Barriers to Using Administrative Data for Evidence-Building[1]

I. Executive Summary

This white paper discusses barriers to the use of administrative data for evidence-building. Evidence-building, for the purposes of this paper, includes the collection, compilation, processing, or analysis of data to better understand the characteristics, behavior, or needs of groups of individuals, or communities. It excludes uses that affect the rights, benefits, or privileges of individuals, but includes a wide range of analytic uses, where only aggregated and de-identified data are made public.

As discussed in the white paper, *Using Administrative and Survey Data to Build Evidence*, both survey[2] and administrative data[3] are valuable resources for building evidence about the health and well-being of our society, economy, and environment. They can help assess the relationship or impact of Federal policies and programs on families, households, and individuals. In many instances, an optimal approach to building such evidence involves linking microdata records for program participants with other administrative or survey microdata records containing variables on critical outcomes of interest. However, a number of barriers hinder such efforts, including explicit statutory prohibitions, policy and legal interpretations, and infrastructure.

Legal barriers. Various Federal and state laws govern the use of administrative data in the U.S. on a program-by-program basis. In some cases, statutes authorizing programs permit use of administrative data for only a narrowly prescribed purpose that does not include evidence-building activities. Other authorizing statutes prohibit the development of certain datasets. More often, the patchwork nature of these statutes at both the state and Federal level inadvertently makes matching administrative data and survey data impractical or impossible.

Policy and legal interpretations. Many statutes governing programs do not explicitly address statistical and research uses of administrative data or are unclear about what uses are permitted, making them subject to interpretation. Their ambiguity, or lack of explicit authority, can lead to a variety of interpretations about permitted uses, ranging from conservative conclusions that a particular use is not allowable to the establishment of conflicting requirements. Public concerns about the use of administrative data can affect how willing agencies are to use them for evidence-building purposes, especially when such use requires legal interpretation of an ambiguous statute. The associated variation in practice can result in delays, inefficiencies, or inconsistent application within a program over time.

[1] This white paper is intended to provide the Commission on Evidence-Based Policymaking with background information on topics relevant to the Commission's work. The paper was prepared by staff from OMB, with assistance from staff at other Federal agencies.

[2] "Survey data," for the purposes of this paper, includes data typically collected through the Federal Statistical System, for statistical purposes, that generate information about broad population characteristics as well as data typically collected and associated with a specific evaluation or other study that may be started and developed directly through agency evaluation functions or indirectly by a funding recipient.

[3] "Administrative data," for the purposes of this paper, refers to administrative, regulatory, law enforcement, adjudicatory, financial, or other data held by agencies and offices of the government or their contractors or grantees (including states or other units of government) and collected for other than evidence-building purposes. Administrative data are typically collected to carry out the basic administration of a program, such as processing benefit applications or tracking services received. These data relate to individuals, businesses, and other institutions.

Office of Management and Budget (OMB) guidance[4] and individual agency efforts[5] have helped to address some of these issues, but more work can be done.

Differing state-level interpretations of both state and Federal laws may further complicate the task of using data for evidence-building. To try to harmonize these interpretations, some agencies have issued guidance to states and others on permissible methods to access certain administrative datasets under current Federal law.

Resource and capacity constraints. Even when statutes and interpretations permit the use of administrative data to build evidence, agencies can lack the capacity to use them and make them available to others. The necessary capacities include:

- The administrative infrastructure required to address data sharing or access requests, develop and implement such agreements, and implement agreements to share funds, as well as potential subject matter expertise to assist researchers and evaluators;
- Data management and curation infrastructure needed to make the data useful for evidence-building activities;
- Information technology and security infrastructure needed to meet operability and data security needs throughout the data lifecycle; and
- An information privacy infrastructure to ensure privacy and confidentiality is fully supported and risks are properly managed throughout the data lifecycle.

There are a variety of solutions to these issues, including the establishment of a formal clearinghouse for administrative and survey data. By providing a coordinated access point for various datasets and subject matter expertise or other similar resources, a clearinghouse can simplify and clarify the process by which qualified researchers may gain permission to use the data. By centralizing the functions necessary to facilitate safe and effective use of the data, a clearinghouse can better match infrastructure resources to valuable pursuits. However, there are a few instances in which large administrative datasets are already being successfully accessed through other systems, and it is not clear that these systems should be subsumed into a single clearinghouse. Designating some agencies as "hubs" may allow many agencies to leverage the significant infrastructure already in place within others, such as the Census Bureau.[6] Such an approach could lead to better use of existing data without creating unnecessary duplication of capacities.

The next section provides a brief background on the role of administrative data in evidence-building. The ensuing sections discuss and highlight in more detail many of the explicit legal barriers to using administrative data for evidence-building; the complications created by this patchwork of laws and possible ambiguity about administrative data use and access; and the constraints imposed by resource and capacity limitations on the functional access and use of data, even when other issues are resolved. Key themes associated with these barriers are summarized in the conclusion. Finally, the accompanying

[4] For guidance and tools to facilitate increased agency use of administrative data for evidence-building, see OMB Memorandum M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes* (Feb. 14, 2014), available at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

[5] See the white paper, *Using Administrative and Survey Data to Build Evidence*, for examples.

[6] The Census Bureau has extensive experience in crafting data-sharing arrangements and interpreting related statutory authorities, technical, and policy infrastructure necessary to curate large, multi-layered datasets, and capacity to provide access to qualified researchers while protecting privacy and confidentiality, as exemplified by its operation of the Federal Statistical Research Data Centers.

appendix provides a case study of how each of these barriers has limited access to wage data for evidence-building purposes.

II. Background

As discussed in the white paper, *Using Administrative and Survey Data to Build Evidence*, administrative data are a critical resource for almost every type of evidence-building. Administrative data can make it possible to track some outcomes for a longer time, more accurately, and at much lower cost than through surveys. These data are especially powerful when linked to other sources of data, given that administrative data may not contain information on crucial items of interest to a particular evidence-building effort. Smart co-locating, linking, and combining of administrative data with survey data can improve the robustness of surveys, while reducing their cost. Moreover, as there as has been a general decline in survey response rates,[7] this benefit may become increasingly important. For researchers, the vast data available in many administrative datasets allow for the possibility of groundbreaking studies.

While administrative data are playing an ever-growing role in evidence-building activities, much of their tremendous potential remains undeveloped, due to a number of barriers that may prevent access or impose heavy costs on would-be data users. As a consequence of these barriers, many valuable uses of administrative data have probably not yet been identified. To date, evidence-building using administrative data by the Federal Government and other researchers has occurred on a limited basis. While some agencies have well-established statistical and research uses for administrative data,[8] most projects have occurred on an isolated basis, conducted at high expense and burden,[9] or at one of a limited number of data enclaves that have the infrastructure to address barriers.

III. Legal Barriers Prevent or Complicate Reusing Administrative and Survey Data to Build Evidence

Various Federal and state laws govern the use of administrative data in the U.S. on a program-by-program basis. Other countries, including Australia and Canada, have explicitly adopted cross-cutting laws and policies that protect privacy while also creating a clear expectation that collected data are a strategic resource for evidence-building to further the public good.[10] Elements from these approaches, which include provisions for data sharing authority and coordination, could be explored in the U.S. context.

Some authorizing statutes explicitly prohibit agencies from sharing certain data with one another, from making such data available for evidence-building purposes, or even from routinely re-using their own data for such purposes. Sometimes these prohibitions reflect long-standing statutes that have not been

---

[7] See, for example, Baruch, Y. and Holtom, B.C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139-1160.

[8] For more information, see the white paper, *Overview of Federal Evidence-Building Efforts*.

[9] See, for example, Prell, M. et al. (2014). Profiles in Success in Statistical Uses of Administrative Data. Federal Committee on Statistical Methodology (FCSM). Available at https://fcsm.sites.usa.gov/files/2014/04/StatisticalUsesofARData.pdf.

[10] The Australian Bureau of Statistics Act (https://www.legislation.gov.au/Details/C2014C00618) says that one function of the Australian Bureau of Statistics is to "ensure co-ordination of the operations of official bodies in the collection, compilation and dissemination of statistics and related information, with particular regard to . . . the maximum possible utilization, for statistical purposes, of information, and means of collection of information, available to official bodies."

The Canadian Statistics Act (http://laws.justice.gc.ca/eng/acts/s-19/fulltext.html) says, "[a] person having the custody or charge of any documents or records that are maintained in any department or in any municipal office, corporation, business or organization, from which information sought in respect of the objects of this Act can be obtained or that would aid in the completion or correction of that information, shall grant access thereto for those purposes to a person authorized by the Chief Statistician to obtain that information or aid in the completion or correction of that information."

These countries also have Federal policy that elaborates on these principles.

updated for modern technology and data analytic techniques. In other cases, they may reflect concerns about privacy and confidentiality that need to be fully addressed in any proposal to expand data access.

Some of the data valuable for building evidence about Federal and state policies and programs reside at the state level. Obtaining these data when there is no direct prohibition still requires negotiation with states on an individual basis when there is no explicit Federal authority and practice of acquiring them routinely. This may mean that using administrative data for evidence-building becomes so complicated as to be prohibitively costly for agencies or other researchers. This situation also makes it costly in time and money for states, since they are often responding to multiple individual requests for administrative data.

1. *Direct Statutory Prohibitions on Access Prohibit the Reuse of Data for Evidence-Building*

For some datasets, access and use are explicitly restricted by statutes that limit authorization to a narrow set of uses. In some cases, concerns about privacy and confidentiality motivate these restrictions, and such concerns would need to be fully addressed in any proposal to expand access to the data. For example, and as described more fully in the appendix, the National Directory of New Hires (NDNH)—a national database of new hire, quarterly wage, and unemployment insurance (UI) claim information—was established to help states administer the Child Support Enforcement program. These data are also useful for a wide variety of evidence-building activities, and several programs are currently using them successfully for such purposes while protecting privacy and confidentiality.[11] However, as explained by the Department of Health and Human Services (HHS), NDNH data may only be used for evidence-building purposes if there is an explicit statutory authority to do so, and existing authorities appear fairly limited. For example, under current statute,[12] NDNH data may be accessed to evaluate the employment and earnings outcomes of non-custodial parents (since increased earnings will result in increased child support collections), but NDNH data cannot be used to evaluate a broader job training program which does not serve populations related to the Temporary Assistance for Needy Families (TANF) or Child Support programs, even though the earnings and employment data are originally collected by the UI system. [13]

2. *Direct Prohibitions on Data Collection Limit the Availability of High-Value Data*

In some cases, the Federal Government is explicitly prohibited from collecting certain types of data, even if such collection would lead to more efficient and effective program administration, reduce burden on states and others, and/or better-facilitate evidence-building activities. In some cases these restrictions reflect out-of-date statutes, while in others they reflect legitimate privacy concerns that should be addressed. For example, the Workforce Investment Act (WIA) did not permit its authorities to be construed "to permit the development of a national database of personally-identifiable information on individuals receiving [WIA] job training services."[14] Under this provision, the Department of Labor cannot create a national database of personally-identifiable information unless it is for program management activities. The 2014 reauthorization of WIA (the Workforce Innovation and Opportunity Act or WIOA) continued this limitation,[15] despite the fact that WIOA recognized that "[performance]

[11] For illustrative examples of research and evaluations that have used NDNH and other sources of quarterly wage data, see Durham, C., & Wheaton, L. (2012). Investigating Alternative Sources of Quarterly Wage Data: An overview of the NDNH, LEHD, WRIS, and ADARE. Urban Institute. Available at http://www.urban.org/sites/default/files/alfresco/publication-pdfs/412688-Investigating-Alternative-Sources-of-Quarterly-Wage-Data-An-Overview-of-the-NDNH-LEHD-WRIS-and-ADARE.PDF.

[12] 42 U.S.C. § 653(j).

[13] *Ibid*.

[14] Workforce Investment Act, sec. 504(b).

[15] Workforce Innovation and Opportunity Act, sec. 501.

reporting and evaluation requirements are important tools in measuring effectiveness, especially for the core [WIOA] programs."[16]

3. *Inconsistencies in Privacy Protections Can Make Evidence-Building Impractical*

Some statutes offering privacy and confidentiality protections, such as the Privacy Act, cover many datasets from many agencies. Some statutes cover data from a specific agency or program. While each set of protections may be entirely appropriate, the sectoral nature of Federal data protections leads to different requirements and practices. Bringing data together from two or more sources can be complicated by their differences. For example, statistical data are collected under a pledge of statistical confidentiality, which promises data will be used only for statistical purposes, including evidence-building, and that access to the data in identifiable form will be limited to agency staff or agents, absent consent.[17] Some administrative datasets, like Federal tax data, have similar prohibitions on transferring their data to another agency, or such transfers would be impractical given their size.[18]

4. *Variations in State Law May Make It Infeasible to Acquire or Use Data to Build Evidence*

Much of the highest-value administrative data associated with Federally-funded programs are collected and maintained at the state level. These data are subject to state laws, which may vary between states and between different programs within a state. For example, while many states have similar statutory definitions of confidential data, the associated confidentiality laws vary. Some states include very explicit language regarding to whom and for what purposes confidential data may be disclosed, while others use language that may leave greater room for interpretation.[19] Similarly, within a single state, some administrative datasets may be available for evidence-building purposes, while others are not. These state-level barriers may not apply when there is explicit Federal statute stating that these data must be available for statistical or other evaluative purposes.[20]

IV. Policy and Legal Interpretations

Many program authorizing statutes do not explicitly address statistical and research uses of administrative data and could address permissible uses with greater clarity; those statutes thus could be subject to interpretations that could vary across states and programs. This variation in interpretation ranges from conclusions that statistical and research uses are not permissible to the implementation of

---

[16] 160 Cong. Rec. S3982-S3990, *Statement of the Managers to Accompany the Workforce Innovation and Opportunity Act*, daily ed. June 25, 2014.

[17] Title V of the E-Government Act of 2002, known as the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), Pub. L. 107-347, § 502(8).

[18] 23 U.S.C. § 6103.

[19] For a more complete discussion of these factors as they apply to state-held education and wage data, see Poole, K., & Springer, H. (2015). *Balancing Confidentiality and Access: Sharing Employment and Wage Data for Policy Analysis and Research.* Center for Regional Economic Competitiveness, available at http://www.lmiontheweb.org/download/2015-05/Report--_Data_Confidentiality_and_Sharing_-_CREC-LMI_Institute_-_May_2015.pdf.

For a discussion of these issues in the context of health and vital records, see Lee, H., Warren, A., & Gill, L. (2015). *Cheaper, Faster, Better: Are State Administrative Data the Answer? The Mother and Infant Home Visiting Program Evaluation-Strong Start Second Annual Report*. OPRE Report 2015-09. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, available at http://www.acf.hhs.gov/programs/opre/resource/cheaper-faster-better-are-state-administrative-data-the-answer-the-mother-and-infant-home-visiting-program-evaluation-strong-sta.

[20] Note that Federal statutes or regulations that allow *permissive* disclosures leave the decision about whether to allow data to be used up to state discretion (as is the case for evidence-building purposes under the Unemployment Insurance statute and associated confidentiality regulations). Federal statutes that apply to federally held data preempt state law, as do Federal statutory *requirements* for state-level disclosure. Requirements that states provide data unless such provision interferes with program operation are more flexible.

different access and use requirements among programs. The Federal Government has already taken steps to address some of these barriers, for example by issuing OMB Memorandum M-14-06 (described in further detail below). However, Executive Branch guidance occurs within the existing legal framework, so its impact will be limited in scope relative to a comprehensive legal framework.

1. *Ambiguities in Statute Lead to Inconsistent Access Policies*

In many cases, Federal and state laws are silent or imprecise about authority to provide access to data for evidence-building purposes. In these cases, agencies may not have policies related to when data may be accessed for evidence-building purposes, or they may develop regulations, policies, and legal interpretations that further restrict access. These structures may be inconsistent among agencies and programs and may vary over time.

Default to "no." Even when data sharing is legally allowed, agencies may be unsure about the rules and may therefore assume that data cannot be shared. Changing this default position can be a substantial undertaking. For example, it took many years of discussion with the Census Bureau for the Department of Agriculture (USDA) to formally determine that it could use individual-level Supplemental Nutrition Assistance Program (SNAP) program records for research and statistics under its "program administration" authority. In 2011, USDA concluded that since statistical reports and analyses furthered effective program administration, such uses were permitted.

In this instance, USDA interpreted "program administration" in a way that allowed the Census Bureau to approach the states to obtain the administrative data. This interpretation provides one example of how to resolve what may seem to be an initial inconsistency between "administrative purposes" and "statistical purposes." A comprehensive interpretation of "program administration" includes not just the program's day-to-day administrative operations but also the far-ranging data analyses and policy discussions undertaken by program officials and managers as they develop policies and recommendations to administer the program. Thus, effective program administration rests on having the best evidence available, including the type of analytical products produced by the Census Bureau and the other statistical and evaluation agencies and offices, about social and economic conditions, program operations, and impacts of past policies and proposed policy changes.

OMB issued M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes*, in February 2014 to help both program and statistical agencies better leverage administrative data for statistical purposes that can inform evidence-based policymaking.[21] The memorandum builds on earlier guidance designed to increase the value of existing data by creating "a presumption in favor of openness to the extent permitted by law and subject to privacy, confidentiality, security, or other valid restrictions."[22] 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 sets forth policy designed to assist agencies in overcoming barriers to using administrative data for statistical purposes created by inertia and confusion by addressing a host of legal interpretation, policy,

[21] OMB Memorandum M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes* (February 14, 2014), available at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

[22] OMB Memorandum M-13-13, *Open Data Policy – Managing Information as an Asset* (May 9, 2013), available at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

interagency agreement, and data quality challenges.[23] OMB has been working with the Federal evidence-building system to facilitate new agency collaborations under the M-14-06 framework.[24]

Ad hoc reviews. In some cases, agencies or programs may not have established clear regulations, policies, or procedures to share data for evidence-building when statutes are not clear or precise enough. In these cases, it is common for each data access request to be handled as an isolated case, which could lead to extensive delays in decision-making and inconsistent interpretations of statute and policy. These cases can require extensive and time-intensive negotiations between the data requestors and the data owners, thereby creating a great deal of uncertainty in outcomes.

Some agencies have taken steps to standardize policies and adopt streamlined structures to facilitate data access for evidence-building purposes. In such cases clarity about access rules can be increased, thereby significantly reducing the time it takes to access data. For example, the Department of Housing and Urban Development (HUD) Office of Policy Development and Research (PD&R) has established a data licensing process that allows qualified research organizations to access HUD data for the explicit purpose of conducting innovative research that will inform HUD's policies and programs. This process is standardized, streamlined, and includes privacy, confidentiality, and data security protections.[25] With such a standardized process, requests for HUD data access are typically reviewed and approved (or denied) within two months.

However, given the large datasets often involved in statistical work and numerous laws governing access to data, even agencies with standardized processes often must perform case-by-case reviews of requests to access agency data. Building the capacity to manage these processes is a challenge. This is touched upon below in the discussion on administrative policy infrastructure.

2. *State-by-State Variations in Legal Regimes*

States may vary in their interpretations of Federal and state laws affecting administrative data access and their ability and willingness to allow access to their administrative data. For example, even with USDA's agreement that individual-level SNAP data, received from states, can be used for statistical research, the Census Bureau (or any other agency seeking SNAP data) must devote significant staff resources to obtain the data. Since USDA does not require states to submit their full individual-level databases—there is no centralized repository for these data—the Census Bureau must negotiate separate access arrangements on a state-by-state basis. Moreover, individual states may vary in the extent to which they are amenable to such agreements. As a result, assembling a dataset providing standard and complete coverage of all states will require significant resources spanning several years and utilizing thousands of person hours. This cost, in both time and resources, often makes it infeasible for researchers and evaluators to utilize administrative data.

Similarly, differing state-level interpretations of Federal privacy and confidentiality statutes can lead to confusion about what is permissible. For example, the Family Educational Rights and Privacy Act (FERPA) has long confused states, localities, and researchers about whether and how it is permissible to use

---

[23] The policy includes: (1) fostering collaboration across program agencies and statistical agencies and components, (2) implementing data stewardship policies and practices that anticipate statistical uses of program data, (3) creating and making available for statistical agencies and components well-documented information on quality control measures and key attributes of the data, and (4) creating IAAs to designate responsibilities and practices between the program agencies and agencies serving statistical purposes.

[24] See *Building the Capacity to Produce and Use Evidence*, FY 2017 Analytical Perspectives, available at https://www.whitehouse.gov/sites/default/files/omb/budget/fy2017/assets/spec.pdf.

[25] For more information on the PD&R data licensing process, see https://www.huduser.gov/portal/research/pdr_data-license.html.

personally-identifiable information (PII) from student education records for evidence-building purposes. Without a clear understanding of how these data may be used in compliance with Federal law, some states and localities have decided to avoid using such data for evidence-building purposes. To address this challenge, the Department of Education (ED) continues to develop and disseminate a number of tools to help states, localities, and researchers better understand the permissible use of PII under FERPA.[26] Additionally, the office in ED that administers FERPA, the Family Policy Compliance Office, routinely responds to inquiries from officials from state departments of education, local school districts, and individual schools on FERPA's application to researcher use of student data.

3. *Public Acceptance*

Program agencies are appropriately concerned that they maintain public acceptance and trust in their handling of data collected for program administration purposes. For this reason, particularly where statutes are not explicit, they may be inclined to interpret data provision statutes with caution. Similarly, studies have documented that data owners may be wary of the potential political consequences resulting from external research if they make their data available and thus limit access to it.[27] This posture can limit access to important sources of administrative data, especially to those identifiers that facilitate linking and combining data, in instances where such access may be appropriate and feasible.

Federal evidence-building efforts are affected by heightened privacy concerns, and the associated offices take seriously the mission to protect privacy in the collecting and handling of information. Because these agencies and offices collect most information from the public on a voluntary basis, they depend critically on public trust to fulfill their missions. The production of credible statistics to support program and policy decisions is critically dependent on the willingness of potential respondents to participate in a survey. Any loss in this trust could harm public perception of the accuracy, objectivity, or integrity of Federal evidence-building efforts and the validity of measures used to monitor and assess performance, progress, and needs. Along these lines, data breaches in the public and private sectors, increased surveillance in an age of concern about safety and national security, general decreases in civic engagement, and reduced trust in the government create challenges in maintaining public trust when acquiring or using identifiable or other sensitive information, even when it is used for legally authorized, socially beneficial purposes.

The Census Bureau, in coordination with the Interagency Council on Statistical Policy, conducted a series of studies on public attitudes towards the use of administrative records. Consistent with prior research, the results indicate that the public's support for greater use of administrative data is somewhat contradictory, highlighting the importance of appropriately explaining how the data will be used. Context-neutral questions garner relatively negative responses, but questions highlighting the benefits of administrative data use elicit solid support for the practice. This difference is perhaps understandable given the relative lack of knowledge about the benefits among a majority of respondents. The benefits eliciting the most support were the potential to save the government money and the advancement of

---

[26] See, for example:

U.S. Department of Education. (2016). *Data Sharing Toolkit for Communities: How to Leverage Community Relationships While Protecting Student Privacy, Washington, D.C*. http://www2.ed.gov/programs/promiseneighborhoods/datasharingtool.pdf.

U.S. Department of Education. *Privacy Technical Assistance Center*. http://ptac.ed.gov/toolkit.

[27] See Poole, K., & Springer, H. (May 2015). *Balancing Confidentiality and Access: Sharing Employment and Wage Data for Policy Analysis and Research*. Center for Regional Economic Competitiveness, available at http://www.lmiontheweb.org/download/2015-05/Report--_Data_Confidentiality_and_Sharing_-_CREC-LMI_Institute_-_May_2015.pdf.

the public good. The biggest concerns were privacy and confidentiality. Overall, this research suggests a key role for public education and information dissemination as administrative data use expands.[28]

V. Resource and Capacity Constraints

A third type of barrier that agencies face is the size and complexity of the capacity (i.e., infrastructure) required to support the use of administrative data.[29] In particular, agencies need legal, administrative, data management/curation, IT, and security infrastructure to carry out evidence-building activities coherently, effectively, and securely. Because providing access to data is integral to the missions of principal statistical agencies, each invests in statistical access and dissemination tools and modalities, such as data enclaves. Yet, they also need to develop infrastructure to acquire administrative data. Many program agencies, by contrast, have not made comparable investments given their different mission focus.

Inadequate infrastructure caused by aging legacy systems, bottlenecks, inefficiencies, or misaligned incentives can hinder data access to both initial and ongoing users. While this section focuses on agency capacity, capacity is also a concern for external data users. They need capacity to develop successful data access requests, safeguard data once they are obtained, conduct rigorous analyses, and properly destroy the data.

1. *Administrative Policy Infrastructure*

Agencies need infrastructure, such as trained personnel, formal written procedures, and information technology tools, to consider data sharing requests and to develop and implement data sharing agreements. Some agencies lack formal procedures for handling data requests, while others are able to improvise or leverage infrastructure designed for other purposes. Agencies that process large volumes of data sharing requests may require more sophisticated infrastructure than those who only process occasional agreements.

A few agencies have invested in infrastructure to process requests for researcher access to administrative data. For example, the Centers for Medicare & Medicaid Services (CMS) processes thousands of data use agreements annually, providing valuable data access to Federal agencies, Federal grantees, and other approved researchers. The Social Security Administration (SSA) created an Office of Data Exchange to standardize and centralize incoming requests for access to its data. As described above, HUD's PD&R has developed a process for data licensing agreements with researchers who conduct projects that inform HUD's work. Finally, due to the high volume of requests for access to administrative data that it receives, the Census Bureau has an entire staff who negotiate agreements for data access with a variety of Federal, state, and other sources. Other tools that these agencies have developed to reduce the burden associated with processing data requests include preliminary assessment tools to determine if a request is eligible, standard data sharing agreement templates, agreement duration extensions (as opposed to renegotiations), and tools to enhance data user knowledge about the data sharing request process and requirements. Other agencies currently lack comparable infrastructure.

---

[28] See Hunter-Childs, H., King, R., & Clark-Fobia, A. (2015). Confidence in US Statistical Agencies. *Survey Practice* 8(5). Available at http://www.surveypractice.org/index.php/SurveyPractice/article/view/314/html_38 and King, R. et al. (2013). Attitudes Toward Using Administrative Records. *AAPOR* 2013. Available at http://www.amstat.org/sections/srms/proceedings/y2013/files/400285_500783.pdf.

[29] For the purposes of this paper, infrastructure includes everything related to the ability of data owners to use and make data available, including staff resources, software and hardware, physical data security, policies and procedures governing administrative data use and accessibility, etc.

Before data move from one agency or office to another or to an external entity, a set of policy and procedural safeguards must be established (typically via written agreement) to certify appropriate data protection and acceptable use. Negotiating the permissions for a given use and user of an administrative dataset can be cumbersome, often requiring a significant investment of staff resources over a period of years.[30] These negotiations are necessary because there are a host of critical obligations that must be addressed as part of careful data stewardship, such as specific methods of information transmission; acceptable levels of encryption and the availability of encryption tools; record keeping, retention, and disposition requirements of the data; and protocols in the event of a breach of confidentiality or a violation of stated disclosure policies.

This negotiation capacity is just one part of a larger legal and administrative infrastructure that is necessary to manage administrative data use on an ongoing basis. In cases where these requirements can be standardized (as in the case of the PD&R data licensing agreement structure), negotiations may be significantly expedited. Monitoring of agreements and associated reporting and record-keeping demand ongoing attention and, sometimes, a dedicated staff (as is the case of the CMS data use agreement structure). Requests, notifications, and other communications need to be coordinated, and consistent governance policies need to be developed and implemented.

2. *Data Management and Curation Infrastructure*

Because administrative data were not originally collected for evidence-building purposes, agencies need to perform a number of actions to make them suitable for these uses, including extracting and transferring files, "cleaning" data and converting them to standardized formats, assessing data quality, conducting preliminary analyses, and integrating data from many sources. These actions, which require a significant technical infrastructure, all fall under the category of *data curation*. Data quality barriers can sometimes be severe, in particular, because some applications of administrative data may not be related to the originally intended use. Consequently, assessments of data quality for a particular application can require a great deal of specialized skill.

Data curation also includes the maintenance of relevant meta-data and other documentation. Prospective data users need at least minimal documentation to assess fitness for a given use, and more detailed documentation enables more informed and appropriate use. For administrative data, documentation must often be generated from scratch, as it is not always available from the provider. When two datasets are linked, new documentation is needed to adequately explain the statistical properties of the linked dataset, such as what percentage of the records could be successfully linked from the incoming files. This task falls on the statistical agency or research team that creates the new, linked dataset.

As with legal and administrative infrastructure, the availability of data curation infrastructure varies across the government. Some Federal agencies operate on a small scale and are able to improvise or leverage infrastructure designed for other purposes, while a few agencies have developed robust infrastructure to address the large-scale provision or acquisition of data. Some Federal agencies, in particular the Census Bureau, have built requisite information technology and security infrastructure to ensure the proper management of large and complex universe (i.e., whole population) files and programs that have sufficient computing power for statistical matching of data. However, even agencies with adequate or robust physical infrastructure may not have the appropriate staff resources to execute requests on a timely basis. In addition, they may not consider these resources for this purpose as a critical part of their missions.

[30] See *Building Evidence with Administrative Data,* FY 2016 Analytical Perspectives, available at https://www.whitehouse.gov/sites/default/files/omb/budget/fy2016/assets/spec.pdf.

When administrative data are acquired from multiple states, the need for data curation infrastructure is especially acute. The challenges in acquiring data state-by-state and integrating files are multiplied by the number of separate sources. CMS is a rare example of an agency that requires all states to provide individual-level administrative data on Medicaid program recipients to the Federal Government in order for states to be reimbursed. Even so, there has historically been much variation in the timeliness, format, and comparability among states, necessitating significant "back-end" curation by CMS to make the data fit for analytical purposes.[31] Similarly, the National Center for Health Statistics has worked with state vital registrars for decades, providing grants and other payments to facilitate more consistent recordkeeping and automation practices in the handling of birth and death records. Where state-by-state acquisition of data is not a routine Federal activity, there will often be even wider variability by state.

3. *Information Technology/Security Infrastructure*

Agencies supporting administrative data use also need an information technology infrastructure that fulfills operability and data security needs throughout the data lifecycle. Specifically, agencies need the capacity to acquire, utilize, and link data and infrastructure to make data accessible to others for evidence-building purposes. These information technology infrastructure needs vary with the sophistication of the data linkages, the amount of data, and how sensitive the data are. All programs need the core IT capacity to safely maintain their own administrative data.

For example, agencies may need to maintain specialized data acquisition interfaces for data from multiple sources, such as states. As the volume of data and the sophistication of data linkages increase, agencies will need increased server capacity in order to effectively use the data. In addition, to facilitate data provision to internal and external researchers, agencies need to establish secure and controlled modes of access such as those described in the next section. Finally, working with sensitive individual identifiers presents a particular challenge, and some agencies have developed processes to replace them with random identification keys. This requires a substantial investment to develop and sustain a linkage infrastructure that can support accurate linkages from multiple and varied sources on a continuous basis.

Due to the demands of meeting these information technology requirements, few agencies have made such investments. Access to administrative data and related linked files is therefore quite limited. However, several agencies have developed effective and robust procedures to provide restricted access to a variety of datasets through data enclaves, including licensing agreements.[32] In particular, the Census Bureau has invested in the Federal Statistical Research Data Center (FSRDC) infrastructure to permit researcher access to data maintained by the Census Bureau, the Agency for Healthcare Research and Quality, and the National Center for Health Statistics via 23 physical locations around the country. The Census Bureau is also helping expand the capability of the FSRDCs to accommodate access to datasets from additional agencies. Use of this common infrastructure provides an increased capacity for internal and external researcher access to data, including to those that have never before been co-located or linked.

4. *Information Privacy Infrastructure*

Agencies supporting administrative data use also require an information privacy infrastructure to ensure that privacy and confidentiality are fully supported, and that privacy risk is properly managed throughout the data lifecycle. Specifically, agencies need the capacity to ensure compliance with all

[31] Additionally, this system – the Transformed Medicaid Statistical Information System – may cause states to incur IT infrastructure and work force costs.

[32] Note that this solution is only available when the legal authority exists for data access to occur.

applicable privacy laws and policies regarding the proper creation, collection, use, processing, storage, maintenance, disclosure, dissemination, and disposal of personally-identifiable information when using administrative data.

The use of administrative data to build evidence may involve, among other things, the collecting, sharing, co-locating, linking, and combining of data. When the use of administrative data involves personally-identifiable information and/or confidential business information, these activities raise sophisticated, and sometimes unique, privacy and confidentiality issues, ranging from questions regarding maintenance responsibilities, to access and disclosure restrictions/limitations, to de-identification and potential re-identification of data. Protecting individual privacy and confidentiality in the handling of administrative data promotes public trust and is essential to the success of the evidence-building mission. A sound information privacy infrastructure is therefore essential.

Recently, the President issued an Executive Order establishing a Federal Privacy Council,[33] which has already begun to meet.[34] The Federal Privacy Council is the principal interagency forum to improve the Government privacy practices of agencies (and entities acting on their behalf) and support the work of agency privacy officials. Its responsibilities include developing recommendations for OMB on Federal Government privacy policies and requirements; coordinating and sharing ideas, best practices, and approaches for protecting privacy and implementing appropriate privacy safeguards; assessing and recommending how best to address the hiring, training, and professional development needs of the Federal Government with respect to privacy matters; and performing other privacy-related functions, consistent with law, as designated by the Chair. In the course of its work, the Privacy Council will coordinate closely with the Interagency Council on Statistical Policy (ICSP)[35] with regard to intersections between privacy practices and statistical data protection.

VI. Summary of Themes

One of the most fundamental barriers to using administrative data relates to statutory challenges resulting from the variety of statutes and interpretations that govern the use of administrative data. Different restrictions apply to different datasets, making it difficult to create linkages between them. In some cases, the agencies that are best equipped to perform and facilitate data analysis are not permitted access. Moreover, there is often no affirmative authorization for building evidence, and in cases where authorization is articulated, it may not always be well-matched to the need or aligned with other evidence-building provisions. As a result, the path to gain access for a given use can be circuitous and resource-intensive. In projects where collaboration between different programs, agencies, governments, or other entities is necessary, a lack of standardization can compound the difficulty. Efforts are underway to streamline the complex and onerous processes needed to navigate the existing legal landscape in a decentralized manner, but more can be done.

---

[33] Exec. Order No. 13719, *Establishment of the Federal Privacy Council*, 81 Fed. Reg. 7685 (Feb. 12, 2016), available at https://www.gpo.gov/fdsys/pkg/FR-2016-02-12/pdf/2016-03141.pdf.

[34] More information about the inaugural meeting of the Federal Privacy Council is available at https://www.whitehouse.gov/blog/2016/03/12/federal-privacy-council-holds-inaugural-meeting-0.

[35] The ICSP is made up of the heads of the 13 principal statistical agencies spanning nine cabinet departments and two other agencies, as well as representatives from other evidence-building programs. It provides advice and counsel to OMB on relevant statistical matters and is a primary vehicle for coordinating cross-cutting statistical work and information exchange about agency programs and activities. For more information about the ICSP, see Chapter 4 of OMB's annual report to Congress, *"Statistical Programs of the United States Government, Fiscal Year 2016,"* available at https://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2016.pdf.

As described in this paper, agencies need expertise and costly infrastructure to perform activities that enable the most sophisticated data use. Such activities include performing all of the necessary administrative functions, curating accessible databases, and facilitating researcher access in a way that protects privacy and confidentiality. Each is subject to at least some economies of scale—multiple projects can share resources, and tasks are more easily carried out as they are repeated. In fact, the Fiscal Year 2016 enacted budget invested $10 million in the Census Bureau for the development of a "more comprehensive infrastructure for linking, sharing and analyzing key datasets,"[36] including improving the infrastructure by which researchers beyond the Census Bureau access data.[37]

There are a variety of potential solutions to these issues, including the establishment of a formal clearinghouse for administrative and survey data. By providing a coordinated access point for various datasets, subject matter expertise, or other similar resources, a clearinghouse can simplify and clarify the process by which qualified researchers may gain permission to use the data. By centralizing the functions necessary to facilitate safe and effective use of the data, a clearinghouse can better match infrastructure resources to valuable pursuits. However, there are a few instances in which large administrative datasets are already being successfully accessed through other systems, and it is not clear that these systems should be subsumed into a single clearinghouse. Designating some agencies as "hubs" may allow many agencies to leverage the significant infrastructure already in place within others. These are important factors to consider as lessons are learned from the Census Bureau's infrastructure build out and as the Commission on Evidence-Based Policymaking ("the Commission") considers the statutory and policy frameworks affecting the allocation of responsibilities for maintaining data.

As the Commission studies the barriers to building evidence and devises recommendations, it will undoubtedly pay careful attention to the issues of data security and privacy protection, and the ways in which the public perceives and understands them. When the public lacks knowledge about the benefits of evidence-building or is worried about privacy and security concerns, it can exacerbate one or more of the barriers described above (e.g., by encouraging agencies to default to conservative data use policies). Increased public trust about data protection and belief in the value of using data could make the public more receptive to any legislative and policy recommendations that the Commission may make.

The questions before the Commission require complex analysis and creative problem-solving. Fleshing out the concept of a clearinghouse provides one promising way forward.[38] There is a wealth of resources available in the accumulated experience of Federal evidence-builders, including principal statistical agencies, evaluation offices, and program offices.

[36] See *Building Evidence with Administrative Data,* FY 2016 Analytical Perspectives, available at https://www.whitehouse.gov/sites/default/files/omb/budget/fy2016/assets/spec.pdf.

[37] The evolution of the Census Bureau Research Data Centers to the FSRDCs foreshadows some initial work to understand these considerations as the Federal evidence-building system explores a shared governance model.

[38] Sec. 4(b) of the Evidence-Based Policymaking Commission Act of 2016, Pub.L. 114-140.

Appendix

Federal Data on Personal Income: An Illustration of Barriers to Effective Use of Administrative Data

This appendix provides a case study of how barriers have limited access to a particular type of data for evidence-building purposes. Earnings and income records are among the administrative data in highest demand. The universal (or near universal) coverage of certain sources make these records an ideal outcome variable for studies of a wide range of job training, education, criminal justice, and place-based interventions.

Six main administrative sources of Federal income data exist, three through the Unemployment Insurance (UI) program and three through Federal Tax Information (FTI). Each administrative data source has unique limits on access, quality, and completeness that affect eventual use in evidence-building. These six sources are summarized in the table below. Use of UI data with wage information is primarily limited by access constraints to repositories of data collected by states. Establishing systems and protocols to report and maintain nationally-consistent UI data may require new legislative authority and/or administrative processes. In contrast, FTI use and disclosure is limited by statutory restrictions.

I. UI Wage Data

Administrative data for state UI programs are available in multiple forums, each with unique barriers: originating state systems, the Census Longitudinal Employer-Household Dynamics (LEHD) Program, and the National Directory of New Hires (NDNH).

1. *State Unemployment Insurance Program*

Each state maintains its own UI data system, subject to state laws and Federal regulations that protect the data from unauthorized access. These state-held data are limited to those persons or employers with claims or tax data in that particular state, but the wage data do not include information on Federal employees, active duty service members, postal workers, or the self-employed. The Department of Labor (DOL) does not have a central repository of state-level data, including for research purposes. DOL cannot compel states to share the existing administrative data for most purposes because the UI statute allows permissive disclosure, but leaves the decision about whether to allow data to be used up to state discretion.[39] The Workforce Innovation and Opportunity Act includes a provision that requires states to cooperate with evaluations led by DOL and the Department of Education, including through the provision of data when practicable.[40] Under Federal regulations, a state is only permitted to disclose these data when it would not interfere with the efficient administration of its UI law, and when the state's law allows for disclosure.[41] In order to access these data for evaluations, DOL develops agreements with individual states, a burdensome and costly task in the absence of a centralized and accessible repository for non-UI evaluations. This process is especially burdensome, as each new project requires a new agreement with the affected states. For large-scale evaluations, the costs of obtaining wage data from states can exceed $1 million per study;[42] these costs recur with each project as each negotiation is treated as an ad hoc study. External researchers may also access data in this way on a state-by-state and case-by-case basis. When states provide access to UI data, they recoup their costs of making the disclosure from the data requestor.[43]

---

[39] 20 C.F.R. Part 603.5.
[40] Sec. 116(e)(1) of Workforce Innovation and Opportunity Act (WIOA), Pub.L. 113-128.
[41] 20 C.F.R. Part 603.5.
[42] See, for example, Government Accountability Office Publication No. 06-766, "New Hires Data Has Potential for Updating Addresses of Convicted Sex Offenders" (July 31, 2006), available at http://www.gao.gov/new.items/d06766.pdf.
[43] 20 C.F.R. Part 603.10.

Major Administrative Sources of Income Data

Data Source	Measure	Universe	Content[1]	Years of Available History	Frequency	Prepared Analytical Files	Permitted Users/Uses	Current Use in Evidence-Building[2]
UI data								
(1) State UI	Wages only	self-employed and Federal workers excluded	Comprehensive	Few	Quarterly	N	State-specific laws restrict access for most purposes.	Limited
(2) NDNH	Wages only	self-employed workers excluded	Comprehensive	Few	Quarterly	N	Law restricts mainly to ACF/HHS program administration[3] (including some evaluation).	Limited
(3) LEHD	Wages only	self-employed workers excluded	Limited	Many	Quarterly	Y	State-specific agreements govern use.	Regular
Federal tax information								
(1) IRS	Most earnings[4]	All	Comprehensive	Many	Annual	Y	Restricted to use by IRS for tax administration (including related statistical activities).	Limited
(2) SSA	All earnings	All	Limited	Many	Annual	Y	Restricted to use by SSA for its administration of the Social Security Act (including related statistical activities), tax administration purposes, and other purposes authorized by the Internal Revenue Code or IRS.	Limited
(3) Census	All earnings	All	Limited	Many	Annual	Y	Restricted to use by Census Bureau for Census Bureau programs.	Limited

[1] Comprehensive content contains most or all of the collected variables; Limited content contains only a subset of collected variables.

[2] Limited use denotes a sporadic or *ad hoc* history of use for Federal evidence-building beyond the named agency; Regular indicates that such use is routine.

[3] Note: other Federal agency research projects may be granted access if the work can be shown to further a child support enforcement or TANF purpose.

[4] Earnings for workers whose incomes are below the filing threshold are not comprehensive.

2. *Census Bureau Longitudinal Employer-Household Dynamics Program*

Longitudinal Employer-Household Dynamics (LEHD) program data are the result of a partnership between the Census Bureau and states to provide high-quality local labor market information and to improve the Census Bureau's economic and demographic data programs. LEHD data are based on different administrative sources, primarily UI data and the Quarterly Census of Employment and Wages, censuses, and surveys. Firm and worker information are combined to create job-level quarterly earnings history data, data on where workers live and work, and data on firm characteristics, such as industry.

The Census Bureau acquires UI data from 49 states, the District of Columbia, and Puerto Rico for the LEHD program. As there is no Federal requirement for states to provide these data to the Census Bureau, the agency had to negotiate with each individual state, a process that took more than ten years.[44] In addition, each agreement has to be negotiated with each state every five years. These agreements and data, which are subject to state laws in addition to Federal laws, permit the Census Bureau to use the data for the LEHD program and other approved statistical uses. Researcher access to LEHD data is available through the network of secure Federal Statistical Research Data Centers (FSRDCs), although less than one half of all states permit their data to be accessed for general research purposes through these centers. Such access requires that the project demonstrate a likely benefit to Census Bureau programs, which cover a broad range of demographic and economic measurements, as well as the development of methodologies to support Census Bureau programs.[45]

The output from all research projects must undergo and pass a disclosure review.[46] In addition to the restricted-use datasets available through FSRDCs, LEHD creates public-use datasets (the Quarterly Workforce Indicators, LEHD Origin-Destination Employment Estimates – OnTheMap, and Job-to-Job Flows) and online tools. Under the LEHD partnership, states do not receive financial compensation from the Census Bureau for access to UI data. Rather, they receive state and sub-state public-use data products in return for the provision of the data.

3. *Health and Human Services (HHS) National Directory of New Hires Program*

The NDNH is part of the Federal Parent Locator Service maintained by the Office of Child Support Enforcement (OCSE) within the HHS Administration for Children and Families (ACF). NDNH includes information on newly hired employees, quarterly wages, and unemployment insurance. NDNH does not contain information on self-employed individuals. The data are obtained from state workforce and employer new hire databases.

Access to NDNH data are explicitly limited by statute.[47] Data may be accessed without identifiers for research purposes found to be likely by the HHS Secretary to contribute to achieving the mission of Temporary Assistance for Needy Families (TANF) and Child Support Enforcement.[48] Among other restrictions, the statute also limits the length of time over which OCSE can maintain state data in the

[44] For additional information, see http://lehd.ces.census.gov/.

[45] 13 U.S.C. § 23. and U.S. Census Bureau. (2009). *The Researcher Handbook*. Washington, DC: Census, Center for Economic Studies. https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/Researcher_Handbook_20091119.pdf.

[46] 13 U.S.C. § 9.

[47] 42 U.S.C. § 653.

[48] 42 U.S.C. sec. 653(j)(5).

system to 24 months,[49] but permits the Secretary to retain samples of data beyond this period as necessary to assist in carrying out approved research purposes.[50] The statute requires that users pay a fee to reimburse costs.[51]

OCSE conducts NDNH data matching programs for a number of primarily HHS/ACF program evaluation and research purposes. Examples of studies using NDNH data include evaluations and studies that examine the employment and earnings patterns of TANF and former TANF recipients, assess the effectiveness of various employment and training programs, and analyze labor market outcomes for noncustodial parents. NDNH data have also been used by HHS to measure employment outcomes for adult TANF recipients, and to determine which state TANF programs were awarded high performance bonuses.

II. Federal Tax Information

Statutory restrictions on accessing data are also a major barrier to the use of FTI. The Federal tax code designates FTI as confidential—disclosure to any party is prohibited, except according to explicit statutory exceptions.[52] Issues specifically faced by the Internal Revenue Service (IRS), Social Security Administration (SSA), and the Census Bureau in navigating this barrier are discussed below.

1. *Internal Revenue Service and Treasury*

Under current law, in some cases FTI at IRS can be accessed by certain Treasury Department employees and external researchers for the express purpose of tax administration, which may include "statistical gathering functions."[53] Within Treasury, this authority enables employees whose official duties require access to IRS data to use FTI to support policy-relevant analyses and research. Specifically, Treasury's Office of Tax Analysis uses FTI data to produce budget estimates and conduct analyses of tax policy. In addition, the IRS Statistics of Income Division (SOI) regularly creates statistical analyses from FTI to inform Treasury policy decisions and disseminates select statistical information to government agencies, external researchers, and the public.

Under narrow exceptions to current law, non-Treasury employees are permitted to access the FTI housed at IRS to conduct statistical analyses in support of tax administration if formally engaged as contractors by Treasury.[54] Contractors are thus subject to the same disclosure rules as Treasury employees and are similarly barred from further disclosure of FTI except in anonymous form, such as through the publication of research papers. Components of Treasury have engaged researchers in this capacity in recent years. For example, SOI used its contracting authority to partner with academic researchers Raj Chetty, Nathaniel Hendren, and Lawrence Katz to study the effect of tax expenditures on intergenerational income mobility. Their "Moving to Opportunity" study, which relied on FTI data on tax credits and earnings, illustrates how researchers, with Treasury support, have used the existing

[49] 42 U.S.C. § 653(j)(2)(A).

[50] 42 U.S.C. sec. 653(i)(2)(C).

[51] 42 U.S.C. § 653(k)(3).

[52] 26 U.S.C. § 6103.

[53] Section 6103(h)(1) provides for FTI disclosure to employees of the Department of the Treasury for tax administration purposes. Section 6103 (b)(4) defines tax administration purposes to include "statistical gathering functions." Additionally, 6103(n) authorizes IRS to hire contractors to support the agency's mission.

[54] 26 U.S.C. § 6103.

statutory authority to access FTI at IRS to make important contributions to our understanding of tax and economic policy.[55]

While this study provides one demonstration of the potential value for FTI to be used in policy analysis, the requirements imposed by the tax code significantly limit the scope of permissible research. The data may only be used for tax administration purposes (which includes tax policy-relevant research and analysis) and only by employees and contractors.[56] Permitting researcher access through a contracting mechanism is also costly and burdensome for the IRS and Treasury. SOI is a small organization, and the division's size and limited budget present challenges in scaling access beyond a relatively modest number of projects. Additional scaling at SOI would require a more robust infrastructure to collaborate with and provide oversight of external researchers. Further, without a change to current law restrictions, such researchers are limited topically to projects that inform tax program administration, which is distinct from providing evidence about the breadth of other government programs that could be considered.

2. *Social Security Administration*

At SSA, the Office of Research, Evaluation, and Statistics (ORES), the Office of Research, Demonstration, and Employment Support (ORDES), and the Office of the Chief Actuary (OCACT) all make extensive use of FTI in support of SSA policy and program development and management. ORES creates statistical analyses and reports to support decision-making and public accountability, but has limited infrastructure and experience necessary to support contracted data access by external researchers. Supported by its new Office of Data Exchange, SSA is routinizing the interagency agreement processes for access to data housed at the agency. OCACT uses FTI data to develop estimates of the dollar amounts for Treasury to transfer to the Social Security and Medicare trust funds on an interim basis, as well as estimates for Congress and others on the revenue effects of proposed legislation and administrative modifications. While these data prove to be an important resource for OCACT, additional FTI data could substantially improve OCACT's estimates.

One example of an authorized re-disclosure of FTI by SSA is a data exchange between SSA and the Census Bureau, under which SSA discloses subsets of FTI for limited research and statistical purposes. In this case, SSA's re-disclosure of FTI is authorized based on express permission from the IRS, which is documented in the IRS' regulations.[57] Under the Census Bureau's authority to create sworn agents, ORES employees then use the linked FTI and Census Bureau confidential data as a primary source for analysis.[58] This arrangement is possible due to the Census Bureau's legal and researcher access infrastructure as described in the white paper, *Privacy and Confidentiality in the Use of Administrative and Survey Data*. This access is separate and distinct from the Census Bureau's access to FTI from IRS, and special sworn SSA employees access only FTI furnished by SSA to the Census Bureau. While this arrangement is possible under the law and regulations, it is not a likely exemplar due to the unique objectives and distinctive authorities of these agencies.

---

[55] Chetty, R., Hendren, N., & Katz, L. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Project. *American Economic Review* 106 (4). http://scholar.harvard.edu/hendren/publications/effects-Exposure-Better-Neighborhoods-Children-New-Evidence-Moving-Opportunity.

[56] 26 U.S.C. § 6103.

[57] 26 U.S.C. § 6103(p)(2)(B) and 26 C.F.R. Part 301-6103(p)(2)(B)-1.

[58] 13 U.S.C. § 9(a)(3).

3. *Census Bureau*

Experiences with FTI housed at the Census Bureau illustrate how researcher access to data can be conditioned by statutory limits, legal interpretations, and the policies adopted by different agencies. The Census Bureau makes extensive use of FTI in its statistical programs, such as the Population Estimates Program, the Small Area Income and Poverty Estimates program, and the Economic Censuses and other surveys. When appropriate, the microdata are accessed by researchers, including in evaluations of Federal policies and programs. For example, researcher Ron Jarmin used FTI on businesses to link census and survey data with client identifiers provided by the Manufacturing Extension Partnership (MEP) to determine whether establishments remained active in the years after receiving MEP assistance.[59]

Current law, regulation, and policy govern the Census Bureau's access to FTI.[60] The statute limits the Census Bureau's access to "only to the extent necessary" to perform its work.

In practice, access to specific variables is provided through Treasury regulations,[61] and the Census Bureau's use of those variables is governed through an interagency agreement known as the "Criteria Agreement," which describes how tax data are used to benefit Census Bureau programs.[62] Treasury and IRS exercise great care in considering additional regulatory actions as part of their responsibility to maintain public confidence in the fairness and integrity of the U.S. tax system, which is highly dependent on voluntary compliance. As the Census Bureau has identified potential additional uses of FTI over time, it has identified specific additional variables and defined the use of these variables. Treasury then considers these uses when determining whether to issue new regulations to provide access to those additional variables. Through this process, the variables for which the Census Bureau has access has expanded a few times over the decades. In the matter of using those variables, the Census Bureau then presents proposed projects to the SOI, describing how each relates to the benefits specified in the Criteria Agreement, and after approval from SOI, the Census Bureau or researchers who are special sworn Census Bureau agents may then carry out those projects.

III. Summary

The case study highlighted in this appendix illustrates the various dimensions of statutory barriers to using administrative data for evidence, and state-by-state variance in policy. The interaction of these barriers with infrastructure demands is also evident in the costs expended by agencies to manage the concomitant legal, policy, and administration issues, and in the limitation of access to a very small number of outside researchers who are able to overcome many hurdles. Additional case studies may show similar dynamics with respect to other valuable sources of administrative data in many fields, including healthcare; elementary, secondary, and post-secondary education; and business dynamics.

---

[59] Jarmin, R. S. (1999). *Government Technical Assistance Programs and Plant Survival: the Role of Plant Ownership Type*. Center for Economic Studies Working Paper CES-99-2.

[60] 26 U.S.C. § 6103(j)(1)(A) permits access by the Census Bureau to FTI for statistical purposes, specifically "the structuring of censuses…and conducting related statistical activities authorized by law."

[61] 26 C.F.R. 301.6103.

[62] For a listing of the relevant criteria, see United States Census Bureau. (2015) *Census Research Data Centers Research Proposal Guidelines*, p. 7, https://www.census.gov/ces/pdf/Research_Proposal_Guidelines.pdf.