7. BUILDING THE CAPACITY TO PRODUCE AND USE EVIDENCE

> *"We usually do better when we're on the side of facts and evidence and science. Just as a general rule, that's proved to be our strength as Americans."*
>
> **—President Obama, "*Remarks by the President to the Business Roundtable*" September 16, 2015**

The President has made it clear that policy decisions should be driven by evidence—evidence about what works and what does not, and evidence that identifies the greatest needs and opportunities to solve great challenges. The Administration is committed to living up to this principle through a broad-based set of activities to better integrate evidence and rigorous evaluation in budget, management, and policy decisions, including through: (1) making better use of data already collected by government agencies; (2) promoting the use of high-quality, low-cost evaluations and rapid, iterative experimentation in addition to larger evaluations examining long-term outcomes; (3) adopting more evidence-based structures for grant programs; and (4) strengthening agency evidence-building capacity and developing tools to better communicate what works.[1]

There is a growing momentum for these evidence-based approaches at all levels of government, as well as among nonprofits, foundations, faith-based institutions, and community-based organizations. The Administration's embrace of these approaches has resulted in important gains in areas ranging from reducing veterans' homelessness, to improving educational outcomes, to enhancing the effectiveness of international development programs. The 2017 Budget advances these approaches through a range of investments in evidence building, as well as by increasing investment in programs with strong evidence of effectiveness. These proposals are described in the main budget volume and accompanying documents.[2]

All of these efforts embody the simple guiding principle that: "Where evidence is strong, we should act on it. Where evidence is suggestive, we should consider it. Where evidence is weak, we should build the knowledge to support better decisions in the future."[3] In order to integrate this guiding principle in all aspects of government

decision-making, it is essential that Federal agencies develop the capacity to credibly build and use evidence and implement a culture that supports doing so—and many agencies are making progress in doing so. Given the centrality of agency capacity to these efforts, this chapter focuses on a few specific components of capacity, especially the principles and practices that support credible evaluation functions.

What is Evidence and How Should it be Used?

The best government programs use a broad range of analytical and management tools, which collectively comprise an "evidence infrastructure," to learn what works (and what does not) for whom and under what circumstances, as well as improve results. Broadly speaking, "evidence" is the available body of facts or information indicating whether a belief or proposition is true or valid. Evidence can be quantitative or qualitative and may come from a variety of sources, including performance measurement, evaluations, statistical series, retrospective reviews, and other data analytics and research.

Evidence cannot be separated from the purposes for which it is being used, and the credible use of evidence in decision-making requires an understanding of what conclusions can and, equally important, cannot be drawn from the information. For example:

- Multiple rigorous impact evaluations, in particular randomized experiments, may provide strong evidence that a particular intervention is effective with a particular population in a particular setting. However, they may be less definitive on how effective that intervention may be in other settings or with other populations.

- Quasi-experimental evidence from large, diverse samples of administrative data may make it easier to generalize across a range of circumstances, but they could lack definitive evidence on causality or be silent on important outcomes not captured in the administrative data.

- Descriptive analyses from Federal statistical series provide context to examine societal and economic trends over time.

- Studies of the observed behavior of individuals, groups, businesses, and other entities can provide

---

[1] Several Administration documents lay out this "evidence agenda," including previous versions of this chapter, the "*Evaluation as a Tool for Improving Federal Programs*" chapter of the Council of Economic Advisers' 2014 Economic Report to the President, and the OMB Memorandum M-13-17, "*Next Steps in the Evidence and Innovation Agenda*," May 2012, jointly signed by the Office of Management and Budget, the Domestic Policy Council, the Office of Science and Technology Policy, and the Council of Economic Advisers. Many of these documents are available on the OMB website at *http://www.whitehouse.gov/omb/evidence*. In addition, note that *OMB Circular A-11* has been updated to be consistent with many of these principles.

[2] See *http://www.whitehouse.gov/omb/evidence*.

[3] See OMB Memorandum M-12-14, "*Use of Evidence and Evaluation in the 2014 Budget*," May 2012.

important insights into the dynamics that policies and programs may be designed to address, whether through the use of administrative, survey, or linked datasets. However, they do not provide direct tests of specific policies, nor account for the exigencies of program administration.

- Qualitative and quantitative implementation studies can complement other evidence by providing insight into how programs and practices can be successfully implemented.

- High-quality performance measures can provide valid, reliable, and useful information on program inputs and/or outputs.

Evidence has varying degrees of credibility, and the strongest evidence generally comes from a portfolio of high-quality evidence rather than a single study or data point, i.e., from multiple sources and/or multiple studies covering different aspects and nuances of the topic. While many of these forms of evidence are complementary, some evidence that is useful for one purpose may not be useful for another. For example, performance measures are an essential resource for agencies to understand ongoing, real-time program performance so they can use that information to build a culture of continuous improvement, but they often do not tell us a lot about some key questions, including the effects of programs. Evaluations provide context for the performance measures and help us better understand what can and cannot be learned from them. In particular, rigorous impact evaluations, particularly randomized experiments, can provide the most credible information on the impact of the program on outcomes, isolated from the effects of other factors. Thus combining both performance and evaluation information, and using the results of one to inform the design of the other, can be very powerful in understanding program performance and ensuring that the program is maximizing performance and impact on an ongoing basis.

Examples of Progress

The Administration encourages agencies to generate more high quality evaluations, place greater attention on goal-setting and measuring performance through the implementation of the Government Performance and Results Modernization Action of 2010, increase the use of existing "administrative" data for evidence building, and strengthen the capacity of statistical agencies to build objective and quality evidence. The Administration is committed to acting on available evidence and has proposed to invest in, scale up, or change a variety of programs on the basis of strong evidence that they are effective. These proposals cover a broad range of policy areas including ending homelessness, improving employment outcomes, reducing crime and recidivism, and reducing global poverty and improving global health.

The Administration has also introduced a number of grant program innovations that embed evidence more fundamentally into their structures. Among the most notable advances in this area are "tiered-evidence" or "innovation fund" grant designs that focus resources on practices with strong evidence while also promoting innovation and further evaluation. The Administration has adopted "tiered evidence" grant programs in multiple areas, such as K-12 education interventions, teenage pregnancy prevention, social innovations for communities, voluntary home visitations for parents, and international assistance efforts.

The Administration is also promoting the Pay for Success financing model in a wide range of programs, including for workforce, education, recidivism, housing, and environmental interventions. Pay for Success financing leverages philanthropic and private dollars to fund preventive services and other interventions, which are provided by nonprofits and other non-governmental entities up front, with the Government paying only after the interventions generate sufficient measurable results. Since as early as 2012, agencies including the Departments of Education (ED), Housing and Urban Development (HUD), Justice (DOJ), and Labor (DOL) as well as the Corporation for National and Community Service (CNCS) have been working to implement Pay for Success. So far, Federal agencies have made awards supporting roughly 50 efforts, and still more agencies continue to explore possible applications of this model.

In addition, the Administration created Performance Partnership Pilots for Disconnected Youth that allow States, tribes, and localities to blend funding from various programs and receive waivers under multiple youth-serving programs in order to improve education, employment, and other key outcomes and build evidence about more effective ways to help vulnerable youth. In FY 2015, a consortium of six agencies awarded the first cohort of nine pilots to give State, local, and tribal communities customized flexibility to make a difference in the lives of local youth. Each pilot is conducting a site-specific evaluation of local outcomes, and the Department of Labor is leading an evaluation to look at cross-site implementation of the initiative overall. In FY 2016, the interagency consortium has grown to include HUD alongside ED, DOL, DOJ, the Department of Health and Human Services (HHS), CNCS, and the Institute for Museum and Library Services. In the coming year, this consortium expects to select up to 20 new pilots under two competitions.

At the same time, the Administration has pursued innovative approaches to improve federal capacity to identify more effective strategies. For example, the Administration established the Social and Behavioral Sciences Team (SBST)—a cross-agency group of experts in applied behavioral science—to help agencies translate findings and methods from the social and behavioral sciences into improvements in Federal policies and programs. Due to SBST projects, more military service members are saving for retirement, more students are going to college and better managing their student loans, more Veterans are taking advantage of education and career counseling benefits, more small farms are gaining access to credit, and more families are securing health insurance coverage.[4]

---

[4] For additional information on SBST's work please see their website (*https://sbst.gov/*) and their *2015 Annual Report*.

One common component of SBST's work is making better use of the administrative data that government already collects in order to learn which approaches work best. The Administration encourages all Federal agencies to make better use of these data to identify effective practices, facilitate day-to-day performance measurement, and inform the public about how society and the economy are faring. As discussed in the "*Building Evidence with Administrative Data*" chapter in the *2016 Analytical Perspectives* volume,[5] the ability to access and make better use of these data—while protecting privacy and confidentiality—has played a pivotal role in a range of policy areas, including some of the most innovative grant reforms and increased accountability and transparency across a range of programs. For example, multiple studies on student aid simplification showed the feasibility and importance of simplifying the Free Application for Federal Student Aid (FAFSA), using Federal administrative records as well as survey data. This research influenced the steps the Administration has already taken to simplify the FAFSA and motivated both Administration and Congressional proposals to make further legislative progress.

Principles and Practices that Support Credible Evidence Development

In order for government to make credible use of evidence, the evidence itself must be credible—meaning that it must be objective and of sufficient quality, utility, and integrity. Informed by national and international professional practice, the Federal Government has taken a number of steps to foster the credibility of evidence. For example, pursuant to the Information Quality Act[6] and OMB guidelines,[7] agencies are required to establish procedures to ensure the objectivity, utility, and integrity of information provided to the public, and to match the quality of the study with its intended use.

Similarly, a central theme of the Presidential Memorandum on the Preservation and Promotion of Scientific Integrity[8] and the associated implementation guidance[9] is that the public must be able to trust the science and scientific processes informing public policy decisions. These documents articulate and provide guidance on principles and procedures integral to the preservation and promotion of scientific integrity, including those related to strengthening the actual and perceived credibility of Government research, communicating scientific and technological information to the public, and the importance of shielding scientific data and analysis from undue political influence.

Most recently, OMB issued Statistical Policy Directive 1, Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units,[10] which affirms the Federal Statistical System's responsibility to produce and disseminate relevant and timely information; conduct credible, accurate, and objective statistical activities; and protect the trust of information providers by ensuring confidentiality and exclusive statistical use of their responses. The framework articulates principles and practices that support these responsibilities and requires Federal statistical agencies and recognized statistical units to adopt policies, best practices, and appropriate procedures to implement these responsibilities. These guidelines and policies provide a common foundation for core statistical agency functions to ensure that the information they provide adheres to a high standard of quality and utility to evidence based decision making.

Under this Administration, several evaluation offices have also established agency-specific statements of evaluation policy—for example *The Administration for Children & Families (ACF) Evaluation Policy* (in the Department of Health and Human Services) and *The Department of Labor Evaluation Policy*. These resources have generated useful conversations and agreements within agencies about their evaluation-related practices and principles.

Many Federal evaluators believe that establishing a common set of government-wide principles and practices for evaluation offices could help to ensure that Federal program evaluations meet scientific standards, are designed to be useful, and are conducted and the results disseminated without bias or undue influence. Establishing these standards is an important building block in furthering agencies' capacity to routinely build and use high-quality evidence to improve program performance, and help evaluation offices maintain standards for their programs across administrations and changes in personnel.

While the process for developing such a set of standards is ongoing, a few fundamental principles emerge as common themes in the established U.S. frameworks discussed above as well as in international frameworks. These principles include: Rigor, Relevance, Independence, Transparency, and Ethics.

Rigor

The accuracy and quality of evaluation results are dependent on the design and implementation of the underlying studies. All forms of evaluation should use the most rigorous methods as appropriate, and should use the most appropriate type of evaluation to answer the specific question(s) being asked. Rigor is not restricted to impact evaluations. It is also necessary in implementation or process evaluations, assessments, descriptive studies, outcome evaluations, and formative evaluations, as well as in both qualitative and quantitative approaches.

There are several practices that agencies use to support the rigor of evaluations. One of the most important is recruiting and maintaining an evaluation workforce with

---

[5] See "*Building Evidence with Administrative Data*," chapter 7 in the *2016 Analytical Perspectives* volume.

[6] See Section 515 of the Treasury and General Government Appropriations Act, 2001 (Pub. L. No. 106-554, 44 U.S.C. § 3516 note).

[7] See *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, February 22, 2002 (67 FR 8452).

[8] See *Memorandum for the Heads of Executive Departments and Agencies 3-9-09*, May 2009.

[9] See OSTP Memorandum, "*Memorandum for the Heads of Executive Departments and Agencies*," December 2010.

[10] See *Federal Register*, Volume 79, Number 231, Part III, "*Statistical Policy Directive 1, Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units*," December 2014

training and experience appropriate for planning and overseeing a rigorous evaluation portfolio. To accomplish this, agencies seeking to maintain or increase the rigor of their evaluation functions recruit staff with advanced degrees and experience in a range of relevant disciplines and provide professional developmental opportunities so that staff can keep their skills current. Agencies have cited hiring and retaining a skilled evaluation workforce as an area of difficulty. The Administration is working on how best to address these issues.

Another practice that helps agencies ensure rigor is the development and implementation of quality review and control procedures. Examples of strong procedures that maintain the integrity of evaluations include technical reviews of all aspects of evaluation designs (methodological design, data collection instruments and procedures, statistical and analytic plans); minimization of the burden of data collection; and external peer reviews by third-party independent technical experts and technical working groups.

Finally, evaluation guidelines and/or frameworks that indicate the standards for high quality evaluations—and how different types of evaluations and studies contribute to the evidence base—facilitate both the production and use of rigorous evidence. OMB encourages agencies to establish such frameworks and make them available as a technical resource for in-house and external evaluators when designing evaluations. These guidelines are also useful when assessing the quality of the evaluation as it was actually implemented.

These guidelines can be particularly powerful when they apply to more than one agency. For example, ED and National Science Foundation (NSF) issued Common

Guidelines for Education Research and Development in 2013.[11] These guidelines clarify how different types of studies contribute to the evidence base, including basic research and impact evaluations, and set expectations for the evidence that different types of studies seek to generate. Other agencies, such as DOL and components of HHS, are using the same guidelines for their evaluation activities. Research experts from Federal agencies, States, and academia are working with the National Academy of Sciences on ways to build consensus on standards for benefit-cost analysis of preventive interventions for children, youth, and families that would help government compare the benefits and costs of multiple strategies focused on similar target populations and outcomes. Common research standards and evidence frameworks across agencies can facilitate evaluation contracting, information collection clearance, and the strengthening or creation of research clearinghouses and repositories about "what works."

Relevance

Evaluations that do not inform decision-making have little applied value. For that reason this Administration has made integrating evidence into all types of budget, management, and policy decision-making a priority. Performance, evaluation, and other research evidence plays an important role in annual agency strategic review processes. Agencies seeking to increase the relevance and eventual use of evaluation findings take into account the viewpoints of a variety of stakeholders when establishing their research agendas, including

---

[11] See Institute of Education Sciences, Department of Education and the National Science Foundation, "*Common Guide-lines for Education Research and Development*," August 2013.

INCREASING RELEVANCE: THE "LEARNING AGENDA" APPROACH

The Administration encourages agencies to adopt "learning agenda" approaches in which agencies collaboratively identify the critical questions that, when answered, will help their programs work more effectively and develop a plan to answer those questions using the most appropriate tools. The key components of this learning agenda approach are that agencies:

- Identify the most important questions that need to be answered in order to improve program implementation and performance. These questions should reflect the interests and needs of a large group of stakeholders, including program office staff and leadership, agency and Administrative leadership, program partners at state and local levels, and researchers, as well as legislative requirements and Congressional interests.

- Strategically prioritize which of those questions to answer within available resources, including which studies or analyses will help the agency make the most informed decisions.

- Identify the most appropriate tools and methods (e.g. evaluations, research, analytics, and/or performance measures) to answer each question.

- Implement studies, evaluations, and analysis using the most rigorous methods appropriate to the context.

- Develop plans to disseminate findings in ways that are accessible and useful to program officials, policy-makers, practitioners, and other key stakeholders—including integrating results into performance measurement and strategic planning activities.

Several agencies have successfully implemented learning agendas. For example, the Department of Housing and Urban Development (HUD) has been a leader in integrating its performance measurement, evaluation, and research efforts into a HUD-Stat process that engages its leadership in evidence-driven discussions of key priorities. These HUDStat processes not only discuss evidence but identify areas where more evidence is needed. The Millennium Challenge Corporation (MCC) organizes itself around its learning agenda. Its decisions to enter into compacts with developing countries are based upon evidence on the effectiveness of particular types of interventions in particular types of countries and typically embed evaluations into those compacts in order to inform future decisions.

Congress, Administration leadership, agency leadership, the implementing program, and other partners and stakeholders. While evaluations can be expensive and lengthy undertakings, taking these viewpoints into account when developing evaluations or research agendas increases the likelihood that the eventual results will be relevant to those implementing programs and making important policy decisions. Taking these viewpoints into consideration also ensures that federal evaluation offices continue to be connected to programs as they are implemented, and ensures that they are addressing the most pressing questions, rather than those that may be of interest academically but have little practical impact. Such an approach requires agencies to develop strong partnerships and collaborations among evaluation staff, program staff, policy makers and service providers. It also requires agencies to effectively disseminate evaluation findings in formats that are easier to interpret and apply.

Credibility and Independence

Actively engaging stakeholders in identifying evaluation priorities and questions and assessing the implications of findings increases relevance. However, developing and maintaining a widely acknowledged position of independence from political or other undue external influences is critical in order for evaluation offices and their work to be credible. Credible evaluations are not constructed or intended to deliver a predetermined or politically expedient result—rather they seek to develop the most accurate evidence practicable to answer a specific question. Just as federal statistics, such as the calculation of the unemployment rate, do not change based on which political party is in power, credible evaluation methods are not be altered due to undue external influences. Credible evaluation offices produce products that are methodologically sound, impartial, clear, and readily perceived to be so by the users of their products and the general public. Through dem-

onstration of rigor, these offices establish their authority to determine the appropriate designs, data, and methods to use to conduct their work. They establish the capacity to make persuasive arguments for their chosen methods based on scientific principles and not by fiat. This capacity requires assurance that the selection and promotion of candidates for evaluation positions is based primarily on each candidate's scientific and technical knowledge, credentials, experience, and integrity.

Given the potential for political or other undue pressures, credible evaluation offices must demonstrate that their efforts can withstand critical review. Any personal or professional biases are stated and made transparent through the scientific testing of hypotheses, and both significant and null results are made transparent so that users understand the full array of hypotheses that were tested. The objectivity of the information released to the public is maximized by making information available on an equitable, policy neutral, transparent, and timely basis. As such, directors of credible evaluation offices have the authority to approve the design of evaluation projects and analysis plans, and the authority to approve, release, and disseminate evaluation reports, subject to legal, judicial, and security restrictions.

In this way, evaluation offices can demonstrate their independence from undue influences that may attempt to sway their work. The credibility that comes from independence and the independence that comes from credibility are both essential for users to maintain confidence in the accuracy and objectivity of evaluation results and for programs to be willing to cooperate with evaluation entities' requests.

Transparency

This Administration has placed a particular emphasis on increasing the transparency of federal evaluation

---

INCREASING THE TRANSPARENCY AND RELEVANCE OF EVALUATIONS THROUGH RESEARCH CLEARINGHOUSES

At the Federal level, many agencies have moved to increase the transparency and relevance of their evaluation findings by making them publicly available through "what works" repositories, sometimes referred to as research clearinghouses. "What works" repositories synthesize evaluation findings in ways that make research useful to decision-makers, researchers, and practitioners in the field. They also make evaluation results easily accessible to the public, and improve the transparency of evaluation results. Information in the repositories also indicates the implementation contexts of programs and strategies evaluated, and areas where more innovation or more evaluation is needed. Examples of Agency "what works" repositories include the:

- Administration for Children and Families (HHS) *Home Visiting Evidence of Effectiveness* (HomVEE),

- Administration for Children and Families (HHS) *Employment Strategies for Low-Income Adults* (ESER),

- Office of the Assistant Secretary for Planning and Evaluation (HHS) *Teen Pregnancy Prevention Evidence Review*,

- Substance Abuse and Mental Health Services Administration (HHS) *National Registry of Evidenced-based Programs and Practices* (NREPP),

- Department of Justice *CrimeSolutions.gov*,

- Department of Justice *What Works in Reentry Clearinghouse*,

- Department of Education *What Works Clearinghouse*, and

- Department of Labor *Clearinghouse for Labor Evaluation and Research* (CLEAR).

activities.[12] Transparency increases public awareness of ongoing and planned evaluation work, and of evaluation results, regardless of their findings. Public awareness of evaluation findings increases the likelihood that evaluation results will be used to inform a wide range of decision making; thus transparency is related to relevance.

Transparent evaluation offices make information about planned and ongoing evaluations/assessments easily accessible, typically through posting online information about the contractor or grantee conducting the work, descriptions of the evaluation/assessment questions, methods to be used, and the expected timeline for reporting results.

In addition, except in cases where there are legal, judicial, or security restrictions, transparent evaluation offices make evaluation plans, progress on ongoing evaluation work, and evaluation findings easily accessible to the public and release them in a timely way regardless of the findings. When these evaluations are released, the reports describe the methods used, including strengths and weaknesses, and discuss the generalizability of the findings. The released reports present comprehensive results, including favorable, unfavorable, and null findings.

Ethics

Evaluations/assessments should be conducted in an ethical manner and safeguard the dignity, rights, safety, and privacy of participants. Individuals and entities that participate in evaluations and the custodians of administrative data that may be used in support of evaluations must be able to trust that the information they provide as a part of an evaluation will be used only for the purposes that the agency has described. Thus, they must be able to trust that information collected for evaluation purposes will not be used for another purpose, such as law enforcement or regulation, directed at specific individuals or organizations. Evaluation offices should further build trust by minimizing the intrusiveness of questions and the time and effort required to respond to such questions, consistent with the agency's requirements for information. This can be accomplished through informing respondents of the expected time required to participate in the data collection, whether the collection is mandatory or voluntary, and any additional uses of the information. Multiple laws are in place in order to protect the rights, safety, and privacy of participants, and evaluations should comply with both the spirit and the letter of the relevant requirements.

Operationalizing an Effective Evidence Infrastructure

Operationalizing an effective evidence infrastructure requires a wide variety of capacities in addition to the principles and practices for evaluation offices articulated above. Developing and supporting the use of evidence and evaluation in decision-making requires a coordinated effort between those charged with managing the operations of a program, including administrative data collection and maintenance, and those responsible for using data and evaluation to understand a program's effectiveness. It requires consistent messages from leaders at different levels of an agency—policy officials, program and performance managers, strategic planning and budget staff, evaluators, and statistical staff—to ensure that data and evidence are collected or built, analyzed, understood, and appropriately acted upon.

No one individual in an agency has the knowledge and skills necessary to develop research designs that address actionable questions; collect, maintain, curate, and analyze administrative data; understand different types of evidence; interpret evidence; and develop and implement effective, evidence-based practices. Rather, it takes an agency leadership team to oversee these efforts and to build and sustain a commitment to learning. It also takes a team of "implementers" at the program level to encourage the use of evidence and data so that it reaches program management.

This section highlights two of the many capacities that support an effective evidence infrastructure: making better use of administrative data, and the establishment of centralized evaluation offices.

Building Evidence with Administrative Data

As described in last year's version of this chapter,[13] making better use of the "administrative data" that the government already collects to build evidence is an incredibly promising strategy. Administrative data are data collected by government entities for program administration, regulatory, or law enforcement purposes. Federal and state administrative data include rich information on labor market outcomes, health care, criminal justice, housing, and other important topics, but they are often greatly underutilized in evaluating programs' effects as well as in day-to-day performance measurement and for informing the public about how society and the economy are faring.

Over the course of this and previous Administrations, Federal agencies have steadily made progress improving the use of administrative data for evidence building. Some agencies are creating capacity to support research and evaluation in a particular policy area. For example, since 1995, the Bureau of Justice Statistics (BJS) has administered the National Criminal History Improvement Program which, among other accomplishments, helped all states achieve full participation in the Federal Bureau of Investigation's Interstate Identification Index. This critical operational network allows criminal justice agencies in the United States to exchange automated criminal history records (records which chronicle offenders' contacts with the justice system —i.e., "rapsheets"). Recently, BJS constructed an automated process which standardizes these variable federal and state records and creates unified researchable databases which can support a variety of research and evaluation of recidivism patterns and sentencing.

Similarly, ED has improved public understanding of how well colleges serve their students by matching administrative federal student loan and grant data to Department of

---

[12] See, for example, OMB Memorandum M-10-01, "*Increased Emphasis on Program Evaluation*," October 2009.

[13] See "*Building Evidence with Administrative Data*," chapter 7 in the *2016 Analytical Perspectives* volume.

Treasury tax data to create the new College Scorecard. The new College Scorecard provides students with the clearest, most accessible, and most reliable national data on college cost, graduation, debt, and post-college earnings to enable them to make better informed choices about colleges that fit their educational and career aspirations. The matched administrative data are also integrated into an open API that allows researchers and policymakers to customize analysis of college performance and allows other organizations to build tools to help students make more informed college choices.

However, most Federal agencies could make greater use of administrative data to build evidence. In addition, many agencies have data that would be useful to other agencies, other levels of government, or outside researchers for these same purposes. At the same time, not all agencies have the technological infrastructure or the expertise needed to utilize, share, or link data themselves, nor does it make sense to fully duplicate these capacities at every agency.

Federal statistical agencies already play a leading role in bringing together data from multiple sources, protecting privacy and confidentiality and ensuring data security, using data to create a wide variety of statistical products, and providing secure access to researchers inside and outside of government to conduct a broad array of policy- and program-relevant analyses. There are several examples where high-capacity statistical agencies have partnered with other Federal agencies to link and analyze administrative and survey data for evidence building purposes. Such partnerships build on the critical capacities that statistical agencies already have in order to make better use of existing data without creating unnecessary duplication.

Some agencies are leveraging capacity across the statistical system to build evidence in a particular policy area. For example, HUD has collaborated with the Census Bureau, the Centers for Medicare and Medicaid Services, and most recently National Center for Health Statistics to combine data and expertise to study relationships between housing, health risk behaviors, and health in order to use housing as a platform to improve quality of life. One outcome of these collaborations is new availability of linked survey and administrative datasets for researchers. Similarly, the Census Bureau and the Bureau of Labor Statistics are leading a multiagency effort to improve the Supplemental Poverty Measure (SPM). The SPM is designed to complement the official poverty measure, which is based on outdated assumptions and does not take into account most government transfer programs, and hence, cannot be used to evaluate their impact. In order to estimate the effectiveness of targeting resources toward the disadvantaged, the SPM integrates household income and expenditure information from national survey data with administrative data from a variety of Federal programs that help families, households, and individuals meet their basic needs.

The Budget proposes to expand this successful collaborative model. The Census Bureau is a leader for the often highly technical work of bringing together data from multiple sources, protecting privacy and confidentiality and ensuring data security, using data to create a wide variety of statistical products, and providing secure access to researchers inside and outside of government to conduct a broad array of policy- and program-relevant analyses. The Budget requests $10 million in funding for the Census Bureau to build on these existing strengths and start developing a more comprehensive infrastructure to prepare and share administrative data. This investment would help the Census Bureau work with States to obtain access to data from State-administered programs, such as the Supplemental Nutrition Assistance Program or the Special Supplemental Nutrition Program for Women, Infants, and Children, allowing new analysis of how these programs are used and their effects. Census would also improve its infrastructure for processing and linking data sets, as well as for providing data to researchers outside the Census Bureau.

Administrative Data Legislative Proposals

The above examples illustrate some of the exciting progress that agencies have made to better use administrative data. However, some significant barriers remain, including legislative barriers. The Budget continues many of the Administrative data legislative proposals included in the 2016 Budget, including the package of proposals designed to facilitate greater use of employment and earnings information and ease implementation of the Workforce Innovation and Opportunity Act. Employment and earnings data are among the most valuable Federal administrative data. Because many Federal (as well as State and local) programs are intended, in whole or in part, to increase employment and earnings, accurate employment and earnings data are needed to measure performance or conduct rigorous evaluations across a range of programs. The National Directory of New Hires (NDNH) is a database of employment and Unemployment Insurance information administered by the Office of Child Support Enforcement within HHS. Access to this data is tightly controlled by statute, and HHS implements strong privacy, confidentiality, and security protections to protect the data from unauthorized use or disclosure—there has never been a breach of the national NDNH data. Currently several programs are successfully using this data for program integrity, implementation, and research purposes.

The Budget proposes to build on this strong history of data stewardship and protection and allow additional programs and agencies to access this valuable data to learn what works and improve program implementation, while continuing to protect the privacy, security and confidentiality of that data. Specifically, the Budget proposes a package of proposals,[14] each of which is designed to clearly specify the purpose for which the data may be used, require that the minimum data necessary be used to achieve the purpose, and include strong penalties for the unauthorized access, use, disclosure, or redisclosure of the data. In order to streamline access to the data by authorized agencies for program integrity purposes, the package includes a proposal which would allow the authorized agencies to access the NDNH data through the Do Not Pay Business Center at the Department of the Treasury.

[14] See Budget Chapter 5, "A Government of the Future," and HHS's Administration for Children and Families Congressional Justification for additional information on the full package of NDNH access proposals and the criteria for considering access to NDNH data.

In addition, each component of the package is designed to satisfy the Administration's criteria for when authority to access NDNH data should be considered. The package also requires HHS to review each agency's data security before allowing that agency to access the data, prohibits HHS from granting access to the data for any purpose not authorized in statute, and requires HHS to publicly report on the use of NDNH data.

Centralized Evaluation Offices

Centralized or chief evaluation offices play an important role in developing and sustaining agency capacity to build and use evidence. A recent General Accountability Office (GAO) report[15] found that Federal agencies with a centralized evaluation authority reported greater evaluation coverage of their performance goals and were more likely to use evaluation results in decision making. However, the GAO report also found that only half of the existing centralized evaluation offices reported having a stable source of funding.

Centralized or chief evaluation offices are often a key component of implementing evaluation policies reflective of the core principles discussed above. Indeed the establishment of a centralized evaluation function and an evaluation policy re-

---

[15] Government Accountability Office Publication No. 15-25, "*Program Evaluation: Some Agencies Reported that Networking, Hiring, and Involving Program Staff Help Build Capacity*," November 2014.

flective of these core principles is a particularly strong and mutually reinforcing combination. Establishing a centralized office allows the agency to credibly establish the independence and transparency of its evaluation work, develop the specialized expertise required to implement rigorous evaluations, and creates a centralized entity responsible for coordinating and disseminating research findings.

These offices also play a central role in implementing effective learning agendas. Learning agendas cross program and agency boundaries and thus are difficult to implement well without an office with the responsibility to look across programs and work across agencies. Often data from one program may be useful in the learning agenda for another program; centralized evaluation offices can play an important role in that cross-program and cross-agency collaboration. Centralized evaluation offices also develop expertise about evaluation, about how to integrate evaluation and other research evidence with performance measurement and strategic review processes, and about how to help decision-makers use evidence. In several cases these offices, working in partnership with other evaluation offices and statistical agencies, also play a crucial role in using administrative data to build evidence about what works. Several agencies have successfully implemented centralized or chief evaluation offices. CNCS has successfully used its centralized evaluation office to coordinate its learning agenda across

---

CHIEF EVALUATION OFFICE AT THE DEPARTMENT OF LABOR

Over the last six years, the Department of Labor (DOL) has made significant progress in institutionalizing a culture of evidence and learning. The Chief Evaluation Office (CEO), established in 2010, plays a critical role in developing and maintaining this culture within DOL. As a part of its primary responsibility to manage DOL's evaluation program, CEO maintains a strong commitment to conducting rigorous, relevant, and independent evaluations. CEO is also committed to identifying and funding research and evaluation priorities established through a collaborative learning agenda process with DOL's various agencies. These agencies cover a broad range of topics, from employment and training programs to worker protection and enforcement activities. CEO plays an important role in initiating research that cuts across these agency and program silos.

CEO also serves as an "honest broker" on evidence issues within DOL, and its work is not limited to implementing evaluations. CEO actively participates in the performance management and strategic planning processes of the Department, and disseminates the results of their evaluations in formats that enable use by programs and policy makers.

Some of the key capacities that DOL developed to support this important work include:

- **Hiring staff with sufficient expertise** for CEO to manage rigorous evaluations of various methodologies. For example, using behavioral insights, CEO worked with Occupational Safety and Health Administration to implement a large random assignment study that identified an effective way to support establishments that have injury and illness rates above the national average.

- Launching the *Clearinghouse for Labor Research and Evaluation (CLEAR)*, which makes research on labor topics more accessible to practitioners, policymakers, researchers, and the public more broadly, thus increasing the transparency and relevance of the Department's evaluation efforts.

- Implementing a **learning agenda process** for the Department, in which CEO collaborates with each agency to identify key evaluation and research priorities.

- Securing the budget authority to **set aside a portion of specified program funds** (.75% in 2016) to support these evaluations.

- Creating a **data analytics unit** to support and complement agencies on their analytic needs and work; build data, statistical, and analytical expertise and capacity for the Department; and promote and innovate DOL administrative and public use data.

- Establishing a *Departmental of Labor Evaluation Policy* to institutionalize and guide the Department's evaluation efforts.

programs and to work with other agencies. NSF has recently instituted a centralized evaluation office that is in the process of developing a learning agenda for the first time. The Chief Evaluation Office at DOL has become a leader both at DOL and across the Federal government in advancing the role of evidence in decision-making. See the box above for more on DOL's Chief Evaluation Office.

Conclusion

The evidence capacity building efforts outlined in this chapter fit into the Budget's broader emphasis on tack-ling challenging but important issues that are integral to making government work better. This chapter articulates common principles that support the development of credible evidence, as well as the capacity required to operationalize an effective evidence infrastructure. This Budget makes substantial investments in programs based on evidence in addition to further building the capacity to produce and use evidence.