

7. BUILDING EVIDENCE WITH ADMINISTRATIVE DATA

Introduction

“We’ve got Democratic and Republican elected officials across the country who are ready to roll up their sleeves and get to work. And this should be a challenge that unites us all. I don’t care whether the ideas are Democrat or Republican. I do care that they work. I do care that they are subject to evaluation. . .”

-- President Obama, “Remarks on Promise Zones,” January 9, 2014

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 already-collected data within government agencies; (2) promoting the use of high-quality, low-cost evaluations and rapid, iterative experimentation; (3) adopting more evidence-based structures for grant programs; and (4) building agency evaluation capacity and developing tools to better communicate what works.

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 of the President*, and the July 2013 memo, “Next Steps in the Evidence and Innovation Agenda,” 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. The 2016 Budget moves the agenda forward 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.¹

This chapter focuses on the Administration’s efforts to address one especially important next step in the evidence agenda: making better use of “administrative data.” 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.

Administrative data have played a key role in some of the most important evidence agenda accomplishments to date. As described below, insights obtained from adminis-

trative data are already influencing education and health policy, among other areas. Access to administrative data has been pivotal in some of the most innovative Federal grant reforms and in increasing accountability and transparency across a range of programs; it has also played an important role in innovation and experimentation at the State and local levels. Meanwhile, as the evidence agenda matures, lack of access to appropriate data is increasingly a key obstacle to progress along a number of dimensions. Whether the objective is to facilitate more rapid, low-cost evaluations, to base more grant decisions on strong evidence, to adopt program structures that permit greater innovation and flexibility in exchange for greater accountability for results, or to provide more and better performance information to the public, administrative data are often a crucial untapped resource.

A significant focus in this year’s Budget is improving access to administrative data for purposes of evaluation, accountability and transparency, performance management, and other research and analytic purposes. (While not discussed in this chapter, the Budget also includes separate proposals to improve the use of administrative data to protect program integrity, for example to combat identity theft.) The Budget proposes a number of specific access and infrastructure improvements across multiple programs and agencies, efforts that build on the long-standing use of Federal statistics to describe the condition of the economy and society and inform Federal policy decisions (see Chapter 5, “Social Indicators,” and Chapter 16, “Strengthening Federal Statistics”). The Budget also embraces Representative Paul Ryan and Senator Patty Murray’s proposal to create a commission that would make recommendations about how to fully realize the potential of administrative data to improve Federal programs. That proposal exemplifies the high-level and bipartisan momentum for doing more to tap this important resource.

This chapter explains the importance of improving access to administrative data, describes some of the key barriers, and outlines the Administration’s agenda, including both Budget proposals and ongoing work. The chapter also explains the strong framework of privacy, confidentiality, and data security protections that governs current uses of administrative data for research purposes, and it explains how these protections would extend to the Budget’s new proposals.

Background

Administrative data are data collected by government entities for program administration, regulatory, or law enforcement purposes. Examples include: data on employment and earnings collected through the Unemployment Insurance (UI) program, data on medical conditions and

¹ See <http://www.whitehouse.gov/omb/evidence>.

payments collected through Medicare and Medicaid, data on local pollution levels collected to administer the Clean Air and Clean Water Acts, and criminal histories maintained as part of police records or arrests. Such data are usually collected on the universe of individuals, businesses, or communities affected by a particular program, in contrast to survey data that are collected for samples of broader populations, typically for research or other statistical purposes.

Administrative data are used for a wide range of purposes, such as reimbursing service providers, determining benefit eligibility, and ensuring compliance with safety or environmental regulations. Sometimes data collected to administer one program can also be useful for administering another. For example, employment and earnings data collected through the UI system could be used to determine eligibility for a means-tested program. Administrative data can also be useful for program integrity efforts to minimize waste, fraud, and abuse. Linking data across programs for administrative purposes can sometimes make government more efficient, but it is not the focus of this chapter.

This chapter is focused on reusing administrative data for “statistical” purposes: the use of data to better understand the characteristics, behavior, or needs of groups of individuals or communities. Statistical purposes *exclude* uses that affect the rights, benefits, or privileges of individuals: indeed, one of the defining characteristics of statistical use is that data about an individual are never made public, and are never used to make decisions about that individual. But statistical purposes *include* a wide range of analytic uses, where only aggregated and de-identified data are made public. For example, statistical use encompasses both traditional program evaluations and the newer “rapid-cycle” experimentation and other data analytics techniques increasingly employed by innovative private-sector firms. It also encompasses transparency and accountability efforts, such as scorecards, that provide Federal agencies, State and local governments, and the public with information on the relative performance of different hospitals, training programs, or other service providers. And it encompasses efforts to quantify how housing, health care, education, or other needs vary across communities, as well as other analysis of patterns and trends for groups of individuals.

Using administrative data for these purposes, alone or in combination with survey data (data gathered from a sample population gathered specifically for research or other statistical purposes), can have a number of advantages over survey data alone, which is typically the alternative. First, because administrative data are collected through the normal administration of programs, they can often be obtained at much lower cost than fielding a new survey. Second, administrative data are sometimes more accurate than survey self-reports, especially with respect to information directly used to administer the program (e.g. earnings in the case of UI or health care utilization in the case of Medicare records). Third, administrative data, especially when linked across multiple programs, are often available for long time periods, permitting study of

long-term impacts that would be prohibitively expensive with a survey. For example, recent studies have used administrative records to look at the effect of being assigned a smaller class size in elementary school on college-going and earnings and at the effect of losing one’s job on mortality over the subsequent 20 years.²

Perhaps most important, reusing administrative data often allows for much larger sample sizes than surveys. Surveying program participants requires tracking them down, getting each to agree to participate in a survey, and constructing and administering the survey itself. Since each of these activities is expensive, evaluations, performance measurement, and other research based on surveys typically draw on small samples. Sometimes, the sample sizes are so small that the resulting studies lack the statistical power to reliably detect policy-relevant effects of programs, even when these effects exist. For example, consider a randomized controlled trial of a job training program with 1,000 participants (and 1,000 non-participants in the control group), where a 2 percentage point increase in employment would be enough to justify the cost of the program. With samples that size, the trial would need at least a 4.4 percentage point increase in employment to be statistically significant (meaning statistically distinguishable from 0) and a 6.4 percentage point increase to have confidence that it was cost-effective.³ In fact, even with a 2.4 percentage point *reduction* in employment, one could not rule out the possibility of cost-effectiveness, illustrating the lack of precision that results from small (or even medium-size) samples. Greater use of administrative data can reduce the costs of collecting data on large samples, helping reduce the number of underpowered studies that misdiagnose programs as “not working” when the problem is actually with the small-sample studies, not the programs. Meanwhile, large administrative datasets also allow for quasi-experimental studies that would be impossible in most survey datasets, particularly research designs that depend on detecting small differences in outcomes based on small but near-random variation in program participation.

That said, administrative data are no panacea. Since administrative data are collected to meet the needs of the relevant program - not the needs of the research design - they will sometimes lack information important for a given evaluation or other statistical use, such as demographic details needed to understand how policies and programs affect different groups within the population. Administrative data usually provide information only on participants and not on those eligible but not participating, who are sometimes the most relevant comparison group for a study. In addition, it may be costly to make administrative data usable for statistical purposes, especially if the original data are incomplete, inconsistent, or

² Chetty, Raj et al, “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star,” *The Quarterly Journal of Economics*, March 2011, 126(4), pp. 1593-1660. Sullivan, Daniel and Till von Wachter, “Job Displacement and Mortality: An Analysis using Administrative Data,” *The Quarterly Journal of Economics*, August 2009, 124(3), pp., 1265-1306.

³ These calculations assume a standard 95 percent confidence threshold.

poorly documented. Neither administrative nor survey data are perfect, but together they can yield greater insight than either can alone.⁴

Examples of Successful Reuse of Administrative Data

In certain areas, including education and health care, reusing administrative data is comparatively common and is already having a large impact on policy. In part as a result of No Child Left Behind and other Federal efforts to improve State education data, some individual States have developed high-quality longitudinal data systems for kindergarten through college (although others remain weak, and there is no national system letting researchers track or compare students across States). These State data and related school district administrative data have been used for important and influential research on topics ranging from teacher value-added to disparities in educational outcomes by family income to the effects of universal pre-kindergarten, charter schools, intensive tutoring programs, and community college remediation programs.⁵ Research on student aid simplification - showing the feasibility and importance of simplifying the Free Application for Federal Student Aid (FAFSA) - also relied on administrative records.⁶ This research has in-

⁴ For a more extensive discussion of the advantages and disadvantages of administrative and survey data, see: Blank, Rebecca M., Kerwin Kofi Charles, and James M. Sallee, "A Cautionary Tale about the Use of Administrative Data: Evidence from Age of Marriage Laws," *American Economic Journal: Applied Economics*, 2009, 1(2): pp. 128-49; Prell et al., Working Paper: "Profiles in Success of Statistical Uses of Administrative Data," Federal Committee on Statistical Methodology, 2014; National Research Council, "Reengineering the Survey of Income and Program Participation," Panel on the Census Bureau's Reengineered Survey of Income and Program Participation, 2009; Cito, Constance F. and John Karl Scholz, editors, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press; and Prewitt, Kenneth, "Science Starts Not after Measurement, but with Measurement," *The ANNALS of the American Academy of Political and Social Science*, September 2010, 631(1), pp. 7-16.

⁵ See Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, March 2005, 73(2), pp. 417-458; Papay, John P., Richard J. Murnane, and John B. Willett, "Income-based Inequality in Educational Outcomes: Learning from State Longitudinal Data Systems," NBER Working Paper No. 20802, December 2014; Andrews, Rodney J., Jargowsky, Paul, and Kuhne, Kristin, "The Effects of Texas's Targeted Pre-Kindergarten Program on Academic Performance," NBER Working Paper No. 18598, December 2012; Abdulkadiroğlu, Atila, et al., "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *The Quarterly Journal of Economics*, 2011, 126(2) pp. 699-748; Fryer, Roland G., Jr., "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments," *The Quarterly Journal of Economics*, April 2014, 129(3), pp. 1355-1407; and Calcagno, Juan C. and Bridget T. Long, "The Impact of Postsecondary Remediation Using a Regression Discontinuity Approach: Addressing Endogenous Sorting and Noncompliance," *The National Center for Postsecondary Education Working Paper*, April 2008.

⁶ Dynarski, Susan M. and Judith E. Scott-Clayton, "College Grants on a Postcard: A Proposal for Simple and Predictable Federal Student Aid," Hamilton Project Discussion Paper, February 2007 utilized the National Postsecondary Student Aid Study, which combines administrative and survey records. Bettinger, Eric P., et al., "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment," *The Quarterly Journal of Economics*, April 2012, 127(3), pp. 1205-1242 utilized Federal administrative records to

influenced steps the Administration has already taken to simplify the FAFSA and motivated both Administration and Congressional proposals to make further progress through legislation.

State education data systems have also contributed to the success of the Department of Education's Investing in Innovation ("i3") tiered evidence program, one of the Administration's most successful grant reform efforts. In a tiered evidence grant program, grantees can receive smaller grants under the "Development" (proof of concept) tier to begin testing new models that have high potential; larger grants under the "Validation" tier to further test interventions that have emerging evidence of effectiveness; or the highest level of funding under the "Scale-up" tier when they have strong evidence that their proposed approach delivers impact. Grants in each tier include funding for rigorous evaluations. To date, i3 has funded over 130 innovative Development and Validation studies, as well as six larger Scale-up projects, and it has helped uncover successful interventions in the areas of teacher and principal effectiveness, turning around low-performing schools, and implementing college- and career-ready standards and assessments. Nearly all of i3's Scale-up grantees have used administrative data for their evaluations, as have many of the Development and Validation grantees.

In the health arena, the Centers for Medicare and Medicaid Services (CMS) has a strong and longstanding infrastructure for making administrative data available for statistical uses, including by outside researchers, with strong privacy protections. Medicare claims data have been the basis for important and influential research on regional variation in health care utilization and costs, payment policies, and other topics.⁷ The availability of these data is essential to ongoing Center for Medicare and Medicaid Innovation (CMMI) delivery system demonstrations that are testing innovative payment and service delivery models to reduce expenditures while preserving or improving quality. CMS and the Department of Housing and Urban Development (HUD) are also collaborating to match Medicare and Medicaid data with HUD tenant data to evaluate the impact of housing assistance on health care utilization and costs over time. Preliminary findings from 12 jurisdictions were released in 2014,⁸ with results feeding into a new demonstration on the cost effectiveness of Federal housing and services interventions that assist seniors who wish to age in place, avoiding the costs of institutionalization.

CMS has also been a leader in using administrative data from Federal programs to provide the public with actionable information about different service providers. In 2014, the Administration for the first time released Medicare utilization and cost data summarized at the physician-level, letting Americans compare their own doctors' practice patterns with national norms. Likewise, the

measure take-up of financial aid.

⁷ See, for example, the Dartmouth Atlas of Health Care project.

⁸ Department of Health and Human Services and the Lewin Group, "Picture of Housing and Health: Medicare and Medicaid Use Among Older Adults in HUD-Assisted Housing," March 2014.

CMS “Hospital Compare” feature provides information on how almost every U.S. hospital performs with respect to clinical quality metrics (such as whether patients receive appropriate care in a timely fashion) and hospital readmission rates.

Barriers to Effective Use of Administrative Data

The examples above highlight the potential gains from making greater use of administrative data, including the gains from being able to link administrative data from one program to administrative data or survey data from another (for example, linking education to earnings data). A number of agencies, including those assisting businesses as well as social service programs, are exploring new ways to use and to link administrative data to study program impacts. Unfortunately, there are significant barriers to doing so, which generally fall into three categories.

- **Legislative barriers to access.** Some authorizing statutes explicitly prohibit Federal agencies from sharing data with one another, or even from routinely reusing their own data for statistical purposes. Sometimes, these legislative barriers are oversights, resulting from out-of-date statutes that have not been updated for modern technology and data analytic techniques. In other cases, they may reflect legitimate concerns about privacy and confidentiality that need to be fully addressed in any proposal to expand data access (and are discussed in detail below).
- **Policy and legal interpretations.** Even where data sharing is legally allowed, agencies may be unsure about the rules and may therefore default to the assumption that data cannot be shared. Meanwhile, many agencies do not have established policies and processes for receiving, reviewing, and approving requests for administrative data and for negotiating the agreements typically required before data are provided. As a result, it can take years for other agencies (or even offices within the same agency) to obtain access to needed data, a major barrier to providing timely information to inform policy debates. Anticipated difficulties with negotiating access to data can also discourage agencies from seeking access in the first place.
- **Resource and capacity constraints.** Even when access barriers are cleared, other challenges remain. As noted above, administrative data are often poorly documented, to the point where individuals not involved in administering the relevant program may have no way to know what key variables mean. Data can also be inconsistent and incomplete. Many agencies lack the technological infrastructure and appropriate personnel to make their data interpretable and usable by researchers outside the program, or to conduct their own analysis using administrative data. Moreover, it would be inefficient for every agency to build this technological infrastructure. Meanwhile, agencies that do have these capacities

(including Federal statistical agencies,⁹ such as the Census Bureau) may lack the resources needed to negotiate agreements to obtain or share data, to address external researcher requests for access to data, and to assist other agencies in utilizing the data.

Ongoing Efforts and Initiatives

Despite the legal and operational barriers discussed above, OMB and other Federal agencies are making notable progress to improve, expand access to, and better utilize administrative data for statistical purposes. Examples include:

- **OMB guidance on using administrative data for statistical purposes.** Earlier this year OMB issued path-breaking guidance to help both program and statistical agencies better leverage administrative data for statistical purposes. The memo builds on earlier guidance designed to increase the value of existing data and creates “a presumption in favor of openness to the extent permitted by law and subject to privacy, confidentiality, security, or other valid restrictions.”¹⁰ The guidance encourages agencies to develop strong data stewardship and data management processes so that statistical use of administrative data is “designed in” from the start. It also assists agencies in overcoming barriers created by inertia and confusion by addressing a host of legal interpretation, policy, interagency agreement and data quality challenges. Since the guidance was issued last year, agencies have reported to OMB what datasets they would like to acquire and indicated both successes and barriers in acquiring them. OMB has been working with agencies on these priorities and continues to assess how to promote agency progress in using administrative data for statistical purposes.
- **Increased utilization of key administrative data resources and improvements in data quality.** Given the potential of administrative data to complement surveys as the foundation of the nation’s evidence base, Federal statistical agencies continue to pursue additional uses of these data for statistical purposes. For example, the Bureau of Justice Statistics (BJS) is using crime reports from local law enforcement agencies to develop the first-ever time-

⁹ “Statistical agency” refers to “an agency or organizational unit of the executive branch whose activities are predominantly the collection, compilation, processing, or analysis of information for statistical purposes” [PL-107347, Title V—CIPSEA, Section 502 (8)]. The statistical agencies within the executive branch of the Federal Government are: the Bureau of Economic Analysis; the Bureau of Justice Statistics; the Bureau of Labor Statistics; the Bureau of Transportation Statistics; the Census Bureau; the Economic Research Service; the Energy Information Administration; the National Agricultural Statistics Service; the National Center for Education Statistics; the National Center for Health Statistics; the National Center for Science and Engineering Statistics; the Office of Research, Evaluation, and Statistics at SSA; and Statistics of Income at IRS.

¹⁰ See OMB Memorandum M-14-06, “Guidance for Providing and Using Administrative Data for Statistical Purposes,” February 2014.

ly, detailed, and accurate national measures of crime incidents, which will increase the Nation's ability to monitor, respond to, and prevent crime. BJS is also using corrections data to produce better longitudinal statistics on offender re-entry and re-integration patterns and costs. Likewise, the National Center for Education Statistics (NCES) is making greater use of student financial aid and enrollment records to increase what the nation knows about the costs of college attendance and student debt. And the National Center for Health Statistics (NCHS) is helping States improve the quality and timeliness of vital birth and death records to help track priority health initiatives in prevention, cancer control, and teenage pregnancy prevention. NCHS is also providing increased secure researcher access to linked survey and administrative data to examine the factors that influence disability, chronic disease, health care utilization, morbidity, and mortality.

- **Improving the use of data at the Department of Labor (DOL).** In FY 2014, DOL established a data analytics unit within the Office of the Chief Evaluation Officer. That office focuses on supporting agencies in their efforts to improve administrative data quality, access data, use data to conduct both simple and complex analysis to answer important program and policy questions, and improve program operations. Agencies within the Department, working independently and in conjunction with the DOL analytics unit, have greatly expanded their data analytics activities in the past year. For example, the Occupational Safety and Health Administration (OSHA) has made great strides in using administrative data from a variety of sources to better identify and target enforcement on severe violators of health and safety standards. The 2016 Budget would establish an OSHA-specific Data Analytics Unit that will support more in depth analysis of data on the effectiveness and impact of OSHA's enforcement, consultation, outreach, and whistleblower protection strategies and activities.

Budget Proposals

While the Administration is already taking steps to realize administrative data's potential to contribute to evidence building, fully unlocking that potential will require legislative changes. The 2016 Budget includes a package of proposals that aim to overcome the statutory and operational barriers discussed above, making additional administrative data from Federal agencies and programs legally and practically available for policy development, program evaluation, performance measurement, and accountability and transparency efforts. The package reflects two guiding principles:

1. Consistent with the philosophy behind the Administration's "Open Data" Initiative,¹¹ Federally-funded data should be available to the public and for public purposes to the greatest extent consistent with strong privacy, confidentiality, and data security protections.
2. Federal statistical agencies should be equipped to facilitate reuse of administrative data, including by other Federal agencies and, where safe and appropriate, State and local governments and outside researchers, to answer policy-relevant questions.

As discussed below, the 2016 Budget includes proposals to: (1) improve access to specific administrative data sets; (2) invest in the infrastructure needed to support more effective use of Federal and State administrative data; and (3) lay the groundwork for further progress going forward.

Expanding Access to Data

Employment and earnings data are among the most valuable Federal administrative data. Because many Federal (and State and local) programs are intended, in whole or in part, to increase employment and earnings, accurate employment and earnings data are needed to construct meaningful performance measures or conduct rigorous evaluations across a range of programs. These data can also shed light on local labor market dynamics and on how different groups are succeeding or failing in the labor market, informing program design.

As noted above, timely and reliable quarterly employment and earnings data are collected as part of administering the Unemployment Insurance (UI) program. While UI data are collected and held by individual States, the Federal government maintains two national databases of quarterly UI records compiled from State data: the Department of Health and Human Services' National Directory of New Hires (NDNH) and the Census Longitudinal Employer-Household Dynamics (LEHD) program. However, both databases are subject to stringent access restrictions, and, despite the fact that Federal government funds support UI data collection, neither is available for most Federal statistical uses. For example, even though the Department of Labor administers the UI system, neither database can generally be used to evaluate the impact of Federally-funded job training programs.

On top of these restrictions on data sharing, the Workforce Investment Act (WIA) has since 1998 prohibited the "development of a national database of personally identifiable information on individuals receiving [WIA] job training services."¹² The implication of the ban is that even if the Department of Labor obtained consent to collect personal identifiers for participants in Federal job training programs, and even if it could then obtain access to UI earnings records, it still could not take advantage of these data to create a national database that could be used to streamline and standardize performance report-

¹¹ See OMB Memorandum M-13-13, "The Open Data Policy - Managing Information as an Asset," May 2013.

¹² Workforce Innovation Act, Section 504b.

ing and transparency efforts for the workforce system or to evaluate which job training programs work best for what types of beneficiaries. This ban was reiterated in the 2014 reauthorization of WIA (the Workforce Innovation and Opportunity Act or WIOA), despite the fact that WIOA recognized that “[performance] reporting and evaluation requirements are important tools in measuring effectiveness, especially for the core [WIOA] programs.”¹³

The 2016 Budget includes three proposals that would facilitate greater use of employment and earnings data.

First, consistent with bipartisan Congressional proposals, the Budget would allow select Federal statistical and evaluation units to access the NDNH for statistical purposes, subject to strong privacy and confidentiality protections. The proposal would allow NDNH data to be used to evaluate Federal job training and other programs intended to increase employment and earnings, as well as to construct job training service provider “scorecards” based on participant employment and earnings outcomes, consistent with WIOA and the goals laid out as part of the Administration’s review of job-training programs.¹⁴ The proposal would also permit the use of NDNH data to improve the completeness and efficiency of the Census LEHD program and the 2020 decennial census. Access to NDNH could help Census reduce the cost of the decennial census by \$1.2 billion or more by using administrative records to identify who resides in non-responding households. The NDNH access proposal would prohibit the Federal statistical and evaluation units from releasing personally identifiable information, and it includes strong criminal penalties for individuals if they willfully make an unauthorized disclosure. A version of this proposal passed the House of Representatives with bipartisan support in 2013 and the Senate Finance Committee in 2014.¹⁵

Second, the Budget proposes to eliminate the WIOA database ban. WIOA laid out a vision for a streamlined workforce system that improves outcomes through standardized performance requirements, integrated service delivery, and stronger evaluation requirements. However, it did not include provisions to allow DOL to further streamline, standardize, and more accurately capture the outcome information essential to these goals. Eliminating the WIOA database ban, in combination with granting DOL access to UI records, would greatly simplify ongoing efforts to evaluate job training programs while reducing State burden associated with WIOA performance reporting and transparency requirements. These authorities would also help improve the accuracy and completeness of performance and transparency efforts, by simplifying State efforts to capture outcomes for WIOA participants

¹³ 160 Cong. Rec. S3982-3990, “Statement of the Managers to Accompany the Workforce Innovation and Opportunity Act,” daily ed. June 25, 2014.

¹⁴ The White House, “*Ready to Work: Job-Driven Training and American Opportunity*,” July 2014.

¹⁵ The Budget also proposes to allow NDNH data sharing for certain non-statistical (administrative) purposes that will help make programs more efficient and effective. For principles governing these other NDNH access proposals, see the Department of Health and Human Services’ Administration for Children and Families FY 2016 Justification of Estimates for Appropriations Committees.

who move to another state and making it easier to identify participants who take advantage of multiple WIOA programs.

Third, as part of a broader UI solvency and reform package, the Budget would require States that receive new Federal funding for UI modernization to allow broader statistical use of the UI earnings records they already provide to the Census Bureau for the LEHD. The UI reform package would also provide incentives for States to improve UI data quality and to take advantage of these data to provide UI recipients with better information on workforce opportunities.

Beyond wage data, the Budget includes a number of other proposals that would improve access to important administrative data resources. For example:

- **Expanding access to Medicare data to spur improvements in health care quality.** The Affordable Care Act allowed the Centers for Medicare and Medicaid Services to make certain Medicare claims data available to qualified healthcare research organizations for the purpose of performance evaluation. The Budget expands this authority to allow the data to be used for a broader array of purposes, such as fraud prevention activities and value-added analysis for physicians to enable better care coordination and practice improvement.
- **Providing targeted access to business tax data to improve economic statistics.** Current law authorizes access to business tax data by the Census Bureau, and these data are important for developing timely and accurate economic statistics. However, the Bureau of Labor Statistics (BLS) lacks access to business tax data, and the Bureau of Economic Analysis (BEA) has only limited access. These restrictions prevent sharing of business information for statistical purposes among these agencies, especially for the large and growing non-corporate sector. The Budget proposes to augment BEA’s current access to business tax data and permit BLS to receive Census Bureau data for businesses with limited tax information. This would allow the agencies to collaborate in producing and verifying business datasets, reducing costs and correcting errors that can degrade the quality of key economic statistics.

Investing in Data Infrastructure

Almost all Federal agencies could make greater use of their own or other agencies’ 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. The Census Bureau in particular already has much of the infrastructure and capacity needed to serve as a leader for this often highly technical work.

The Budget requests \$10 million in additional funding for the Census Bureau to build on its existing strengths and start developing a more comprehensive infrastructure for linking, sharing, and analyzing key datasets. Specifically, the additional funds would allow the Census Bureau to:

- **Accelerate the process of acquiring and processing additional data sets.** Census has explicit legal authority to request data from any public or private entity, but it generally needs to negotiate access, often a time-consuming and resource-intensive process. Census is already in the process of acquiring Supplemental Nutrition Assistance Program (SNAP) data from States, and additional funding would allow it to accelerate the process of acquiring other Federal and Federally-sponsored program data.
- **Expand and improve infrastructure for processing and linking data.** As discussed above, because they are collected for program administration rather than statistical purposes, administrative data are often poorly documented, inconsistently formatted, and otherwise difficult to work with. The proposed investment will provide the Census Bureau with resources to document, link, and anonymize additional data sources to make them usable for analysis.
- **Improve the infrastructure for providing data to non-Census researchers.** Most Census data can be accessed through the Census Bureau's network of Research Data Centers (RDCs), which allow non-Census researchers, including both staff from other Federal agencies or levels of government and outside experts, to access these data. However, limited Census resources sometimes lead to long delays in reviewing and approving RDC project proposals. With additional funding, the Census Bureau would be able to improve and expedite the process of approving proposals to use RDC data. Census would also expand capacity in RDCs and, building on existing models, would offer other statistical and non-statistical agencies the opportunity to make their data available through the RDC network. Finally, the Census Bureau would explore the feasibility and desirability of creating secure virtual access to select datasets, as some other agencies have done.

In addition to the major Census investment, the Budget also includes other investments in making administrative data usable and available for statistical purposes, including proposals that would:

- **Improve higher education data and increase the Department of Education's capacity to utilize these data to inform policy and management.** The Budget provides \$11.6 million for the second phase of the Department's Enterprise Data Warehouse project, which will allow for deeper analysis of the Federal student loan portfolio and borrower behavior. The Budget also provides increases at NCES for more frequent administration of the National Postsecondary Student Aid Study (NPSAS). The NPSAS, which integrates student aid administrative data with robust survey data on demographics and student experiences, is a leading example of using administrative and survey data together to provide greater insight than either can alone. The Budget proposal will make more timely data on educational costs, financial aid, enrollment, and student progress available to policymakers and the public.
- **Help States improve their workforce and education data systems.** The Budget includes \$107 million through the Workforce Data Quality Initiative and the State Longitudinal Data Systems grant programs (at the Labor and Education Departments, respectively) to help states build and use integrated and longitudinal data systems across their workforce and education programs. The Budget also includes \$60 million to support state consortia as they modernize their UI tax and benefit systems, which will improve both the claimant experience and the quality of the UI data. Finally, to help address some of the policy and legal interpretations that states grapple with when trying to make better use of their own workforce and education data, the Budget includes funding for a joint DOL and Education team that will serve as the central point of contact and technical assistance for States, Federal programs, and researchers on issues related to accessing, collecting, and using workforce and education data.
- **Support linking Bureau of Justice Statistics (BJS) data with the Department of Justice's grants management system data.** This proposal would enable the BJS to explore the feasibility of linking its statistical collections to the Department of Justice's grants management system data. Linking these data would shed light on the variation in funding across geographies and over time and on the effects of Federal justice system funding (both by amount and type) on recidivism and other outcomes.

Setting the Stage for Future Progress

In November, 2014, Representative Paul Ryan and Senator Patty Murray jointly introduced legislation (H.R. 5754/S. 2952) that would create a Commission on Evidence-Based Policymaking. The Commission would be charged with reviewing "the inventory, infrastructure, and protocols related to data from Federal programs and

tax expenditures while developing recommendations for increasing the availability and use of these data in support of rigorous program evaluation.”¹⁶ In particular, the Commission would advise Congress on whether and how to create a “clearinghouse” for administrative and survey data that would facilitate accessing and linking data to evaluate program effectiveness and inform domestic policymaking.

The Budget embraces the Ryan/Murray approach, putting forward a similar proposal to create a commission that would make recommendations for how to better utilize administrative data for evidence building. Specifically, the commission would be tasked with recommending how to make additional administrative data available for evaluation and other statistical uses by Federal and outside researchers, what legislative changes are needed to facilitate such access, how to expand access while ensuring data security and fully protecting privacy and confidentiality, and how to improve data quality. As in the Ryan/Murray proposal, the commission would include executive and legislative branch appointees, selected based on their expertise in program evaluation, data analytics, data management, statistics, and privacy. The Budget proposes to fund the commission as part of the Census investment described above and to base the commission at Census to take advantage of the Census Bureau’s extensive expertise in utilizing, linking, and sharing sensitive data while protecting privacy.

Privacy, Confidentiality, and Data Security

Proposals to expand access to data can raise concerns about privacy, confidentiality, and data security. In this context, the term “privacy” includes limiting the collection of personally identifiable information to only that which is required for specific purposes. “Confidentiality” refers to protecting information against unauthorized disclosure by limiting the access and use of personally identifiable information, and “data security” refers to protecting information systems from unauthorized access, use, disclosure, disruption, modification, or destruction. Privacy, confidentiality, and data security are all high priorities for the Administration, and the Administration would not support data access changes without strong technical, legal, and policy protections to mitigate risk and prevent unauthorized use and disclosure of the data. The Census Bureau and other Federal statistical agencies adhere to a robust framework of privacy, confidentiality, and security protections governing the use of personally identifiable information, and these agencies have a long-standing and successful history of collecting, protecting, and making available in secure environments some of the Nation’s most sensitive information (including data on personal health status, immigration status, income, and proprietary business data). Federal statistical agencies are required to follow strict rules and protocols based in Federal law (described below) that include rigorous access and usage protections and other requirements to safeguard personally identifiable information and ensure its appropriate use for statistical purposes. Their

track record shows that it is possible to make extensive use of sensitive data to inform and improve public policy, while also protecting privacy, confidentiality, and security. Notably, response rates observed for Federal surveys administered by the Census Bureau and other statistical agencies have far surpassed private sector surveys, one indicator of the public’s confidence that Federal statistical agencies are able to appropriately safeguard personally identifiable information.

The statistical agencies’ successful record reflects a strong data stewardship framework,¹⁷ key features of which include:

- **Limiting access to authorized statistical use.** As explained above, statistical uses are those that, by definition, do not affect the rights, benefits, and privileges of individuals. Thus, they are inherently protective of individual privacy, since the goal is to learn about groups (e.g. participants in a given program or residents of a given community) and to release only aggregate information. Statistical agencies further limit data access to only those staff with a need to know and to authorized and approved projects.
- **Minimizing direct access to personally identifiable information.** Existing rules and protocols minimize access to directly personally identifiable information, with all users using datasets that have been anonymized to the greatest extent compatible with the intended use. For example, at the Census Bureau, a small specialized unit receives administrative data from other agencies, strips off directly personally identifiable information (e.g. names or Social Security Numbers) and replaces such information with a “protected identification key” before even other Census Bureau employees can use the file.
- **Disclosure review and severe penalties for inappropriate disclosure.** Statistical agencies require expert review of any results that will be made public to protect against inadvertent disclosure of individual information. Existing statutes also provide severe penalties for disclosure of personally identifiable information.
- **Data security standards.** Existing laws and regulations require strong technological and other safeguards for personally identifiable information. Statistical agencies implement stringent confidentiality laws and policies with a comprehensive set of physical and information technology data security practices that protect data throughout the entire chain of custody, including training for everyone who touches the data, even incidentally, firewalls within firewalls

¹⁷ For additional discussion about privacy, confidentiality, and data security protections see the “OMB implementation guidance issued for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002” and “OMB Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units.”

¹⁶ See the *Evidence-Based Policymaking Commission Act of 2014*.

to protect against unauthorized access, and secure data enclaves to tightly control the process of authorized access. As cybersecurity challenges continue to affect Federal agencies, the protection of the Government's information and information systems has become more critical in safeguarding the confidentiality of personally identifiable information. To further strengthen technological and other safeguards to protect data, the Budget funds key investments to enhance the Federal Government's cybersecurity posture including the Continuous Diagnostics & Mitigation Program, the EINSTEIN intrusion detection and prevention system, government-wide testing and incident response training to mitigate the impact of evolving cyber threats, and investments in cyber research and development to strengthen our cybersecurity defenses.

The Budget would extend this same data stewardship framework to any newly available administrative data. In particular, to the extent data would be made available to or through agencies that are not part of the Federal statistical system, these agencies would be held to the same core standards and would have to demonstrate their ability to meet them. For example, in the case of NDNH, the Department of Health and Human Services (HHS) has robust procedures to ensure that any other agency accessing the data has the required privacy and data security protections in place and has never experienced a data breach

at the Federal level. In particular, HHS reviews the other agency's security posture and those of its contractors, including computer system controls, safeguarding and oversight procedures, and administrative structure.

The commission discussed above would also be charged with making recommendations regarding new privacy, confidentiality, or data security protocols and standards that should accompany further expansions in access to administrative data.

Conclusion

The administrative data package outlined in this chapter fits into the Budget's broader emphasis on tackling challenging but important reforms that are integral to making government work better. Harnessing the full potential of administrative data can improve transparency and support efforts to hold programs and service providers accountable; allow Federal agencies to adopt private-sector best practices for using data analytics to improve performance and customer service; support ongoing innovation and experimentation, coupled with evaluations to learn what works; and permit a greater understanding of the different needs of different groups and communities. The proposals in this chapter represent a first step in this agenda and point the way to uncovering what more needs to be done.

