BARRIERS TO USING GOVERNMENT DATA:

Extended Analysis of the U.S. Commission on Evidence-Based Policymaking's Survey of Federal Agencies and Offices

October 2018

BIPARTISAN POLICY CENTER

AUTHORS

Nick Hart, Ph.D.
Director of the Evidence Project

Kody Carmody
Research Analyst

ACKNOWLEDGMENTS

The authors would like to thank the Laura and John Arnold Foundation, the William T. Grant Foundation, and Schmidt Futures for their generous support of the Bipartisan Policy Center's Evidence Project.

DISCLAIMER

This technical paper is the product of the staff of the Bipartisan Policy Center's Evidence Project. The findings and conclusions expressed by the authors do not necessarily reflect the views or opinions of BPC, its founders, its funders, or its board of directors, nor do the views reflect those of the co-chairs or advisory members of BPC's Evidence-Based Policymaking Initiative.

Executive Summary

Policymakers face many demands from constituents, budgetary processes, and their commitment to providing good services for the American people. This last concern is made easier when policymakers have access to reliable information to guide their decisions. But access to data and the ability to turn those data into evidence to inform decisions can be hampered by legal restrictions on access to sensitive data, institutional constraints, and the availability of resources.

In 2016, Congress and the president established the U.S. Commission on Evidence-Based Policymaking and charged it with developing a strategy for addressing these barriers. During the commission's fact-finding efforts, it launched a survey of agencies and units across the federal government to better understand existing barriers to data access and use. The data collected in the survey then provided initial evidence that the commission considered in making its recommendations.

Extended analysis of the commission survey confirms much of what the commission concluded in its final report, validating identified legal and regulatory barriers to using data. The extended analysis also leads to new findings:

1. Federal offices perceive that their roles in evidence-building activities are in niches and largely do not perceive their data collection as for a broad range of purposes like evaluation that would require better coordination across an agency.
2. Units within federal agencies exhibit wide variation in their capacity for data sharing and linkage.
3. Challenges to using data for evidence building are distributed across virtually every policy domain. Respondents identify federal tax information as especially difficult to access and use.
4. Despite some offices reportedly lacking resources to conduct evidence-building activities, it is still quite common for offices to conduct at least some data sharing and linking. However, agencies still indicate substantial gaps in developing metadata, sharing with third parties, conducting disclosure reviews, and engaging in disclosure avoidance protocols to protect data. Statistical agencies were by far better positioned for this work than other agencies.

While the commission's survey was conducted in 2017, the issues identified through an extended analysis of the survey data are likely still relevant, as little policy change has occurred since the survey to wholly address the identified issues.

Collectively these findings offer relevant insights for development of the commission's recommended National Secure Data Service to fill identified capacity gaps as a shared service center for federal agencies. There are many activities already underway to analyze, link, and use data for evidence-building activities, in addition to other purposes. These survey results suggest that government needs to improve capacity to engage in this work, with better abilities and an organizational infrastructure and legal framework that supports securely accessing and analyzing data.

Introduction

Researchers and evaluators have for years asserted that government data are difficult to access, and thus difficult to use for generating evidence about government policies and programs. Researchers and evaluators seeking to use government-collected data have suggested numerous problems and barriers that affect their ability to use those data. Many barriers are unintentional and result from complicated, unclear processes for delineating how data can be accessed and used.

Many barriers are unintentional and result from complicated, unclear processes for delineating how data can be accessed and used.

In 2016, Congress and the president established the U.S. Commission on Evidence-Based Policymaking to develop a strategy for comprehensively addressing barriers to accessing and using government data for evidence building. During the commission's fact-finding efforts, it launched a survey of agencies and units across the federal government to better understand existing barriers to data access and use. Agencies were asked to identify different types of challenges and areas of potential gains in their evidence-building capacity. The commission then considered this information in developing its final recommendations.

The commission's survey was an attempt to systematically identify key issues experienced by federal agencies. From January 2017 to March 2017, the commission invited more than 200 units within federal agencies to complete a survey asking about their data, research, and evaluation activities as well as barriers faced when seeking to access and use government data for evidence building.

The survey collected detailed information on existing evidence-building activities and barriers respondents experienced, all of which supplemented other input to the commission's recommendations. For example, the commission also heard testimony from policymakers and federal employees, held public hearings, and solicited written input from the American public. This feedback was collectively compiled and distilled into a set of findings and resulting recommendations produced by the commission in its final report, presented to Congress and the president in September 2017.[1]

Given the relatively short timeframe the commission had to complete its fact-finding and develop final recommendations, an exhaustive analysis of the survey results was not possible. In its final report, the commission presented initial results from the survey that informed the recommendations.[2] This technical paper presents additional findings from an extended analysis of the commission's survey, delving deeper into the barriers and priorities acknowledged by surveyed federal offices. As work proceeds in implementing the commission's recommendations, further analysis of the survey results can continue to offer insights for strategies to more effectively address real and perceived barriers to using government data.

This technical paper identifies four key new findings from the extended survey analysis:

1. Responding federal offices perceive that their roles in evidence-building activities are in niches and largely do not perceive their data collection as for a broad range of purposes. The perceptions of various roles in evidence-building activities suggests that coordination across functions and units may be essential for successful and useful evidence-building capacity to emerge and be sustained in government.
2. Units within federal agencies exhibit wide variation in their capacity for data sharing and linkage. An identifiable capacity gap appears especially pronounced in offices that are not statistical agencies, including evaluation offices.
3. Certain types of data, such as federal tax information, were perceived by responding agencies to be particularly challenging to access and use. However, challenges to using data for evidence building are distributed across virtually every policy domain.
4. Despite some offices reportedly lacking resources to conduct evidence-building activities, it is still quite common for offices to conduct at least some data sharing and linking. The frequency with which these activities occur, combined with gaps in capacity, often means a lack of best practices in terms of sharing with third parties, providing metadata, and conducting disclosure reviews. The survey results suggest clear gaps in privacy protections and disclosure avoidance protocols that could be addressed.

While the commission's survey was conducted in 2017, the issues identified through an extended analysis of the survey data are likely still relevant as little policy change has occurred since the survey to wholly address the identified issues.

Survey Background

The commission's survey of federal agencies asked a targeted series of questions to offices involved in generating data, building evidence, or using evidence for making policy.[a] The commission prioritized the offices invited to participate in order to efficiently conduct non-response follow-up. The stratification of offices was also a strategy for maximizing the validity and reliability of the survey results, ensuring that the offices with the most relevant experiences about challenges and barriers could appropriately inform the commission's findings and recommendations.

Offices designated as high priority were those deemed to engage in more substantial evidence-related activities. Evaluation offices, principal statistical agencies, and other offices with a primary focus on research and evaluation activities were considered especially high priority for the survey. Low priority offices were expected to be less engaged in the issues under consideration by the commission. The designation of low or high priority was not intended to reflect or judge the quality or value of any particular office's overall activities or mission.

Of the 209 offices invited to participate in the survey, 90 offices in 26 federal agencies responded, including 79 percent of the high-priority units.[3] The response rates for evaluation offices and principal statistical agencies, which were classified as high priority and received additional non-response follow-up, were higher than in the other category (see Table 1).

Table 1. Survey Response Rates by Type of Office and Priority Category

	Number of Eligible Offices	Offices Responding	Response Rate
Type of Office			
Evaluation Office	12	10	83%
Principal Statistical Agency	13	13	100%
Other	184	67	36%
Priority Category			
High	57	45	79%
Medium	63	18	29%
Low	89	27	30%
Total	**209**	**90**	**43%**

[a] See: CEP 2017b for more details on the survey, including methods, instruments, and full responses to the open-ended questions.

Table 2. Reported Spending for Evidence-Building Activities by Type of Office

Type of Office	% of Budget for Evidence-Building Activities				
	0% of Budget	1 to 10%	11 to 50%	51 to 90%	91 to 100%
Evaluation Office (n = 10)	0%	10%	0%	40%	50%
Principal Statistical Agency (n = 13)	0%	0%	15%	8%	77%
Other (n = 57)	7%	49%	19%	5%	19%
Total (n = 90)	**5%**	**36%**	**16%**	**10%**	**33%**

Validating the prioritization of offices used by the commission, evaluation units and principal statistical agencies reported that a higher percentage of their budgets were for "statistics, evaluation, research, and analysis activities," or evidence-building activities (Figure 2).[b] Offices in the other category, in contrast, reported that smaller percentages of their budgets were for evidence building. Offices that had a primary purpose related to evidence-building activities tended to dedicate greater percentages of their budgets than those that embed the activities within other administrative responsibilities.

While the preliminary analysis conducted by the commission identified a positive relationship between the percentage of an office's budget used for evidence-building activities and the likelihood that other agencies access that office's data, this relationship appears to primarily be related to office mission. Evaluation offices and principal statistical agencies, by virtue of their missions, are more likely to both have their data accessed and to dedicate more resources to evidence-building activities than other offices.

This is not to say that resources are not an important factor in enabling the use of government data. In fact, the commission recognized in one of its recommendations that resources—including funding and personnel—were necessary to implement its recommendations (see recommendation 5-5).[4] However, this technical paper does not rely on office estimates of the percentage of their budgets used for evidence-building activities to analyze core questions in the survey, even though part of the commission's exploratory analysis did in the final report. Instead, office type is used for this extended analysis because it highly correlates to resources while also offering an intuitive framework for informing future policy actions.

[b] The commission survey specifically asked about "statistical, evaluation, research, and analysis activities." For the purposes of this technical paper, these activities are collectively referred to as "evidence-building activities."

Collecting and Analyzing Data

Collecting and analyzing data that identify individual units such as people, households, and businesses—referred to as microdata—is foundational for engaging in evidence-building activities. While access to microdata is not necessary for many types of analyses, identifiable microdata may be required to connect different datasets to generate requested descriptive statistics or to permit research for specific types of questions. The commission recommendations largely addressed making this type of data available for evidence-building activities.

Government agencies collect data for a variety of purposes, including to run programs, determine service eligibility, monitor performance, and engage in research. Among federal office survey respondents, the largest share reported they collect data for conducting analysis of existing policies or programs to assist decision making (see Table 3). This suggests that the intent for many agencies collecting data is to use them for some beneficial management purpose.

Table 3 presents a snapshot of how offices view their purposes for data collection. Principal statistical agencies—because of their direct, specific missions—all agreed that they have a role in producing statistics about the economy, society, or the environment, and in supporting research. However, they were less likely to view the purpose for which they collect data to be as a resource for evaluating programs, supporting program operations, or monitoring program performance. This perspective is consistent with the legal framework within which many of those agencies operate, such as the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) which prohibits the disclosure of identifiable, individual-level records.

Similarly, evaluation offices primarily viewed their data as collected to support evaluation and research and to provide analysis for decision making. They were less likely to report their purpose as being for general knowledge production or for program operational activities.

Of the other surveyed offices, data were reportedly collected largely for evaluation, analysis, and support of program operations. For these units that are generally embedded in parts of agencies running programs, this suggests a more proximate relationship to the day-to-day program decisions of the agency.

CHALLENGES IN DATA COLLECTION

Before government agencies can collect data, they must receive approval of their information collection requests (ICR) from the U.S. Office of Management and Budget. These requests are subject to public notice and comment periods, required by the Paperwork Reduction Act (PRA). The PRA is intended to "ensure the greatest possible public benefit from and maximize the utility of information created, collected, maintained, used, shared and disseminated by or for the federal government" and to "improve the quality and use of federal information to strengthen decision-making, accountability, and openness in government."[5] A centralized review process in the federal executive branch was designed as a means to achieve these goals prescribed in law. Today, the PRA is recognized to have a substantial role in supporting efforts to transparently engage the public about data uses.[6]

While some offices in the commission survey identified PRA requirements as a significant barrier, the vast majority did not indicate they were a substantial impediment to evidence-building activities. But, in the offices where PRA requirements were perceived as a significant barrier, the approval process and corresponding time delays were frequently cited by respondents in open-ended responses (see Table 4). The process for how data are collected affects whether and to what extent they will be useful for analysis, linking, or any other aspects of evidence-building activities.

Several agencies noted in open-ended responses that they perceive PRA barriers to their evidence-building activities as so burdensome and restrictive on what information can be collected that they may not pursue those activities at all. For example, one agency indicated that "we come across questions that would require surveys and we don't even seriously contemplate requesting a [sic] ICR due to the time commitment." While Information Collection Request (ICR) approval requirements are not a consistent barrier for agencies, the identification of challenges that may inhibit further data collection altogether despite otherwise being perceived as necessary and useful by an agency suggests additional efforts to further understand this barrier would be worthwhile.

ANALYZING CONFIDENTIAL DATA

Some data collected by the government contain identifiable information about individuals, households, or businesses. Depending on the legal authority under which they are collected, those data may require stringent confidentiality protections and restrictions for access or use.[7] Of the 65 offices that reported routinely collecting data, nearly all (92 percent) indicated they collect confidential microdata. This includes all 13 principal statistical agencies and all eight evaluation offices that reported primary data-collection activities in the survey.

Table 3. Reported Reasons That Offices Collect Data on Individual Units (Percent by Type of Office)

Type of Office	# of Offices:		Purpose for Collecting:					
	Total	That Collect Data	Produce Statistics	Evaluate a Program's Impact or Implementation	Support Research on a Group or Policy	Analyze Policy to Assist Decision Making	Conduct or Support Program Operations	Monitor Program or Grant Performance
Evaluation Office	10	8	38%	100%	100%	100%	50%	50%
Principal Statistical Agency	13	13	100%	46%	92%	62%	23%	8%
Other	67	49	27%	61%	45%	63%	69%	51%
Total	**90**	**70**	**41%**	**63%**	**60%**	**67%**	**59%**	43%

Table 4. Respondents' Explanations for Information Collection Requests Being Their Most Critical Barrier

- "Often we come across questions that would require surveys, and we don't even seriously contemplate requesting a [sic] ICR due to the time commitment."
- "We are often asked to enter into user agreements to obtain the data, which our [general counsel] does not allow. Often in working with tribal entities, we need to get tribal IRB or tribal board approval, which takes forever and is often denied."
- "Too many information collection requests are delayed at [the Office of Management and Budget] for reasons beyond the control of [Centers for Disease Control and Prevention] programs."
- "…The PRA process also is very cumbersome and time consuming."
- "…The burdens of the centralized PRA clearance process, exacerbated by the limited capacity of the [Office of Information and Regulatory Affairs] to address this comprehensive government-wide workload, has [sic] dramatically increased the cost and delay in getting critical information to program administrators. While the government has an interest in ensuring that data collected is valid for its purpose, high-quality, and non-duplicative, all but the most potentially burdensome collections could be appropriately managed at the agency or sub-agency level."
- "The information collection request process under PRA has posed a significant barrier to our office's ability to use data. Challenges include:

  (1) "The length of time the process takes can exceed 12 months. The delay alone is problematic but also has a negative impact on our ability to award and manage contracts to conduct evaluations, research, or measure program performance…The delays also impact the ability to collect baseline data for new programs and initiatives and for those with short life cycles."

  (2) "…Testing of data-collection tools (e.g., reliability and validity), if conducted with more than nine individuals, must go through a full ICR process…The process does not accommodate a rapid cycle of question testing, revision, and testing. An evaluation or research study will likely undergo two ICR approvals—one for testing and the second for implementation, which further delays the research."

As suggested in Table 5, confidential data are most frequently analyzed by office staff (88 percent of offices) and/or contractors (73 percent). Analysis by other actors is substantially lower: 23 percent of respondents reported that grantees analyzed their confidential data, though evaluation offices were more likely than other government units to rely on grantees to conduct evaluation activities (38 percent). Similarly, analysis by researchers outside the office was generally low (45 percent), except for principal statistical agencies, 85 percent of which reported that researchers analyzed their confidential data. This suggests that, while access is difficult for researchers seeking to use statistical agency microdata, these data also tend to be more accessible than many other parts of government.

Respondents with data that are not analyzed by other offices or agencies, researchers, or grantees were more likely to cite inadequate funding, legal limitations, and privacy or data security requirements as the most critical barriers to using their data (see Table 6). These perceived barriers may also reflect the underlying causes of these offices not engaging in analysis of their data. Further, the number of offices that largely do not analyze these data seem to validate that certain barriers are perceived as exceedingly limiting for robust data analysis.

Sharing Data

Once data are collected by agencies, they may be relevant and useful to addressing issues in other programs or policy areas. However, as the commission identified in its final report, agencies face challenges in sharing data even when strong privacy and confidentiality safeguards are in place.[8] To achieve the commission's vision that data are used to conduct rigorous analyses that inform government decisions, improved access to these data across agencies is critical, especially when data are needed to understand a wide range of outcomes.

Offices across government reported critical barriers to analyzing their own data. These barriers are also closely related to protocols for sharing data. Inverting the perspective to focus less on data owners, agencies were also asked in the survey about data that they would like to access, or do access, from other agencies. Respondents identified numerous data sources they would like to analyze but are collected by other parts of government and are difficult to access (see Table 7).

The most frequently sought data source that was reportedly difficult to access was federal tax information. Title 26 of the U.S. Code restricts access to federal tax information. The individual-level records can be accessed under certain conditions and when the resulting analysis provides a benefit for federal tax administration. For example, both the U.S. Census Bureau and the Bureau of Economic Analysis are authorized to use federal tax information, but the restrictions under Title 26 transfer to any data products and activities related to the use of those data at the other agencies.

The Bureau of Labor Statistics (BLS) was one office that listed legal limitations as their most critical barrier to using data for evidence building. The BLS elaborated in its response that "as a result [of statutes restricting IRS data sharing], Census, [Bureau of Economic Analysis], and BLS cannot achieve the full benefits of data sharing. Specifically, Title 26, Section 6103(j) needs [to be] edited to allow for BLS access . . . to fully achieve the data sharing benefits that [CIPSEA] sought to achieve." Other offices and agencies also shared this sentiment in the survey. When listing types of data that they considered using but ultimately did not because of difficulty in accessing them, federal tax data were by far the most common among respondents.

Table 5. Who Analyzes Respondents' Confidential Data[c]

Type of Office	Staff	Contractors	Grantees	Other Agency	Other Office	Researchers	Other
Evaluation Office (n = 8)	88%	88%	38%	38%	13%	50%	0%
Principal Statistical Agency (n = 13)	100%	77%	23%	62%	46%	85%	0%
Other (n = 39)	85%	69%	21%	23%	28%	31%	5%
Total (n = 60)	**88%**	**73%**	**23%**	**33%**	**30%**	**45%**	**3%**

Table 6. Most Critical Barrier to Using Data, by Analyst Group

Is your confidential data analyzed by…		Most Critical Barrier (as a % of all data-collecting offices):					
		Inadequate Funding (n = 18)	Legal Limitations (n = 13)	Privacy and Data-Security Requirements (n = 6)	Lack of Ability to Hire Appropriate Staff (n = 8)	ICR Requirements (n = 4)	Other (n = 14)
Other Office	*No*	61%	77%	100%	50%	75%	64%
	Yes	39%	23%	0%	50%	25%	7%
Other Agency	*No*	61%	69%	100%	50%	50%	50%
	Yes	39%	31%	0%	38%	50%	21%
Researchers	*No*	72%	31%	83%	50%	100%	29%
	Yes	28%	69%	17%	50%	0%	43%
Grantees	*No*	72%	77%	83%	88%	50%	50%
	Yes	28%	23%	17%	0%	50%	21%

[c] The universe for Tables 5 and 6 only includes the 60 offices that collect confidential data.

Table 7. Count of Data Types Offices Want but Have Trouble Accessing

Office Type	Data Type:						
	Criminal Justice	Census	Fed. Edu.	State Edu.	Fed. Business	Establishment	Fed. Tax
Evaluation Office	2	0	1	3	0	0	4
Principal Statistical Agency	1	4	0	2	1	5	8
Other	9	3	3	6	5	3	10
Total	**12**	**7**	**4**	**11**	**6**	**8**	**22**

Office Type	Data Type:						
	Social Security	Human Services	Medicare/ Medicaid	Military	Wage	Vital Records	Other
Evaluation Office	4	0	2	1	1	1	3
Principal Statistical Agency	3	3	2	1	5	3	1
Other	4	6	8	5	5	8	25
Total	**11**	**9**	**12**	**7**	**11**	**12**	**29**

Recognizing the frequency with which tax records were identified as a relevant data source, the commission specifically recommended that consideration be given to modifying access provisions for administrative data that are critical for evidence-building activities (see recommendation 2-4). But Title 26 was not the only identified barrier for using government-collected data. For example, the Office of Justice Programs at the Department of Justice suggested that the varying "statutes among the different agencies make it difficult to come to an agreement." This statement suggests that the matrix of statutory authorities poses complex challenges for potential data users to navigate in identifying what eligibility and restrictions may be required for data sharing, if permissible at all.

Linking Confidential Data

Combining and linking data can enable the creation of new datasets that may offer additional insights or knowledge about important policy-relevant questions. For example, a researcher might have two disparate datasets on business activity and individual wages that can be independently analyzed to provide information about the business activity or wages. When the information is combined, however, additional insights can be gained from generating statistics about workers' wages for specific employers and business activities.

Conducting data linkages can be time and resource intensive. Federal agencies face several resource constraints in linking data, even when such linkages may be necessary and permissible for assessing program missions. Of the offices that reported in the commission survey that they link confidential data, 58 percent (21 offices) did not have dedicated staff or contractors to perform the tasks necessary (see Table 8). Among principal statistical agencies, only the Census Bureau reported it had more than six dedicated staff or contractors to perform data linkages. Collectively this suggests that across the federal government there may be large workforce and skills gaps in applying best practices to link or combine confidential data.

Table 9 shows that the most frequent method reported to cover the costs associated with data linkage is through program funding (58 percent of offices). Covering the costs with user fees (11 percent) and funds from other offices (17 percent) are relatively rare, while funding through program allocations and budget set-asides (31 percent) are slightly more frequent. While the survey did not ask agencies to distinguish between the types of appropriations they receive (mandatory or discretionary), most are likely recipients of annual discretionary appropriations. Given the prevalence of agencies that reported funding as a barrier for conducting data linkages, an increase in the prevalence of data linkage activity to support evidence-building activities may not be expected without first increasing direct appropriations, establishing new sources of funding (e.g., user fees), or implementing a shared service for linkages, such as the National Secure Data Service proposed by the commission.[d]

For most survey respondents, data linkages tend to be for ad hoc projects; relatively few are for routine processes (see Table 10). This may partially explain the lack of dedicated resources reported by most offices in the survey. Linking datasets can require significant time investment for personnel, not only in data-sharing processes, but also in gaining knowledge about key attributes and variables in datasets. This can be made more difficult by a lack of detailed metadata and a lack of harmonization across government data-documentation practices, as noted by the commission.[9] Table 11 suggests that while many offices do provide information widely recognized as useful on the data they generate, many do not document changes in definitions, imputation procedures, or other changes over time. Evaluation offices and principal statistical agencies reportedly provide metadata more frequently than other offices, though gaps remain there as well. Similarly, Table 12 shows that six of 26 offices that allow researchers to analyze their confidential data do not report having a process for reviewing outside researchers' work to prevent disclosures of sensitive information and breaches of privacy, including one principal statistical agency.

Table 8. Number of Staff That Respondents Have Dedicated to Data-Linkage Activities[e]

Type of Office	Number of Staff					
	Did Not Answer	0	1 to 2	3 to 5	6 to 10	More than 10
Evaluation Office (n = 6)	1	5	0	0	0	0
Principal Statistical Agency (n = 10)	0	4	3	2	0	1
Other (n = 20)	1	12	2	1	1	1
Total (n = 36)	**2**	**21**	**5**	**3**	**1**	**2**

[d] See recommendations 2-1, 2-2, 4-1, 4-2, 4-3, and 4-4 in CEP, 2017a.

[e] The universe for Tables 8-10 only includes the 36 offices that report linking data.

Table 9. How Respondents Pay for Data Linkage

Type of Office	Budget Allocation	Programs Absorb	Set Aside Budget	Other Offices	User Fees	Other
Evaluation Office (n = 6)	17%	50%	50%	0%	0%	0%
Principal Statistical Agency (n = 10)	30%	70%	30%	30%	20%	0%
Other (n = 20)	30%	55%	25%	15%	10%	5%
Total (n = 36)	**28%**	**58%**	**31%**	**17%**	**11%**	**3%**

Table 10. How Regularly Respondents Link Data

Type of Office	All Are Part of Our Regular Processes	Most Are Regular; Few Ad Hoc Projects	About Evenly Split	Mostly Ad Hoc; Few Regular Processes	All Are For Ad Hoc Projects
Evaluation Office (n = 6)	0	0	1	1	4
Principal Statistical Agency (n = 10)	0	4	2	1	3
Other (n = 20)	2	4	4	4	5
Total (n = 36)	**2**	**8**	**7**	**6**	**12**
Subtotals	**10**		**7**	**18**	

Table 11. Provision of Different Types of Metadata[f]

Type of Office	Variable Definitions, Units of Measurement, Response Ranges	Changes to Definitions, Scope, or Data Collection Processes	Imputation Procedures or Other Edits
Evaluation Office (n = 8)	100%	63%	88%
Principal Statistical Agency (n = 13)	92%	85%	85%
Other (n = 44)	57%	55%	39%
Total (n = 65)	**69%**	**61%**	**53%**

[f] The universe for Table 11 only includes the 65 offices that collect data.

Table 12. Respondent Processes for Disclosure Review[g]

Type of Office	Process for Disclosure Review		Process for Documenting Review Results		Staff for Doing Disclosure Review	
	Yes	No	Yes	No	Yes	No
Evaluation Office (n = 5)	2	3	0	0	0	2
Principal Statistical Agency (n = 13)	12	1	8	1	6	5
Other (n = 8)	6	2	3	1	4	2
Total (n = 26)	**20**	**6**	**11**	**2**	**10**	**9**

Table 13. Where Linkage Activities Occurs

Type of Office	Other Office	Our Office	Third Party	Outside Parties On-Site	Academic Research Institutions	Total
Evaluation Office (n = 6)	13	5	3	0	0	21
Principal Statistical Agency (n = 10)	12	36	2	9	3	62
Other (n = 20)	13	48	1	0	0	62
Total (n = 36)	**38**	**89**	**6**	**9**	**3**	**145**

Statistics reported here are summed across the 13 types of data.

As mentioned above, the Census Bureau is the only principal statistical agency that reports having more than six dedicated staff for data-linkage activities. At the time of the survey, these staff supported activities in the Census Bureau's existing data-linkage infrastructure.[10] This infrastructure in the Center for Administrative Records Research and Applications and the Center for Economic Studies supports small teams of researchers linking Census Bureau data to other administrative or survey data. These activities are limited to those that relate to the Census Bureau's mission.[11] Some outside researchers are also able to take advantage of this infrastructure through the Federal Statistical Research Data Centers. These physical locations across the country restrict access to data to authorized researchers for approved purposes, ensure that data are used for exclusively statistical activities, and apply disclosure avoidance protocols.[12]

Only a few offices, such as the Census Bureau, report working with outside parties or academics to perform data linkages (see Table 13). The commission noted that many agencies lack the necessary authority to use tools like grants and cooperative agreements for such activities. This is unfortunate, as these tools can be important for building evidence capacity through external partnerships, especially for smaller agencies or those in need of extra staff capacity or expertise.

Overall, offices face significant barriers in accessing and linking data based on restrictions from other agencies. Table 14 summarizes the frequencies of these numerous barriers reported by survey respondents. Agencies suggested that statutes, regulations, and policies in other agencies are by far the most widespread barrier, followed by in-office difficulties establishing data-sharing agreements and a lack of requisite expertise to perform data linkages. The lack of information technology (IT) systems capacity in-office and concerns from other offices and agencies about data security were also identified as notable barriers by multiple offices.

[g] Due to item nonresponse, Table 12 only includes 26 of the offices whose confidential data is accessed by outside parties.

Agencies further elaborated on information about barriers to linking confidential data in open-ended responses provided to the commission (see Table 15). They suggested substantial concerns related to a lack of consistent procedures, a lack of resources, and restrictive laws and policies preventing effective data sharing and linkage.

Despite the significant barriers discussed above, as reflected in Table 16, the majority of offices surveyed that collect confidential data also reported that they engage in at least some linking of confidential data (53 percent). Linking data appears to be even more common for evaluation offices (60 percent) and principal statistical agencies (77 percent).

Table 14. Types of Data Respondents Found Hard to Access

Barriers to Accessing Data for Evidence-Building Activities																			
Type of Data	# of offices reporting data are hard to access	Regulations or Policies That Make It Difficult to Link Data			Statutes Prohibiting Data Sharing			Lack of Staff, Policies, and Procedures for Data-Sharing Agreements			Lack of Staff with Technical Expertise Needed to Link Data			Lack of Capacity in IT Systems			Concerns About Keeping the Linked Data Secure		
		Own Office	Other Office or Agency	State	Own Office	Other Office or Agency	State	Own Office	Other Office or Agency	State	Own Office	Other Office or Agency	State	Own Office	Other Office or Agency	State	Own Office	Other Office or Agency	State
Criminal Justice	12	2	5	3	1	5	1	4	5	1	5	3	2	1	2	2	0	3	0
Census/ACS	7	0	4	na	0	3	na	0	2	na	0	0	na	0	0	na	0	0	na
Federal Education	4	0	2	na	0	2	na	0	1	na	0	0	na	0	0	na	0	1	na
State Education	11	0	5	2	0	4	3	0	0	0	0	0	0	0	0	0	0	0	2
Federal Business	6	0	3	na	0	2	na	0	0	na	2	1	na	2	0	na	0	0	na
Federal Establishment	8	0	3	na	0	4	na	1	0	na	0	0	na	0	0	na	0	0	na
Federal Tax	22	1	15	na	2	12	na	7	2	na	2	2	na	3	2	na	2	6	na
Social Security	11	0	4	na	0	3	na	3	2	na	2	3	na	1	0	na	0	2	na
Human Services	9	0	6	2	0	3	1	2	0	0	1	0	0	1	0	0	0	2	0
Medicaid/ Medicare	12	1	5	1	0	3	0	2	1	1	2	1	0	0	0	0	1	2	0
Military	7	1	4	na	0	0	na	1	3	na	1	1	na	1	2	na	0	1	na
Wage	11	0	4	4	0	4	3	1	2	3	0	0	0	1	0	0	1	1	0
Vital Records	12	0	6	2	0	2	1	2	0	2	1	0	0	0	0	0	1	4	1
Total	**79**	**5**	**66**	**14**	**3**	**47**	**9**	**23**	**18**	**7**	**16**	**11**	**2**	**10**	**6**	**2**	**5**	**22**	**2**

Table 15: Respondents' Explanations of Most Critical Barriers related to Data Linkage Activities

- "The lack of standard procedures or guidelines for sharing data across federal agencies that fund research makes efforts to link and share data difficult or inefficient."
- "Lack of planning for linking by program agencies is a factor."
- "Finding Agency point person to coordinate on data-sharing agreements and put in substantial leg work to troubleshoot, execute, and get senior-level support is very challenging."
- "Procuring or licensing data, whether federal or commercial, is expensive. The procurement costs are high, and the costs for staff time to learn about and use acquired sources is also high."
- "The most critical barrier to data exchange is legal and disclosure limitations … Statutes and systems of records notices (routine uses) generally (with certain exceptions) limit our authority to disclose … data to research use."
- "We have great authority to access administrative data, yet there are some files which would greatly advance our mission that we either cannot access or have great difficulty accessing. The [National Directory of New Hires] and SNAP/WIC are examples of these two categories."
- "Many agencies have restrictive requirements or restrictive interpretations of confidentiality laws and regulations that make it difficult to access valuable supplemental data."

Table 16. How Many Offices Collect and Link Data?

Type of Office	Collect Data	Collect Confidential Data	Link Data
Evaluation Office (n = 10)	8	8	6
Principal Statistical Agency (n = 13)	13	13	10
Other (n = 67)	44	39	20
Total (n = 90)	**65**	**60**	**36**

Table 17. Which Offices Keep Linked Data?

Type of Office	No	Yes, but not Regularly	Yes, Routinely	Total
Evaluation Office (n = 6)	1	2	3	6
Principal Statistical Agency (n = 10)	2	1	7	10
Other (n = 20)	4	5	10	19
Total (n = 36)	**7**	**8**	**20**	**35**

Table 17 suggests that the majority of offices that link data also regularly retain data (28 offices out of 35). The survey did not, however, ask respondents to clarify how long they maintain data, as some of these could be for relatively short periods of time or for purposes prescribed in law. Many agencies retain data for defined periods of time consistent with the Fair Information Practices Principles. These principles encourage a concept called data minimization, through which data are only retained in an identifiable form until the purpose for which they were collected is achieved.[13]

Barriers to Using Data

Legal limitations and restrictions related to data access and linkage were identified as barriers to evidence-building activities, as presented in the previous sections. Table 18 suggests that these barriers are largely in specific data-providing offices, such as Title 26 and the Internal Revenue Service. But the respondents to the commission survey also suggested there are significant barriers within offices trying to access data. Notably, although the commission identified contract management as one of these barriers,[14] it appears to be the least significant of those listed by survey respondents, with only 4 percent of offices reporting it as a major barrier and only one office identifying it as their most critical barrier.

The most prevalent barrier by far appears to be inadequate funding—it is the most critical barrier reported by nearly a third of all offices. This is consistent with the commission's fact-finding, during which resources "repeatedly emerged" as a major challenge for federal evidence-building.[15] Several open-ended responses expand on the issues and indicate that their offices' primary barrier is an all-around lack of funding (see Table 19). For example: "Most of our resources go to meet basic reporting requirements, leaving limited time/personnel for research." Many others, however, highlight that their evidence funding difficulties are due to differing priorities:

- Some highlight a trade-off between evaluation and providing program services: "We are primarily a health system for impoverished people. Any dollar misspent results in less medical care" and "There are resource trade-offs involved in spending our limited budget on studies to evaluate the effectiveness of our STEM education programs, rather than on the programs themselves."
- Other offices point to political considerations. For example: "We would use data to analyze and allocate budget funds. So many are already earmarked or are political initiatives that very little latitude for analysis is left."

Several offices, in their open-ended responses, noted that a lack of funding leads to other resource-related difficulties, such as in staffing and IT infrastructure.

Table 18. Reported Barriers to Using Data for Evidence-Building

Severity of Barrier	Type of Office	Barriers to Using Data for Evidence-Building Activities								
		Inadequate Funding	Legal Limitations	Lack of Ability to Hire Appropriate Staff	Privacy and Data Security Requirements	ICR Requirements	IT Systems or Requirements	Political Processes	Lack of Ability to Execute and Manage Contracts	Other
Most Critical	**Total (n = 79)**	**32%**	**19%**	**13%**	**11%**	**8%**	**4%**	**5%**	**1%**	**8%**
	Evaluation Office (n = 10)	*20%*	*20%*	*0%*	*30%*	*0%*	*10%*	*10%*	*10%*	*0%*
	Other (n = 56)	*38%*	*7%*	*14%*	*11%*	*11%*	*4%*	*5%*	*0%*	*11%*
	Principal Statistical Agency (n = 13)	*15%*	*69%*	*15%*	*0%*	*0%*	*0%*	*0%*	*0%*	*0%*
Major	**Total (n = 79)**	**34%**	**20%**	**16%**	**15%**	**16%**	**14%**	**8%**	**4%**	**3%**
	Evaluation Office (n = 10)	*10%*	*20%*	*10%*	*30%*	*30%*	*20%*	*10%*	*10%*	*10%*
	Other (n = 56)	*39%*	*13%*	*18%*	*13%*	*18%*	*14%*	*2%*	*4%*	*2%*
	Principal Statistical Agency (n = 13)	*31%*	*54%*	*15%*	*15%*	*0%*	*8%*	*31%*	*0%*	*0%*
Moderate or Major	**Total (n = 79)**	**61%**	**52%**	**54%**	**47%**	**46%**	**44%**	**27%**	**16%**	**9%**
	Evaluation Office (n = 10)	*40%*	*80%*	*50%*	*90%*	*60%*	*50%*	*20%*	*10%*	*10%*
	Other (n = 56)	*66%*	*41%*	*54%*	*41%*	*50%*	*45%*	*23%*	*14%*	*9%*
	Principal Statistical Agency (n = 13)	*54%*	*77%*	*62%*	*38%*	*15%*	*38%*	*46%*	*31%*	*8%*

Table 19. Respondents' Elaborations on Funding as Their Most Critical Barrier and How That Can Cause Other Barriers

- "Office is understaffed, thus the time and attention necessary for analysis suffers."
- "The approval processes for IT transformation in our bureau are subjected to a team approach, and people on the team often have other funding priorities."
- "Funding is often not available for historical data digitization and necessary investments in information technology infrastructure."
- "Having sufficient resources to hire, put IT systems in place and making statistics, research a priority."
- "Funding affects our ability to hire adequate staff to support outside researchers, develop new products on our own, explore new data sources, etc. We have been under a hiring freeze for several years and have not hired from outside our agency in at least 4 [years]. Existing staff have been frozen at relatively low journeyman pay levels due to a freeze on promotions. All this limits our ability to bring in staff with new analytical skills to help the organization embrace newer technologies and processes, including tools associated with Big Data. Budgets have been flat while expenses have increased steadily. We have had some success leveraging outside academics to help innovate. Just a few new hires would make a significant difference in our program."

Foundations for Evidence-Based Policymaking Act

The commission made recommendations to strengthen overall federal evidence-building capacity, which includes privacy protections and data access. It focused on strengthening capacity without relying on recommendations that would require vast resources with new appropriations or major organizational expansions.

Several aspects of the commission's recommendations that could begin to address barriers identified in the commission survey are included in the Foundations of Evidence-Based Policymaking Act (H.R. 4174). H.R. 4174 includes provisions that would require both Chief Evaluation Officers (recommendation 5-1) and learning agendas (recommendation 5-2). In the case of Chief Evaluation Officers, the act encourages designating an already present senior official rather than hiring new staff, when possible, to minimize the expansion of personnel.

H.R. 4174 could begin to address some legal barriers identified by respondents with modifications to CIPSEA that would encourage data use within a privacy-protective framework. The bill includes a provision that would establish a presumption of access for evidence-building activities under CIPSEA. A July 2018 paper from the Bipartisan Policy Center, reviewing the legal and ethical framework under which federal agencies collect and use confidential data, points out that a lack of clarity among program administrators on permissibility of certain uses of data constrains evidence-building activites even beyond the actual restrictions laid out in law.[16] For example, in the commission survey the Department of Veterans Affairs, which cited legal limitations as their most critical barrier to using data and suggested that the key issue they face is "a risk-averse application of law or policy."

H.R. 4174 includes provisions that would bolster metadata, data standardization, and data inventories available from government agencies. Yet another provision would provide a portal for researchers to have a common interface for beginning data requests to agencies. Collectively, the recommendations that are reflected in the legislation address 10 of the commission's 22 recommendations.[17]

Next Steps

The original commission survey and, by extension, this paper focus on a broad range of issues related to data collection, access, sharing, and use. Absent are concerns at other parts of the evidence-using process, such as ensuring that data analysis is rigorous and that agencies are identifying useful research questions to prioritize. The survey was also not designed to address the broader culture of evidence in the federal government for evidence production and use, which may not be meaningfully assessed by a survey sent to only one senior official per office. This extended analysis of the commission survey, however, suggests that the issues discussed by the commission are real and substantial barriers to evidence-building activities in government. Fielding an extended or updated survey to federal offices and staff could prove productive in the future as continued progress is made in implementing the commission's recommendations.

★ ★ ★

While many agencies are capable and manage their data well, there are also pockets of government today – largely outside the statistical agencies or evaluation units – that lack the capacity for executing this work well.

Beyond H.R. 4174, this extended survey analysis bolsters the case the commission presented for the need and design of the National Secure Data Service. The commission recommendations included potential solutions beyond those included in H.R. 4174, meaning additional efforts will be needed within Congress and the executive branch to fully realize the commission's vision. Chief among areas for future consideration is the commission proposal for establishment of a National Secure Data Service (recommendation 2-1), which would directly address many of the concerns identified by the agencies in the survey. While H.R. 4174 includes the establishment of an advisory committee to further consider how a data service would be implemented, it does not take steps to establish the shared service unit.

Central to the commission's design of the National Secure Data Service was a capacity to facilitate and assist with data linkages by establishing shared expertise and resources. Agency survey responses suggested a notable gap in expertise and resources to support data linkage of confidential data. A shared-service operation such as the National Secure Data Service would be useful in addressing this gap. The National Secure Data Service would increase the evidence-generating capacity of other offices, especially those that are resource-constrained when it comes to accessing and linking data.

Recognizing the gaps that exist today in securely performing such activities with transparency, the nominal investment could produce invaluable results for the American public. For example, the commission recommended (recommendation 4-3) that the National Secure Data Service maintain a searchable inventory of approved projects using confidential data and that these projects be audited regularly to ensure that privacy is being fully protected. This system promotes both efficiency and accountability, reducing redundancy while making it easier for the public to see how linked data are used and protected.[18] The commission also called on the National Secure Data Service to use its expertise to research and model data privacy best practices and technologies.

Developing the National Secure Data Service, as recommended by the commission, could viably address the major challenges facing the evidence-building community and help meet the demand for information from policymakers. Stringent requirements for project approval and access to confidential data could remain in place—and in some cases be made stronger—as a result of establishing such a service.

While the National Secure Data Service holds promise as a trusted way to steward sensitive information, it must be coupled with transparency about what happens, technical capacity for evidence building, and the legal authorities to ensure that strong privacy protections and safeguards are put in place. If designated as a statistical agency, such protections would be made available under CIPSEA with criminal and civil penalties for violations of confidentiality protections.

The absence of such an ability to address capacity gaps within a strong legal framework means that opportunities to meet the needs of policymakers and the American public in supporting effective and efficient government will inevitably be lost. In turn, taxpayer dollars may be needlessly wasted when small investments in data infrastructure could otherwise help address these issues.

In sum, as the respondents to the commission survey identified, many data-linkage activities are already underway across government. While many agencies are capable and manage their data well, there are also pockets of government today—largely outside the statistical agencies and evaluation units—that lack the capacity for executing this work well. The survey results strongly suggest, in line with the commission's final recommendations, that government needs an improved capacity to engage in this work with better capabilities and an organizational infrastructure for securely accessing and analyzing confidential data.

Appendix: Respondent Characteristics

Agency	Number of Eligible Offices	Offices Responding	Response Rate	Percent of Total Responses
Department of Agriculture	19	4	21%	4%
Department of Commerce	19	5	26%	6%
Department of Defense	4	2	50%	2%
Department of Education	13	6	46%	7%
Department of Energy	10	4	40%	4%
Department of Health and Human Services	39	24	62%	27%
Department of Homeland Security	9	1	11%	1%
Department of Housing and Urban Devel..	6	1	17%	1%
Department of Justice	12	5	42%	6%
Department of Labor	8	4	50%	4%
Department of State	3	1	33%	1%
Department of Transportation	11	2	18%	2%
Department of Veterans Affairs	5	3	60%	3%
Department of the Interior	10	5	50%	6%
Department of the Treasury	8	4	50%	4%
Environmental Protection Agency	8	2	25%	2%
National Science Foundation	3	3	100%	3%
Small Business Administration	2	2	100%	2%
Social Security Administration	7	5	71%	6%
Agency for International Development	2	0	0%	0%
Single Bureaus and Commissions	11	7	64%	8%
Corporation for National and Community Service		1		1%
Equal Employment Opportunity Commission		1		1%
Nuclear Regulatory Commission		1		1%
General Services Administration		1		1%
Millennium Challenge Corporation		1		1%
National Aeronautics and Space Administration		1		1%
National Archives and Records Administration		1		1%
Total	**209**	**90**	**43%**	**100%**

Offices were classified through the following processes: evaluation offices through the Interagency Council on Evaluation Policy and through the expertise of commission staff, and principal statistical agencies through the *Statistical Programs of the United States Government: Fiscal Year 2016*. "Other" comprises all offices not in the two preceding groups, including several programmatic offices.

Endnotes

1 Commission on Evidence-Based Policymaking (CEP). *The Promise of Evidence Based-Policymaking: Final Report of the Commission on Evidence-Based Policymaking.* 2017a. Available at: https://bipartisanpolicy.org/wp-content/uploads/2018/07/Full-Report-The-Promise-of-Evidence-Based-Policymaking-Report-of-the-Comission-on-Evidence-based-Policymaking.pdf.

2 CEP. *The Promise of Evidence Based-Policymaking: Final Report of the Commission on Evidence-Based Policymaking.* "Appendices E-H." 2017b. Available at: https://bipartisanpolicy.org/wp-content/uploads/2018/09/Appendices-e-h-The-Promise-of-Evidence-Based-Policymaking-Report-of-the-Comission-on-Evidence-based-Policymaking.pdf.

3 The complete data file for the survey is available at: https://bipartisanpolicy.org/wp-content/uploads/2018/07/CEP-Survey-Analysis-File.xlsx.

4 CEP, 2017a.

5 44 U.S.C. § 3501.

6 Nicholas Hart and Katherine Wallman. Transparency, Accountability and Consent in Evidence Building: How Government Ethically and Legally Uses Administrative Data for Statistical Activities. Bipartisan Policy Center, 2018.
Available at: https://bipartisanpolicy.org/library/transparency-accountability-and-consent-in-evidence-building/.

7 Ibid.

8 CEP, 2017a.

9 CEP, 2017a, 78-79.

10 CEP, 2017a, 81.

11 CEP, 2017a.

12 See: U.S. Census Bureau. "Data Linkage Infrastructure." Last revised July 5, 2017. Available at: https://www.census.gov/about/adrm/linkage/guidance.html.

13 CEP, 2017a, 49.

14 CEP, 2017a, 98.

15 CEP, 2017a.

16 Hart and Wallman, 2018.

17 Nick Hart and Sandy Davis. "FACT SHEET: Foundations for Evidence-Based Policymaking Act." Bipartisan Policy Center, 2017.
Available at: https://bipartisanpolicy.org/blog/fact-sheet-foundations-for-evidence-based-policymaking-act/.

18 CEP, 2017a, 85.

Notes

Notes

BIPARTISAN POLICY CENTER

The Bipartisan Policy Center is a non-profit organization that combines the best ideas from both parties to promote health, security, and opportunity for all Americans. BPC drives principled and politically viable policy solutions through the power of rigorous analysis, painstaking negotiation, and aggressive advocacy.

bipartisanpolicy.org | 202-204-2400
1225 Eye Street NW, Suite 1000
Washington, D.C. 20005

@BPC_Bipartisan
facebook.com/BipartisanPolicyCenter
instagram.com/BPC_Bipartisan

BPC POLICY AREAS

Economy

Education

Energy

Evidence

Finance

Governance

Health

Housing

Immigration

Infrastructure

National Security

BIPARTISAN POLICY CENTER

1225 Eye Street NW, Suite 1000 | Washington, D.C. 20005
202-204-2400| bipartisanpolicy.org