In the CEP Survey of Federal Offices, 40 percent of responses identified lack of ability to execute and manage contracts as a barrier to their ability to use data for evidence building, including 17 percent who identified this as a moderate or major barrier.²⁴

One challenge with contracts relates to how to best structure a procurement at the outset of a project to support evidence building. This challenge can be addressed through better application of existing flexible guidance related to procurement applied to research and development contracts.²⁵ The Commission finds that existing regulations acknowledge that certain types of services, specifically those identified as research and development, may be more appropriately procured with a more flexible, yet riskier, contract. The Federal Acquisition Regulation states "because of the importance of technical considerations in research and development, the choice of contract type should be made after obtaining the recommendations of technical personnel."²⁶ Applying this standard consistently throughout the evidence-building community holds the potential to improve the government's ability to issue contracts that are properly structured to support evidence-building activities.

A second challenge relates to staff familiarity with and ability to manage contracts that support evidence-building activities. Acquisition staff who may be unfamiliar with contracts that are not structured as the more typical fixed-price variety and with the flexibilities available under research and development contracts may seek to limit an agency's preference to use a riskier contract type in support of evidence-building activities. Similarly,

20. The Performance Improvement Council was established by Executive Order 13450 in 2007 and was later codified in the GPRM Modernization Act of 2010. The Interagency Council on Evaluation Policy was established administratively in 2015. The Privacy Council was established by executive order in 2016; <https://obamawhitehouse.archives.gov/the-press-office/2016/02/09/executive-order-establishment-federal-privacy-council> (accessed August 10, 2017).

21. Interagency Council on Evaluation Policy, Comments to the Commission, "'Top-Five List' of Issues and Solutions Related to Federal Evaluation Activity," November 22, 2016.

22. Federal Acquisition Regulation, 48 C.F.R. § 2.101 (2005).

23. For the total value of awarded contracts (as well as grants, loans, and other assistance) made by the Federal government by fiscal year, go to www.usaspending.gov.

24. Includes offices that reported that they collect or use data for statistics, evaluation, research, or policy analysis or spend a portion of their budget for such purposes.

25. Federal Acquisition Regulation, 48 C.F.R. § 35.000 (2005).

26. Federal Acquisition Regulation, 48 C.F.R. § 35.006 (2005).

technical staff responsible for overseeing the day-to-day management of the contract may be unfamiliar with proper techniques for managing the performance of a riskier contract type. Differing goals between acquisition professionals and technical staff may lead to difficulties in identifying the right contract type and raise concerns about the capacity of the government to properly manage the contract.

Information Collection Reviews

The PRA directs OMB to coordinate requests for information collections initiated by Federal departments.²⁷ The law requires that any collection for ten or more individuals from a survey, questionnaire, or form undergo a review to ensure collected information achieves public benefit and maximum utility. The vast majority of data collections undertaken by the evidence-building community fall within the scope of the PRA, as do administrative data collections undertaken, sponsored, or required by Federal departments. The purposes embodied by the PRA are important for ensuring that Federal information collections are necessary, useful, and of high quality. However, the Commission finds that the current structure of the information collection review and approval process can be inefficient. These inefficiencies lead to problematic delays in data collection and challenges to OMB's ability to maximize the coordination and transparency of Federal information collections.

By law, there are three primary actors involved in processing an information collection for approval. The agency wishing to sponsor the collection prepares an Information Collection Request, which must meet the standards outlined in the PRA, including elements such as the need, the plan, and the public burden estimated for the collection. The departmental CIO is responsible for certifying that the collection meets these standards *prior* to submitting the package to OMB. OMB is responsible for review and approval of the package.²⁸ The Information Collection Request, along with the proposed data collection instruments to

be administered under the collection, also must be made available for two public comment periods of 60 and 30 days during this sequential process. Completing these steps requires four months at a minimum, but six to nine months is considered a more realistic time frame for approval, including incorporation of public comments as well as necessary departmental and OMB clearances.²⁹

During the Commission's fact-finding phase, members of the evidence-building community from both inside and outside the Federal government mentioned that PRA requirements are burdensome and time-consuming. The Laura and John Arnold Foundation noted that the PRA processes cause "long delays and bottlenecks for agencies seeking approval of evaluations that have stymied efforts by agencies to increase the number of high-quality studies of important research questions."³⁰ Six of the 10 evaluation offices that responded to the CEP Survey of Federal Offices noted that information collection requirements acted as a moderate or major barrier to their evidence-building activities.

There are two primary areas of concern for those who identify the Information Collection Request review and approval process as a barrier to engaging in evidence building with maximum efficiency. The first issue relates to the broad net cast by the PRA regarding the size of information collections requiring OMB review and approval. The low threshold number of respondents that triggers the Information Collection Request clearance process (10 or more) means most collections the Federal government wishes to undertake require OMB review and approval; few collections for evidence building can be implemented with fewer than 10 respondents. The broad definition of what constitutes an "information collection" means that almost every effort to collect standardized information from the public, including information from grantees spending Federal funds, triggers the Information Collection Request review and approval process.

The second issue is the length of time required to gain approval for an information collection. Delays prior to starting data collection may be

27. Paperwork Reduction Act, 44 USC § 3501-3520 (1995).

28. The 1995 amendments to the PRA used the phrase "senior official," which was later changed to Chief Information Officer in the Clinger-Cohen Act (Public Law 104-106, February 10, 1996; Public Law 104-208, September 30, 1996).

29. Stuart Shapiro, "The Paperwork Reduction Act: Benefits, Costs, and Directions for Reform," *Government Information Quarterly* 30 (2013): 204-210.

30. Letter submitted to the Commission by the Laura and John Arnold Foundation.

compounded by Institutional Review Board reviews also required for certain types of research. These delays often mean that the window of opportunity passes for collecting critical baseline data from participants in a new program and also can inhibit the ability of agencies to gather real-time information about program operations. As Katherine O'Regan, former Assistant Secretary at HUD, noted in her testimony to the Commission, if the collection of such information cannot take place in a timely manner, agencies are "left to make policy decisions and program changes with very little information; often reacting to anecdote rather than a more complete picture."³¹

Within OMB, OIRA is tasked with the review and approval of new information collections. Because of the high priority of the regulatory function within OIRA, the swift review of new collections often is not a priority. The statutory language of the PRA intentionally casts a broad net to ensure that the majority of information collections undertaken by the Federal government are included in the required review and approval process. However, flexibilities currently exist within the PRA statute that could be appropriately applied for evidence building activities. For example, the PRA provides OMB with the ability to delegate some authority to approve proposed collections of information.³² Exploring opportunities to delegate authorities could create an opportunity to focus additional attention within OMB on other purposes and benefits of the PRA, such as coordinating information collections across government and improving quality.

Challenge #5: The Federal evidence-building community has insufficient resources and limited flexibilities that restrict the ability to expand evidence-building activities.

Through the course of the Commission's research and deliberations, the topic of resources repeatedly emerged as a major perceived need for the evidence-building community and a challenge for improving the volume and quality of evidence

produced. The Commission believes that a responsible investment of resources in more and better evidence holds the potential to yield substantial savings in the longer term as programs that are improved become more cost-effective, and as programs that are not effective are discontinued.

“Using evidence to improve government is what taxpayers expect—smart and careful use of limited resources to best address national priorities.”

– Analytical Perspectives, Budget of the United States Government, Fiscal Year 2018

The Commission recognizes that resource prioritization is essential to ensuring the goals of the Commission are achieved, but it is not recommending an infusion of large sums of funding to create new agencies or to launch massive new evidence-building endeavors. Instead, throughout this report, the recommendations of the Commission balance the need to prioritize evidence building while recognizing fiscal constraints. In some departments, sufficient resources already exist to enable evidence building, though such resources may have use restrictions that inhibit the most cost-effective approach for evidence building. In several instances shared with the Commission, restrictive funding procedures appear to have hampered or terminated otherwise positive evidence-building practices. Specifically, with regard to the procurement challenges discussed previously, the lack of funds that are available across multiple years is one considerable barrier to carrying out long-term projects. Evidence-building activities are generally today considered “non-severable services” under Federal appropriations law, which means that they must be fully funded upfront or structured as a multi-year contract. Departments seldom have sufficient funding to fully fund the entire cost of large or complex studies up front; rather, departments must break those contracts into phases. Doing so, however, requires considerable extra staff time and additional costs to the government and its contractors. Forcing departments to issue multiple contracts to conduct a single evaluation or research project distracts from efficiently implementing projects aimed at generating more and better evidence.

31. Katherine O'Regan, HUD, Commission Meeting, Washington, D.C., November 4, 2016.


32. Paperwork Reduction Act, 44 USC § 3507(i)(1) (1995).

Some Federal departments struggle to identify resources that can be made available for evidence-building purposes. Flexibilities in funding have allowed other departments to pursue evidence-building activities within existing resources. Several departments have the legal authority to set-aside, or allocate, a set amount of funding for their evidence-building budgets that can be transferred across budget accounts. The explicit transfer authority in appropriations is one approach to reducing the need for burdensome efforts to justify budget transfer under general authorities, such as the Economy Act. DOL, for example, receives funding available for multiple years for evaluation through the set-aside and transfer authority in its appropriation, with a requirement to notify appropriations committees about evaluations to be conducted.³³ The Department of Justice and HHS also have programs that include small funding set-asides for evidence building. Agencies like the Social Security Administration create a reliable funding stream for information technology improvements through set-asides. The same strategy could be used to support the full suite of evidence-building activities, including data collection and curation, policy-relevant research, and evaluation.

Recommendations

The Commission identified a series of opportunities to *build and maintain a strong Federal infrastructure for the sustained production and use of evidence*. To maximize the generation of evidence, the Federal evidence-building community must support a full range of analytic approaches, including statistics, evaluation, and policy research. Such expertise may be established within Federal departments, and may be strengthened through partnerships with the non-Federal evidence-building community. To achieve the greatest gains, evidence building must be well-coordinated both within and across departments. Strong leadership that prioritizes evidence building and creates the demand for evidence is vital for institutionalizing these functions within departments and ensuring coordination across the evidence-building community. In addition, administrative functions must be aligned and tailored to support, rather

than hinder, evidence-building and sufficient resources and flexibilities must be provided to support the entire enterprise. To build capacity across the Federal evidence-building community, the Commission identified five recommendations:

 **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.**


Ensuring the routine and continuous production of evidence to support policy decisions and program administration will require the capacity to leverage each evidence-building function—statistics, evaluation, and policy research—within Federal departments. Today, some combination of these functions exists in all major departments, including through collaborations with non-governmental partners, but these functions are not necessarily operating at the level needed. This recommendation directs Federal departments to establish the capacity to undertake the full range of evidence-building activities through internal human resource strategies and by leveraging partnerships with external partners. Federal departments should also conduct a regular inventory of units that perform evidence-building functions across the department as part of their strategic planning process.

To advance the goal of strengthening the government's more nascent program evaluation function, the Commission recommends that Federal departments identify or establish a Chief Evaluation Officer. A Chief Evaluation Officer can help lead efforts to coordinate the department's evidence-building activities. Further, OPM should take immediate steps to support the growth and development of the program evaluation field, including the establishment of occupational structures to address critical needs regarding technical expertise and contract management. One approach to address these staffing challenges would be for OPM to collaborate with agencies that already have high-functioning program evaluation offices to identify human capital strategies to support the development of the evaluation field. For newer evaluation offices that may be struggling to grow

33. Division H, Section 107, Consolidated Appropriations Act, 2017 (Public Law 115-31).

their capacity, OPM could identify hiring strategies focused on recruiting and retaining professionals with the specialized skills needed to directly execute evaluations on behalf of the Federal government or manage evaluation contracts. Such strategies might include the establishment of a targeted occupational series that would allow departments flexibility to hire based on specific criteria most relevant for the position and the function.

Tools and mechanisms to leverage external partnerships and build agency capacity should also be made available to Federal agencies in the evidence-building community, including grant-making authority and cooperative agreement authority, to encourage the use of partnerships with foundations, universities, and others. The President should also encourage the increased use of programs, such as the IPA program, that allow agencies to bring on temporary personnel from outside the Federal government to help expand capacity and advance the department's learning agenda.

 **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.**


The supply of evidence to support policymaking is more likely to increase when there are consistent signals from policymakers that the production of evidence is a priority. The Congress and the President can take steps to encourage departments to develop more and better evidence. As a starting point, the Congress and the President should encourage Federal departments to inventory the units responsible for various evidence-building activities to ensure that each unit is recognized and operating in concert with other units engaging in evidence-building activities. Such an inventory could be developed through the regular quadrennial strategic planning process.

The President should encourage Federal departments to develop multi-year learning agendas in consultation with program and evidence-building units. The learning agenda should be updated on a regular basis and contain both short- and long-term evidence-building priorities for each department. Through the development of learning agendas, departments should proactively identify where the need for more evidence is greatest. Developing

learning agendas through routine strategic planning reduces the burden for identifying knowledge gaps and helps align evidence building with departmental strategic priorities. The resulting document also helps to clearly communicate the high-priority research and policy questions of a department both to the public and to the many different actors within the evidence-building community.

The Congress and the President should provide sufficient and appropriate authority for departments to design programs and policies that enable a portfolio of evidence to support continuous learning and information needed to ensure accountability. The Commission acknowledges the value of research and evaluation that takes place at various points throughout the lifecycle of a program, as well as the broad variety of research methods that may be employed to address different types of research questions. When appropriate and feasible, randomized controlled trials should be conducted to understand program impacts.

Federal departments should be encouraged to routinely evaluate programs and policies, including their cost effectiveness. Continuous improvement can be facilitated by supporting a cycle of first pilot testing a new program, policy, or regulation, then conducting research to learn from the pilot test, and finally, adapting the program, policy, or regulation based on what was learned through the research. In establishing new programs in law or reauthorizing existing programs, the Congress and the President should strongly encourage that pilot and demonstration projects are evaluated, and provide flexibility to design policies that allow experimentation, including phased implementation options.

 **REC. 5-3: The Congress and the President should direct the Office of Management and Budget (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.**

The Commission recognizes the President's and OMB's prerogative in organizing and optimizing OMB's structure and resources. Implementing a government-wide vision for evidence building

consistent with the Commission's vision, though, may require substantial changes in how government operates. The Commission believes that having this effort championed at OMB is critical. Many of the Commission's recommendations rely implicitly on the integration of evidence building functions within departments. Different components of the evidence-building community, including statistical agencies, evaluation offices, privacy offices, and performance management units must increasingly work together to realize the Commission's vision. Because many department structures for implementing evidence-building activities mirror the structures at OMB, careful consideration should be given to whether a consolidation of activities necessary for evidence building should occur in the near term at OMB. A focus on evidence building can get crowded out by other priorities. The President and OMB should carefully consider how a greater commitment to foundational, critical information policy setting and coordination responsibilities and resources can be achieved, in light of the significant, high stakes and fast-moving regulatory responsibilities within OIRA, responsibilities that currently comprise the overwhelming majority of OIRA's work. This includes any reorganizations or required statutory changes that could improve the impact of OMB's investment in evidence-based policymaking.

Efficiently implementing evidence-building activities across government requires a strong coordination function to address cross-cutting research and policy questions, minimize duplicative efforts, and reduce the burden on the public. For the Federal government to maximize resources available for evidence building across government, OMB must provide leadership that spans the Executive Branch. In addition, OMB's role in leading many interagency councils must be strengthened to ensure these efforts are appropriately visible and influential in improving how government agencies collaborate.



REC. 5-4: The Congress and the President should align administrative processes to support evidence building, in particular by streamlining the approval processes for new data collections and using existing flexibilities in procurement policy.

To generate a greater volume of evidence in a more efficient manner, foundational administrative processes must be aligned and tailored to better support evidence building. The Commission identified a specific set of actions related to procurement and streamlining the review and approval processes for new data collections that would require little cost, but offer substantial benefits and savings while making it easier to produce evidence. Specifically, OMB should clarify the applicability of the research and development procurement policies to contracts that support evidence-building activities, and should support the establishment of training for technical staff who are responsible for managing research and development contracts that support evidence building. In partnership with this approach, training should be developed for technical staff responsible for managing evidence-building contracts, including effective management techniques for oversight and accountability of riskier service contracts. An additional strategy would be to establish a Government-wide Acquisition Contract (GWAC) for use in issuing contracts for evidence-building activities and services.

The Congress and the President also should take steps to streamline the PRA's review and approval process for Information Collection Requests, while ensuring that information coordination, transparency, and data quality remain central features of PRA. OMB should pilot approaches for better meeting the needs of the evidence-building community in conducting ICR reviews, including expanding the use of the delegated authority already authorized under the PRA. The process of review and approval internal to Federal departments might also be strengthened by removing the departmental CIO from the ICR review and approval process and instead assigning this responsibility to the senior data policy official proposed in Recommendation 3-3. At a minimum, OMB should issue guidance to Federal agencies describing the flexibilities under the PRA, including the process for acquiring delegated approval authority, and strengthening departmental capabilities to design rigorous studies to advance evidence building.

The Congress should consider additional statutory changes to the PRA to streamline the review process. These include assessing the rationale for the current standard under the PRA that requires

the review and approval of all new data collections of 10 or more respondents, and considering shortening the initial public comment period for new collections from 60 days to 30 days.



REC. 5-5: The Congress and the President should ensure sufficient resources to support evidence-building activities about Federal government programs and policies.

Insufficient resources for evidence building can be perceived as an insurmountable limitation to the pursuit of more and better evidence about government programs and policies. There are several actions that should be taken by the Congress and the President to signal the importance of evidence building within Federal departments, and set the expectation for the production and application of evidence through the provision of flexibilities or a commitment of resources.

The Congress, through the appropriations process, and the President should provide departments in active pursuit of a learning agenda access to multi-year funding to pursue articulated evidence-building priorities. The availability of multi-year funds for departments that establish learning agendas both creates an incentive for the generation of learning agendas and enables departments to more efficiently contract for multiyear studies.³⁴

Appropriators can further support evidence building that often spans the domain of multiple policies and jurisdictional silos by enabling the transfer of funding across the evidence-building community. The Congress and the President should enable and encourage transfers across budget accounts that support multi-departmental evidence-building needs. Explicit transfer authority in appropriation bills will enable departments to spend less time justifying transfers under general authorities (e.g., the Economy Act), and enable them to focus on the production of evidence.

In establishing new programs in law or reauthorizing existing programs, the Congress and the President should enable the use of new set-aside authorities of up to 1 percent of program administration resources to support the full suite of evidence-building activities, including data collection and curation, policy-relevant research, and evaluation. In addition, the Congress and the President should establish and grant Federal departments access to Evidence Incentive Funds to supplement the production of future research, evaluation, and related activities identified in departmental learning agendas. Evidence Incentive Funds in each department are conceptualized by the Commission to operate similarly to Working Capital Funds or Salary and Expense accounts. The funds could be created by taking up to 10 percent of unobligated balances at the end of a fiscal year to be allocated for future evidence-generating activities. ■

34. Interagency Council on Evaluation Policy, Comments to the Commission, “‘Top-Five List’ of Issues and Solutions Related to Federal Evaluation Activity,” November 22, 2016.

6

Conclusion: Possibilities with More and Better Evidence

The Commission's recommendations present a comprehensive strategy for addressing the greatest problems facing evidence building today: unintentional limits on data access, inadequate privacy practices, and insufficient capacity to generate the amount of quality evidence needed to support policy decisions. The Congress, the President, and the American people are ill-served by this state of affairs. The Commission believes that fully implementing the Commission's recommendations will lead to substantial progress in addressing these challenges, enabling more and better evidence for our society, generated in a more secure fashion.

Over the past century, numerous commissions and panels offered strategies to improve the U.S. evidence-building system. The changes made in response to their recommendations have not kept pace with the need. Twenty years ago, many of the recommendations included in this Commission's strategy could not have been feasibly implemented. In this report, the Commission proposes a modernization of the country's evidence-producing capacity that uses available approaches and that incorporates new technologies and methods as they come on line. This strategy will enrich the capabilities for producing evidence in the U.S. and ensure a formidable ability to use evidence to improve government policies.

The Commission is confident that, with full and ongoing implementation of the Commission's recommendations, privacy-protective approaches will continuously improve to ensure the confidentiality of individuals' information

in all evidence-building activities. The risks of re-identifying individuals in data used for evidence building will be continuously reviewed and mitigated. The evidence-building community will have avenues for secure access to the information they need to answer questions from policymakers, program administrators, and the public. The institutions of government will further enhance their systems for stewarding data and enabling evidence building. Principal Statistical Agencies will take on increased leadership within the Federal government in facilitating the secure statistical use of administrative data. Program offices within the Federal government, state officials, and the public will demand and use statistical analysis to support routine decision-making. Perhaps most importantly, there will be greater accountability and transparency about evidence-building approaches and uses of data for the Congress, the President, and the American public.

In developing recommendations, the Commission relied on five guiding principles related to the concepts of privacy, rigor, transparency, humility, and capacity (described in Chapter 1). Each of these principles is reflected in the recommendations developed by the Commission for building and using evidence to improve our society. The principle for respecting individual privacy and confidentiality is captured in virtually every recommendation in this report, particularly those in Chapters 2, 3, and 4 that seek to improve privacy protections. The principle of rigor encourages well-designed and well-implemented methods; improved access to administrative data to answer

important research questions, as recommended in Chapters 2 and 3, advances this principle. The recommendations in Chapter 5 concerning learning agendas and portfolios of evidence as tools for addressing policymakers' questions are grounded in the principle of humility in the development and use of evidence. In all of the recommendations, the Commission encourages increased transparency about evidence-building efforts, particularly with the new transparency portal recommended in Chapter 4. In addition to the creation of the National Secure Data Service which will enable greater capacity for the entire evidence-building community, Chapter 5 includes numerous recommendations that seek to improve government's capacity for evidence building while encouraging the use of information.

The Commission's recommendations recognize the complexity of improving evidence about government programs and policies. The Commission's recommendations seek to create an environment that enables routine production of evidence to meet the country's informational needs. And the Commission's recommendations embody the spirit of bipartisan cooperation.

People need credible information to inform their actions. We review ratings on websites before making online purchases. We inspect homes before moving in. We seek second opinions before major surgeries. Businesses rely on data and analysis to make their decisions. Why should govern-

ment spending be treated differently? Taxpayers and policymakers should receive credible information to know and understand how well the programs and policies they fund achieve their intended goals.

Generating and using evidence to inform policymaking and program administration is not a partisan issue. The strategy described in this report offers a responsible approach to improving how government officials, private researchers, foundations, non-profits, the business community, and the public interact to make sure government delivers on its promises.

All of these recommendations will depend on the leadership of the President and the Congress in calling for credible evidence to support policy decisions throughout government. Whether making decisions on funding allocations, assessing new regulations, or understanding how to improve processes for efficiently providing services, evidence is needed in every decision made by government officials—career civil servant, political appointee, or elected official. Without the use of evidence in our democracy, we are only guessing at whether government programs and policies are achieving their intended goals.

This is a milestone moment. The Congress and the President should seize it by working together to enact the laws and develop the regulations necessary to implement the Commission's recommendations. ■

Acronyms

ACF: Administration for Children and Families (U.S. Department of Health and Human Services)

ASPE: Office of the Assistant Secretary for Planning and Evaluation (U.S. Department of Health and Human Services)

CARRA: Center for Administrative Records Research and Applications (Census Bureau, U.S. Department of Commerce)

CASD: Centre d'Accès Sécurisé aux Données ("Secure Access Data Center" – France)

CEP: U.S. Commission on Evidence-Based Policymaking

CES: Center for Economic Studies (Census Bureau, U.S. Department of Commerce)

CIO: Chief Information Officer

CIPSEA: Confidential Information Protection and Statistical Efficiency Act of 2002

CMS: Centers for Medicare and Medicaid Services (U.S. Department of Health and Human Services)

DARE: Drug Abuse Resistance Education

DOL: U.S. Department of Labor

ED: U.S. Department of Education

ED-DRB: disclosure review board at the U.S. Department of Education

eMOU: Enterprise Memorandum of Understanding

FIPPs: Fair Information Practice Principles

FITARA: Federal Information Technology Acquisition Reform Act

FNS: Food and Nutrition Service (U.S. Department of Agriculture)

FSRDC: Federal Statistical Research Data Center

FTI: Federal tax information

GPRA: Government Performance and Results Act of 1993

GWAC: Government-wide Acquisition Contract

HHS: U.S. Department of Health and Human Services

HUD: U.S. Department of Housing and Urban Development

ICR: Information Collection Request

IES: Institute of Education Sciences (U.S. Department of Education)

IPA: Intergovernment Personnel Act

LEHD: Longitudinal Employer-Household Dynamics Program

NCES: National Center for Education Statistics (U.S. Department of Education)

NDNH: National Directory of New Hires

NSDS: National Secure Data Service (proposed by recommendations)

NSF: National Science Foundation

OIRA: Office of Information and Regulatory Affairs (Office of Management and Budget, Executive Office of the President)

OMB: U.S. Office of Management and Budget (Executive Office of the President)

OPM: U.S. Office of Personnel Management

OPRE: Office of Planning, Research, and Evaluation (Administration for Children and Families, U.S. Department of Health and Human Services)

PII: Personally identifiable information

PRA: Paperwork Reduction Act of 1995

PSA: Principal statistical agency

SMC: Secure Multiparty Computation

SNAP: Supplemental Nutrition Assistance Program

TANF: Temporary Assistance for Needy Families

UI: Unemployment Insurance

USDA: U.S. Department of Agriculture

VA: U.S. Department of Veterans Affairs

WIC: Special Supplemental Nutrition Program for Women, Infants, and Children

WIOA: Workforce Innovation and Opportunity Act of 2014

This page intentionally left blank

Appendix A: Evidence-Based Policymaking Commission Act of 2016



PUBLIC LAW 114–140—MAR. 30, 2016

130 STAT. 317

Public Law 114–140
114th Congress

An Act

To establish the Commission on Evidence-Based Policymaking, and for other purposes.

Mar. 30, 2016
[H.R. 1831]

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the “Evidence-Based Policymaking Commission Act of 2016”.

Evidence-Based
Policymaking
Commission Act
of 2016.

SEC. 2. ESTABLISHMENT.

There is established in the executive branch a commission to be known as the “Commission on Evidence-Based Policymaking” (in this Act referred to as the “Commission”).

SEC. 3. MEMBERS OF THE COMMISSION.

(a) NUMBER AND APPOINTMENT.—The Commission shall be comprised of 15 members as follows:

- (1) Three shall be appointed by the President, of whom—
 - (A) one shall be an academic researcher, data expert, or have experience in administering programs;
 - (B) one shall be an expert in protecting personally-identifiable information and data minimization; and
 - (C) one shall be the Director of the Office of Management and Budget (or the Director’s designee).
- (2) Three shall be appointed by the Speaker of the House of Representatives, of whom—
 - (A) two shall be academic researchers, data experts, or have experience in administering programs; and
 - (B) one shall be an expert in protecting personally-identifiable information and data minimization.
- (3) Three shall be appointed by the Minority Leader of the House of Representatives, of whom—
 - (A) two shall be academic researchers, data experts, or have experience in administering programs; and
 - (B) one shall be an expert in protecting personally-identifiable information and data minimization.
- (4) Three shall be appointed by the Majority Leader of the Senate, of whom—
 - (A) two shall be academic researchers, data experts, or have experience in administering programs; and
 - (B) one shall be an expert in protecting personally-identifiable information and data minimization.
- (5) Three shall be appointed by the Minority Leader of the Senate, of whom—

President.

130 STAT. 318

PUBLIC LAW 114–140—MAR. 30, 2016

(A) two shall be academic researchers, data experts, or have experience in administering programs; and

(B) one shall be an expert in protecting personally-identifiable information and data minimization.

(b) EXPERTISE.—In making appointments under this section, consideration should be given to individuals with expertise in economics, statistics, program evaluation, data security, confidentiality, or database management.

President.

(c) CHAIRPERSON AND CO-CHAIRPERSON.—The President shall select the chairperson of the Commission and the Speaker of the House of Representatives shall select the co-chairperson.

Deadline.

(d) TIMING OF APPOINTMENTS.—Appointments to the Commission shall be made not later than 45 days after the date of enactment of this Act.

(e) TERMS; VACANCIES.—Each member shall be appointed for the duration of the Commission. Any vacancy in the Commission shall not affect its powers, and shall be filled in the manner in which the original appointment was made.

(f) COMPENSATION.—Members of the Commission shall serve without pay.

(g) TRAVEL EXPENSES.—Each member of the Commission shall be allowed travel expenses, including per diem in lieu of subsistence, at rates authorized for employees of agencies under subchapter I of chapter 57 of title 5, United States Code, while away from their homes or regular places of business in the performance of services for the Commission.

SEC. 4. DUTIES OF THE COMMISSION.

Recommendations.

(a) STUDY OF DATA.—The Commission shall conduct a comprehensive study of the data inventory, data infrastructure, database security, and statistical protocols related to Federal policymaking and the agencies responsible for maintaining that data to—

(1) determine the optimal arrangement for which administrative data on Federal programs and tax expenditures, survey data, and related statistical data series may be integrated and made available to facilitate program evaluation, continuous improvement, policy-relevant research, and cost-benefit analyses by qualified researchers and institutions while weighing how integration might lead to the intentional or unintentional access, breach, or release of personally-identifiable information or records;

(2) make recommendations on how data infrastructure, database security, and statistical protocols should be modified to best fulfill the objectives identified in paragraph (1); and

(3) make recommendations on how best to incorporate outcomes measurement, institutionalize randomized controlled trials, and rigorous impact analysis into program design.

(b) CLEARINGHOUSE.—In undertaking the study required by subsection (a), the Commission shall—

(1) consider whether a clearinghouse for program and survey data should be established and how to create such a clearinghouse; and

Evaluation.

(2) evaluate—

(A) what administrative data and survey data are relevant for program evaluation and Federal policy-making and should be included in a potential clearinghouse;

PUBLIC LAW 114–140—MAR. 30, 2016

130 STAT. 319

(B) which survey data the administrative data identified in subparagraph (A) may be linked to, in addition to linkages across administrative data series, including the effect such linkages may have on the security of those data;

(C) what are the legal and administrative barriers to including or linking these data series;

(D) what data-sharing infrastructure should be used to facilitate data merging and access for research purposes;

(E) how a clearinghouse could be self-funded;

(F) which types of researchers, officials, and institutions should have access to data and what the qualifications of the researchers, officials, and institutions should be;

(G) what limitations should be placed on the use of data provided;

(H) how to protect information and ensure individual privacy and confidentiality;

(I) how data and results of research can be used to inform program administrators and policymakers to improve program design;

(J) what incentives may facilitate interagency sharing of information to improve programmatic effectiveness and enhance data accuracy and comprehensiveness; and

(K) how individuals whose data are used should be notified of its usages.

(c) REPORT.—Upon the affirmative vote of at least three-quarters of the members of the Commission, the Commission shall submit to the President and Congress a detailed statement of its findings and conclusions as a result of the activities required by subsections (a) and (b), together with its recommendations for such legislation or administrative actions as the Commission considers appropriate in light of the results of the study.

Recommendations.

(d) DEADLINE.—The report under subsection (c) shall be submitted not later than the date that is 15 months after the date a majority of the members of the Commission are appointed pursuant to section 3.

(e) DEFINITION.—In this section, the term “administrative data” means data—

- (1) held by an agency or a contractor or grantee of an agency (including a State or unit of local government); and
- (2) collected for other than statistical purposes.

SEC. 5. OPERATION AND POWERS OF THE COMMISSION.

(a) EXECUTIVE BRANCH ASSISTANCE.—The heads of the following agencies shall advise and consult with the Commission on matters within their respective areas of responsibility:

Consultation.

- (1) The Bureau of the Census.
- (2) The Internal Revenue Service.
- (3) The Department of Health and Human Services.
- (4) The Department of Agriculture.
- (5) The Department of Housing and Urban Development.
- (6) The Social Security Administration.
- (7) The Department of Education.
- (8) The Department of Justice.
- (9) The Office of Management and Budget.
- (10) The Bureau of Economic Analysis.
- (11) The Bureau of Labor Statistics.

130 STAT. 320

PUBLIC LAW 114–140—MAR. 30, 2016

Deadline. (12) Any other agency, as determined by the Commission.
 (b) MEETINGS.—The Commission shall meet not later than 30 days after the date upon which a majority of its members have been appointed and at such times thereafter as the chairperson or co-chairperson shall determine.

(c) RULES OF PROCEDURE.—The chairperson and co-chairperson shall, with the approval of a majority of the members of the Commission, establish written rules of procedure for the Commission, which shall include a quorum requirement to conduct the business of the Commission.

(d) HEARINGS.—The Commission may, for the purpose of carrying out this Act, hold hearings, sit and act at times and places, take testimony, and receive evidence as the Commission considers appropriate.

(e) CONTRACTS.—The Commission may contract with and compensate government and private agencies or persons for any purpose necessary to enable it to carry out this Act.

(f) MAILS.—The Commission may use the United States mails in the same manner and under the same conditions as other agencies of the Federal Government.

(g) GIFTS.—The Commission may accept, use, and dispose of gifts or donations of services or property.

SEC. 6. FUNDING.

(a) IN GENERAL.—Subject to subsection (b) and the availability of appropriations—

(1) at the request of the Director of the Census, the agencies identified as “Principal Statistical Agencies” in the report, published by the Office of Management and Budget, entitled “Statistical Programs of the United States Government, Fiscal Year 2015” shall transfer funds, as specified in advance in appropriations Acts and in a total amount not to exceed \$3,000,000, to the Bureau of the Census for purposes of carrying out the activities of the Commission as provided in this Act; and

(2) the Bureau of the Census shall provide administrative support to the Commission, which may include providing physical space at, and access to, the headquarters of the Bureau of the Census, located in Suitland, Maryland.

(b) PROHIBITION ON NEW FUNDING.—No additional funds are authorized to be appropriated to carry out this Act. This Act shall be carried out using amounts otherwise available for the Bureau of the Census or the agencies described in subsection (a)(1).

SEC. 7. PERSONNEL.

Appointment. (a) DIRECTOR.—The Commission shall have a Director who shall be appointed by the chairperson with the concurrence of the co-chairperson. The Director shall be paid at a rate of pay established by the chairperson and co-chairperson, not to exceed the annual rate of basic pay payable for level V of the Executive Schedule (section 5316 of title 5, United States Code).

(b) STAFF.—The Director may appoint and fix the pay of additional staff as the Director considers appropriate.

(c) EXPERTS AND CONSULTANTS.—The Commission may procure temporary and intermittent services under section 3109(b) of title 5, United States Code, at rates for individuals which do not exceed the daily equivalent of the annual rate of basic pay for a comparable position paid under the General Schedule.

PUBLIC LAW 114–140—MAR. 30, 2016

130 STAT. 321

SEC. 8. TERMINATION.

The Commission shall terminate not later than 18 months after the date of enactment of this Act.

Approved March 30, 2016.

LEGISLATIVE HISTORY—H.R. 1831 (S. 991):**HOUSE REPORTS:** No. 114–211 (Comm. on Oversight and Government Reform).**SENATE REPORTS:** No. 114–151 (Comm. on Homeland Security and Governmental Affairs) accompanying S. 991.**CONGRESSIONAL RECORD:**

Vol. 161 (2015): July 27, considered and passed House.

Vol. 162 (2016): Mar. 16, considered and passed Senate, amended.

Mar. 17, House concurred in Senate amendment.



This page intentionally left blank

Appendix B: Commissioner Biographies

Katharine G. Abraham

University of Maryland

Commissioner and Chair

Katharine G. Abraham is Professor of Economics and Survey Methodology at the University of Maryland. She has written extensively on the effects of labor market policies and institutions on firm and worker behavior, labor market adjustment over the business cycle, and the measurement of economic activity. Abraham first joined the University of Maryland faculty in 1987, after holding prior positions at the Brookings Institution from 1985 to 1987 and at the Massachusetts Institute of Technology's Sloan School of Management from 1980 to 1985. From 2011 to 2013, Abraham served as a Member of the Council of Economic Advisers and from 1993 to 2001, she served as Commissioner of the Bureau of Labor Statistics. Abraham is a Fellow of the Society of Labor Economists and of the American Statistical Association and has served as a Vice President of the American Economic Association. Abraham received a Bachelor's degree from Iowa State University and a Ph.D. from Harvard University.

Ron Haskins

Brookings Institution

Commissioner and Co-Chair

Ron Haskins is a Senior Fellow and holds the Cabot Family Chair in Economic Studies at the Brookings Institution, where he co-directs the Center on Children and Families. Haskins is also a senior consultant at the Annie E. Casey Foundation and is Past-President of the Association for Public Policy Analysis and Management. Prior to joining Brookings and Casey, he spent 14 years on the staff of the House Ways and Means Human Resources Subcommittee, serving as the subcommittee's Staff Director after Republicans became the majority party in the House after the 1994 elections. He holds a Bachelor's degree in History, a Master of Arts in Teaching, and a Ph.D. in Developmental Psychology from the University of North Carolina at Chapel Hill. In his Washington career, he has focused on evidence-based policy, early childhood education, marriage and family formation, poverty, equal opportunity, abused and neglected children, and budget issues.

Sherry Glied

New York University

Commissioner

In August 2013, Sherry Glied became Dean of New York University's Robert F. Wagner Graduate School of Public Service. From 1989-2013, she was Professor of Health Policy and Management at Columbia University's Mailman School of Public Health. She was Chair of the department from 1998-2009. On June 22, 2010, Glied was confirmed by the U.S. Senate as Assistant Secretary for Planning and Evaluation at the U.S. Department of Health and Human Services, and served in that capacity from July 2010 through August 2012. She has previously served as Senior Economist for health care and labor market policy on the Council of Economic Advisers in 1992-1993, under Presidents Bush and Clinton. Glied's principal areas of research are in health policy reform and mental health care policy.

Robert M. Groves

Georgetown University

Commissioner

Robert M. Groves is the Provost of Georgetown University and the Gerard J. Campbell, S.J. Professor in the Math and Statistics Department as well as the Sociology Department. He served as director of the U.S. Census Bureau during the conduct of the 2010 Census of Population. Groves studies how to improve large-scale sample surveys and censuses. His research has focused on the impact of mode of data collection on responses in sample surveys, the influences on survey participation, the use of adaptive research designs to improve the cost and error properties of statistics, and public concerns about privacy affecting attitudes toward statistical agencies. Groves is an elected member of the National Academy of Sciences and the National Academy of Medicine. He is also the chair of the Committee on National Statistics at the National Academies of Sciences, Engineering, and Medicine. He is an appointed member of the National Science Board, overseeing the National Science Foundation, as well as the Board of the Pew Research Center.

Robert Hahn

University of Oxford

Commissioner

Robert Hahn is Professor and Director of Economics at the Smith School of Enterprise and the Environment at the University of Oxford; a senior fellow at the Institute for New Economic Thinking at Oxford; a senior fellow at the Georgetown University Center for Business and Public Policy, and a non-resident senior fellow at the Brookings Institution. Hahn worked at American Enterprise Institute for two decades, where he co-founded and directed the AEI-Brookings Joint Center. Previously, Hahn worked for the Council of Economic Advisers and was the chief economist on the White House drafting team for the 1990 Clean Air Act Amendments. He also has served on the faculties of Harvard University and Carnegie Mellon University. Hahn is currently conducting several behavioral economics experiments aimed at encouraging the conservation of energy and water resources, improving labor productivity, improving health outcomes, and understanding the welfare benefits of new technologies. In addition, he co-founded the Community Preparatory School, an inner-city middle school in Providence, Rhode Island, that provides opportunities for disadvantaged youth to achieve their full potential.

Hilary Hoynes

University of California, Berkeley

Commissioner

Hilary Hoynes is a Professor of Economics and Public Policy and holds the Haas Distinguished Chair in Economic Disparities at the University of California, Berkeley. From 2011 to 2016, she was a co-editor of the leading journal in economics, the *American Economic Review*. Hoynes specializes in the study of poverty, inequality, food and nutrition programs, and the impacts of government tax and transfer programs on low-income families. Current projects include evaluating the effects of access to the social safety net in early life on later life health and human capital outcomes, examining the effects of the Great Recession on poverty, and the role of the safety net in mitigating income losses. Professor Hoynes is a member of the American Economic Association's Executive Committee and a panel member of the Committee on National Statistics at the National Academies of Sciences, Engineering, and Medicine on Building an Agenda to Reduce the Number of Children in

Poverty by Half in 10 Years. Previously, she was a member of the Advisory Committee for Social, Behavioral and Economic Sciences for the National Science Foundation and the National Advisory Committee of the Robert Wood Johnson Foundation Scholars in Health Policy Research Program.

Jeffrey Liebman

Harvard University

Commissioner

Jeffrey Liebman is the Malcolm Wiener Professor of Public Policy at the Harvard Kennedy School where he directs the Taubman Center for State and Local Government and the Rappaport Institute for Greater Boston. Liebman teaches courses in social policy, public sector economics, American economic policy, and public sector management. In his research, he studies tax and budget policy, social insurance, and public sector innovation. During the first two years of the Obama Administration, Liebman served at the U.S. Office of Management and Budget, first as Executive Associate Director and Chief Economist and then as Acting Deputy Director. From 1998 to 1999, Liebman served as Special Assistant to the President for economic policy and coordinated the Clinton Administration's Social Security reform technical working group. For the past five years, his Harvard Kennedy School Government Performance Lab has been providing pro bono technical assistance to state and local governments interested in improving the results they achieve for their citizens.

Bruce D. Meyer

University of Chicago

Commissioner

Bruce D. Meyer is the McCormick Foundation Professor at the Harris School of Public Policy at the University of Chicago where he has been since 2004. From 1987 to 2004, Meyer was a professor in the Economics Department at Northwestern University. He is also a Research Associate of the National Bureau of Economic Research and a Visiting Scholar at the American Enterprise Institute. He studies poverty and inequality, tax policy, government safety net programs such as unemployment insurance, workers' compensation, food stamps, and Medicaid, and the accuracy of household surveys. His most recent work includes research on trends in poverty and inequality, the consequences of disability, the effects

of Medicaid, and errors in household surveys. He currently serves on the Technical Advisory Committee to the Bureau of Labor Statistics, the American Economic Association Committee on Government Relations, and as an officer of the Business and Economic Statistics Section of the American Statistical Association. Meyer received his Bachelor's degree and Master's degree in economics from Northwestern University and his Ph.D. in economics from the Massachusetts Institute of Technology.

Paul Ohm

Georgetown University

Commissioner

Paul Ohm is a Professor of Law at the Georgetown University Law Center. He specializes in information privacy, computer crime law, intellectual property, and criminal procedure.

Allison B. Orris

U.S. Office of Management and Budget

Commissioner

(Departed the Commission on January 20, 2017)

Allison B. Orris served as the Associate Administrator of the Office of Information and Regulatory Affairs at the U.S. Office of Management and Budget until January 2017. Orris served in several roles at the U.S. Department of Health and Human Services Centers for Medicare & Medicaid Services from 2009 to 2014, including Senior Policy Advisor, Acting Director of the Division of State Demonstrations and Waivers, and Director of the Low Income Programs Analysis Group in the Office of Legislation. Before joining CMS, Orris was a Senior Legislative Associate at the Center on Budget and Policy Priorities from 2006 to 2009. From 2002 to 2006, she practiced law at Powell Goldstein LLP in Washington, D.C. Orris received a Bachelor's degree in History from Columbia University and a J.D. from Yale University.

Nancy Potok

U.S. Office of Management and Budget

Commissioner

(Appointed to the Commission on March 10, 2017)

Nancy Potok is Chief Statistician of the United States and Chief of the Statistical and Science Policy Branch in the U.S. Office of Management and Budget. Prior to January 2017, she served as Dep-

uty Director and Chief Operating Officer of the U.S. Census Bureau. She has also served as Deputy Under Secretary for Economic Affairs at the U.S. Department of Commerce; Principal Associate Director and CFO at the U.S. Census Bureau; Senior Vice President for Economic, Labor, and Population Studies at NORC at the University of Chicago; and Chief Operating Officer at McManis & Monsalve Associates, a business analytics consulting firm. She is an adjunct professor at the Trachtenberg School of Public Policy and Public Administration at The George Washington University. Potok is an elected Fellow of the National Academy of Public Administration. She currently serves on the Board of Directors for the Institute of Pure and Applied Mathematics at the University of California Los Angeles; the Board of Trustees for the Arthur S. Flemming Award, which recognizes outstanding achievement by federal employees with fewer than 15 years of service; and The George Washington University Trachtenberg School Advisory Board. Potok received her Ph.D. in public policy and public administration at The George Washington University, with an emphasis on program evaluation.

Kathleen Rice Mosier

Faegre Baker Daniels, LLP

Commissioner

Kathleen Rice Mosier is a Counsel at Faegre Baker Daniels, LLP, where she provides advice and assistance to organizations on issues relating to privacy, data security, risk management, and compliance with applicable laws and regulations. Before joining FaegreBD, she spent nearly 20 years in law enforcement and national security. She served as a Counsel on the Senate Select Committee on Intelligence, as an Assistant General Counsel for the Federal Bureau of Investigation, and as an Assistant U.S. Attorney for the Southern District of Florida. She is a frequent author and speaker on privacy and data security matters. She is also a Certified Information Privacy Professional, International Association of Privacy Professionals, and an Adjunct Professor of Law at Notre Dame Law School.

Robert Shea

Grant Thornton, LLP

Commissioner

Robert Shea is a principal and a member of the Public Sector practice at Grant Thornton, LLP. He

leads Strategy and Communications for Grant Thornton Public Sector and provides performance improvement services to international, federal, and state and local government agencies. Before joining Grant Thornton, Shea served as Associate Director for Management at the U.S. Office of Management and Budget (OMB). Before joining OMB, Shea served as counsel to the Senate Committee on Governmental Affairs, legislative director for Congressman Pete Sessions (TX), and professional staff member for the House Committee on Government Reform and Oversight.

Latanya Sweeney

Harvard University

Commissioner

Latanya Sweeney is Professor of Government and Technology in Residence at Harvard University. Her mission is to create and use technology to assess and solve societal, political and governance problems, and to teach others how to do the same. Her primary area of focus is the scientific study of technology's impact on humankind. She serves as the Editor-in-Chief of the Journal of Technology Science, Director of the Data Privacy Lab at Harvard, and Faculty Dean at Currier House. Sweeney formerly served as the Chief Technology Officer, also known as the Chief Technologist, at the U.S. Federal Trade Commission.

Kenneth R. Troske

University of Kentucky

Commissioner

Kenneth R. Troske is the Associate Dean for Graduate Programs and Outreach and Richard W. and Janis H. Furst Endowed Chair in Economics at the Gatton College of Business and Economics at the University of Kentucky, as well as a Research Fellow with the Institute for the Study of Labor (IZA) in Bonn, Germany. Troske served as a member of the Congressional Oversight Panel whose task was to assess the existing condition of America's financial markets and the regulatory system as well as to monitor the actions of the Treasury Department and financial institutions to determine if their actions are in the best interest of the American economy. His primary research areas are labor and human resource economics. He received his undergraduate degree in economics from the University of Washington in 1984 and

his Ph.D. in economics in 1992 from the University of Chicago.

Kim R. Wallin

D.K. Wallin, Ltd.

Commissioner

Kim R. Wallin is currently in private practice at her CPA firm, D.K. Wallin Ltd., which she founded in 1984. She took eight years off to be the Controller for the State of Nevada. Wallin was the first CPA to be elected to the office of Nevada State Controller in 50 years and the first Certified Management Accountant to ever hold this office. As Controller, she received the Association of Government Accountants Presidents Award and Excellence in Government Award and the National Association of Comptrollers President's Award for her work to improve transparency in government with the use of technology. She also received the National Association of State Auditors, Comptrollers and Treasurers President's Award for her work on various U.S. Presidential work groups and for her continued efforts to improve efficiency and accountability in Government. Wallin is a graduate of the University of Nevada, Las Vegas with a degree in Business Administration with a major in accounting.

Commission Staff

Rochelle (Shelly) Wilkie Martinez,

Executive Director

H. Lucas Hitt, Deputy Executive Director

Sharon A. Boivin, Senior Policy and Research Analyst

Anne Fletcher, Senior Policy and Research Analyst

Nicholas Hart, Policy and Research Director

Michael Hawes, Privacy Consultant

Kristy L. Howell, Senior Policy and Research Analyst

Mary D. McKoy, Chief, External Affairs

Kathryn McNamara, Librarian and Records Coordinator

Sara Stefanik, Policy and Research Analyst

Robin L. Wyvill, Meetings and Events Coordinator ■

Appendix C: Commission’s Fact-Finding and Deliberative Processes

Fact-Finding Process

Following the initial appointments of commissioners, the Commission convened its first public meeting in July 2016. Over the course of the following eight months, the Commission studied issues about the current state of evidence production and use in the Federal government as well as the Federal government’s policies and practices to protect data confidentiality. The Commission’s fact-finding process included public meetings, public hearings, and meetings with organizations to which Commissioners or staff were invited or that staff initiated, a survey of Federal offices, a Request for Comments in the *Federal Register*, and other public input received through email correspondence.

Public Meetings

The Commission’s seven public meetings included a total of 49 invited witnesses.

- **July 22, 2016: Introductory Meeting**
 - o Devin O’Connor—Office of Management and Budget, Executive Office of the President (former)
 - o Jeri Mulrow—Bureau of Justice Statistics, U.S. Department of Justice
 - o John Righter—Committee on Health, Education, Labor & Pensions, U.S. Senate
 - o Mary Bohman—Economic Research Service, U.S. Department of Agriculture
 - o Nancy Potok—Census Bureau, U.S. Department of Commerce (former)
 - o Raj Chetty—Stanford University
 - o Ted McCann—Office of the Speaker, U.S. House of Representatives
- **September 9, 2016: Key Considerations in Privacy Relevant to the Commission’s Charge**
 - o Aimee Guidera—Data Quality Campaign (former)
 - o Cynthia Dwork—Microsoft Research (former)
- o Justin Erlich—California Attorney General’s Office
- o Katherine Wallman—Office of Management and Budget, Executive Office of the President (former)
- o Marc Groman—Office of Management and Budget, Executive Office of the President (former)
- o Marc Rotenberg—Electronic Privacy Information Center
- o Michael Basil—Illinois Department of Innovation & Technology
- **November 4, 2016: Considerations for the Commission Related to Evaluation**
 - o Adam Gamoran—William T. Grant Foundation
 - o Demetra Nightingale—U.S. Department of Labor (former)
 - o Evelyn Kappeler—Office of Adolescent Health, U.S. Department of Health and Human Services
 - o Jim Sullivan—University of Notre Dame
 - o Katherine O’Regan—U.S. Department of Housing and Urban Development (former)
 - o Kelly Fitzsimmons—The Edna McConnell Clark Foundation
 - o Matthew Klein—New York City Center for Economic Opportunity
 - o Naomi Goldstein—Administration for Children and Families, U.S. Department of Health and Human Services
 - o Tanya Beer—Center for Evaluation Innovation
- **December 12, 2016: Considerations for the Commission Related to Federal Models**
 - o Barry Johnson—Statistics of Income Division, Internal Revenue Service, U.S. Department of Treasury

- o David Grusky—Stanford University
 - o Erica Groshen—Bureau of Labor Statistics, U.S. Department of Labor (former)
 - o Marilyn Seastrom—National Center for Education Statistics, U.S. Department of Education
 - o Niall Brennan—Centers for Medicare and Medicaid Services, U.S. Department of Health and Human Services (former)
 - o Ron Jarmin—Census Bureau, U.S. Department of Commerce
- **January 13, 2017: International and State Models for Managing Data**
 - o Charles Rothwell—National Center for Health Statistics, U.S. Department of Health and Human Services
 - o David Mancuso—Washington State Department of Social and Health Services
 - o Domenico Parisi—Mississippi State University
 - o Ivan Thaulow—Statistics Denmark
 - o Kenneth Dodge—Duke University
 - o Robert Goerge—University of Chicago
 - o Roxane Silberman—Secure Data Access Centre, France
 - o Shawna Webster—National Association for Public Health Statistics and Information Systems
 - o Stefan Bender—Deutsche Bundesbank
 - o Tanvi Desai—Administrative Data Research Network, United Kingdom
 - **February 24, 2017: Role of Legal Standards and Technology in Maximizing Data Security and Privacy**
 - o Alexandra Wood—Harvard University
 - o Bradley Malin—Vanderbilt University
 - o Daniel Goroff—Alfred P. Sloan Foundation
 - o Jerome Reiter—Duke University
 - o Lars Vilhuber—Cornell University
 - **March 13, 2017: Federal Statistical and Partnership Infrastructure for Evidence Building – Opportunities and Limitations**
 - o Andrew Reamer—The George Washington University
 - o Brian Moyer—Bureau of Economic Analysis, U.S. Department of Commerce
 - o Erin Ulrich—Colorado Department of Public Health and Environment
 - o Maria Cancian—University of Wisconsin
 - o Seth Harris—U.S. Department of Labor (former)

Public Hearings

The Commission convened three open public hearings—in Washington, D.C., Chicago, and San Francisco—during which any member of the public who requested to testify before the Commission was allowed to present. A total of 39 members of the public presented information during the three hearings.

- **October 21, 2016: Washington, D.C.**
 - o Amanda Janice Roberson—Institute for Higher Education Policy
 - o Carrie Wofford—Veterans Education Success
 - o Christine Keller—Association of Public & Land-grant Universities
 - o Clyde Tucker—American Statistical Association
 - o Daniel Crowley—National Prevention Science Coalition to Improve Lives
 - o David Medina—Results for America
 - o Emmett McGroarty—American Principles Project
 - o Erin Knowles—United States Parents Involved in Education
 - o George Grob—American Evaluation Association
 - o Kelleen Kaye—The National Campaign to Prevent Teen and Unplanned Pregnancy
 - o Mark Schneider—American Institutes for Research
 - o Quentin Wilson—Public Performance Improvement Researcher
 - o Rachel Fishman—New America

- o Rachel Zinn—Workforce Data Quality Campaign
- o RK Paleru—Booz Allen Hamilton
- o Sara Dube—Pew-MacArthur Results First Initiative
- o Tiffany Jones—The Education Trust
- o Tom Allison—Young Invincibles
- **January 5, 2017: Chicago, IL**
 - o Cassie Creswell—Raise Your Hand Action/Parent Coalition for Student Privacy
 - o Donna K. Ginther—University of Kansas
 - o Lint Barrage—Rhode Island Innovative Policy Lab
 - o Margaret Levenstein—ICPSR
 - o Matthew Stagner—Mathematica Policy Research
 - o Quentin Palfrey—J-PAL North America
 - o Timothy Slaper—Indiana Business Research Center
 - o Tom Schenk, Jr.—City of Chicago
 - o V. Joseph Hotz—Duke University
 - o Virginia Knox—MDRC
- **February 9, 2017: San Francisco, CA**
 - o Andrew Wiegand—Social Policy Research Associates
 - o Cindy Guy—Annie E. Casey Foundation
 - o David Johnson—University of Michigan
 - o Jim Hill—Proofpoint Systems, Inc.
 - o Joy Bonaguro—City and County of San Francisco
 - o Karen Levesque—RTI International
 - o Karen R. Effrem—Education Liberty Watch
 - o Mary Ann Bates—J-PAL North America
 - o Maryann Feldman—University of North Carolina—Chapel Hill
 - o Sandra Torosian—United Parents Involved in Education
 - o Susan Dreyfus—Alliance for Strong Families and Communities

Additional Meetings

In addition, Commissioners and staff met with experts, and participated in conference panel presentations to solicit additional input from more than 40 organizations. These meetings supplemented information gathered during the Commission's fact-finding phase.

- o Actionable Intelligence for Social Policy Public—Academic Research Colloquium
- o American Evaluation Association
- o American Public Human Services Association National Summit
- o Amy O'Hara, Census Bureau, U.S. Department of Commerce (former)
- o Anna McDowell, Statistics New Zealand
- o Army Analytics Group
- o Association for Public Policy Analysis & Management
- o Association of Public Data Users Webinar
- o Bipartisan Policy Center
- o Bureau of Economic Analysis Advisory Committee
- o Census Scientific Advisory Committee
- o Center for Regional Economic Competitiveness
- o Center for the Study of Social Policy
- o Committee on Population Statistics
- o Council of Professional Associations on Federal Statistics
- o Cross Agency Learning Community on Research and Evaluation
- o Defense Manpower Data Center, U.S. Department of Defense
- o Defense Personnel and Security Research Center, U.S. Department of Defense
- o Federal Committee on Statistical Methodology Policy Conference
- o Federal Economic Statistics Advisory Committee
- o Federal Evaluators
- o Federation of Associations in Behavioral & Brain Sciences

- o Frauke Kreuter, University of Maryland
- o Health and Human Services Administrative Data Technical Expert Panel
- o Interagency Council on Evaluation Policy
- o Jenny Hunter Childs, Census Bureau, U.S. Department of Commerce
- o John Abowd, Census Bureau, U.S. Department of Commerce
- o John Mitchell, Stanford University
- o Julia Lane, New York University
- o Kathy Stack, Laura and John Arnold Foundation
- o Laura and John Arnold Foundation and Bill and Melinda Gates Foundation Workshop on the use of data
- o Liz McPherson, Statistics New Zealand
- o Mark Mazur, U.S. Department of the Treasury (former)
- o MDRC
- o Michael Duff, Stanford University
- o Michigan Retirement Research Center, Financial Research Symposium
- o Minister Hiro Matsumoto, Embassy of Japan
- o National Association of State Workforce Agencies Labor Market Information Committee
- o National Conference of State Legislatures
- o National Governors Association
- o National Institute of Standards and Technology, Privacy Risk Assessment Meeting
- o Naval Postgraduate School
- o Open Data Center Korea
- o Pew Charitable Trusts
- o Pew Results First & State Data Project Team
- o Results for America
- o Ross Young, United Kingdom Statistics Authority
- o Society for Benefit Cost Analysis Annual Conference

- o Transportation Research Board Special Task Force on Data for Decisions and Performance Measures
- o U.S. Census Bureau and Economic Research Service Administrative Data Conference
- o U.S. Department of Health and Human Services Data Council
- o Workforce Data Quality Campaign: Fly-In Conference

Other Stakeholder Input

The Commission issued a Request for Comments in the Federal Register and accepted comments by email, which generated over 350 responses from the public (see online Appendix G). Finally, to understand the current or potential capacity of Federal agencies to engage in aspects of evidence-based policymaking, the Commission administered a survey to 209 offices of the Federal government that the Commission identified as likely to be generating or using evidence (see online Appendix E).

Deliberative Process

Following the formal stakeholder input processes, the Commission compiled and reviewed the body of information collected. A series of 12 memoranda that compiled existing literature, stakeholder input, and options for recommendations were developed. Each memorandum was initially discussed in small workgroups then formally discussed by Commissioners during closed session meetings as the Commission deliberated. Multiple memoranda involved discussion during more than one meeting; the list below reflects the initial discussion points:

- **April 3, 2017**
 - o Vision Statement and Evidence Principles
 - o National Secure Data Service Implementation
 - o Data Access and Use—Legal and Statutory Issues
 - o Data Access and Use—Federal Policy and Standards
 - o Increasing Access to State—Collected Administrative Data
 - o Tiered Access with a National Secure Data Service

- **May 15, 2017**
 - o Enhancing Collaboration in the Federal Evidence Ecosystem
 - o The Role of the Federal Statistical System within the Evidence Ecosystem
 - o Protecting Privacy and Increasing Confidentiality
 - o Protecting Privacy by Increasing Transparency About Evidence Generation and Benefits
 - o Establishing a National Secure Data Service
- **June 7, 2017**
 - o Role of the National Secure Data Service and the PSAs in the Federal Evidence Ecosystem
 - o Institutionalizing and Incentivizing the Evaluation and Policy Research Functions Across Government

- **June 27, 2017**
 - o Collaborating for Evidence Building
 - o Incorporating Evaluation and Policy Research in Program Design
 - o Administration

Once the Commission developed tentative recommendations, the final two closed session meetings of the Commission were dedicated to discussing report text.

- **July 11, 2017**
- **July 20, 2017**

Commissioners collaborated on draft report text during this time. Commissioners cast digital votes of approval from July 28 to August 2. All 15 Commissioners cast votes in the affirmative approving the findings and recommendations in the report. ■

This page intentionally left blank

Appendix D: Examples of Data Productive for Evidence Building

During the course of the Commission's fact-finding phase, numerous experts and members of the public suggested specific sources of data that could be useful for evidence-building activities. Statistical uses of administrative data collected through program operations—such as enrollment and utilization information—as well as statistical data collected directly through topical surveys intended to support evidence building offer opportunities to enhance evidence-based policymaking in the United States. Secure access to data, including data in the domains listed below, is addressed throughout the Commission's report.

The data described below are not intended to reflect a comprehensive inventory of government data, but rather reflect suggestions provided to the Commission about some data that have the potential to be useful for evidence building. Numerous other administrative and survey data sources are likely to be no less valuable for evidence building in their respective policy domains. In some cases, the purposes for which data may be used are defined narrowly, limiting their use for evidence building. The Commission includes recommendations to reconsider these limitations. Recommendation 2-4 proposes a review of such statutes to ensure that limitations that preclude the use of administrative data for evidence building are applied only when the Congress and the President deem the limitations still to be necessary. In some cases, existing laws specifically prohibit the collection or analysis of information to support evidence building. Recommendation 2-5 calls for a reconsideration of such bans and restraint in the enactment of future bans. Data relevant for implementing Recommendations 2-4 and 2-5 include those indicated with an asterisk.

Administrative Data

Income, Wages, and Earnings

- **Federal Tax Information*** is collected by the Internal Revenue Service (IRS) in the Department of the Treasury and includes business and individual information. *Statutory Citation: 26 USC § 6103.*
- The **National Directory of New Hires***, maintained by the Administration for Children and Families (ACF) Office of Child Support Enforcement in the Department of Health and Human Services (HHS), includes quarterly wage information, Unemployment Insurance benefit information, and information on newly hired employees. *Statutory Citation: 42 USC § 653(j)(5).*
- **Unemployment Insurance (UI)*** quarterly wage records are collected by states as part of their administration of state UI programs operated through the Federal-state UI partnership.
- The **Wage Record Interchange System (WRIS)***, maintained by the Department of Labor, facilitates the exchange of wage data between states for performance accountability purposes, enabling improved reporting on the outcomes experienced by participants in Federally funded state employment and training programs. It allows researchers access to wage data on a state-by-state approval basis for projects with a direct benefit to these programs. The WRIS is being revised to reflect changes introduced by the Workforce Innovation and Opportunity Act, passed in 2014, and will be called the State Wage Interchange System. *Statutory Citation: 42 USC § 1111.*

Health

- **Health Services Research Program** data are collected by the Substance Abuse and Mental Health Services Administration (SAMHSA) in HHS. Data include research related to substance abuse treatment, access to mental health care, healthcare costs, and patient outcomes. *Statutory Citation: 42 USC § 290aa.*
- **Medicaid, Medicare, and Children's Health Insurance Program (CHIP)** data are collected

by the Centers for Medicare and Medicaid Services (CMS) in HHS. Data include information on Medicare claims, beneficiaries and providers, Medicaid eligibility, and claims. *Statutory Citations: 42 USC § 1396 (Medicaid and CHIP), 42 USC § 1395 (Medicare).*

- **Vital Records** administrative data are collected by the National Center for Health Statistics (NCHS) in HHS from vital registrars in all states and U.S. territories. Vital Records files include information about births and deaths. *Statutory Citation: 42 USC § 242k(h).*

Human Services

- **Child Welfare Program** data include statistics collected through the Adoption and Foster Care Analysis and Reporting System (AFCARS) and through the National Child Abuse and Neglect Data System (NCANDS). AFCARS collects case-level information on all foster children and those who have been adopted in the United States. NCANDS includes data on child abuse and neglect cases in the United States. Both data collection systems are administered by the Children's Bureau of ACF. *Statutory Citations: 42 USC § 679 (AFCARS) and 42 USC § 5119 (NCANDS).*
- **Head Start Program*** data are collected by local jurisdictions through a program funded by the ACF in HHS about program participants. *Statutory Citation: 42 USC § 9836a and 9844.*
- **Social Security Disability Insurance (SSDI)** data are collected about program length of service, benefits, eligibility, appeals, and denials through the Social Security Administration (SSA) and Disability Determination Services in each state. *Statutory Citation: 42 USC § 401.*
- **Social Security Old Age and Survivors Insurance (OASI) Program** data includes information on retired and survivor beneficiaries and benefit amounts received. *Statutory Citation: 42 USC § 401.*
- **Supplemental Nutrition Assistance Program (SNAP)** data are collected by all states through a program funded by the Department of Agriculture's Family and Nutrition Services. SNAP data include information about program eligibility, length of services, and amounts of benefits. *Statutory Citation: 7 USC § 2011 et seq.*
- **Supplemental Security Income (SSI) Program** data collected by states and SSA includes income assistance information for those participants who are needy, aged, blind, or disabled. *Statutory Citation: 42 USC § 1383.*
- **Temporary Assistance for Needy Families (TANF)** is administered by ACF's Office of Family Assistance in HHS. TANF program data includes caseloads, financial information, workforce participation rates and other program participant characteristics. *Statutory Citation: 42 USC § 611-611a.*
- **Unemployment Insurance (UI) payments** (benefits) provide temporary financial assistance to unemployed workers who meet the requirements of State law. Information about payments includes the benefits paid and duration. *Statutory Citation: 29 USC § 49b(b).*

Other

- **Business Program Participation** data, such as the small business loan program at the Small Business Administration includes information on lenders, borrowers, and loan amounts. Data collected by the Economic Development Administration reflects information about grantees use of funds to support workforce development and other regional economic opportunities. *Statutory Citations: 15 USC § 638 (Small Business Administration) and 42 USC § 3192 (Economic Development Administration).*
- **Criminal justice** data include crime incident reports and criminal history record information, or "rap sheets." Crime incident data are collected through the Federal Bureau of Investigation (FBI) National Incident-Based Reporting System (NIBRS). Data include the nature and types of specific offenses in the incident, characteristics of the victim(s) and offender(s), types and value of property stolen and recovered, and characteristics of persons arrested in

connection with a crime incident. Rap sheets are created and maintained by law enforcement agencies and accessible via the FBI Interstate Identification Index for law enforcement and other purposes. They contain information about an individual's arrests and subsequent dispositions, such as jail sentences. *Statutory Citations: 42 USC § 3732 and 3735 (NIBRS), 28 USC § 534 (criminal records).*

- **Federal Student Aid*** data are collected by the Department of Education. Data include information about the Federal assistance given to students and families under Title IV of the Higher Education Act of 1965, including student aid applications, student loan applications and receipt, and characteristics of applicants. *Statutory Citation: 20 USC § 1092b(d)(2).*
- **Housing Assistance Data** are collected by public housing agencies funded by the Department of Housing and Urban Development (HUD) and maintained in shared systems, including the Public and Indian Housing Information Center and the Tenant Rental Assistance Certification System. Data collected include characteristics of properties and units as well as information about the households receiving subsidized housing assistance. Data regarding participation in homeless assistance programs are collected at the local level through Homeless Management Information Systems. *Statutory Citations: 42 USC § 3535 and 3543.*
- **Military and Veteran Program** data are collected by a variety of sources, including the Department of Defense and the Department of Veteran Affairs, among other departments. These data include demographic, health, and other information about members of the active duty military and their families as well as services provided to veterans.

Although Federal statistical agencies produce many statistical data products that are available to the public without restriction, access to data files produced by these agencies that contain confidential information about individual people, businesses, or organizations is, quite appropriately, tightly controlled. This is true even when access is sought

to serve a statistical purpose. In some cases, the statistical purposes for which a statistical agency may grant access to confidential data under applicable law are defined very narrowly; the Census Bureau's law, for example, permits the use of sworn agents such as researchers only to help the agency carry out its work, and projects must specifically advance the mission of the Census Bureau irrespective of the broader value that access to data might have for evidence building. Examples of statistical data from the Principal Statistical Agencies are listed below.

Statistical Agency Data¹

- **Bureau of Economic Analysis**, in the Department of Commerce, collects data on multinational enterprises and international services transactions. *Statutory Citations: 22 USC § 3101-3108.*
- **Bureau of Justice Statistics**, in the Department of Justice, collects data on crime, offenders, crime victims, and the operations of the criminal justice system at all levels of government. *Statutory Citations: 42 USC § 3732, 42 USC § 3735 and 3789g.*
- **Bureau of Labor Statistics**, in the Department of Labor, collects data on employment and unemployment, projections of the labor force and employment by industry and occupation, prices and inflation, consumer expenditures, wages and employee benefits, occupational injuries and illnesses, and productivity and technological change. *Statutory Citations: 29 USC § 2, 2b, 5, 6, 8, and 9; 19 USC § 2393; 29 USC § 673.*
- **Bureau of Transportation Statistics**, in the Department of Transportation, collects data on transportation and transportation systems, including freight and travel statistics. *Statutory Citations: 49 USC § 6302(c); 49 USC § 6306; 49 USC § 6307; 49 USC § 6313.*
- **Census Bureau**, in the Department of Commerce, conducts the Decennial Census and

1. U.S. Congressional Research Service, Selected Information Regarding Statistical and Evaluation Entities in the Executive Branch [Memo] (February 28, 2017), by Clinton T. Brass and Jennifer D. Williams; <https://www.cep.gov/library/testimony.html>.

administers surveys that collect information on the demographic and economic health of the United States, such as the American Community Survey and the Survey of Income and Program Participation. *Statutory Citation: 13 USC § 9.*

- **Economic Research Service**, in the Department of Agriculture, collects data on the economic and social science aspects of agriculture, rural development, food, commodity markets, and the environment. *Statutory Citation: 7 USC § 2276.*
- **Energy Information Administration**, in the Department of Energy, collects data and information which is relevant to energy resource reserves, energy production, demand, and technology, and related economic and statistical information. *Statutory Citation: 42 USC § 7135(d).*
- **National Agricultural Statistics Service**, in the Department of Agriculture, conducts the Census of Agriculture and collects data on agricultural production and the economic and environmental status of the farm sector. *Statutory Citations: 7 USC § 2276, 7 USC § 2204g.*
- **National Center for Education Statistics**, in the Department of Education, collects data related to education in the United States and in other nations. *Statutory Citations: 20 USC § 9541-9544, 20 USC § 9546), 20 USC § 9573.*
- **National Center for Health Statistics**, in HHS, collects data on the extent and nature of illness and disability of the population of the United States and several other specified metrics. *Statutory Citation: 42 USC § 242k.*
- **National Center for Science and Engineering Statistics**, in the National Science Foundation, conducts periodic surveys and data collections on science, engineering, technology, and research and development. *Statutory Citation: 42 USC § 1862p.*
- **Office of Research, Evaluation, and Statistics**, in SSA, is responsible for conducting policy research and evaluation, providing statistical data on SSA programs, sponsoring special-purpose survey data collections and studies, and other activities related to SSA's mission.
- **Statistics of Income Division**, in the IRS of the Department of the Treasury, collects and disseminates income, financial, and tax information and also provides periodic reports on items from other returns and schedules. *Statutory Citation: 26 USC § 6108(c).* ■

The Commission's collection of public input, survey data, and related materials are available via the National Archives and Records Administration. The materials available from the Archives include:

Appendix E. Report on the CEP Survey of Federal Offices

Appendix F. CEP Public Meeting Materials and Presentations

Appendix G. CEP Public Input-Hearing Testimony and Other Public Comments

Appendix H. Prior Commissions Related to Evidence Building

This page intentionally left blank



The Promise of Evidence-Based Policymaking

Report of the Commission on Evidence-Based Policymaking

September 2017