

Data Sharing in the Federal Statistical System: Impediments and Possibilities

Amy O'Hara, U.S. Census Bureau

Amy.b.ohara@census.gov

4600 Silver Hill Road, Washington DC, 20233

Carla Medalia, U.S. Census Bureau

Carla.medalia@census.gov

Abstract:

While federal agencies have engaged in data sharing for decades, current systems of exchange must be transformed in order to meet the growing needs of analysts in public policy, program evaluation, and basic research. Primary barriers include both perceived legal barriers and actual financial barriers. To address these problems, the potential path forward could include mandating data sharing for key sources, increasing transparency, and improving efficiency by making it easier for agencies to share data (including developing standards, templates and incentives). An agency within the Federal Statistical System should serve as a clearinghouse for centralizing this work. The Census Bureau, with its experience and uniquely broad authority to seek data from any public or private entity for statistical purposes, is well-positioned to fill this role.

Keywords:

Administrative data; infrastructure; evidence-building; program evaluation

DISCLAIMER: The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Policymakers, researchers, program administrators, and government employees need to access more data to understand program performance and measure outcomes. The most recent Federal budget dedicates a chapter to “Building evidence to improve government effectiveness,” defining the need for evidence for an effective and efficient Federal government (Office of Management and Budget 2017). To use evidence to answer important policy-relevant questions, an evidence-based infrastructure is required. The infrastructure will enable shared and linked administrative data across programs, and build longitudinal panels that reveal how programs have affected participants’ lives over the long-term. The Census Bureau, situated at the forefront of the Federal Statistical System (FSS), should coordinate these efforts. To a limited degree, some of this work is already happening. The Census Bureau currently provides data acquisition, integration, and access services for projects that achieve the agency’s mission of improving statistics measuring the U.S. population and economy. The Census Bureau should expand this role, acting as a hub for data integration across the FSS, coordinating best practices in data acquisition and agreements, and taking a lead role in data curation, data integration, and secure data access. The Census Bureau should engage with the state and local clearinghouses, academic and research warehouses, and private sector data enclaves to support evidence building. To succeed, the Census Bureau would embrace this role with a mandate from Congress. In multiple meetings over the past year, the Commission for Evidence Based Policymaking has discussed the need for an entity to lead this work with a transparent and service-oriented approach. The Census Bureau would expand and improve its current methods to serve more users in a responsible fashion, maintaining the trust of data providers and continuing to build trust with the public. To achieve this vision, key stakeholders must work together to address the core impediments to sharing administrative data. Specifically, **perceived barriers** and actual **financial constraints** limit administrative data sharing and use. Affirming the findings of Laurie and Stevens (2016) from the United Kingdom context, this paper describes these problems for the United States and proposes ways to move forward.

Background

Government agencies generate data through their administrative, regulatory, and enforcement activities. Such administrative data, referring to entities including persons, families, businesses, and institutions, were not collected for statistical purposes, but are valuable for measurement purposes.

Federal agencies use administrative data to monitor programs and produce reports for policymakers. The FSS¹ uses administrative data to generate statistics, including population estimates, economic indicators, and national benchmarks. Data are produced and used for statistical purposes, never for enforcement, surveillance or marketing purposes. Statistical purposes are defined as follows: “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” (Budget, Overview of the Federal Performance Framework 2016, 200-18).

Data sharing is essential for access and use of administrative data to produce federal statistics. Agencies often use a combination of data collection, data acquisition, and procurement. Examples of administrative data uses for a few statistical agencies include:

- The Bureau of Economic Analysis (BEA) uses data from the Census Bureau, Bureau of Labor Statistics (BLS), Internal Revenue Service (IRS), Centers for Medicare and Medicaid Services (CMS), Economic Research Service (ERS), National Agricultural Statistical Service, Institute of Education Sciences, Energy Information Administration, and commercial vendors to measure the economy.²
- The Bureau of Justice Statistics (BJS) releases crime data compiled from the Federal Bureau of Investigation (FBI) from the Uniform Crime Reporting program, prisoner data from state

correctional facilities and federal facilities through the Bureau of Prisons (BOP), data on arrests from U.S. Marshals Service and Drug Enforcement Administration.

- The Census Bureau uses data from many sources, including the Social Security Administration (SSA), Department of Housing and Urban Development (HUD), IRS, CMS, and BOP. The Census Bureau uses administrative data as frames and content for many surveys and censuses. For example, data from the U.S. Postal Service and local governments form the master address list that is the backbone of the Decennial Census, and data from IRS form the basis of the business register and economic census.³

Impediments

What is preventing more agencies from sharing data, and how can more analysts access the data? It is not typically an *actual* legal barrier that curtails sharing and access. Rather, there are *perceived* legal barriers and actual financial constraints limiting progress.

Perceived barriers include long-standing practices based on experience instead of law or policy, narrow and/or irregular interpretations of statute and regulations; and resistance to change. These cognitive limits prevent the establishment of new data access pathways and collaborations not because of technical obstacles or legal prohibitions but because people think it cannot be done, that laws in general prevent sharing administrative data (Petrila 2017), or that the administrative data can only be shared to assist in program administration (Goerge 2017).

When approached about a data sharing opportunity, agencies often demur or decline because data sharing has not happened before. They rely on long standing practices, concluding that their data never leave their agency and that they cannot produce an extract. Irregular interpretation of laws and policies can also halt progress. A frustrating example of this limitation involves federal statutes being interpreted differently by counsel in different states. This happened when the Census Bureau approached states to

share food stamps and welfare data. Federal agencies at the Department of Agriculture (USDA) and the Department of Health and Human Services (HHS) were interested in linked data analyses using data from these federally sponsored programs. Some states read Title 7 for the Supplemental Nutrition Assistance Program data (from USDA) and Title 42 for Temporary Assistance for Needy Families data (from HHS) and believed that sharing was permitted, while other states thought sharing was prohibited. A memo from the Office of General Counsel at both USDA and HHS clarified that data sharing with Census for statistical purposes was both legal and encouraged. The attorneys took this position after seeing that data sharing would enable evidence building through statistics that could help the states administer their programs. Despite this guidance, some states are still resistant to sharing their data, potentially for actual financial reasons.

Agencies lack staff to document, extract, and transmit data even when sharing is advantageous. Programs and agencies have varied or missing metadata. There are few interoperability requirements, and none funded across departments. Another financial impediment is the distinctness of funding streams. Across agencies, dollars appropriated for food security cannot be spent on other programs. Given these silos of funding – on top of the silos of data – the establishment and enforcement of data standards is challenging. The funding of a federal clearinghouse to serve the needs of the research community and state/local programs and governments would be similarly difficult to fund. It is unclear whether the base infrastructure would be funded through appropriations and user fees could address the variable costs. Also lacking are coordinated investments in technology to permit secure networks across existing clearinghouses. At present, no agency or organization exists to support the governance around data sharing and access of this magnitude.

Data Sharing: Current Practices and Possible Paths

Data sharing is taking place on a mandatory or voluntary basis, and data requests are managed through a designated staff/process or diffusely through an organization. Below are examples of ways agencies are currently sharing data despite the impediments listed above, followed by possible paths forward. This is an informal summary of data sharing from the bureaucrat's perspective, as opposed to formal models such as those reviewed by Jeng, He, and Oh (2016) from the Information Science perspective.

Some data owners *want* to share data, others *have* to share data. In the former situation, a voluntary desire to comply with a request or to generate new measures drives the arrangement. Voluntary agreements often rely on language in statute that authorizes (but does not compel) data sharing. The Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) authorizes BLS, BEA, and Census to share confidential business data for the purposes of improving the Nation's economic statistical system. Many voluntary data sharing or access arrangements are possible only when the parties can achieve mutual benefits. Examples include the IRS Statistics of Income Joint Statistical Research Program, which permits data sharing only to achieve a tax administration benefit, and the SSA Retirement Research and Disability Research Consortia, which allows data sharing only to understand and improve SSA program administration. States are encouraged to share food security program data with Census to support a joint research agenda with USDA's ERS and the Census Bureau.

Required data sharing arrangements are driven by statute, often ad hoc, preventing reuse beyond the specific purposes mandated. Mandated data sharing is common for government enforcement functions. For example, the Do Not Pay Portal at the Department of the Treasury aggregates data from the Department of Justice, HUD, Veterans Affairs, and SSA for eligibility verification purposes. Required data sharing in the FSS includes the Census-IRS arrangement in which the Internal Revenue Code compels IRS to share data with Census.

Data sharing arrangements are handled through a designated office or process at some agencies, but through extemporaneous processes at many others. Some agencies have professionalized the data

exchange process. They post request and access procedures on their websites, and staff offices to handle data exchanges. SSA's Office of Data Exchange has documented and streamlined the process of accessing retirement and disability program data. CMS has a clear process and assistance available (Research Data Assistance Center) to support data access requests. The Defense Manpower Data Center also has a transparent process for data requests (DMDC Data Request System). At the same time, many other agencies lack such dedicated resources and struggle to process data sharing requests. Most agreements rely heavily on interpersonal relationships and informal quid pro quo arrangements, handling data requests in a less centralized fashion. Data sharing agreements within a department, such as HUD or Education, are handled within business units, each with different routing channels and legal teams. This can confuse reviewers when multiple data requests between the same parties are routing simultaneously but separately.

Agencies currently support data sharing with mixtures of the characteristics above: with required vs. voluntary data sharing, and designated resources to handle requests vs. informal administrative processes. It is possible to continue with current circumstances, but it is not optimal, and certainly not suitable for expanded uses of administrative data. The following suggestions aim to affect the willingness to share data, and to alter the way such arrangements are managed.

- Congress should mandate data sharing for key sources. Many evaluations seek data on outcomes – tackle the perceived barriers and actual constraints to have broader use of income and earnings data currently housed at HHS, IRS, and/or SSA. Dedicated funding must accompany requirements for data sharing, providing resources to document and arrange secure access.
- Incentivize data sharing when not mandated – make more agencies want to share data. Offer them something they want or need, perhaps production of metadata; data harmonization or

standardization; or hosting their analysts in a data facility that has great tools, data structures, and security protocols.

- Make it easy for agencies to share data with a clearinghouse. Pursue templates that clarify terms for different types of uses and users. Pursue standards for data transport. Pursue methods to allow data to stay in place (e.g., privacy protecting record linkage, secure multiparty computation).
- Be transparent about costs and benefits – share information on how long data linkage projects take (from idea to manuscript), how much it costs, and how the results addressed a measurement gap. Use this information to market opportunities for similar projects.
- Congress should establish a clearinghouse to centralize these efforts. Provide acquisition, linkage, and access services to agencies. Have sherpas at the central hub that assist with agreement development, data documentation, data extract and transport, and reporting/monitoring on uses.
- Have agencies improve their existing approaches. Develop and implement more efficient processes, rely on templates that have been cleared through Offices of General Counsel, have greater clarity on precedents, and institute common descriptions of uses and users. Such actions could be coordinated through the Interagency Council on Statistical Policy (ICSP) or Data Cabinet.

Pursuing the Possibilities

Improving the data sharing model for federal and federally sponsored state programs requires investment in methods and tools, and coordinated implementation. Below are proposed specific investments in standards, templates, and incentives to induce more and better managed data sharing.

Standards

Government agencies use many different information systems, even within departments or agencies. Standards make data sharing and integration more efficient. We can improve data structures, data transport, data security, and access/services through application program interfaces (APIs) through investments in interoperability. Federal agencies, including those that fund state programs, should support the establishment and adoption of standards to unlock data from the large federal and federally sponsored state program silos. Buy-in from the data source (or their agency or department) is critical, understanding that it will be simpler to engage with a federal agency, with one information system or set of systems, than with the state-level programs administering federally sponsored human services benefits, with many information systems. One potential solution is the establishment of “integrated data systems,” which integrate data on individuals across agencies and programs to improve understanding of how services are used (Culhane, et al. 2017).

Data documentation and schema harmonization would greatly improve data sharing. Many agencies lack financial and human capital resources to improve metadata. Despite the fact that researchers and evaluators benefit from this infrastructure, the costs are often imposed most heavily on the provider (Foster 2017). A data clearinghouse can work with data sources, offering tools and technical assistance to improve data documentation. This can also be offered as a service to data sources, shifting the role and burden to the hosting entity. The Census Bureau is striving to improve documentation for data the Census Bureau collects and administrative data that have been acquired by the Census Bureau. In fiscal year 2016, the Census Bureau launched the Data Linkage Infrastructure website (www.census.gov/datalinkage). A work in progress, this site provided the first public-facing information about the Census Bureau’s administrative data holdings, policies, and projects using linked data. The Census Bureau is piloting several methods to improve the documentation of metadata. For example, the Census Bureau is working with the National Science Foundation-Census Bureau Network (NCRN) funded

Cornell node to test their Comprehensive Extensible Data Documentation and Access Repository (CED²AR) on administrative data. Recently, an internal team was awarded funds from the Census Bureau's Improving Operational Efficiency program to discover ways to document and share metadata via the University of Michigan's open metadata platform through the Inter-University Consortium for Political and Social Research (ICPSR). The Census Bureau is also working with academic partners from New York University and the University of Chicago on a new approach to enable data curation, including data documentation, provenance, and version control, within a cloud-based data facility that enables data discovery and secure access.

Better metadata will aid schema harmonization, which will improve the accuracy and efficiency of data integration. Again, many data agencies lack the financial and human capital to test and implement schema harmonization tools and methods. The Census Bureau is striving to improve its processes to ingest, link, provision, and analyze administrative data. Partnerships with external organizations allow testing and implementation of innovative approaches, but there are challenges working within the centralized information technology (IT) environment of a federal agency. Policies and practices limit the use of microdata to test applications, restricting opportunities for researchers to develop new methods, such as data linkage (Foster 2017). The Census Bureau is working on approaches that enable the agency to modernize while maintaining a strong security posture. These include: schema alignment to assess and align the characteristics of multiple datasets, including filenames, titles, layouts, and variable and category names, values, and format (e.g., string, numeric, length); models to assess the trustworthiness of a source, a compilation, or a transformation to create provenance and other tags that support data discovery and auditing needs; and testing of algorithms for classification to improve data cleaning and de-duplication.

Standards will improve data sharing and use. But who will decide which standards are needed and how adherence will be enforced? It is possible that Congress could direct an agency or a clearinghouse-

type unit to propose, implement, and enforce standards. Or perhaps OMB could direct such action through budget language or via memoranda providing guidance and direction. Alternatively, OMB could direct the actions through the ICSP. Chaired by the Chief Statistician of the United States, the ICSP coordinates activities across the FSS and could determine and pilot standards that could be implemented across the agencies.

Standards are absolutely essential to deal with the increasing number of administrative data sources, growing data volumes, and the need for high frequency data. Standards to produce adequate documentation will help with data management, discovery and analysis. Standards affecting data structures, transport, and encryption will improve the security posture for sending and receiving institutions. Standards address many of the financial problems described above: once understood and implemented, they help data providers document data, extract data, and transmit data. Dedicated effort and resources to developing standards is the cross-agency step needed towards interoperability. To proceed, the system should encourage the ICSP to catalog the interoperability efforts underway across the FSS, and encourage collaboration across industry, academia, and department CTO offices. The FSS should assess which tools and best practices could be shared across the system (e.g., how Census pilots on schema harmonization tools or the secure cloud based data facility could be tested at other agencies).

Templates

Universal data access templates must be developed to support broad uses by analysts inside and outside of government. Efficient data sharing is not only about getting files “over the transom”, but also about facilitating access for qualified users. Data sharing agreements need to incorporate nimble processes for reviewing requests, approving researchers, monitoring data access and study scope, conducting disclosure avoidance, and reviewing any research output to ensure no confidential

information is disclosed. Efficient data exchange is pointless if it is not also accompanied by clearly specified usage guidelines that are understood by all parties. These operational details need to be incorporated into new agreement templates. The standards described in the previous section will inform and shape content in data sharing templates.

Templates address part of the perception problem: they offer a way forward when an agency stands on past practice, providing examples of how other agencies have entered similar arrangements. Templates also help with irregular interpretation of laws by stating the authority under which sharing can occur for departments and programs.

The Census Bureau's administrative infrastructure includes dedicated staff that identify new data sources, negotiate data sharing arrangements, and coordinate researcher access to data files. These agreement negotiators are versed in the Census Bureau's authority to obtain and use data for statistical purposes. They have a set of templates to facilitate data sharing for Census Bureau needs and sponsored work. They have experience working with lawyers, privacy officers, database experts, and program administrators to identify the most efficient ways to share data between agencies. Parts of this process are more scalable than others: more staff does not necessarily result in more or quicker agreements. Data owners or their attorneys may reject agreement templates. Templates often speed review, but many delays on the data provider's side involve leadership and staffing changes, the data provider's interpretation of their own legal authorities to share data, or a lack of resources to review and approve a template. On the Census Bureau side, a dedicated legal staff and an automated agreement development system could improve efficiency and therefore aspects of the infrastructure.

While the Census Bureau has followed a standardized acquisition path for decades, in fiscal year 2016, the Census Bureau initiated twelve pilot evaluation projects to test and document every aspect of the data sharing process. The pilot projects span housing, health, welfare, education, and labor market

outcomes. Each pilot requires multiple agreements to clarify roles and responsibilities of statistical uses and data sharing. Many of the agreements for the pilot projects are narrowly scoped. A transformed model could have tiered agreement templates, with graduated scope levels. Requests to access the data could have similar graduated uses, permitting users to select the depth and breadth of data needed for their evaluation. The agreement process must be transformed if data sharing and data use/access are to support timely analyses that range from producing performance metrics to scholarly work, spanning the use of near real-time data to longitudinal studies requiring pooling data over years or decades.

To move forward, the Census Bureau must continue to refine templates, seek more support to obtain legal review of agreements, implement its agreement tracking tools, and document the gaps and shortcomings of our current process. An administrative data clearinghouse at the Census Bureau could coordinate such efforts, disseminate templates and best practices, and offer technical assistance.

Incentives

The academic, evaluation, and policy-making communities must incentivize or require data sharing to make data controllers want to share data (or have to share data). Incentives must be appropriate and consistent across agencies, using common and transparent language on uses and retention. Access to high value datasets (e.g., mortality, employment, earnings) should be expedited or required. When data sharing is voluntary, the data quality, availability, and delivery are fragile and subject to change when executive buy-in or budgets shift. Requiring data sharing of high value sources should be accompanied by clear requirements and terms on the treatment of identifiable data and approved uses. To support further data sharing, laws and regulations must clarify and encourage data access for statistical use. For example, legislation or budget language could state that data sharing is required to produce statistics evaluating programs or outcomes.

In any document incentivizing data sharing, the terms of use must describe the treatment of identifying information and retention terms. What will be kept, for how long, and why? The Australian public's reaction to the 2016 Census planned compulsory collection and retention of names and addresses was instructive. The public was surprised by the planned uses of the data and threatened boycott or fraudulent responses. Transparency is required to gain and maintain the trust of data sources and the public.⁴

Properly designed incentives could address many of the challenges we face today. Incentives can help agencies saying "we have never done that" and resistant-to-change agencies see the need for or advantage of data sharing. Incentives can shift burden away from data controllers who often lack resources to clean or document their data to agencies with more experience and tools. For example, states may lack the resources to produce metadata but could see production of documentation and harmonized data as an incentive to share files. Incentives or mandates to share data could also address the distinct-funding-stream problem by directing data sharing to evaluate programs and measure outcomes.

The FSS, or a federal clearinghouse, should design and pursue a research agenda to create data sharing incentives. Beyond the basics, including enabling new measurements and confirmation of previously published results, we need to determine how to incentivize researchers and agencies to share data. Best practices should be drawn from the natural, social, and health sciences. The United Kingdom Data Archive has also conducted studies on the topic.⁵ The research agenda should assess the parties who can affect behavior such as funders, academic institutions, publishers, and OMB.

The following table summarizes how standards, template and incentives can address some of the impediments to greater data sharing across the FSS. Together, these efforts can make more parties want to or have to share data through uniform processes.

Table 1. Use of standards, templates and incentives to address barriers to sharing data

Impediment Example	Standards	Templates	Incentives
Long standing practices (e.g., we can't do that)		✓	✓
Irregular interpretations of laws	✓	✓	
Resistance to change			✓
Inadequate resources to document data/poor or missing metadata	✓		✓
Inadequate resources to extract data	✓		
Inadequate resources to transmit data	✓		
Inadequate investment in interoperability	✓		✓
Distinct funding streams			✓

Examples of Programs Seeking, Establishing, or Enforcing Standards

The 2014 Farm Bill (Public Law 113–79) included data exchange standards for food security and family self-sufficiency legislation. As stated in Section 4016, states must incorporate a widely accepted, non-proprietary, searchable, computer readable format such as the eXtensible Markup Language, and contain interoperable standards developed and maintained by intergovernmental partnerships such as the National Information Exchange Model. These requirements cover data exchanges across state agencies for federal reporting and data exchanges required by applicable law. USDA was to issue a proposed rule to amend the Food and Nutrition Act (7 U.S.C. 2020) regarding data exchanges. In summary, this bill had good intentions, but lacked either requirements or resources.

The Administration for Children and Families (ACF) sought investment at the state level.⁶ This unfunded proposal would create a Statewide Human Services Data System through grants for states to develop data systems in order to analyze program investments and outcomes across multiple programs

and over time, and a Systems Innovation Technical Assistance Center that would develop standards/architecture and provide reusable technology components and shared services so states and tribes can address common requirements, including eligibility and enrollment. The ACF proposal was developed in the spirit of the grant-based investment in states to develop statewide longitudinal data systems for education. In summary, this proposal had good intentions and would have provided technical assistance and resources, but lacked sufficient investment.

The Digital Accountability and Transparency Act of 2014, (DATA Act) at Treasury makes government-spending data available as open data. Treasury designed a data platform to enable harmonization and ease of use. In 2015, they released 57 data standards (final standards were available one year later) and established a standard data exchange. They identified 30 key data elements that will be standardized in their system. In summary, this act had success in establishing standards, and had sufficient investment and enforcement.

The National Directory of New Hires (NDNH) at HHS contains data on unemployment insurance wages and benefits, new hires, and the federal parent locator program data. NDNH requires that states submit the same elements in a set format. States had set up State Directories of New Hires, which roll up to the federal collection that is used for child support enforcement. In summary this initiative had successful implementation, with both sufficient investment and enforcement.

Conclusion

Data sharing is a critical component of our decentralized federal statistical system. While nearly all units in the system participate in data sharing, they face common burdens (real and perceived) in their efforts to acquire and use data from other federal government agencies and states. This paper has described

paths to more efficient data sharing through implementation of standards, the need for legal templates and protocols, and the need to incentivize data sharing. Improvements in these areas will promote not only the efficient and legal exchange of data but also the credibility of analyses based on the shared data.

This paper calls for a centralization of these efforts and the establishment of a federal data-sharing clearinghouse. A coordinating unit is needed to manage investments in standards and template development. Such a unit should be situated within the FSS to leverage the system's abilities to protect data and support analyses free from political influence, and thereby to build the trust of data providers, the research community, and the public. Federal statistical agencies are ready partners for other federal agencies, federally funded state program agencies, local governments, research organizations, academics, and policymakers that form the evidence-producing ecosystem. Our path forward requires modernizing the infrastructure and creating governance processes that support a much broader analytic capacity.

References

- Budget, Office of Management and. 2016. *Overview of the Federal Performance Framework*. OMB Circular No. A-11, Washington: U.S. Government Publishing Office.
https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/a11_current_year/s200.pdf.
- Culhane, Dennis, John Fantuzzo, Matthew Hill, and TC Burnett. 2017. "Maximizing the Use of Integrated Data Systems to Serve State and Local Governments through Advancing Core IDS Components." *Annals of the American Academy of Political and Social Science*.
- Foster, Ian. 2017. "Research infrastructure for the safe analysis of sensitive data." *Annals of the American Academy of Political and Social Science*.
- Goerge, Robert M. 2017. "Barriers to accessing state data and approaches to addressing them." *Annals of the American Academy of Political and Social Science*.

- Jeng, Wei, Daqing He, and Jung Sun Oh. 2016. "Toward a conceptual framework for data sharing practices in social sciences: A profile approach." *Proceedings of the Association for Information Science and Technology* 53: 1-10.
- Laurie, Graeme, and Leslie Stevens. 2016. "Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom." *Journal of Law and Society* 43: 360–392.
- Office of Management and Budget. 2017. *Analytical Perspectives: Budget of the U.S. Government, Fiscal Year 2018*. Washington: U.S. Government Publishing Office.
- Petrila, John. 2017. "Legal Issues in Using Administrative Data." *Annals of the American Academy of Political and Social Science*.
-

Notes

¹ The FSS is comprised of thirteen principal statistical agencies and approximately 130 statistical programs housed in agencies across the federal government that support program planning and evaluation functions. These agencies or organizational units are in the Executive Branch, overseen by the Office of Management and Budget, with activities involving the collection, compilation, processing, or analysis of information for statistical purposes. See

https://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2015.pdf for more information.

² See <https://bea.gov/about/pdf/acm/2016/improving-regional-pce-estimates-using-credit-card-transactions-data.pdf> for example of vendor data testing.

³ The Census Bureau protects the confidentiality of data that it acquires from other agencies just as it protects the information it collects directly from individuals and businesses. Data sharing supports operational uses, evaluation uses, and research uses. However, using data collected for non-statistical purposes can be challenging. When using data from other agencies, the Census Bureau reviews data quality, assessing dimensions including relevance, accuracy, timeliness, punctuality, accessibility, clarity, coherence, and comparability.

⁴ Another example of the importance of clarifying retention terms involves Unemployment Insurance (UI) data held at HHS. Federal law requires deletion of data in the National Directory of New Hires (NDNH) system 24 months after the date of receipt. However, the Social Security Act allows HHS to keep samples of data as needed to support research on child support topics if personal identifiers are removed. Other data development projects have maintained panels longer than two years (e.g., the Longitudinal Employer Household Dynamics program at the Census Bureau has decades of data for most states). States themselves may also retain the data, depending on their own statutes. But the mandated, central repository at HHS has limited use and a rolling window of retention. Such issues need to be transparently addressed.

⁵ See <http://www.data-archive.ac.uk/about/projects/incentive>.

⁶ See www.acf.hhs.gov/blog/2016/03/interoperability-at-acf.