TABLE OF CONTENTS

Discussion: Models for a Data Facility

At the December 12 and January 13 Commission meetings, we learned about Federal, state, and international models for acquiring, managing, and making accessible data for research and evaluation purposes. At the January 13 meeting commissioners received a matrix that classified these models according to several dimensions, and commissioners suggested developing several data facility "toy models" to facilitate consideration of options. During the February 24 meeting, commissioners will have the opportunity to begin discussing the dimensions and their implications.

* * * * *

Discussion Goals

1) Become familiar with the framework for evaluating models of a data facility.
2) Consider and discuss how the dimensions relate to each other.
3) Begin to identify pros and cons for different potential models and attributes of a potential data facility.
4) Narrow the list of models to be considered.

Overview of Model Categories and Definitions

The long list of dimensions previously discussed in earlier meetings were classified into three categories:

1) Dimensions of the **structure** of the data facility
2) Dimensions related to the **administration** of the data facility
3) Dimensions related to **data security and privacy protection**

The dimensions within each category are defined below under the heading, *Dimensions of a Data Facility.*

We suggest starting with the question of facility purpose, which we have included within the category of dimensions called **data facility structure**. Dimensions within the other two categories—administration of the data facility and data security and privacy protection—are informed by the data facility structure choices. For the purpose of this discussion, it will be helpful to hold constant the scope of a facility, therefore for this discussion we will limit focus on government demographic data.

There are three key dimensions within the structure of a data facility category prioritized for developing the toy models: **services**, **data architecture**, and **operational centralization**. Under the heading *Toy Models for the Structure of a Data Facility* below, we present the potential combinations of these three key dimensions.

In the development of the models displayed in the table, some assumptions were made about what combinations of dimensions would result in models that were not logical given the dimension definitions. For example, because the network data architecture was defined as one in which the resulting linked datasets would not be maintained, a network architecture would not likely be able to provide analytical services. Also, note that the network and hybrid models could require the application of technology that is still largely theoretical or that has not been implemented on a large scale.

Discussion Session
The goal of presenting these "toy models" is to facilitate **a discussion about what models seem most feasible for achieving the Commission's desired outcomes** in developing recommendations about a data facility, particularly with respect to a facility's utility and privacy features. Starting with the **services** dimension at the top, consider what services are desired for a data facility to provide, and then, what choices about data architecture and centralization can best achieve that outcome. In this way we will collectively consider the pros and cons of the various attributes of the different models.

Consider especially the following criteria in evaluating the toy models:

- Implications for data confidentiality and security
- Public perception and trust
- Capacity and funding
- Availability of state and other data
- Control of data access decisions

During the meeting, there will be an opportunity for small group discussions. After the small group discussions, we will reconvene to debrief, outline next steps, and satisfy the goal to develop a narrower set of potential models (ideally 3-4) for future consideration. From there the staff can further develop the remaining models and provide more information on the legal and policy implications for recommending each model.

Dimensions of a Data Facility

Dimensions of the Structure of a Data Facility (See Toy Models Table)	Dimensions Related to the Administration of the Data Facility	Dimensions Related to Security and Privacy Protection
Primary Dimensions **Services** – What data management and access services are provided by the facility? • Data Access Only – Program Management Office (PMO) provides access. Researcher does own curation and linkage. • Data Access and Linkage Services – PMO provides access and curates and links data. • Access, Linkage, and Analytical Services – PMO curates, links, and performs analysis for the customer. **Data Architecture** – Where are data located and how are they accessed? Are data moved into a central facility (*repository*), connected on-demand (*network*), or some combination (*hybrid*)? • Repository – Data are moved into a central location in order to provide secure access and to allow for linkages (as in the CARRA model). Linked datasets are generally kept in the repository as well. • Network – Data remain siloed in their original location and are accessed on-demand for linking. This could be done by sending encrypted files (as in the UK's ADRN model), or by the application of a technological protocol such as secure multi-party computation (note the latter is still largely theoretical and not yet implemented to scale). Linked datasets are not maintained in the system (although metadata for linking may be maintained). • Repository-Network Hybrid – Some data are kept in a repository (for example, to create a spine); other data are accessed from their original location on-demand (as described under the Network option) to allow linking. Linked datasets are generally kept and secure access is provided to researchers. **Operational Centralization** – Are approvals (person, project, data access) and operational decisions (MOUs, training, etc.) centralized or de-centralized? • Fully centralized – Centralized approvals, centralized operational decisions • Semi centralized – De-centralized approvals, centralized operational decisions • De-centralized – De-centralized approvals, de-centralized operational decisions *Other Dimensions* **Source of Data** – Does the data facility include data from the Federal government, state/local government, private sector? **Breadth** – What number and types of datasets are covered? What policy domains are covered?	**Institution** – Under what entity type is the model classified? • Federal government entity • Quasi-governmental entity • Consortium of government and non-government entities **Governance Structure** – Who has authority over the data facility? What laws and policies govern the data facility? What stakeholders are involved? • A single actor, such as a Chief Statistician • A single actor, with an Advisory Board • An appointed Board of Governors • A stakeholder representatives Board **Funding/Budget** – What are the sources of funding? These could include a combination of: • Federal appropriation • Researcher fees • Agency contributions • State contributions **Staffing** – How is the data facility staffed, i.e. where do staff come from? **Public Engagement** – How does the interaction with the public ensure transparency, promote accountability, and build trust?	**Privacy-Protecting Technology** – What technology is employed to ensure privacy and confidentiality? These could include: • Differential privacy • Query-language approach • Secure multi-party computing • Synthetic data verification model • Physical enclave • Virtual enclave • A combination (via tiered access) **Data Access Control Methods**

Toy Models for the Structure of a Data Facility

	Service: Data Access Only						
Data Architecture	Repository		Network*			Repository-Network Hybrid*	
Operational Centrali-zation	Semi centralized	Fully centralized	De-centralized	Semi centralized	Fully centralized	Semi centralized	Fully centralized

	Service: Data Access and Linkage						
Data Architecture	Repository		Network*			Repository-Network Hybrid*	
Operational Centrali-zation	Semi centralized	Fully centralized	De-centralized	Semi centralized	Fully centralized	Semi centralized	Fully centralized

	Service: Access, Linkage, and Analytical			
Data Architecture	Repository		Repository-Network Hybrid*	
Operational Centrali-zation	Semi centralized	Fully centralized	Semi centralized	Fully centralized

* Network and hybrid options would bring data together on-demand, either by sending encrypted files, or by the application of a technological protocol such as secure multi-party computation (note the latter is still largely theoretical and not yet implemented to scale).

Discussion Document: Data Facility

At the February 24 meeting, we presented a framework for talking about the dimensions of a data facility, including its 1) structure, 2) administration and governance, and 3) data security and privacy protections. During small group discussions in February, we focused on the *structure* of the data facility. Part I of this memo builds on that work by developing a more fully formed description of a *structure* model for discussion, based on the similar models around which each small group coalesced, and which we will call the **Data Facility Network**.

During the March 13 meeting we can discuss any critical open points with the *structure* model, yet plan to devote the majority of the data facility discussion to elaborating on the administration and governance dimensions, described in Part II, including the institutional structure, governance approach, administrative processes/policies, and resource needs. At the meeting, we will present a strawman for the *administration and governance* structure given the Data Facility Network choice and pose a set of case problems for the Commission to consider (to be provided during the meeting).

* * * * *

Part I: Data Facility Structure

The Data Facility Network is a repository-network hybrid architecture. The initial scope involves connecting to core datasets from the Federal government and selected data from Federal-state data partnerships.[1] The components of the Data Facility Network include:

- **Program Management Office (PMO).** PMO provides data access, curation, linkage, and analytical services.
- **Core Repository.** In the short run, the Core Repository is operated by the PMO. Similar to CARRA, it maintains key linking variables from core sources to create a spine for linking other datasets. It also maintains, in linkable form, other priority datasets such as Unemployment Insurance (UI) earnings and wage data. We will assume for now that this is CARRA. As technology matures to support linkages without a spine, the Core Repository should cease to maintain unnecessary datasets.
- **State Repositories.** State Repositories provide access to certain state-level datasets, such as human services data and state education data.
- **Agency Repositories.** Agency Repositories maintain other Federal data. Agency Repositories include:
    - The Centers for Medicare and Medicaid Services (CMS) Chronic Conditions Warehouse, which includes information on Medicare and Medicaid beneficiaries across the continuum of care;
    - The Department of Education's core data sets, including: the Common Core of Data (CCD), the universe of public elementary and secondary schools; the Integrated Postsecondary Education Data System (IPEDS), the universe of postsecondary institutions eligible for Federal student aid; and the National Student Loan Data System (NSLDS), which includes data on Federal student loan recipients;

[1] See, for example the priority datasets list attached.

- The Internal Revenue Service (IRS) Statistics of Income Division (SOI) Compliance Data Warehouse (CDW), which includes microdata for the universe of tax filers; and
- The National Center for Health Statistics (NCHS) National Death Index (NDI) and vital statistics records.

A core function of the PMO is to link datasets from more than one repository "on demand" for a particular project. Initially, State and Agency Repositories send datasets to the Core Repository as encrypted files; as technology advances this could evolve to more advanced applications of privacy-protecting protocols (e.g., secure-multiparty computing). The Core Repository maintains the resulting linked datasets for a set period of time, determined through the governance process and applicable laws. The PMO maintains and keeps indefinitely the metadata required to link datasets together, consistent with policies determined by the government entity. Data Providers from the State and Agency Repositories play a key role in curating data and supporting the technical aspects of data sharing, such as making sure data include appropriate metadata. The PMO works with the Data Providers from the State and Agency Repositories to update the metadata as datasets change.

The Data Facility Network can function with either fully-centralized or semi-centralized approvals. Under a fully-centralized approval system, the central governing body approves projects and researchers (see options below for governance). Under a semi-centralized model, the central governing body or administrator handles operational decisions (such as the application process and the training requirements), but data owners retain the right of approval for individual projects and researchers. Data Providers may choose to delegate their approval authority to the central governing body or a centralized approval authority. Such a model allows Data Providers to remain involved in decision-making, while allowing for efficiencies in the approvals process.

The PMO also provides analytical services for states and agencies, but this is not its primary function. The PMO could provide these services on a fee-for-service basis or as a quid pro quo for data access.

Assumptions about Legal Authority for the Data Facility Network

For the purposes of this discussion, we will assume that the Data Facility Network has statutory authority to operate and that there is a mechanism for compelling agencies to make their data available as part of the network, subject to existing statutory limits of use of the data. If we were to relax this assumption and assume that additional statutory changes were already in place, we could envision some changes to the model laid out above. For example, if the current ban on creating a Federal student-unit level database were lifted, the Department of Education could expand IPEDS into both an institution- and student-level universe data collection, which would provide a mechanism for accessing student-level data through the Data Facility Network. Also, if Title 26 authority to access tax data were broadened, tax data could potentially be incorporated into the Core Repository.

* * * * *

Part II: Data Facility Governance and Administration

Part II of the document is informed by comments provided during a call with Commissioners Glied, Groves, Meyer, Rice, Shea, Troske, and Wallin on March 8.

II.A. Attributes and Objectives for Facility Governance

Building on the structure of a potential facility and articulating what functions (services) a facility would have to support evidence-building, we next consider how a facility would be governed and administered. Much like the evaluation criteria for the design, we begin by positing attributes or objectives that can be addressed in full, or in part, by governance and administration.

- **Objective #1: Support for Evidence-Building.** The purpose of the data facility is ultimately to provide services that support and enhance existing evidence-building endeavors, including by making microdata available across policy domains and jurisdictions to support robust analyses. Thus, enabling the governing entity to consider and modify key policy priorities and areas of focus over time to reflect policymakers' needs for evidence will inevitably relate to the facility's ability to actually support evidence-based policymaking.

- **Objective #2: Positive Perceptions and Public Trust**. A positive perception of the system that aligns with the public's values is a necessary feature for continued public support (funding and other means), as well as continued participation between entities and the facility. Importantly, a failure to maintain a positive perception and the public's trust will likely result in a slower adoption of policies to support data sharing and evidence-building, as well as potentially jeopardize the existing infrastructure that exists to support this work. In addition to features of the (not yet articulated) *privacy and security* dimension, aspects of public engagement, nonpartisan operation, sufficient oversight, apolitical or objective decision-making, and enforcement mechanisms for violations may contribute to satisfying this objective by demonstrating to the public that their values would be protected.

- **Objective #3: Strategic Coordination and Collaboration.** A successful facility will require participation from numerous entities within and among governments, as well as from non-governmental researchers and evaluators. A governance structure will need to have appropriate authority to incentivize, or mandate as appropriate, participation. The need for coordination and collaboration to bridge statutory and operational hurdles suggests that administrative features of a data facility must also seek to address these same issues through continuous partnerships and meaningful roles in appropriate decision-making processes.

- **Objective #4**: **Privacy and Security.** In the 1960s, designers of the National Data Center failed to articulate how such a system could be designed to facilitate a reasonable level of trust and to protect individuals' privacy. The Commission's discussion of principles has reinforced the belief that privacy and security are important features for any future data facility. A governance structure that incorporates privacy-minded executives and protocols and processes that periodically review privacy protections may help ensure this objective is consistently pursued.

- **Objective #5: Authority and Flexibility.** A facility will need sufficient statutory authority to execute the articulated purposes, including acquiring, linking, and providing access to data currently managed under a variety of statutory authorities. It will also need sufficient statutory flexibility to adapt operations and policies over time to changes in government, culture, technologies, and the needs of social science research and evaluation. This is particularly the case for approaches to facilitating privacy by incorporating emerging technologies and protocols into existing processes.

- **Objective #6: Scalable Functionality.** Scalability involves both data acquisition/management and access to a facility. A facility will likely be most successful if it ultimately recognizes existing infrastructure in government, sufficiently pilots changes or expansions, and prepares to expand with a larger infrastructure as needed or desired based on demand. One approach to managing scalability would be to first prioritize government data in a specified set of areas (e.g., income and human services), then build to incorporate more data in future iterations, and potentially include private data in a longer-term plan.

- **Objective #7: Sustainability.** The expected sustainability of a system will have a direct relationship to the potential users' and data suppliers' willingness to participate. One major aspect of sustainability is constrained costs that satisfy the articulated goals. This involves providing for reasonable fees (or other financing mechanisms) that are justified to users, and fostering a perception that a facility will exist long enough for other Federal agencies to deploy scarce resources to support participation.

II.B. Developing Solution Sets for Facility Governance

Based on the articulated attributes in II.A., this section outlines how each relates to the dimensions of institutional structure, governance approach, administrative processes/policies, and resource needs.

Institutional Setting

The institutional setting is defined as the location within or in relationship to government for the entity that oversees and operates a data facility. The institutional setting can address some but not all aspects of the objectives outlined above. For example, an entirely Federal entity may engender a greater degree of public trust than for-profit entities for some aspects of facilitating the provision of PII; but, public institutions themselves are not a silver bullet for public trust. Plausible options include a facility within an existing Federal entity (e.g., CARRA within the Census Bureau), a relocated existing Federal entity (e.g., moving CARRA elsewhere in the Department of Commerce), a completely independent agency within the Federal government, a quasi-governmental entity, or a private entity.[2] Each option has limitations and opportunities. For any of these options, a collaborative governance process could be established to engender support and cooperation of multiple levels of government and the organizations that would contribute to the information available in the facility.

[2] Other options such as placing the facility within the Executive Office of the President are generally not viewed as feasible.

Existing Federal Entity. Using an existing infrastructure has a practical benefit, building on available administrative systems, existing professional staff, established levels of public trust and operational know-how. The "which one" issue of existing entities cannot be ignored, however, as every agency totes its own baggage and exists in its own cultural context. Relying on an existing Census entity, for example, would have implications tied to the nuanced relationships between Census and other agencies. Importantly, an entity nested within an existing infrastructure must essentially assimilate to the culture, tempo, and bureaucratic tendencies of its parent, and may face challenges of maintaining autonomy and independence. It would also be essential to acknowledge the facility's priority within the overall organizational mission. While examples do exist of "independent authorities" within a larger operating unit, this model involves extra levels of complexity (just ask your Commission staff!) and is not ideal for an on-going mission.

Existing Federal Entity, Relocated. An existing entity could also be reorganized within the Federal government. For example, elevating an office like CARRA to "bureau" status within the Department of Commerce (i.e., as a peer to Census) satisfies the need to take advantage of existing infrastructure for basic administrative overhead and processing, while enabling it to develop its own culture and bureaucratic processes. This, however, necessarily also means being further removed from the benefits of expertise and other infrastructure at Census. The public trust factor is also mixed, leaving behind both the positives and negatives associated with an existing agency brand. At the same time, it provides motivation and opportunity for a champion (e.g., the Commerce Secretary) to invest political capital and facilitate the growth of a distinctly new or modified entity. Similar to the existing Federal entity approach, nesting within an existing infrastructure will introduce challenges of maintaining autonomy and independence from the parent entity. Generally speaking, however, a "bureau" will be more autonomous within a department than a "division" would be within a bureau.

New Federal Entity. A completely new entity within the Federal government likely prioritizes the benefits of greater independence, but could suffer from "start up" costs and administrative delays necessitating by establishing new processes and protocols for basic governmental functions and thus delay the focus on mission. Such an enterprise will be restricted in its ability to capitalize on existing infrastructure, and, to a greater extent than leveraging an existing entity, would impose new costs on the government.

Quasi-Governmental Entity. In contrast to a wholly governmental solution, an alternative is to establish a new quasi-governmental entity, blending certain features of government institutions with private sector entities. A classic example is a Federally Funded Research and Development Center (FFRDC). This benefits from mixed participation in the governance structure and an ability to operate some aspects of the organization outside the strict confines of traditional government rules (e.g., collecting donations for exhibits, private facilities). FFRDCs offer some appeal as formal structures that historically work collaboratively with Federal agencies, yet they are designed in law to be sponsored by a Federal agency to support short-term, singular R&D projects rather than sustained long-term programs.

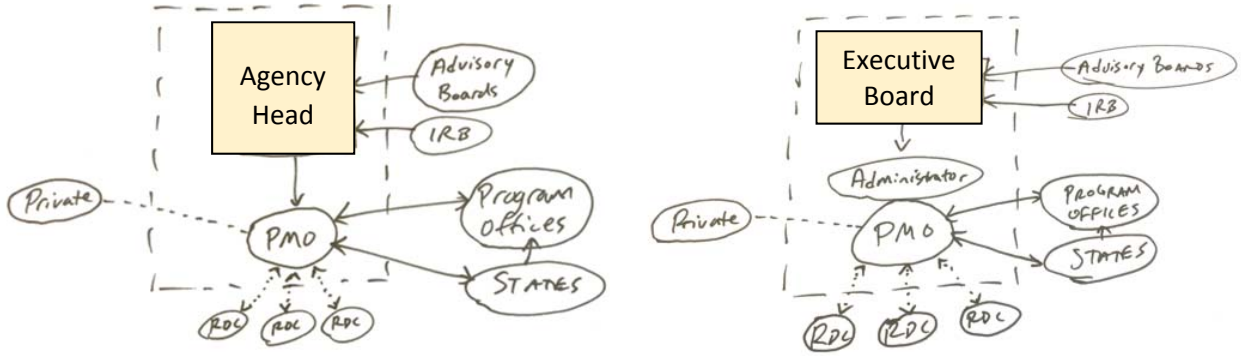
One important challenge for a quasi-governmental entity solution is that sensitive government data would essentially leave the strict control of the Federal government. This presents a new slate of challenges. However, such an arrangement may encourage participation of private entities who would object to placing data within the Federal government. Other important challenges exist for using a quasi-governmental entity. The Congressional Research Service (CRS) has denoted that the

use of quasi-governmental entities highlight a tension between managerial flexibility and a public law approach to management, which is not necessarily designed to maximize performance but to ensure public and political accountability.[3] It is also possible that this option is not exclusive to the others, and that an FFRDC for private data could exist in collaboration with a government entity.

Private Entity. A fifth option is to establish or contract with a private entity to operate and coordinate a facility. Such an organizational structure would not rely on the institutional framework of an official government entity, which may engender a degree of flexibility in operations. However, as with a quasi-governmental entity, in practice, government agencies may be hesitant to share confidential data with a private entity in the absence of accountability and oversight features available for public institutions.

Governance Structure

The governance dimension—defined here as the individual or group of individuals who exercise authority and control over immediately relevant decisions related to the facility—can be thought of separately from the institutional setting. There are at least two distinct models: an agency head (with or without an advisory board) or an executive board with an administrator. Both approaches could exist within any of the five institutional approaches and both would require staff for operations.



Single Executive Approach *(with or without Formal Advisors).* This approach, perhaps the most traditional within government, establishes an individual as the chief executive who operates within the given hierarchical boundaries. That individual could be the U.S. Chief Statistician, the Census Bureau Director, or a new senior level administrative position within the identified institution to oversee the facility, make major policy decisions, and conduct day-to-day operations. This individual could receive feedback from a variety of representative entities through advisory groups, but would have primary decision-making authority. Such an approach would likely create some tension in establishing meaningful partnerships outside the Federal government (thereby limiting overall coordination and collaboration) and may not engender the highest degree of public trust, though this may be mitigated by establishing an advisory board(s).

Example of a Single Executive (with or without Formal Advisors) Approach. Currently CARRA operates within this approach, with the Center Chief (career SES-level civil servant) taking on responsibility for the direction of CARRA as well as day-to-day operational decisions. The model lacks attributes of public accountability and disclosure that may be desirable, though such features could be stacked

[3] Congressional Research Service. (2011). The Quasi Government: Hybrid Organizations with Both Government and Private Sector Legal Characteristics. Washington, D.C.: CRS. https://fas.org/sgp/crs/misc/RL30533.pdf

on top of the existing infrastructure. New Zealand's Government Statistician (head of Statistics New Zealand) operates similarly, but relies on formal advisory committees and other more informal public outreach to inform the government's program for researcher access to microdata.

Executive Board Approach. A second approach could involve an executive board which can reflect a range of relevant interests and expertise in formal decision making, set overall direction and provide oversight of staff-delegated activities. Collectively, these individuals could be selected through a mixed range of mechanisms (i.e., appointment, application), and the Commission may desire to emphasize non-political strategies. The desire to include a broad range of representatives on a governing board must address the tension of having too many decision-makers, which can limit the ability to reach agreement and maintain flexibility. But such an approach may help ensure that varied perspectives from Federal, state, research institutions, privacy and technology experts, or other stakeholders can be represented formally within the decision-making process, not merely in an advisory role.

Examples of the Board Approach. Below are two brief examples of relevant models for consideration of a data facility's governance.

Table 1. Attributes of Select Board Governance Models

	PCORI	**NAGB**
Institutional Setting	Non-profit, independent	Federal, within ED
Reports To	Self	Education Secretary
Appointer	Comptroller General	Education Secretary
Number of Members	21	26
Member Qualifications	Detailed In Statute	Detailed In Statute
Term Length	6	4
Term Limit	2	2
Political	Non-political	Bipartisan
Board Cost	$1m/year[4]	$6m/year[5]
Authority	PL 111-148, Title VI, Subtitle D	PL 107-110

- *Patient-Centered Outcomes Research Institute (PCORI)* – Since 2010, PCORI has operated as an independent, nonprofit organization (created by law) that funds research related to improving patient care and outcomes. The organization is overseen by a 21-member Board of Governors, with members who serve for up to two 6-year terms.[6] Members are appointed by the Comptroller General of the United States based on criteria established in statute, including the designation of Federal positions, members representing patients and consumers, medical professionals, states, and researchers.[7]

[4] Denotes the administrative cost of the Board itself, not the program expenses or grants issued by PCORI. In 2015, PCORI's balance sheet denotes $1m for board administration, $23m for general administration, and $253 in program services.
[5] https://www.gpo.gov/fdsys/pkg/BUDGET-2017-APP/html/BUDGET-2017-APP-1-9.htm
[6] http://www.pcori.org/sites/default/files/PCORI-Bylaws.pdf
[7] http://www.pcori.org/sites/default/files/PCORI_Authorizing_Legislation.pdf

- *National Assessment Governing Board (NAGB)* – This Board is an independent, bipartisan group that establishes policies and procedures for the National Assessment of Education Progress (NAEP), a project of the National Center for Education Statistics (NCES) at the U.S. Department of Education. NAGB is comprised of 26 members, including state elected officials, state and local school officials, educators, researchers, business representatives, and members of the general public. The Board members are appointed by the Education Secretary for up to two 4-year terms.[8]

Other Governance Options. Referenced above, there are multiple other opportunities to provide either a Single Executive or an Executive Board with additional input. For example, the advisory boards for consultation or direction could aid with research or technology, ethical considerations, or priority setting.

Policy Decisions and Administration

Articulating the policies and processes for administration of a facility may be an unnecessary level of detail for the Commission to specify at the outset; however, the Commission may want to explicitly articulate some features that commissioners believe are essential for. Several potential features are highlighted below as illustrative examples.

Policies:

- Compelling or requesting data for linkage
- Transparency procedures
- Privacy and security procedures
- Defining qualified researchers
- Establishment and imposition of penalties for misuse
- Applicability of the Common Rule

Administration:

- Public comment procedures for major policy decisions
- Public notifications of benefits (e.g., providing final studies publicly)
- Data linking and management tasks, including prioritizing datasets for linking and negotiation of data use agreements
- Periodic public and oversight, reporting, auditing
- Approaches to technology or access (e.g., tiered access, secure multi-party computing), including assessing data security

Resources and Sustainability

An important consideration about the institution, governance, and administrative features of the facility is the resource availability and cost. An efficient operation will maximize the benefits while constraining costs. There is a clear tension, for example, between creating a larger governing board or an institution responsible for its own administrative functions (e.g., human resources, procurement) and the goal of constraining costs. Funding for the facility may be absorbed through a variety of mechanisms, including direct appropriations (mandatory or discretionary), reimbursements

[8] https://www.nagb.org/who-we-are/overview.html

from participating agencies, or fees imposed on those who access the facility. The discretionary appropriations process can provide a constraint on costs by ensuring OMB and Congress review the budget each year. Similarly, user fees paid by users of the facility may be set as a representation of the benefits accrued by individuals for the services provided, ensuring the size a facility is a reflection of the demand for use. However, user fees alone may be cost prohibitive and a combination of funding mechanisms may be desired, particularly during system development and start-up as a reflection of the anticipated public benefits of the services.

* * * * *

Part III: Strawman Data Facility Network Governance Strawman

On Monday, March 13, we will present and discuss a strawman proposal that ties together the design features articulated by Commissioners in February and aspects of the governance dimensions described above.

Attachment: Priority Datasets

Administrative data

- **Income, Earnings, or Wages**
  - Tax (IRS, SSA, and Census)
  - Unemployment Insurance (NDNH, LEHD, States)
  - Quarterly Census of Wages (BLS)
- **Health**
  - Medicaid/Medicare/CHIP (HHS/CMS)
  - Vital Records (HHS/CDC/NCHS)
- **Education**
  - Federal Student Aid (ED/NCES)
- **Human Services**
  - SNAP (USDA/FNS)
  - Head Start (USDA/FNS)
  - Child Welfare (USDA/FNS)
  - TANF (USDA/FNS)

Survey Data

- CPS (Census)
- Decennial (Census)
- ACS (Census)
- SIPP (Census)
- Economic Survey Data
- NCHS Surveys (HHS)

Discussion Paper: Tiered Access with a Data Facility

Much of the Commission's discussion thus far has focused on improving agency and researcher access to restricted microdata. However, the Federal government releases far more data than it makes available via restricted access and that is expected to continue with or without a new data facility. With or without employing a formal privacy budget approach, the release of data or results publicly by researchers and agencies creates mutual re-identification risk. Therefore, this memo provides some background on the current state of data dissemination in the Federal government, introduces key considerations for how a data facility might facilitate data dissemination, and discusses the concept of tiered access as a major strategy for addressing re-identification risk.

Data Dissemination by the Federal Government

The U.S. statistical system comprises 13 principal statistical agencies and 115 other programs engaged in statistical activities throughout the Federal government. These agencies collect a wide variety of data on individuals, households, and businesses for statistical, evaluation, research and policy analysis (SERA) purposes. Although underlying microdata are often confidential by law, many statistical programs produce public-use (de-identified and otherwise masked) datasets, aggregate statistics, and summary analyses. Prior to release, agencies use statistical disclosure limitation techniques to protect individual confidentiality. In addition, many statistical programs make restricted microdata available through secure environments, such as a Research Data Center (RDC). Furthermore, agencies that run Federal programs collect data on individuals, households and businesses for program administration, compliance, and regulatory purposes. Some of these agencies also make these data available for SERA purposes.

Many agencies have adopted what is effectively a two-tiered approach to access; they widely disseminate public-use data files and aggregate statistics and release restricted microdata to other agencies and researchers under limited conditions, for example through an RDC. Some agencies have adopted a more nuanced approach, with multiple tiers of access. A few examples are described below, along with some information about the relative volume of users that access data through each tier.

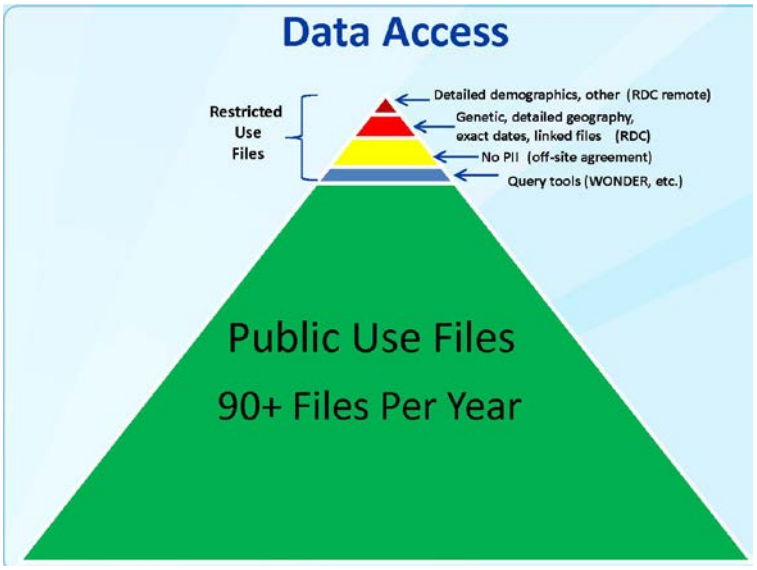
Bureau of Economic Analysis (BEA). BEA produces comprehensive statistics on the domestic and foreign activities of U.S. multinational enterprises including their employment, sales, and expenditures. BEA publishes these statistics—obtained from mandatory business surveys—in tabular form with detail by country and by industry. Prior to release, BEA conducts disclosure review on these tabular presentations, suppressing sensitive cells and creating versions for public release. Through its Special Sworn Employee (SSE) program, BEA also allows researchers to access confidential business microdata. One public-use tabulation on U.S. direct investment abroad has been accessed from the BEA website over 3,000 times since 2013. In contrast, only 34 researchers have used these data as part of BEA's SSE program during that time.

National Center for Education Statistics (NCES). NCES publishes aggregate tabulations of data and also makes restricted use data available through several means. The Commission heard a presentation at the December 12 meeting on the NCES Data Licensing System, which allows users to access restricted microdata. There are currently about 1,200 licenses in place (and each license allows for up to seven users).

NCES also has a set of data tools that allow the public to access those data with less privacy risk. The DataLab suite of tools—QuickStats, PowerStats, and TrendStats—provides the ability to query NCES restricted-use microdata without actually seeing it, therefore without the need to apply for and obtain a restricted-use data license.[1] The DataLab tools have over 20,000 current users and average about 4,000-6,000 reports run per month. Users must agree to the terms of an NCES Data Use Agreement to access the tools. PowerStats and TrendStats also require users to provide their email address. While NCES designed this feature to allow users to store analysis specifications and output, they could also use it to communicate with data users if necessary. Each tool allows users to pick the variables they are interested in and then generates tables and regression results based on user specifications. At no point does the user have access to any microdata or unweighted estimates. The DataLab uses underlying data that have been stripped of direct personal identifers; estimates are automatically suppressed if cell sizes are too small in order to minimize the risk of re-identification.

National Center for Health Statistics (NCHS). NCHS releases data from its surveys and systems in several ways to accommodate users with varying needs. They use a multi-tiered approach to data access, depicted in the graphic below. They disseminate public-use files through interactive tables, online tools, and direct download from their website. Its RDC allows access to restricted data. Restricted data are classified into one of four levels (depicted in the graphic as blue, yellow, red, and crimson), each with varying levels of data handling and access requirements. NCHS recently developed ANDRE, a system that allows researchers to remotely submit code to an automated system to analyze restricted data. Researchers must register for remote access with a secure university, government, or corporate email address and agree to terms of use. As of 2014, NCHS handled 5-10 data requests per week through ANDRE. NCHS has also begun hosting data from other Department of Health and Human Services agencies on this platform.

NCHS Data Access Model [2]



[1] nces.ed.gov/datalab

[2] Source: Meyer, P. NCHS Remote Data Access Systems: Present and Planned, presented at Federal Committee on Statistical Methodology Disclosure Seminar (December 3, 2015), available at https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2016/03/J3_Meyer_2015FCSM.pdf

Dissemination with a Data Facility

In its recommendations, CEP will need to consider the scope of data dissemination that a data facility would handle. The facility would likely focus, at least initially, on providing access to linked microdata. Other agencies would continue to disseminate their data using existing practices, while the Office of Management and Budget (OMB) led a standard-setting practice (See Recommendation Memo #4). However, as the data facility expands it could undertake new efforts to facilitate access to administrative records for evidence building. Eventually, the facility could play a greater role in managing access to individual-level administrative data (whether linked or unlinked).

Formalizing the Concepts of Tiered Access and a Privacy Budget

The creation of a government-wide data facility provides an opportunity to implement a consistent, coordinated approach to data access and dissemination. The data facility would participate in OMB's standard setting and would lead some or all of the implementation efforts. The standard would address specifics around the general concept of "tiered" access. The data facility could employ a model similar to Commissioner Sweeney's "datatags" model to formalize the concept of multiple tiers of access for a given subset of data (e.g., files it links).[3] In that case, the Program Management Office (PMO) of the data facility would be responsible for making determinations about how data are tagged. In consultation with the SERA agencies, it would decide which data fit where on a continuum and, based on its location, would dictate how those data are stored and transmitted and who is eligible to access them. The example below modifies the Sweeney et al. (2015) model to include only 4 color-coded levels ("tags"): blue, green, orange, and red.

Data Tags Model

Tag Type	Description	Security Features	Access Requirements
Blue	Public	Clear storage Clear transmission	Open
Green	Controlled public	Clear storage Clear transmission	Email or other verified registration
Orange	Accountable	Encrypted storage Encrypted transmit	Password, Registered, Approval, Signed Data Use Agreement
Red	Maximally restricted	MultiEncrypt store Encrypted transmit	Two-facter authentification, Approval, Signed Data Use Agreement

As described in the examples above, agency data releases range from highly confidential identifiable data that should be tagged in the red level of security, maximally restricted, to those at the orange, green, or blue levels. In the NCHS example, data query tools allow access at the green level, while public use datasets are at the blue level. In the BEA example, public data files are at the blue level. Finally, in the NCES example, the Data Licensing program grants access at the orange level, while the DataLab facilitates access at the green level.

[3] Sweeney, L., Crosas, M. Bar-Sinai, M. Sharing Sensitive Data with Confidence: The Datatags System. *Technology Science,* 2015101601. (October 16, 2015), available at http://techscience.org/a/2015101601/

Similarly, the data facility would tag a linked microdata file at the red level, but could also create orange, green, or blue versions of the linked dataset to facilitate appropriate use of the data.

If CEP were to recommend that the facility employ the concept of a privacy budget, either for data it releases or more generally, the PMO of the data facility would be responsible for monitoring releases from all sources including the facility and originating agencies. A tiered access approach supports the goal of quantifying the privacy loss from each release by categorizing the types of data according to their privacy risk.

RECOMMENDATION MEMO #2

Data Facility Implementation

Staff Lead: Kristy Howell

In order to facilitate access to and use of administrative and statistical data for evidence-building, CEP must consider three types of inter-related recommendations: those that address (1) law and regulation; (2) executive branch policies and standards; and (3) implementation, via a data facility, other cross-agency measures, agency-specific measures or a combination. This paper focuses on implementation, specifically with regard to the creation of a data facility. Over the last several meetings we have developed a framework for thinking about the dimensions of a clearinghouse, or data facility. This paper expands on how a data facility would be implemented over the short- and long-term within the existing decentralized environment of the U.S. statistical system and in coordination with states and program agencies. Specifically, based on past Commissioner discussions about the features of a data facility, this paper includes the staff's attempt to integrate perspectives about how a data facility would operate in the real world with a set of recommendations and alternative options.

* * * * *

INTRODUCTION

The Federal government engages in a wide variety of evidence-building functions, including data collection, processing, analysis, and dissemination of data for statistical, program evaluation, research, and policy analysis (SERA) activities. Moreover, there are several agencies across the Federal government that are currently undertaking efforts to link administrative records for statistical purposes.[1] To date, these efforts are highly decentralized, and no single agency undertaking these efforts has either the capacity or the authority to undertake all of these functions. A centralized data facility could leverage existing resources to provide improved data access, linkage, and other services.

One of the primary offices currently engaged in these efforts is the Center for Administrative Records Research and Applications (CARRA) within the Research and Methodology Directorate (ADRM) of the Census Bureau. CARRA has established processes for acquiring administrative data from other agencies and a sophisticated data linkage infrastructure. CARRA also houses a research and development (R&D) function that conducts research on record linkage methodology and linkage quality. Taking this into account, a cost-effective approach to creating a government-wide data facility would be to adapt and scale the existing infrastructure of CARRA. Infrastructure could also be drawn from other components of ADRM, including parts of the Center for Economic Studies (CES) that are engaged in similar functions for business data and that are responsible for the management of the Federal Statistical Research Data Centers (FSRDCs). For simplicity, we call this collective infrastructure "CARRA Plus" throughout the paper. The existing link between CARRA and the FSRDCs makes the FSRDCs a logical starting point for a system for providing researcher access.

[1] See OMB background paper, *Overview of Federal Evidence-Building Efforts.*

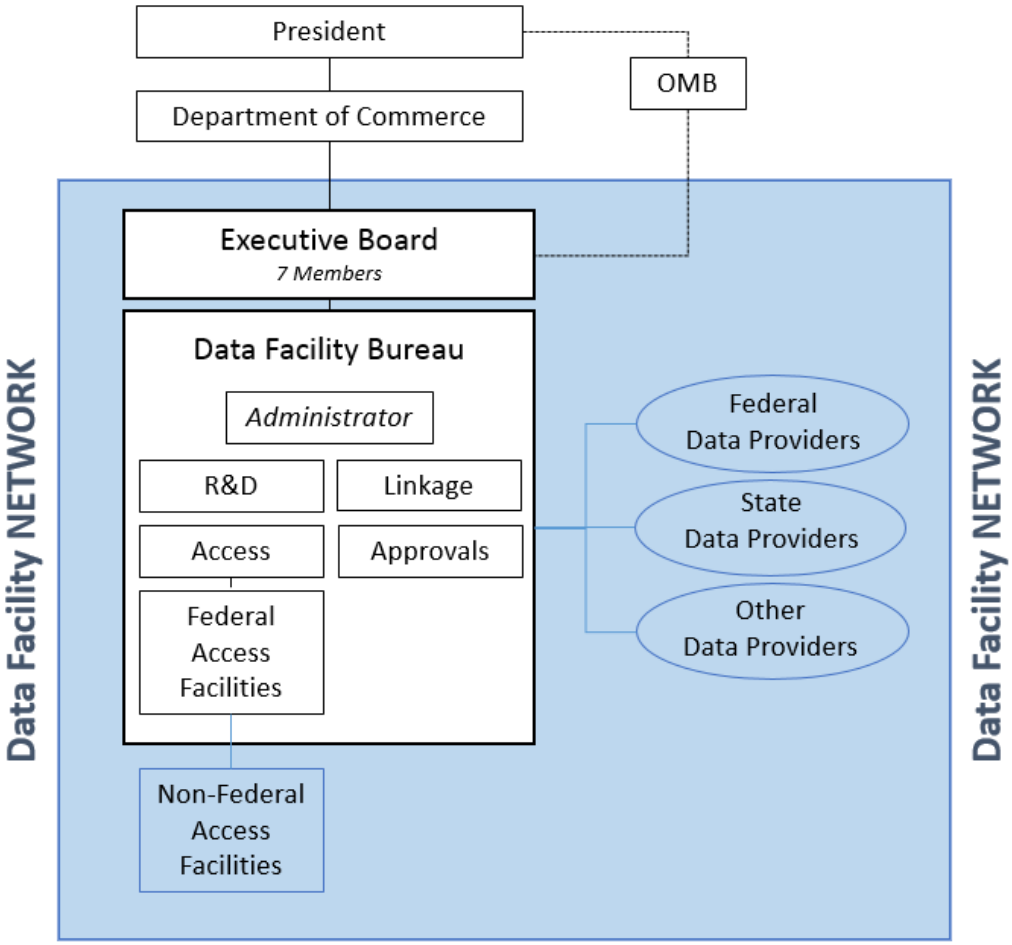
With CARRA Plus as a starting point, the following sections describe the structure, scope, administration, and governance of the data facility. A final section discusses how CARRA Plus could be scaled up to become the government-wide data facility.

DATA FACILITY STRUCTURE AND SCOPE

Structure

The Data Facility Network would be a repository-network hybrid architecture (see Figure 1). This means that some data are kept in a centralized location while other data remain siloed in their original location and are connected on-demand for linking.

Figure 1. The Data Facility Network



Note. The roles and responsibilities of the components of the Data Facility Network are discussed below under the "Governance and Administration" section.

Within the new Data Facility Network, CARRA Plus would transform into the **Data Facility Bureau**, which is the operational unit that provides linkage and access services for approved researchers. The Bureau also would provide analytical services for states and agencies on a fee-for-service basis or as a quid pro quo for access to a state's data (see Table 1). In addition, the Bureau would provide data management services for other government agencies, such as storing data and

making data available to researchers (see Table 2). The Bureau would be headed by an **Administrator**, a senior executive. The Bureau would operate the **Centralized Functions** related to data management (metadata and documentation, data storage, and linkage functions), administration (such as those related to application, qualification, and approval processes), and research.

Federal Data Providers include program offices and SERA agencies that would provide data via the Data Facility Network. **State Data Providers** would provide state-collected administrative data from Federally-funded programs.[2] The FSRDC network would become **Data Access Facilities** for the Data Facility Network. These could be expanded to accommodate both physical and virtual data access.

Table 1. EXAMPLE OF ANALYTICAL SERVICES PROVIDED TO STATES

CARRA has been working to acquire state-level administrative data from the Special Supplemental Nutrition for Women, Infants and Children (WIC) program for evidence-building purposes. In exchange for providing their data, CARRA has provided states with analyses and other statistical products that they can use to improve their program. At the March 13 Commission meeting, Erin Ulric from the Colorado Department of Public Health and Environment described how her state has been able to use the information that Census has provided to track program performance and identify populations that they are not reaching for targeted outreach.

Table 2. EXAMPLE OF DATA MANAGEMENT SERVICES PROVIDED TO A SERA AGENCY

The Department of Housing and Urban Development (HUD) Office of Policy Development and Research (PD&R) partnered with CARRA to store data from its Moving to Opportunity for Fair Housing (MTO) demonstration and the Family Options Study, to make those data available to qualified researchers. The Census Bureau Linkage Infrastructure enables secure access for policy analysis, program evaluation, and research, meeting HUD's need for a central and secure repository from which additional research and policy questions can be addressed. Once available at CARRA, both of these high value datasets will be more readily able to be matched with other government records under strong privacy protections.

Scope

The Data Facility Network will provide data access, linkage, and analytical services. A primary function of the Data Facility Network would be to link administrative and statistical datasets. Therefore, the facility will likely focus, at least initially, on providing access to linked microdata. To enable linkages between administrative and survey data, the Data Facility Network would facilitate access to core datasets from the Federal government and Federal program data collected at the state level. At least initially, program and statistical agencies that hold these datasets would continue to

[2] This paper does not address the approach through which access to data collected by states for Federally-funded programs would be achieved. This is discussed in Recommendation Memo #5.

facilitate researcher access to their data. However, as the Data Facility Network expands, it could undertake new efforts to facilitate access to administrative records for evidence building. Eventually, the Data Facility could provide the majority of access to individual-level administrative and survey data (linked or unlinked).

ADMINISTRATION AND GOVERNANCE

Institutional Arrangement

The data facility (including CARRA Plus) would be elevated to "bureau" status within the Department of Commerce. This Data Facility Bureau would have its own appropriation and budget account within the Department of Commerce's budget. Note that the Data Facility Network more broadly would be comprised of the Bureau, plus other data providers (e.g., CMS) who could interact with the Bureau as appropriate.

OPTIONS FOR INSTITUTION

***Proposed:* "CARRA Plus" is relocated within the Data Facility Network as a new Bureau in the Department of Commerce**

- *Pros:* The location within Commerce would allow the facility to take advantage of existing infrastructure for basic administrative overhead and processing, while enabling it to develop its own culture and bureaucratic processes and essentially elevating the profile of the entity. It also would give the facility a natural champion, the Secretary of Commerce, who should be motivated to invest political capital and facilitate the growth of the new entity. A specific appropriations line for the Bureau affords greater accountability to appropriators in Congress and allows OMB a greater degree of oversight in managing how funds are expended.
- *Cons:* The data facility would be further removed from the benefits of expertise and other infrastructure at Census. Also, nesting the facility within the existing infrastructure of the Department of Commerce would introduce challenges of maintaining autonomy and independence from the parent entity. Locating the facility within Commerce instead of as an independent agency could create some tension with other Cabinet-level departments who would be expected to support the Data Facility Network. Providing the Bureau with its own appropriation line limits the flexibility to absorb cost reductions, such as across-the-board funding cuts sometimes proposed by appropriators, and may also make the Bureau an easier target for political retribution in the unlikely event of a breach or incident.

***Alternative:* "CARRA Plus" remains in Census**

The functions described in the proposed option remain in the Census Bureau but are organized so that they operate as a separate entity. Increased autonomy and independence would be achieved by introducing a separate appropriation for the Data Facility and a separate process for identifying strategic priorities.

- *Pros:* The Data Facility would benefit from the existing infrastructure of the Census Bureau, including administrative systems, information technology infrastructure, and professional staff and expertise. The Data Facility would also benefit from the Bureau's existing public trust. To the extent existing Census infrastructure evolves into the facility, this approach

represents a potentially streamlined avenue for implementation with little disruption to existing processes.

- *Cons:* Relying on an existing Census entity would have implications tied to the nuanced relationships between Census and other agencies, including historic tensions within the Federal Statistical system. The Data Facility would also need to assimilate to the culture, tempo, and bureaucratic tendencies of its parent, and may face challenges of maintaining sufficient operational autonomy. There is also inherent risk in having to compete with other high priority functions within the Census Bureau.

Governance Structure

The Data Facility Network would be governed by an **Executive Board** authorized to dictate the policy direction of the Bureau.

OPTIONS FOR GOVERNANCE STRUCTURE

***Proposed:* Executive Board**
The Board would be comprised of **7 voluntary, part-time members** representing a range of relevant interests and expertise. The **Chair** of the Board would be appointed by the President, with the advice and consent of the Senate. The other members of the Board would include: 2 private citizens (one representing an academic institution, one a privacy expert), 2 state government officials, the Chief Statistician of the United States (from the Office of Management and Budget), and a Federal government official from a program agency. Ex officio members could include the Chief Information Officer, the Privacy Lead for the U.S. government, and a Congressional committee designee.

Private citizen representatives would be selected through a competitive application process modeled after the one used to select career Senior Executive Service appointees. The Bureau would develop a set of minimum criteria for each position. The final selections would be made by the Secretary of Commerce. The appointed private citizens would be appointed without regard to political affiliation.

The state government officials would be appointed by the President from two panels: one panel with nominations submitted by the National Governors Association and one with nominations submitted by the Council of State Governments.[1] The members appointed from this panel must be from different political parties and from different states.

Members would have staggered terms of 5 years. The Board would meet no less than bimonthly or at the discretion of the Chair. Members of the Board who are officers of the executive branch of the Federal government or officers of a state government would serve without compensation, beyond what they receive in their regular public employment (with an exception for allowable travel expenses). Other members would be compensated a nominal rate for each day they are engaged in the performance of their duties as members of the Board, in addition to reimbursement for the cost of travel.

- *Pros:* The executive board approach ensures that varied perspectives from Federal government, state government, research institutions, privacy experts, and other stakeholders are represented formally within the decision-making process, not merely in

an advisory role. The executive board would serve to strengthen the data facility's independence within the Department of Commerce.

- *Cons:* The number of decision makers could limit the ability to reach agreement and maintain flexibility.

***Alternative:* Single Executive**
The Data Facility Network would be governed by the Chief Executive, who operates within the given hierarchical boundaries. Under the proposed option for Institution, this Chief Executive would be a new senior level administrative position within the newly established entity. The Chief Executive would oversee the Data Facility Network, make major policy decisions, and conduct day-to-day operations. This individual would receive feedback from a variety of representative entities through advisory groups, but would have primary decision-making authority.

- *Pros:* Decision-making is streamlined with a single executive.
- *Cons:* Such an approach would likely create some tension in establishing meaningful partnerships outside the Federal government (thereby limiting overall coordination and collaboration) and may not engender the highest degree of public trust.

***Alternative:* Chief Executive with Advisory Board**
The Data Facility Network would be governed by a Chief Executive with an Advisory Board. The Chief Executive would oversee the Data Facility Network and conduct day-to-day operations. The Chief Executive would make major policy decisions in consultation with an Advisory Board.

- *Pros:* Decision making is still streamlined relative to an Executive Board approach.
- *Cons:* This approach results in the inefficiencies of a board approach without fully ensuring that varied perspectives are considered in decision-making, since the Chief Executive retains decision-making authority, though these could be partially mitigated by the criteria used to recruit, evaluate and remove such an official, and by requiring transparency in decision-making.

Oversight
The Board would be required to submit annual reports directly to Congress on the activities of the Data Facility Network, including any data breaches or "near misses" with regard to privacy protection, as well as detailing the investment and cooperation of stakeholder entities.

Funding/Budget/Staffing
If the Data Facility were housed in Commerce, it would likely be funded largely by a shared-services model with support coming from other SERA agencies. Presumably, the current staffing and budget levels of CARRA Plus are largely sufficient for the core initial work currently in progress.[3] Note that these current levels include a combination of appropriated funds and reimbursable expenditures from participating agencies. Additional demand for data access, linkage, and other services would likely need to be funded by an additional direct appropriation for the Bureau, transfers from other government programs, or by user fees. CEP will address costs more completely in a future recommendation memorandum on funding issues.

[3] Forthcoming is an estimate Census is preparing of the budget and staffing for CARRA Plus as it exists today.

Roles and Responsibilities

Executive Board. The Board would be responsible for oversight of the Administrator and staff, and all facility operations, including oversight of day-to-day decisions made by the Administrator regarding access requests. The Board would hire the Administrator in consultation with the Secretary of Commerce. The Board would approve the Administrator's budget request and allocation of available funds, which would be incorporated in the Department's budget request, with parallel authority to notify Congress of the original request level (e.g., such as the authority Inspectors General use). The Board would conduct strategic planning for the facility, including the appropriate nexus with evidence-building activities. The Board would be responsible for implementing regulations or other procedures necessary to acquire data from Federal and state agencies.[4] The Board would review appeals of decisions made by the Administrator, for access or enforcement. The Board could also establish advisory committees such as a privacy advisory board, an ethics review board, or an institutional review board.

Administrator and Bureau Operations. The Administrator would assume full responsibility as the executive of the facility for day-to-day operations and decisions, with direction from the Board. In addition to overseeing and directing the Bureau staff, duties would include functions related to the acquisition of data, such as establishing data sharing agreements with Data Providers. It would also include functions related to physical security and data security such as prescribing data access control methods (including data use agreements, training for researchers, and penalties for inappropriate use or failure to follow procedures).

A core function of the Bureau would be to link datasets from more than one Data Provider "on demand" for a particular project. Initially, state and Federal agency Data Providers would send datasets to the Bureau as encrypted files, similar to CARRA's current operations; as technology advances this could evolve to more advanced applications of privacy-protecting protocols (e.g., secure multi-party computation). The Bureau would maintain the resulting linked datasets for a period of time determined through the governance process and applicable laws. The Bureau would maintain and keep indefinitely the metadata required to link datasets. The Bureau would work with the Data Providers to update metadata as datasets change.

The Administrator would approve individual projects and researchers in consultation with the Data Providers. Once a project was approved, the Bureau would determine the appropriate access for the project, consistent with any data product release standard set by OMB. The data facility could employ a model similar to Commissioner Sweeney's "datatags" model to formalize the concept of multiple tiers of access.[5] In that case, the Bureau would be responsible for making determinations about how data are tagged. In consultation with the SERA agencies, it would decide which data fit where on a continuum and, based on its location, would dictate how those data are stored and transmitted and who is eligible to access them.[6]

---

[4] Recommendation Memo #3 includes a recommendation for legal changes that would provide a clear legal authority for the sharing of administrative data for statistical purposes, including with the proposed Data Facility. Regardless of the approach taken, there would need to be additional regulations to implement new data sharing requirements. Similarly, Recommendation Memo #4 discusses options for facilitating access to state data.

[5] Sweeney, L., Crosas, M. Bar-Sinai, M. Sharing Sensitive Data with Confidence: The Datatags System. *Technology Science,* 2015101601. (October 16, 2015), available at http://techscience.org/a/2015101601/

[6] CEP will address recommendations for privacy protections more completely in a future recommendation memorandum.

The Bureau would oversee the Data Access Facilities. The Bureau would be the point of contact for researchers and would be responsible for undertaking statistical disclosure analysis of researchers' output prior to release. The Bureau would also play a role in the appropriate public release of a limited set of "production" statistical results, created as part of its ongoing analytical capability.

Data Providers. Data Providers would approve individual projects and researchers, in conjunction with the Administrator. Data Providers could choose to delegate their approval authority to the Administrator. This approach allows Data Providers to remain involved in decision-making, while allowing for efficiencies in the approvals process. The Data Providers would also play a key role in curating data and supporting the technical aspects of data access, such as ensuring data include appropriate metadata. Data Providers would also have an active role to play in any standard-setting activities.[7]

HOW TO SCALE

The existing CARRA Plus infrastructure supports internal research and operational use of administrative data as well as externally motivated research projects, largely through the FSRDC network. To support evidence-building on a grander scale, CARRA has told the Commission "the infrastructure would need to be scaled to support increasing numbers of queries, proposals, users, agreements, datasets, processing and documentation, projects, reviews, and outputs."[8]

One can envision a phased approach in which researchers would work with the Bureau to conduct pilot projects that would be used to demonstrate success, develop metadata and linking procedures, and establish relationships. A more formal metadata standardization for priority datasets could facilitate metadata development.[9] Standardized metadata, coupled with advancements in technology and linkage methodologies, will be key to the Data Facility Network's success. As technologies advance, CARRA could move from its current linkage methodology, which requires the use of a spine to link administrative data sets together, to a scenario in which datasets are linked via secure multi-party computation or another technological protocol that does not require datasets to be brought together physically. Together these innovations will help ensure that the "network" aspect of the proposed facility are fully achieved. Success will require a program of continuous research about privacy protections, data security, and statistical approaches through partnerships with agencies, universities, and researchers.[10]

---

[7] See Recommendation Memo #4

[8] See "Capacity and Scalability of the U.S. Census Bureau's Data Linkage Infrastructure for Program Evaluation and Evidence-Building" (12/4/16) available on Max.gov.

[9] See Recommendation Memo #4

[10] CEP will address recommendations related to potential research partnerships in a future recommendation memorandum.

FOLLOW-UP TO RECOMMENDATION MEMO #2

Establishing a Data Facility

Commissioner Lead: Hoynes
Staff Lead: Howell

During the April 3 meeting, the facility discussion focused on the extent to which access to data should be centralized through a single entity or network, and on the associated governance structure and institutional location. Commissioners highlighted the needs of researchers and evaluators inside and outside government, and discussed the extent to which the facility should serve both.

In the discussion, Commissioners expressed contrasting views on the governance of the facility, with several Commissioners concerned about vesting responsibility with a board to oversee a bureau within a department, when the board's relationship to the departmental hierarchy is unclear or in conflict. Commissioners coalesced around the idea of an executive board that is more than advisory and a Chief Executive that had sufficient authority to run the facility, expressing interest in exploring existing Federal models. An earlier paper on the facility described how existing infrastructure in the Census Bureau (CARRA Plus) could be expanded to become the core of a government-wide network.[1] While Commissioners generally agreed with this approach, some expressed concern that moving the data facility to a separate bureau within the Commerce Department might create inefficiencies because of a potential need for Census to retain some linkage capabilities in-house to support the growing use of administrative records in regular statistical production. However, there seemed to be support for initially expanding CARRA Plus within Census and then devolving it into a separate bureau in the near future.

Building on the April 3 discussion and further consultation with Commissioners Groves, Hoynes, Potok, Glied, and Troske, this follow up to Memo #2 introduces recommendations to develop and design a data facility, as specified in Section 4(b)(1), 4(b)(2)(D), and 4(b)(2)(E) of the Commission's charge. The recommendations based on past discussions with Commissioners provide a vision for the facility network design (i.e., architecture, institution, and governance), needed legislative authority, and implementation approach. In the meantime, the Commission has already been working on a view of the broader Federal evidence ecosystem, and is therefore now better able to describe the role of this new "facility" entity in the context of the larger ecosystem.[2]

* * * * *

DRAFT RECOMMENDATIONS

Establishment of a Data Facility with Appropriate Authority

❖ **Finding:** All Federal Departments collect administrative data about programs and policies, and numerous program agencies have made significant investments in acquiring, curating, and making administrative data accessible specifically for evidence building. Principal Statistical Agencies (PSAs) acquire, curate, and use data from surveys and increasingly access administrative data from program agencies to develop statistics and datasets useful for evaluation, research, and analysis. If more effectively coordinated, these resources offer tremendous opportunities for developing greater knowledge about government programs and policies.

[1] See Memo #2 from the April 3 CEP meeting.
[2] See especially the memos on Evidence Ecosystem (Memo #7), Statistical System (Memo #9), and Improving Confidentiality (#10).

❖ **Finding:** Producers of evidence, both internal and external to government, often need to bring together multiple data sources, but face many barriers to access and linkage of datasets. The Commission finds that addressing these barriers is necessary to facilitate increased evidence building and PSAs will be an important part of the government's infrastructure and solution.

► **Recommendation 2.1 (PRIORITY):** The Commission recommends that the Congress and the President enact legislation establishing a [Data Facility], designated as a new PSA, to facilitate access and linkage of administrative and survey data, especially for researchers external to government.

The data facility fills a gap in the Federal Government's current capacity for evidence building. While recommendations in the Evidence Ecosystem (Memo #7) and Statistical System (Memo #9) memos would allow Federal agencies to coordinate better data access with one another, a newly established data facility will serve as the primary coordination point for external researcher access.

To fulfill this role, the data facility will need to operate within the "Federal Statistical Data Network" (described in Memo #9), comprised of the PSAs and other recognized units. PSAs will continue to perform their core statistical functions under their own authorizing statutes, but expansion of legal authorities for data sharing under the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) (see Memo #9) will allow them to operate within a common legal and policy framework for data sharing (the "CIPSEA Island"). By designating the data facility as the Federal Government's 14th PSA, the facility will also be able to protect and share data under CIPSEA and, thus, will be able to facilitate access to CIPSEA-protected data. Similarly, the introduction of a "Yes, unless" prevailing statute and other conforming amendments (discussed in Memo #3 on Legal Frameworks) will clear some of the barriers that agencies face currently in acquiring and linking administrative datasets from other Federal agencies.

Structure

❖ **Finding:** A facility can be most effective at addressing gaps in the current arrangement of data while avoiding duplication by initially focusing on (a) development of best practices for data protection and record linkage, (b) coordination across the [Federal Statistical Data Network] network, and (c) addressing limitations to access and linkage of priority datasets.

❖ **Finding:** As a result of the purpose of administrative datasets in operating programs, their size, periodicity for updates, or other features of their design, "moving" or "copying" these databases may frequently be infeasible, or doing so could introduce increased risks to data security or would raise unnecessary privacy concerns.

❖ **Finding:** An effective way to deal with existing gaps in the current arrangement of data is to have a network, with the [Data Facility] as one member along with the other PSAs as nodes in the network. The Census Bureau's Center for Administrative Records Research and Applications (CARRA), plus other components of the Research and Methodology Directorate (collectively, "CARRA Plus"), have the initial infrastructure in place to become the new member of this network. The 13 PSAs are ideal nodes because they already house much of the data that can be used for evidence building and have appropriate infrastructure and technical expertise to manage and protect data.

► **Recommendation 2.1.1:** The [Data Facility] should facilitate data access and linkage by operating as a coordinator of the nodes in the [Federal Statistical Data Network] and other data providers. The [Data Facility] should minimize the amount of data ingested and centralized within the facility, with most data remaining in their original location and being linked on-demand for an approved project.

To enable linkages between administrative and survey data "on-demand" for a particular project for an external researcher, the facility will need to coordinate with the other PSAs. Since all members of the "Federal Statistical Data Network" will share data under a common legal and policy framework (see Memo #9), which member of the network performs a particular linkage would depend on factors such as whether both datasets are already held by that member, available subject matter expertise, and linkage capacity. Even those other Federal data providers who are not part of the CIPSEA Island will be following new standards from OMB (see Memo #4 on Policies and Standards) and therefore would still be part of the larger ecosystem.

As one of the Commission's key recommendations is to provide a consistent application and approval process for external researchers (Recommendation #4.1.3 in Memo #4), the preferred route would be for the external researcher to contact the data facility to apply for access to the needed data. Regardless of whether the researcher wants to access data that are already curated and linked by the data facility or access a dataset in a single PSA, the facility provides a one-stop shop, or a single point of access to a standard application and approval process for external researchers. If an external researcher's needs involve the use of CIPSEA-protected data, that activity will only be allowed on the "CIPSEA Island," facilitated either by a PSA or the facility. In the case where a researcher wants to access administrative data held by a program agency, the facility would coordinate with the PSA for that department to facilitate access. In the case of a department that does not house a PSA, the facility would coordinate directly with the program agency or with another PSA that the department has designated to serve as its facilitator. Figure 1 below illustrates this from the perspective of a researcher external to government.[3]

Initially, Federal agency data providers most likely would need to send datasets to the appropriate network node as encrypted files; as technology advances this arrangement could evolve to more advanced applications of privacy-protecting protocols (e.g., secure multi-party computation).[4] The facility (or other node) would maintain the resulting linked datasets for a period of time determined through the governance process and applicable laws. The facility would maintain and keep indefinitely the metadata required to link datasets, and work with the data providers to update metadata as appropriate over time.

Much of the detail of how the facility would provide access to researchers is dependent on the Commission's specific determinations in the memo on Improving Confidentiality (Memo #10). But presumably access to "red" or "crimson" level restricted microdata would be provided through data access centers, such as the Federal Statistical Research Data Centers (FSRDCs). These could be both virtual and physical enclaves.

Relationship of Facility to Other Government Entities

❖ **Finding:** Many high-value administrative data associated with Federally-funded programs are collected and maintained at the state level. Because these data are not always acquired by PSAs, the facility will need to engage with other entities to access these data on behalf of external researchers.

► **Recommendation 2.1.2:** The [Data Facility] should develop collaborative partnerships with other data providers to facilitate access to data relevant to Federal programs and policies.

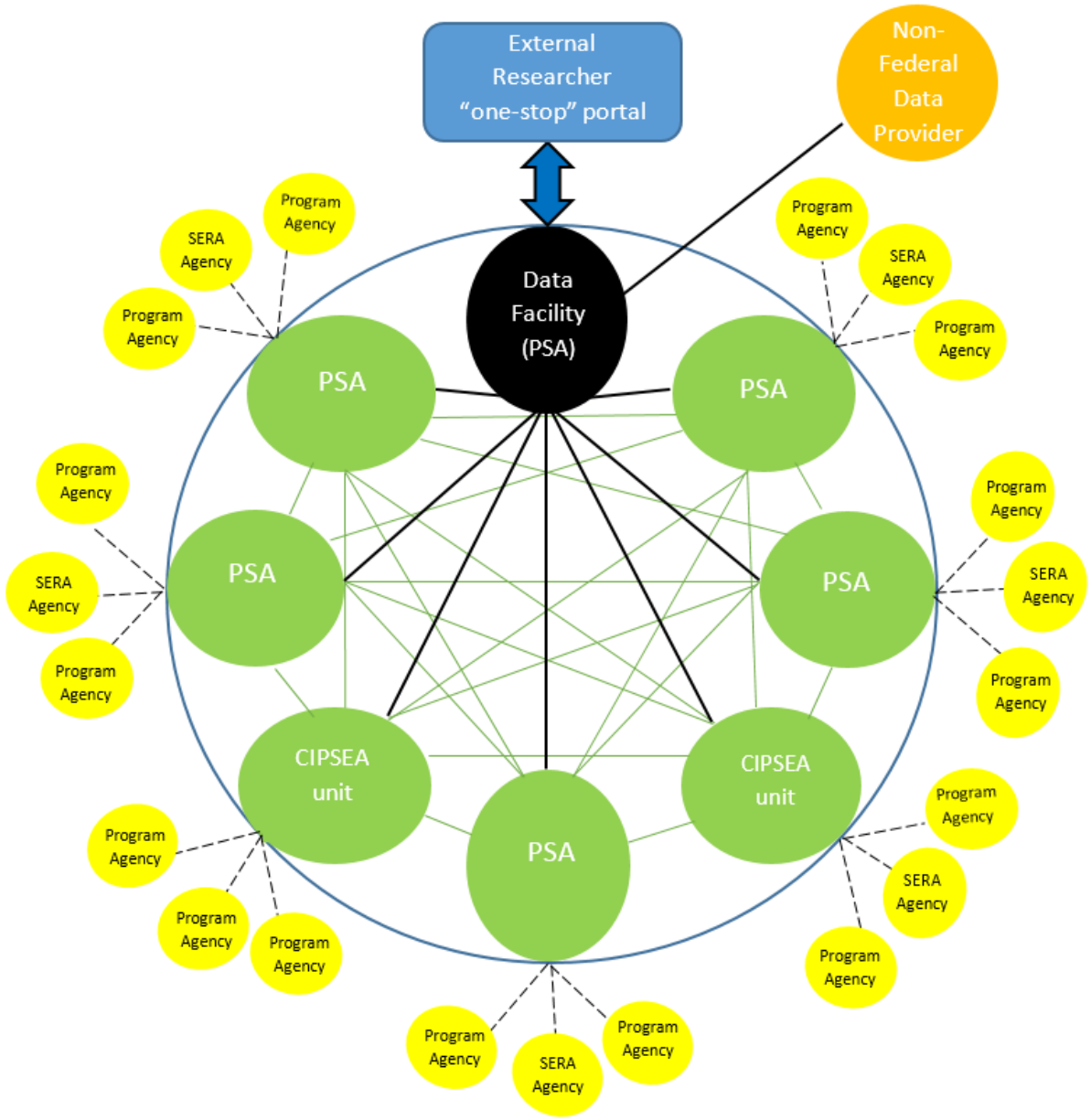
At the April meeting, the Commission agreed to recommend encouraging incentives to increase states' willingness and capacity to make their administrative data available to the Federal government

[3] See also appendix 2 for a table that describes the process for how an external researcher might access different datasets.
[4] See the materials from the February 24 CEP meeting.

and to identify a subset of high-value programs for which to recommend a requirement for states to make their data available (see the follow-up to the States Memo #5).

Because state and other data providers are operating outside of the "Federal Statistical Data Network," they would most likely need to send datasets to the data facility as encrypted files to allow for linkage until technology advances to allow for other more advanced arrangements.

Figure 1. The Data Facility as a One-Stop Shop for External Researcher Access to CIPSEA-Protected Data



Institutional Arrangement of a Facility

❖ **Finding:** Rather than build an entirely new and potentially duplicative infrastructure, a cost-effective approach to implementing the proposed [Data Facility] is to adapt and scale the existing elements of CARRA Plus. However, to achieve an "optimal arrangement" for expanding evidence building, the Commission finds that CARRA Plus needs greater independence (through institutional placement) and an improved governance structure.

❖ **Finding:** The institutional placement and governance of the facility should be evaluated in terms of the implications for the objectives of (a) support for evidence building, (b) transparency and trust, (c) strategic coordination and cooperation, (d) confidentiality and security, (e) authority and flexibility, (f) scalable functionality, and (g) sustainability.

❖ **Finding:** An institutional placement within an existing Federal Department has a practical benefit, building on available administrative and information technology (IT) systems, existing professional staff, established levels of public trust, and operational knowledge. Relocating these functions to a separate entity would provide for additional autonomy and would enable the facility to develop its own culture and bureaucratic processes.

► **Recommendation 2.1.3:** The [Data Facility] should be located within the Federal government as an entity with a level of independence sufficient to set its own strategic priorities. The Commission recommends appropriate levels of technical independence be achieved by establishing the facility under one of two institutional arrangements: 1) as a new separate "bureau" in the Department of Commerce or 2) as an independent entity attached to the Census Bureau.

During the Commission's fact-finding phase, Commissioners heard a great deal about the efforts underway at the Census Bureau to expand capacity for evidence building. CARRA has established processes for acquiring administrative data from other agencies and a sophisticated data linkage infrastructure. CARRA also houses a research and development (R&D) function that conducts research on record linkage methodology and linkage quality. The Commission generally agreed that CARRA could be expanded, with greater independence and a stronger governance structure.

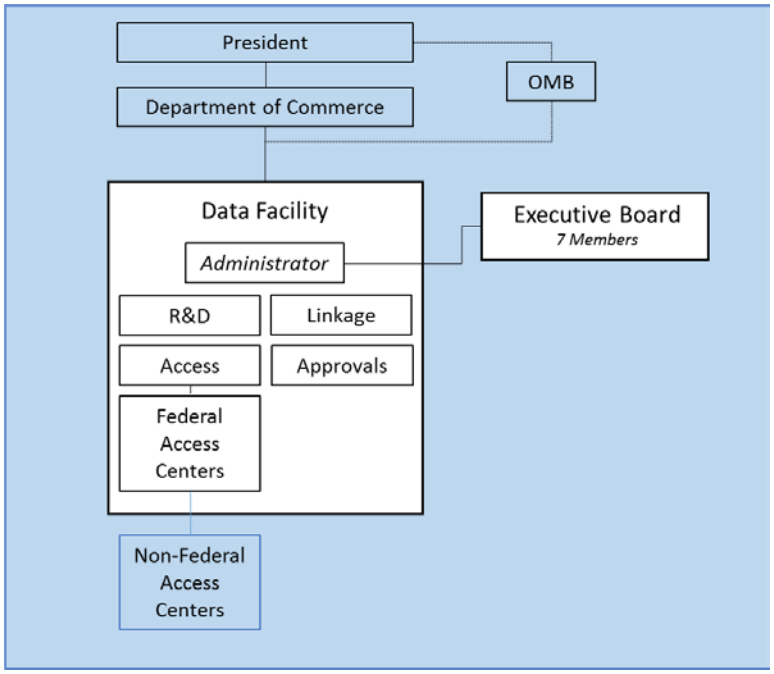
The Commission considered many possible locations that could achieve the independence envisioned, weighing the pros and cons of each. Ultimately, the discussion has narrowed to two options, both within the Department of Commerce. A location within Commerce is a logical choice given that the infrastructure to develop the facility will draw from within the Census Bureau. Census has the strongest legal and technical infrastructure already available. The facility could take advantage of this existing infrastructure for basic administrative overhead and processing as well as IT infrastructure. The Census Bureau has broader authority to acquire administrative records from other government agencies than any other statistical agency. This authority comes from Title 13, U.S. Code—which gives Census broad authority to collect data from public and private sources—and from an exception in the Privacy Act that allows other agencies to provide data to the Census Bureau without obtaining consent for such a disclosure. The Title 13 authority could transfer with the facility under either institutional arrangement. Lastly, there is precedent for an independent agency within Commerce. The National Institute for Standards and Technology (NIST) is an example of a Commerce agency that is a resource government-wide and beyond government, by design. However, the Commission also recognizes that locating the facility within Commerce could create some tension with other Cabinet-level departments, who would be expected to support the data facility. Addressing potential interagency tension means incentivizing these agencies to make their data available through the facility will be important, perhaps best achieved by the "Federal Statistical Data Network" concept, where PSAs may see as an incentive the transfer of some of their responsibilities for supporting researcher access.

There are other pros and cons specific to each of the two options (discussed below); however, the proposed recommendation acknowledges that either arrangement would be adequate to meet the Commission's vision.

Option 1: New Bureau in the Department of Commerce (See Figure 2)
The data facility would operate as a separate bureau in the Department of Commerce with its own appropriation and budget account, within the Commerce budget. Not pictured are the PSAs, which would serve as members of the same data network under either option.

- *Pros:* The location within Commerce would allow the facility to take advantage of existing infrastructure, while enabling it to develop its own culture and bureaucratic processes and essentially elevating the profile of the entity. This option also would give the facility a natural champion, the Secretary of Commerce, who should be motivated to invest political capital and facilitate the growth of the new entity. A specific appropriations line for the Bureau affords greater accountability to appropriators in Congress and allows OMB a greater degree of oversight in managing the utilization of funds.
- *Cons:* The data facility would be further removed from the benefits of expertise and other infrastructure at Census. In addition, nesting the facility within the existing infrastructure of Commerce would introduce challenges of maintaining autonomy and independence from the parent entity. Providing the Bureau with its own appropriation line limits the flexibility to absorb cost reductions, such as across-the-board funding cuts sometimes proposed by appropriators, and may also make the facility an easier target for political retribution in the unlikely event of a breach or incident.

Figure 2. The Data Facility Bureau within Commerce

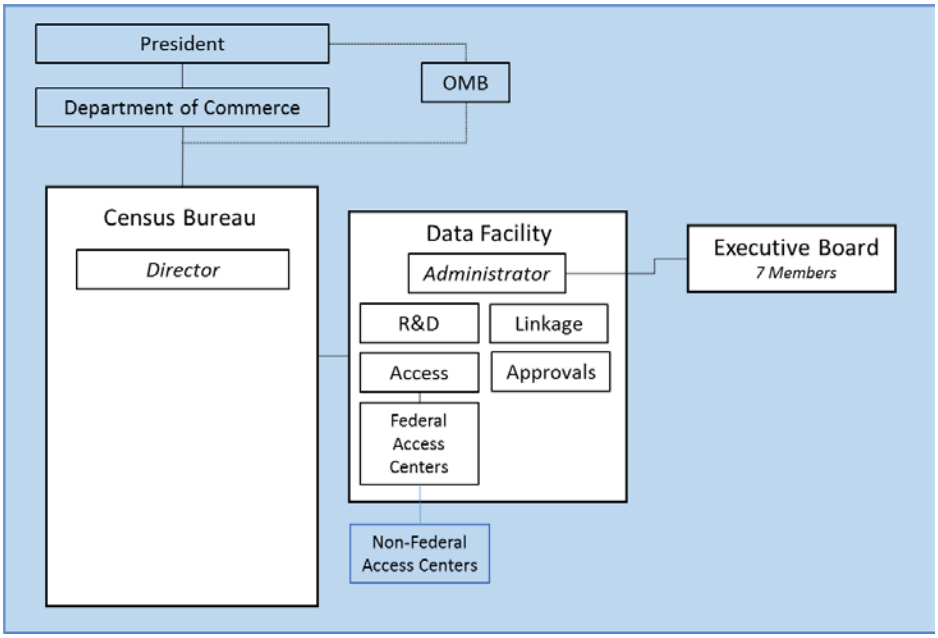


Option 2: An Independent Unit Attached to the Census Bureau (See Figure 3)
An alternative would be for the data facility to operate as an independent unit attached to the Census Bureau. Independence from the Census Bureau would be achieved by introducing a separate appropriation for the facility and a separate process for identifying strategic priorities.

Under this alternative, the role of the Census Bureau Director as it relates to the facility would need further discussion.

- *Pros:* The facility would benefit from the existing Census infrastructure, including administrative systems, IT infrastructure, and professional staff and expertise. The facility would also benefit from the Census Bureau's existing public trust. This approach represents a potentially streamlined avenue for implementation with little disruption to existing processes. Establishing the facility as an independent entity with its own priority-setting processes helps to address the inherent risk of having to compete with other high-priority functions within Census, an issue raised in previous discussions when considering the option of keeping the facility within the Census Bureau hierarchy.
- *Cons:* Keeping the facility associated with Census would have implications tied to the nuanced relationships between Census and other agencies, including historic tensions within the Federal Statistical System. There may still be pressure for the facility to assimilate to the culture, tempo, and bureaucratic tendencies of its parent, and it may face challenges of maintaining sufficient operational autonomy.

Figure 3. The Data Facility Attached to Census



Management and Oversight of the Facility

❖ **Finding:** Oversight by a representative board promotes accountability and transparency. A governing board can help ensure that varied perspectives from Federal and state agencies, research institutions, privacy and technology experts, and other stakeholders are represented formally in the decisions of the data facility. Day-to-day management is best achieved by an empowered executive administrator, who also is responsible for adhering to legal and government-wide policy.

► **Recommendation 2.1.4:** The [Data Facility] should be governed by an Executive Board empowered to set policy direction for the facility, and headed by an Administrator, a senior executive empowered to control day-to-day operations and decisions.

a. To ensure a diversity of perspectives, the Executive Board should include representatives of Federal agencies, state agencies, academia, and the public.

The duties of the Board should include the following:

a. Oversee the Administrator and staff, and all facility operations.
b. Hire the Administrator in consultation with the Secretary of Commerce.
c. Approve the Administrator's budget request to the President and allocation of appropriated funds.
d. Conduct strategic planning for the facility, including the appropriate nexus with evidence-building activities.
e. Review appeals of decisions made by the Administrator, for access or enforcement.

The duties of the Administrator should minimally include the following:

a. Administer, oversee, and coordinate the activities of the [data facility] and [data facility] staff.
b. Implement regulations or other procedures necessary to acquire data from Federal and state agencies.
c. Approve individual projects and researchers in consultation with data providers.
d. Prescribe data access control methods, such as data use agreements, training for researchers, and penalties for inappropriate use or failure to follow procedures (consistent with OMB policies and standards).
e. Ensure compliance applicable Federal laws and policies, especially with regard to data confidentiality and security.
f. Provide leadership across the [Federal Statistical Data Network] on developing and employing state-of-the-art data protection and access methods and technologies.

New legislation establishing the data facility could also establish the executive board. The statute could specify the roles of responsibilities of the board and the administrator. During the April Commission meeting, staff presented a model for an executive board that included details such as the number of board members, appointment processes, and term lengths. The Commission could choose to include some of these additional details in its recommendations to help convey what the Commission had in mind when deliberating on the governance of a facility. However, Congress and other stakeholders may want to modify and add certain details as the legislation is developed.

Implementation

❖ **Finding:** The existing Census Bureau infrastructure (CARRA Plus) supports a modest level of internal research and operational use of administrative data, as well as externally motivated research projects. The infrastructure needs to be scaled to support increased evidence building.

❖ **Finding:** A phased approach to establishing and implementing a data facility would allow the entity to work with researchers to conduct pilot projects to demonstrate success, develop metadata and linkage procedures, and to develop new (or enhance existing) relationships with data providers.

► **Recommendation 2.1.5:** To maximize efficiency, the President should reallocate designated staff and resources from the Census Bureau (CARRA Plus) to form the core infrastructure for the [Data Facility]. Implementation should continue in phases, with CARRA Plus fully devolved from the Census Bureau in 2-3 years.

Potential phases for implementation could be outlined as follows:

Phase I – Legal Authority. New legal authority is enacted to establish the data facility in its new location and to establish an executive board, with a stated transition period of 2-3 years.

Phase II – Relocation and Staffing. The executive board is appointed and begins to convene. The board hires an administrator. The relevant assets, personnel, etc. from Census are transferred to the facility. The administrator hires additional staff as needed.

Phase III – Pilot Projects and Metadata Development. The facility works with researchers and other stakeholders to conduct a series of pilot projects. This phase would need to sync with the metadata standardization process and policies described in Memo #4.

Phase IV – New Development and R&D. The facility continues to develop new relationships with data providers to meet user needs (perhaps to include the incorporation of some private sector data as well). The facility continues to scale as necessary to support additional demand. The facility also expands efforts to develop advanced linkage methodologies and evaluate data linkage quality. It undertakes new partnerships with agencies, universities, and researchers focused on innovations in the areas of data confidentiality and security.[5]

Absent new legal authority, Census could continue to expand CARRA Plus within its existing location and with the existing authority of Title 13. Unless and until the recommended legislative changes are enacted, the Census Bureau should continue efforts to expand the technical and administrative capacity of CARRA Plus to operate as a data access and linkage services provider within its existing legal authority.

Funding

❖ **Finding:** The envisioned data facility would provide a clear public benefit through generated evidence about government policies and programs, with limited direct benefits to data users. Therefore, the Commission finds that user fees for some costs associated with the facility can support "self-funding," but that other funding mechanisms would be necessary to achieve the vision of the Commission, likely including a direct appropriation or reimbursable funding from other departments.

► **Recommendation 2.1.6:** In establishing the [Data Facility] in law, Congress and the President should provide the [facility] with the authority to collect and spend user fees and have sufficient flexibility to adjust based on changes in demand.

The Commission was charged with evaluating how a data facility "could be self-funded" (Sec. (4)(b)(2)(E)). The authority to collect and spend user fees would ensure that the facility is able to be at least partially self-funded. The facility should have flexibility to adjust its pricing model and to waive fees to encourage certain high-priority research or to ensure diversity in the types of eligible researchers that are able to conduct research through the facility.

Other Flexibilities

❖ **Finding:** By encouraging research and development in the areas of privacy-enhancing technologies, the facility and the other PSAs will spur private sector innovation. Development of new technologies related to providing data access, such as virtual data access technologies, will help to address increased demand.[6]

---

[5] An upcoming recommendation memo will address potential research partnerships (Memo #12).

[6] Through its public input period, the Commission heard some concerns that the FSRDCs would be overwhelmed by the additional demands of supporting the increase in access that would come from the facility. Witnesses encouraged the Commission to consider ways to other ways to provide researcher access, including virtual data access technologies. See, for example, Donna Ginther's response to Questions for the Record available on Max.gov.

► **Recommendation 2.1.7:** In establishing the [Data Facility] in law, Congress and the President should provide the facility with the authority to sponsor a Federally Funded Research and Development Center (FFRDC).

❖ **Finding:** The facility should be well positioned to exceed in its intended ethical and transparent project review process all requirements under the newly promulgated regulations on the Protection of Human Subjects. As the intent of the facility is to permit more seamless access to eligible researchers, its review should be deemed to have met the requirements of the Common Rule.

► **Recommendation 2.1.8:** In establishing the [Data Facility] in law, Congress and the President should provide the facility with the authority to satisfy requirements of the Federal Policy for the Protection of Human Subjects ("Common Rule") for projects under the facility's purview by its own review processes.

Recommendation #2.1.7 acknowledges that the board may want to sponsor an FFRDC to use private sector resources to meet its R&D needs. Any agency can sponsor an FFRDC under the same authority that allows them to contract for the purchase of goods and services. However, in some cases Congress has explicitly authorized agencies to establish FFRDCs.[7] Recommendation #2.1.8 clarifies that projects approved by the data facility can satisfy Institutional Review Board (IRB) approvals, though individual institutions may continue to require other approvals for liability purposes.

LIKELY REACTION TO RECOMMENDATIONS:

- ***Congressional:*** As the Commission's charge specifically directs recommendations for a clearinghouse, Congress expects recommendations in the vein of those in this paper. The recommendation to build on existing infrastructure should be appealing from a cost perspective.
- ***Executive:*** Most agencies will welcome the recommendations as a way to expand evidence building. As noted, there may be some concern from departments other than Commerce who might dislike the recommendation to locate the facility within Commerce. However, the proposed "network" design, which allows agencies to retain control over their data, may alleviate these concerns to an extent. Some agencies may raise concerns about the ability to meet the increased demand for their data that the facility could create.
- ***State and Local:*** State and local governments will stand to benefit from the creation of a data facility as data users, but like Federal agencies, will likely raise concerns about any new requirements to provide data to the facility.
- ***Other:*** Researchers should welcome the streamlined access that the facility would provide. As the facility is able to demonstrate success, researchers will gravitate toward it as a primary means for accessing government data.

[7] See GAO, "Opportunities Exist to Improve the Management and Oversight of Federally Funded Research and Development Centers." (November 7, 2008), pp. 4-5, available at http://www.gao.gov/products/GAO-09-15

COST IMPLICATIONS

By building on the existing infrastructure of CARRA Plus, the cost of establishing a data facility is lowered significantly. As the facility will be built over time, the initial setup cost does not need to be substantial. The current funding for CARRA Plus—$6.5 million per year and 40 full-time employees (FTE)—is largely sufficient for the core work currently in progress. There would be some additional cost to establish the board and to hire more staff, in particular to expand capacity for project review and for R&D. There would also be costs associated with providing technical assistance to data providers to make non-statistical data available for linkage.

Some of the fixed costs of the facility could be funded with support from the statistical, evaluation, policy research, and policy analysis agencies, perhaps using a tiered pricing structure tied to the volume of work. As use of the facility expands, some of the additional costs could be covered by user fees. There is a risk that the cost of increased demand would, of course, filter down to the data providers as well. The executive board would need to consider the volume of research that the facility can handle and establish procedures for prioritizing research. Having flexibility to collect user fees and allow researchers to bring in their own funding will allow for more consistent growth.

The data facility could continue to use Census IT infrastructure (regardless of whether it remains located at Census) so, while there is a cost to this infrastructure, it is not a new cost (for example, the cost of setting up its own servers, network, etc.). Likewise, the facility could take advantage of shared services through the Department of Commerce to the extent feasible, thus eliminating the potential cost to establish redundant administrative functions (HR, finance, etc.).

New and expanded data access centers would also require additional funding, although not immediately. Presumably, FSRDCs would continue to be funded by host institutions, so to the extent new data access centers are established as FSRDCs, these would be funded by new host institutions.[8] However, if the executive board decides to develop other data access centers or adopt new technologies, such as virtual data access, the facility may need to secure new funding.

* * * * *

[8] The FSRDC network has grown organically with demand. There are currently 25 operating FSRDCs, the most recent being the Georgetown University FSRDC, which opened in April 2017. In mid-2017, Kentucky and Colorado will become the 26th and 27th. With these additions, the FSRDC network will have doubled in size since 2013. As of 2016, the FSRDC network supported about 900 researchers working on about 250 different projects. The host institutions must establish a funding model that will guarantee that it will be self-funded. However, an organization proposing a new FSRDC can request, from the National Science Foundation, up to $100,000 per year for a three-year term to cover part of the start-up costs and annual operating costs.

APPENDIX 1

Example of an Independent Bureau with a Board within the Federal Government

An example of an independent entity with an executive board within a Cabinet-level department is the Institute of Education Sciences (IES). IES is an independent research institute, housed within the Department of Education and administered by a Director and a board of directors (the National Board for Education Sciences, or NBES). However, this example is not perfectly analogous to the arrangement envisioned if the data facility were to be located in a separate bureau within Commerce because the NBES is an advisory board, not a governing board with policy-setting responsibilities.

Still, the authorizing statute for the IES and NBES does provide a good example of how a board could be established in law. The duties of the IES Director and the board of directors, as prescribed in Title 20 of the U.S. Code, sections 9514 and 9516, are summarized below.

The Director:

- Proposes to the Board priorities for the Institute,
- Ensures methodologies are consistent with standards,
- Coordinates education research and related activities carried out by the Institute with activities carried out by other agencies,
- Advises the Secretary of Education on research, evaluation, and statistics activities,
- Establishes procedures for peer review, ensures that the privacy rights of research participants are protected, and
- Coordinates the dissemination of research.

The NBES:

- Advises and consults with the Director on the policies of the Institute,
- Considers and approves priorities proposed by the Director,
- Reviews and approves procedures for peer review,
- Presents the Director with recommendations for strengthening education research.

Example of an Independent Entity Attached to a Statistical Agency

During the December 2016 Commission meeting we heard about France's Secure Data Access Center (the Centre d'Accès Sécurisé aux Données, or CASD), which was founded in 2009 under the auspices of France's public institution of higher education and research, GENES (Groupe des Écoles Nationales d'Économie et Statistique) under the Ministry of Economy and Finance. GENES conducts its work with technical supervision from the National Institute of Statistics and Economic Studies (Institut National de la Statistique et des Études Économique, or INSEE), one of the agencies in the French statistical system.

APPENDIX 2

Table 1. Mechanisms for External Researchers to Access Restricted Microdata

If an external researcher wants to…	Researcher would…	Data Facility would…	Data Provider/s would…
Link data on student loan recipients from Department of Education with tax data from the Internal Revenue Service	• Contact Data Facility to discuss data needs and requirements for research proposal. • Submit application to become eligible researcher.* • Submit research proposal to the Data Facility. • Upon approval, conduct research at a Data Access Center. If not approved, researcher can appeal to the Board. • Submit completed research to Data Facility for disclosure analysis.	• Assist researcher with determining data needs. • Review and approve research proposal and researcher application in consultation with the Data Providers, based on policies set by the Board. • Work with Data Providers to determine the data needed for the project. • Discuss linkage methodology with researcher to determine if matching criteria meets his/her needs. • Link datasets and assess quality of the linkage results. Discuss results with researcher. • Create a project dataset with direct identifiers removed. • Work with Data Access Center to get researcher set up to work on their project. • Conduct disclosure analysis. • Maintain linking metadata so that data can be used again in the future.	• Work with Data Facility to determine the data needed for the project. • Consult with Data Facility on the research proposal and researcher application. • Provide metadata to facilitate use of the data and linkage. • Prepare and send file with direct identifiers to Data Facility (or facilitate connection in some other way using technology such as secure multi-party computation). • Maintain metadata so that data can be used again in the future.
Access CMS data on Medicaid spending	• Contact the Research Data Assistance Center (ResDAC), CMS's contractor for data request services. • Submit application to become eligible researcher.* • Submit research proposal to CMS. • Upon approval, conduct research at the CMS VRDC. • Submit completed research to CMS for disclosure analysis.	No responsibility in this situation, although the Data Facility would engage with CMS as part of its coordinating responsibilities.	• Determine the data needed for the project. • Provide metadata to facilitate use of the data. • Prepare file for researcher. • Get researcher set up to work on their project. • Conduct disclosure analysis.
Access a new dataset that hasn't been used previously for evidence building	• Contact Data Facility to discuss data needs and requirements for research proposal. • Submit application to become eligible researcher.* • Submit research proposal to the Data Facility. • Upon approval, conduct research at a Data Access Center. • Submit completed research to Data Facility for disclosure analysis.	• Assist researcher with determining data needs. • Review and approve research proposal and researcher application in consultation with the Data Providers, based on policies set by the Board. • Work with Data Providers to determine the data needed for the project and if data can be made available. If dataset has not been accessed previously through the Data Facility, work with Data Provider to ensure proper metadata exist and that determine necessary file formats. • Work with Data Access Center to get researcher set up to work on their project. • Conduct disclosure analysis.	• Work with Data Facility to determine the data needed for the project • Consult with Data Facility on the research proposal and researcher application • Provide/Develop metadata to facilitate use of the data • Prepare and send file to Data Facility (or facilitate connection in some other way using technology such as secure multi-party computation) • Maintain metadata so that data can be used again in the future.

** Assumes a standardized application for "eligible researchers." Memo #4 includes the recommendation to "set standards for a single, streamlined process through which researchers external to the government may apply, qualify, and are approved to access government data not publicly available."*

RECOMMENDATION MEMO #3

Data Access and Use: Legal and Statutory Issues

Staff Lead: Lucas Hitt

Objective: Barriers to data access and use among agencies, a prospective facility, and non-governmental researchers exist in law. This recommendation addresses those barriers by suggesting a clear legal authority for the sharing of administrative data for statistical purposes, and defining qualifications for access, statistical purpose, and confidentiality.

In order to facilitate access to and use of administrative and statistical data for evidence building, CEP must consider three types of inter-related recommendations: those that address (1) legal and statutory issues; (2) executive branch policies and standards; and (3) implementation, via a data facility, other cross-agency measures, agency-specific measures, or a combination. This paper focuses on **legal and statutory issues** to facilitate appropriate and efficient access to and use of data for statistical purposes, specifically to support statistical, evaluation, research, and policy analysis (SERA) activities.

* * * * *

BACKGROUND AND ANALYSIS

As outlined in its authorizing law, the Commission is to "determine the *optimal arrangement* for which administrative data on Federal programs and tax expenditures, survey data, and related statistical data series *may be integrated and made available* to facilitate program evaluation, continuous improvement, policy-relevant research, and cost-benefit analysis by qualified researchers and institutions while weighing how integration might lead to the intentional or unintentional access, breach, or release of personally-identifiable information or records."[1] This memorandum discusses possible remedies to the data access, "statistical use" definition, and qualified researcher issues.

Clarity and consistency of access and allowable use, together, provide for safety and maximized value of data to facilitate evidence building. As measured by these characteristics, the status quo does not measure up, as it is neither clear nor consistent. The Commission heard considerable frustrations regarding existing barriers to data access and use, including in testimony at meetings and hearings, through the Request for Comments (RFC), the Federal Office Survey (CEP Survey), and through the Commission's own research. Additionally, a number of Commissioners themselves have directly raised issues related to access and use.

International and historical models. Both international models and influential prior United States commissions are instructive in determining an "optimal arrangement" within the U.S. The models converge around three primary attributes of an "optimal arrangement:" (1) clear authority for access, (2) clearly defined and appropriately limited uses to ensure safety, and (3) consistency across data assets to ensure stringent implementation.

Internationally, the most common arrangement for an official statistical organization is a single national statistical office (NSO), or agency, with a broad statistical law. This is in line with the model statute proffered by the United Nations Statistical Commission. The UN's model, among other things, includes a clear mandate for a statistical agency to access administrative records, however, only within

[1] Sec. 4(a)(1) of the Evidence-Based Policymaking Commission Act of 2016, P.L. 114-140.

a consistent framework demanding three conditions for data. Such a framework requires that the data are accessible only by the NSO and qualified researchers, used solely for statistical purposes, and handled with strong confidentiality protections (administrative data protected by the same strong confidentiality protections as the NSO's survey data). The UN's model statute also defines these central concepts clearly. This arrangement inextricably binds data access to core privacy principles of use limitations and confidentiality.

Regardless of the challenges, there is substantial precedent in the U.S. supporting the contention that the use of survey and administrative data for statistical purposes has both value and the potential to improve program effectiveness, while also limiting the burden of data collections on people and businesses. The 1977 *Report to the President from the Privacy Protection Study Commission* (PPSC) and the *Report of the Commission on Federal Paperwork* both reached this conclusion, identifying the same three critical conditions of qualification, statistical purpose, and confidentiality. Critically, the PPSC articulated the concept of "functional separation" between the statistical and administrative uses of data, providing a means to operationalize clarity and consistency. ***Functional separation*** is defined as "separating the use of information about an individual for a research or statistical purpose from its use in arriving at an administrative or other decision about that individual."[2] The *Report of the Commission on Federal Paperwork* states that data "collected for administrative or regulatory purposes must be made available for statistical use, with appropriate confidentiality and security safeguards, when assurances are given that the information will be used solely for statistical purposes." An additional contribution of the PPSC is a definition of the term ***statistical purpose***, as "the development and reporting of aggregate or anonymous information" not intended for decision-making about an individual.

These Commissions and others recognized that confidentiality protections are meaningful only if supported by strong laws. In 1993, a National Academy of Sciences panel recommended that "statistical records across all federal agencies should be governed by a consistent set of statutes…including the following features of fair statistical information practices: (a) A definition of statistical data that incorporates the principle of functional separation as defined by the Privacy Protection Study Commission, (b) a guarantee of confidentiality for data…" and "legal sanctions for those who violate confidentiality requirements." The Office of Management and Budget (OMB) then acted administratively to "clarify and make consistent government policy protecting the privacy and confidentiality interests of individuals and organizations who furnish data for Federal statistical programs" in the Confidentiality Order in 1997. In 2002, the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) codified the three concepts of qualification, statistical purpose, and confidentiality – See Appendix A.

Prior to CIPSEA, in 1980, the Paperwork Reduction Act (PRA) codified mindfulness of burden on people and businesses an additional aspect of data's value. A principle objective of the PRA was to limit burden by maximizing the utility of existing data. In 2014, OMB further addressed this objective within existing law via G*uidance for Providing and Using Administrative Data for Statistical Purposes* (M-14-06), stating that program agencies should make every effort to provide administrative data for statistical purposes to the extent allowable under existing law. OMB's memorandum was an effort to tilt the default interpretation where clarity was lacking.

---

[2] See Chapter 15, *The Relationship Between Citizen and Government: The Citizen as Participant in Research and Statistical Studies*, *Personal Privacy in an Information Society: The Report of the Privacy Protection Study Commission*, 1977.

Challenges given the U.S. legal environment: As the U.S. has a decentralized statistical system governed by many statutes, there is no single legal framework for access and use limitations of government data. Rather, the authority to provide administrative data for statistical purposes occurs one of two ways: ***explicitly*** in law or ***implicitly***, derived by interpreting a given program's authorizing statute.[3] Implicit authorities are found where express authorities are not in law, but the agency's general statutory authority is interpreted to grant sufficient authorization to provide administrative data for statistical purposes.[4] In these cases, agency interpretation, rather than statute, governs access and use. Of course, sometimes authority explicitly prohibits any unspecified use and access for statistical purposes cannot occur. Other times, agencies may not interpret existing authority to include statistical purposes, or more likely, just not have an interpretation one way or another. Identifying the legal framework becomes further complicated when efforts utilize data from multiple agencies operating under different authorities or interpretations.

The Commission's working list of "top 20" priority data sources fall across this spectrum, as shown in Table 1. The result is a complex web of statutes, regulations, implementing guidance—or absence thereof—that can drive risk aversion in agencies, cause frustration for researchers, and limit the value of data for statistical purposes. In practice, this can also lead to a bureaucratic tendency to manage access requests to the common denominator, rather than a specifically determined or articulated policy. Therefore, current practice does not meet the characteristics of clarity or consistency of an "optimal arrangement." Solutions to this complex web must address all three scenarios, providing for a clear law for those areas without, and possibly provide for new access authorities within those explicit scenarios.

In the Commission's open meetings, we heard a number of specifics regarding access barriers to specific datasets or classes of data, including student data, workforce data, health data, and Census data. The lack of clarity and consistency do not contribute to maximized safety nor value. Beyond the confusion it engenders, this lack of optimization also affects the efficacy of both government and non-government researchers.

Within government, this condition is evident in several ways, such as the inability of statistical agencies to share their data for statistical purposes with each other or with evaluation offices. Twenty-two responding offices in the CEP Survey reported that they would like to use Federal tax data but did not because it was too hard to access. Of these offices, 12 cited "statutes prohibiting data sharing" and 15 cited "regulations or policies that make it difficult to link data" as factors that made it hard to access the tax data. One respondent added, "Statutes surrounding the IRS data currently prohibit Census from sharing business data that originates from IRS filings with BEA and BLS. As a result, Census, BEA, and BLS cannot achieve the full benefits of data sharing. Specifically, Title 26, Section 6103(j) needs edited to allow for BLS access to the data to fully achieve the data sharing benefits that the [statistical efficiency portion of the] Confidential Information Protection and Statistical Efficiency Act sought to achieve." When addressing access to Census data, the Economic Research Service in USDA specifically cited an "inability to show benefit to Census" as a barrier.

Outside of government, researchers and state and local government officials face similar challenges in accessing and using data. These researchers and officials often confront inconsistent and multi-party approval processes, particularly when seeking to link different datasets together. Comments received via the RFC highlighted barriers to specific datasets, as well as a particularly arduous process for access

[3] A good example of this is Title 26, the tax code, which limits access to tax data with explicit exceptions.

[4] A good example of this is USDA's determination that sharing SNAP data for statistical purposes is allowable.

to individual level vital records data. One CEP Survey respondent noted that, "in working on sub committees (Confidential Data Access Committee and others) we are hopeful that with the support of other agencies, OMB will continue to work on improving the laws surrounding data collection and security - CIPSEA in particular - so that one day we will be able to support researchers…on a cloud environment, which would only help speed and efficiency."

As a practical matter, data sharing between agencies or non-governmental actors is effectuated by an agreement (often an MOU) that must cite the authority for the exchange. Absent a clear citation for the legal authority, creativity can lead to outcomes potentially beyond the intent; hence, the need to ensure clarity of authority in statute. Moreover, several commenters noted that agency interpretations could vary and lead to confusion. For example, the Employer Data Matching Work Group noted, "at the agency level, there can be confusion in the interpretation of policies on data sharing, which may create additional barriers for agencies trying to access and share Federal data sources for matching purposes. Agencies may spend months or years coming to agreement on the proper interpretation of a particular statute or policy, and to develop interagency agreements to allow matching."

Remedies to achieve clarity and consistency must provide both clear authority and standard definitions of qualification, statistical purpose, and confidentiality. CIPSEA provides a strong starting point for standard definitions of all three concepts, potentially needing only minor modifications if the Commission were to recommend using those definitions to apply more broadly to achieve its goals. CIPSEA also provides for consistent application of disclosure limitations and associated penalties.

A potential strategy that achieves clarity and consistency was introduced to the Commission at the March 13 meeting by Maria Cancian in her presentation about Federal-State partnerships in support of evidence-based programs. She recommended an approach called "Yes, unless," discussed further as a legislative strategy below and incorporated into the preliminary recommendation.

> **"Yes, unless" approach:** "Yes, unless" would provide for language clearly stating the intent to allow access to and use of survey and administrative data for statistical purposes, save for those cases where Congress has clearly provided alternative guidance. This approach is comprehensive enough to capture programs and data that may not have the benefit of specific statutory language, while stopping short of overturning those cases where Congress has acted (or may act in the future) to limit access and use. This approach offers the benefit of being both reactive and proactive, providing a mechanism for future programs to be governed by this standard, absent specific Congressional direction to the contrary. A "Yes, unless" approach codifies OMB's M-14-06 direction to default to yes.

PRELIMINARY RECOMMENDATIONS

► The Commission recommends the adoption of a new "Yes, unless" prevailing statute to govern data access and use specific limitations. The language would clearly indicate the intent and authority for data to be accessed and used for statistical, evaluation, research, and analysis (SERA) purposes unless explicitly prohibited by other applicable law;

The Commission further recommends that the new prevailing statute build upon CIPSEA so that in the absence of more specific language in an existing program-specific statute it provides:

- A common definition of statistical purpose and use;
- Specific, strict disclosure limitations and confidentiality protections; and
- Specific qualifications for researcher access to a given dataset.

► The Commission recommends the following [insert list from forthcoming CEP Recommendation Memo #6] amendments to broaden access to specific priority datasets for statistical, evaluation, research, and analytical purposes;

The preliminary recommendation has two essential parts, collectively dubbed the "Yes, unless plus a little" approach: (1) A "Yes, Unless" provision grounded in CIPSEA and (2) additional amendments to address specific identified limitations (i.e., current "unlesses") in law, such as those governing tax data, Census data, the National Directory of New Hires (NDNH), and others. The success of implementation may be complicated by these additional requirements. For example, the tax code is specific in its limitation of both qualifications of access, as well as the uses. To achieve the level of access suggested in input to the Commission, language beyond "Yes, unless" would be required. This is true for Census data, NDNH, and other data as well. Regardless, this approach achieves clarity, as a standard applies in all cases except those with clear exceptions. It similarly provides broad, but not complete, consistency because of the need to address the "unlesses."

Appendix B summarizes the current legal state of affairs.

The "Yes, unless" approach would likely take the form of language embedded in the PRA chapters of Title 44, which is conveniently slated for reauthorization around the time the Commission's report is due. Locating this provision within CIPSEA would clearly signal the centrality and applicability of its definitions of qualification, statistical purpose, and confidentiality as well as of its provisions covering disclosure limitations and penalties for breaches.

The slightly more expansive "Yes, unless plus a little" approach in the recommendation above offers near full compliance with the goals of clarity and consistency, but preserves the prerogatives of Congress to go beyond the default standard when desired. As a result, it is the most appropriate choice in both the policy and political contexts. Further, achieving a clear and consistent default definition for statistical use, disclosure limitations, and qualifications of researchers will serve to increase the management and efficacy of privacy and confidentiality protections, and effectively raise the bar on these protections to a level higher than is otherwise generally true.

LIKELY REACTION TO RECOMMENDATION:

- ***Congressional:*** As the Commission's statute specifically directs recommendations regarding removing access and use barriers, it is reasonable to assume Congress is expecting recommendations in the vein of those in this paper. The "Yes, unless" provision provides a positive response to the Commission's charge while also preserving all Congressional prerogative, which should satisfy most concerns. Further, the application of provisions for allowable use, disclosure limitations, qualifications, etc. should be appealing to Congress, as they reflect current statute, and thus Congressional intent.
- ***Executive:*** Most agencies will likely welcome clarity and specification on how to manage access and use controls. Statistical agencies, in particular, are likely to be supportive, as this provision will resolve a number of long-standing challenges related to data sharing. Likewise,

evaluation offices should be supportive of the provision, as it would substantially expand their access to a wide array of data assets. That said, all agencies would face some level of burden for implementation.

- ***State and Local:*** No negative reactions anticipated, as the provision does not apply directly to states as written (this could be affected by Memo #5 discussing state-held data). Further, this provision would provide a clear access path to Federal data for state and local evaluators.
- ***Other:*** No negative reactions anticipated, as researchers and qualified members of the public would gain clear access paths to federally held data.

COST IMPLICATIONS

While implementation of these recommendations would not have immediate costs, the implication of removing barriers is that there would be increased demand placed on agencies whose data can now be shared more easily. Some agencies may be able to absorb the increased demand within existing budgets while others may need additional resources to expand their capacity to fulfill the additional data requests. Cost could take the form of delays in approving new data sharing agreements. CEP will address implementation and other costs more completely in memo #16 on funding issues.

ALTERNATIVE OPTIONS

- ***Option #1:*** **Notwithstanding clause.** A more simple and comprehensive option would be to recommend statutory language providing for access to and use of administrative data for statistical purposes as defined in CIPSEA "notwithstanding any other provision of law." The advantage of this strategy is its simplicity, comprehensiveness, and clarity. The disadvantage is that it effectively renders void all existing Congressional intention to provide for data-specific limitations and controls on access and use. A notwithstanding clause is fully evergreen, serving both retroactively and proactively, and negating any further legislative action on access and use limitation, which is both a strength and a weakness.

  The notwithstanding approach is politically complicated, and necessitates a likely unrealistically high bar for passage. While there is precedent for notwithstanding clauses in statute, the clause can and should be a difficult level of action to achieve, as it opens the door to unintended consequences given its broad application, particularly in those cases where contradictory language may exist in the code. A notwithstanding clause can conflict with another notwithstanding clause, in which case the tendency is to see the most recent Congressional action as supreme. This is the statutory equivalent of saying "this is the single most important consideration, regardless of what else has been said." For example, a somewhat recently passed notwithstanding provision was related to access to government networks by the Department of Homeland Security (DHS) for the purposes of cyber defenses. In this context Congress was willing to say this mattered above all else, but does statistical use—while highly valuable—rise to the same level.

  A notwithstanding clause would most logically be embedded within the PRA chapters of Title 44 and thus co-located with CIPSEA. With respect to implementation, a notwithstanding clause likely faces the highest hurdle for adoption, but generally would not require additional steps. It is likely the closest one can achieve in statute to comprehensive clarity and consistency.

- ***Option #2:*** **Complementary amendments.** Another option would be to make changes to each individual statute necessary to permit access for statistical purposes. On this subject, the UN suggests, "an ideal state of affairs is one of reciprocity, where the statistical legislation lays down the rights and conditions of access, and the specific legislation that protects administrative holdings, wherever they may be within Government, recognizes as an exception the right of access … for statistical purposes."[5] However, the UN guidance assumes a single statistical law exists. In the U.S., with its collection of statutes governing the components of the decentralized statistical system, this approach would require amendments to most, if not all titles of the U.S. Code. Each SERA agency's authorizing legislation would need to be amended to include an explicit mandate to access administrative records (with the exception of the Census Bureau statute, Title 13, which already provides such authority, but in practice can be limited by the lack of clear provision authority by the counterparty).[6] In addition, each statute that protects administrative records across the government would require modification to recognize an exception for statistical use.

  The complementary amendments option provides the benefit of a "case-by-case" analysis and intentional choice by the Congress as it relates to each SERA agency and use. However, the approach also entails the risk of being less than comprehensive, should some of the necessary amendments not be enacted. The systematic review of each agency's authorization could also invite additional scrutiny of various programs for political reasons unrelated to data access and use. On its own, this approach is neither reactive nor proactive, as each authorization would require action. Further, this approach fails to speak to data access issues for future provisions, the lack of which, in context, should be viewed as an intentional exclusion. Most problematically, this approach carries substantial risk that it does not achieve clarity or consistency. Indeed, it furthers the risk of legislating even less clarity and consistency. For these reasons, this option would be the most complex and risky to implement.

  The complementary amendments approach could not be achieved solely by amending the PRA, but its various provisions could reference the common definitions as desired.

- ***Option #3: Yes, Unless.*** This options mirrors the preliminary recommendation except, critically, it does not seek to rectify existing specific barriers for high priority data sets identify by the Commission.

- ***Option #4: Status Quo.*** Do not pursue statutory changes, and instead recommend increased resources and attention for implementation of M-14-06.

---

[5] United Nations. *Handbook on Statistical Organization.* (New York: United Nations, 2003), para. 146, available at https://unstats.un.org/unsd/publication/SeriesF/SeriesF_88E.pdf

[6] 13 U.S.C. Section 6

FOLLOW UP TO RECOMMENDATION MEMO #3

A Prevailing Statute and Amendments to Address Prohibitions to Data Access

Commission Lead: Rice
Staff Leads: Hitt, Howell

Issue: Equip the evidence-building community with legal authorities to acquire, protect, and utilize data in order to accelerate generation of evidence about government programs and policies.

This paper expands on the Commission's discussion and initial conclusions regarding Legal and Statutory Issues (memo #3) during the April 3 meeting. The Commissioners generally agreed with the draft recommendation (#3.1), to propose a default to "yes" statute for data access and use for statistical purposes unless explicitly prohibited by other applicable law. Remaining outstanding were two questions, (1) how to reconcile misaligned statutes (previously "unlesses") and (2) if the statutes should require access (shall) or only allow access (may). The shall vs. may issue was, in-part, resolved during discussions regarding access to state data, where the Commissions generally viewed requiring access to Federal data as appropriate. This paper addresses the question of shall vs. may, and presents a specific recommendation, referred to herein as the "CIPSEA Exception," to build upon an existing legal framework to address the remaining question of the "unless" issue.

* * * * *

Recommendation memo #3 on Legal and Statutory Issues addressed legal barriers to data access and use among agencies, including the proposed data facility, and nongovernment researchers. The paper summarized the current statutes as either explicitly authorizing access (rare), explicitly prohibiting or strongly limiting access (moderately common), or vague, not addressing access for evidence building (common). Via its conclusions on memo #3, the Commission addressed the vague condition by suggesting a prevailing statute that provides authority for data to be accessed and used for exclusively statistical purposes unless explicitly prohibited by other applicable law. Notably, this recommendation perfects and reiterates an existing law, 44 USC 3510, which provides explicit authority to the Director of OMB to require the sharing of data between two agencies, unless inconsistent with applicable law. This is effectively an existing "Yes, Unless" authority, albeit not limited to statistical purposes.

Further, Commissioners requested staff provide additional information on the issue of a requirement versus a general authority (shall vs. may), as well as proposed solutions to the cases of explicit prohibition or tightly limited access (the "unlesses"). As the Commission continued to address other issues in scope, several other determinations became relevant to this question:

(1) Ensuring that authorized data sharing is not a "free for all,"

(2) Ensuring that newly accessible data is protected at the highest reasonable standard,

(3) Ensuring that access is monitored and reviewed appropriately; and,

(4) Respecting Congressional prerogative to limit access.

These four items represent additional, refined objectives beyond those originally noted in memo #3, reflecting the Commission's desire to respond to the most prevalent barrier of evidence-generation by meeting the legitimate needs for data access, while also strengthening privacy. An elegant solution to these combined goals is to provide authority for the Principal Statistical Agencies (PSAs), including the proposed Data Facility, to acquire and share data under an existing limited common authority, the Confidential Information Protection and Statistical Efficiency Act (CIPSEA).

The "CIPSEA Exception" solution meets each of the four requirements noted above, by limiting the "who" and the "why," raising or maintaining penalties for disclosure, and including new clear qualifications and monitoring requirements. This is accomplished by building upon an existing, road-tested legal framework, which Congress originally created for essentially this purpose. In comparison to the alternative – ad hoc amendments to a number of statutes to add specifically allowable disclosures and uses – this solution is both simple in construction, as well as powerful in application.

Avoid a free-for-all by limiting the "who"

In the discussion of memo #3, commissioners agreed implicitly that the "who" would be anyone with a "statistical purpose," but remained concerned that in practice this may be unreasonably broad, and include actors without the capacity to protect data properly. In memo #10, the commissioners developed a privacy framework built around CIPSEA, and memos #7 and #9 provided for the ecosystem and PSA-network concept. Combining those concepts can reasonably narrow the "who" because the authority to acquire and access data under CIPSEA would be limited to those agencies and agents qualified under CIPSEA.

Currently, CIPSEA Subtitle A provides to OMB the authority to delegate the authority to protect data to agencies or units based on the determination of the Director.[1] CIPSEA Subtitle B authorizes the sharing of business data by three "Designated Statistical Agencies" listed in the statute.[2] The "CIPSEA Exception" solution would align the "who" in both subtitles, but would also shift the authority to designate CIPSEA status from an OMB policy into law via two prongs.

First, Congress could explicitly name agencies when it updates CIPSEA. Second, Congress could strengthen the current mechanism by which OMB identifies agencies into a formal rulemaking process.[3] In this construction, OMB would be directed to determine and propose in regulation the requisite qualifications for an agency to gain CIPSEA authority, and through a rulemaking process subject to public comment under the Administrative Procedure Act would nominate agencies for that status.[4] The qualifications would reflect an agency's ability to demonstrate functional separation from non-statistical agencies or activities, including from policy, law enforcement, and regulatory functions, and a demonstrated ability to maintain IT systems separate and secure from those of policy, law enforcement, and regulatory functions. The rulemaking process would provide transparency, opportunity for public input, and the opportunity for Congress to reject OMB's nomination. This formal process would expressly limit CIPSEA authorities to agencies meeting the

[1] This is delineated in the OMB CIPSEA Implementation Guidance.
[2] These are the Bureau of the Census, the Bureau of Economic Analysis, and the Bureau of Labor Statistics.
[3] Commissioner Rice, the memo lead, noted that her implementation preference is for the amended statute to specifically list the 13 PSAs, and the facility, and direct OMB to develop the qualifications and bear the responsibility to ensure all 13+1 meet those within a specified timeframe.
[4] P.L. 79-404

determined criteria and receiving due consideration from the public and Congress, while still allowing for growth in qualified agencies as needed to meet future needs. In particular, the Commission hopes that some larger evaluation and research offices may seek to achieve CIPSEA status in the future, further enabling the routine production of evidence. However, as the PSAs are all currently CIPSEA qualified (under today's OMB standards), they would serve as the core group, and facilitate access as necessary.

Avoid a free-for-all by limiting the "why"

As written, CIPSEA addresses the "why" by expressly limiting disclosure of CIPSEA-protected data to exclusively statistical purposes, or the use of data for purposes of describing, estimating, or analyzing the characteristics of groups rather than individuals or organizations. In some ways "statistical purposes" as defined in CIPSEA could be considered broader than the limiting purposes identified in other statutes, such as the "Census benefit" implied in Title 13 and "tax administration purposes" defined in Title 26. However, the "Census benefit" standard is effectively exclusively statistical, given the mission of the Census Bureau, and thus this does not meaningfully change the "why." "Tax administration," however, far exceeds the narrow scope of statistical purpose, and thus the application of the "CIPSEA Exception" solution would expand the access to tax data, but only within the narrow context of statistical purposes.

Protect it all to the highest reasonable standard

Protecting data under CIPSEA, particularly for those data not expressly protected by an organic statute, also ensures data are protected at the highest statutory level, and discourages unauthorized disclosure of information with rather severe penalties. Under CIPSEA today, the penalty for any officers, employees, or agents of a CIPSEA agency who willfully discloses protected data is a felony, punishable by up to 5 years and/or fines up to $250,000. For those statutes vague on access, this creates specific penalties for the first time, and for those statutes with existing penalties, CIPSEA either matches or exceeds those penalties.

These strong penalties mean that when Congress chooses to apply the "CIPSEA Exception" solution, those data will be protected from disclosure at a level at or beyond those penalties in individual statutes, including Title 26.

Provide for monitoring and review

The "CIPSEA Exception" solution also could reasonably include increased accountability by including a specific monitoring and auditing element, along the lines of a Special Inspector General (SIG) for Privacy and Confidential Information Protection. This SIG would have authority to review and audit data use, handling, procedures, and processes, as well as to monitor agency compliance with CIPSEA qualifications, including providing an annual report to Congress and President. The SIG would be, essentially, a specialized law enforcement unit staffed and resourced specifically to guard the application and use of data protected under CIPSEA.

Respecting Congressional Prerogative

This leaves the issue of Congressional prerogative. Since the existing statutory "unlesses" exist because Congress saw fit to include them in law, it is not the Commission's place to recommend their elimination. As with the Commission's determinations on data bans (Memo #6), here the Commission is recommending that Congress review and consider the inclusion of the "CIPSEA

Exception" in the various statutes that limit data access. This solution provides Congress a meaningful tool to enable access to data for exclusively statistical purposes, including certain aspects of evidence building, but to do so within an established framework subject to qualifications, regulation, and oversight.

DRAFT RECOMMENDATION

To accomplish this vision, the Commission recommends the following actions:

► **Recommendation 3.1 (replacement): Recommend that the Congress and the President enact legislation establishing a prevailing legal authority for protective data sharing and review prohibitions on data access and use.**

Action Item 1: The Commission recommends that Congress and the President act to amend CIPSEA to provide for, and accomplish the following:

- Direct OMB to establish, through a rulemaking process, the specific requirements, and qualifications necessary to gain CIPSEA authority.
- Establish a Special Inspector General for Privacy and Confidential Information Protection, with all appropriate authorities, and an annual reporting requirement.
- Expressly authorize CIPSEA-designated agencies/units to acquire and protect data from Federal agencies for exclusively statistical purposes.
- Expressly authorize CIPSEA-designated agencies to disclose minimally necessary data with other CIPSEA authorized agencies, units, and agents.[5]

These amendments, in combination with existing CIPSEA text, would create a specific legal structure to ingest data into the CIPSEA-protected environment and share that data within that environment, as well as a means to limit access to that environment in concert with monitoring and auditing.

The data acquisition elements above (i.e., the "Yes, Unless" prevailing statute) could be constructed as follows:

> "Unless otherwise specified in law, a 'Designated Statistical Agency/Unit' may request data held by an agency of the Federal government, which shall be provided for exclusively statistical purposes."

Action Item 2: Congress should review statutes that include prohibitions against data access and use and determine if it deems limited sharing acceptable within the envisioned CIPSEA environment. Where it deems it appropriate, Congress should act to amend relevant statutes to include language authorizing a "CIPSEA Exception" to access and use data for statistical purposes.

[5] Note this is not a "shall" requirement; per existing CIPSEA text, disclosure between two CIPSEA agencies will still require a Memorandum of Understanding, including details of the allowable use and specific variables, and the agreement of the head of the protecting CIPSEA agency.

While respecting Congressional prerogatives, the Commission recommends that Congress specifically consider applying the CIPSEA Exception to Title 7 (SNAP), Title 13 (Census), Title 20 (FSA), Title 22 (BEA), Title 26 (IRS), Title 28 (FBI), Title 29 (BLS), Title 42 (NDNH, Head Start).

This recommendation allows Congress to review its prior determinations in the context of the desire for increased use of evidence, and a newly constructed legal environment. By acting to include the CIPSEA Exception, Congress demonstrates specific intent, ensuring full transparency and clarity and representing public input.

The CIPSEA Exception could be constructed as follows:

> "Notwithstanding any other provisions of law, data collected under this chapter shall be made available under 44 USC XXX [the CIPSEA statute] to an appropriately qualified agency for exclusively statistical purposes."

Many of the statutes that will require such an amendment are those statutes that authorize the PSAs, including Title 13 (Census Bureau statute) and Title 22 (which includes authority for BEA to collect survey data) (See Appendix A). By acting to include the CIPSEA Exception, Congress is also specifically acting to "deem" data collected under a different authority as now covered by CIPSEA.

Additional Options and Enhancements for Consideration

- Establish a CIPSEA Advisory Committee, or assign this responsibility to the transparency board discussed in memo #10, to provide express public input and transparency to all agencies qualified under CIPSEA, as well as to advise OMB on implementation and qualifications.
- Establish a tiered access structure, whereby not all CIPSEA-designated agencies would have access to all CIPSEA-protected data, and a regulatory mechanism would stipulate the specific data sharing authorities between each designated agency. This could be modeled after existing regulatory structures that further limit access to tax data beyond the limitations stipulated in Title 26.
- Include explicitly authority for a CIPSEA-designated agency to acquire, protect, and share Federal data held at the State level under the similar parameters.

Appendix B: Preliminary and Partially Complete Summary of CEP "Top 20" Data Sources

Datasets	Source	Legal Authority	*Comparison of Existing Laws to CIPSEA-like scope of evidence-building. Authorizes…*		
			CIPSEA-like "Who (agencies/agents)?[1]	CIPSEA-like "what" (statistical purposes)	CIPSEA-like confidentiality protections?
Income/Earnings/Wage					
FTI	IRS	26 USC § 6103 et seq.			High
UI	States	CFR 603.5	??	??	??
NDNH	HHS/ACF	42 USC § 653			Low
LEHD	Census	13 USC § 9			High
QCEW	DOL/BLS	29 USC § 2			High
Health					
Medicaid	HHS/CMS	42 USC § 1396			Low
CHIP	HHS/CMS	42 USC § 1396	?	?	?
Vital Records	HHS/NCHS	42 USC § 242k(h)			High
	SSA	42 USC § 405			Low
Health Surveys	HHS/NCHS	42 USC § 242k			High
	HHS/SAMHSA	42 USC § 290aa	?	?	?
Human Services					
SNAP	USDA/FNS	7 USC § 2011 et seq.			Low
Head Start	HHS/ACF	42 USC § 9836a			Low
Child Welfare (AFCARS)	HHS/ACF	42 USC § 679	?	?	Low
TANF	HHS/ACF	42 USC § 611-611a	?	?	?
Education					
Federal Student Aid	ED/FSA	20 USC § 1092b(d)(2)			?

[1] Referencing CIPSEA definitions of agency and agent, not existing disclosure authority

Datasets	Source	Legal Authority	CIPSEA-like "Who (agencies/agents)?	CIPSEA-like "what" (statistical purposes)	CIPSEA-like confidentiality protections?
Crime					
Crime reports and rap sheets	DOJ/FBI				Low
Criminal justice	DOJ/BJS	42 USC § 3732 & 3735			High
Demographic and Economic					
CPS	Census	13 USC § 9			High
Decennial	Census	13 USC § 9			High
ACS	Census	13 USC § 9			High
SIPP	Census	13 USC § 9			High
Economic Survey Data	Census	13 USC § 9 but for 26 USC § 6103(j)			High
Other Lightly or Not Explored					
Business Program Participation Data	various		?	?	?
Military/Veterans Data	DOD and VA		?	?	?
Transportation	DOT		?	?	?

Key

CIPSEA Who

(Full( )/Some( )/Little( )/None( ))

CIPSEA What

(Full( )/Some( )/Little( )/None( ))

CIPSEA Confidentiality
(High/Low)

RECOMMENDATION MEMO #4

Data Access and Use: Federal Policy and Standards

Objective: Some barriers to data access and use among agencies, a prospective facility, and non-governmental researchers exist because of variability or discordance in Executive Branch policy and practice. This recommendation will address those barriers by recommending Executive Branch standardization of policies and practices affecting access to administrative data for statistical purposes.

In order to facilitate access to and use of administrative and statistical data for evidence building, there are three types of inter-related recommendations: those that address (1) law and regulation; (2) executive branch policies and standards; and (3) implementation, via a data facility, other cross-agency measures, agency-specific measures or a combination. This paper focuses on executive branch **policies and standards** to facilitate appropriate and efficient access to and use of data.

* * * * *

BACKGROUND

The Commission is to conduct a comprehensive study of the government's data inventory and analyze barriers to linking the data in this inventory. It is to determine the optimal arrangement by which administrative and statistical data should be made available to facilitate evidence building and make recommendations on how data infrastructure should be modified to best fulfill the objectives.

Some data access and use barriers require legislation to address. With or without new legislation, the Commission has heard about a number of actions the Executive Branch could take to address others. These actions generally boil down to standardizing the way agencies do things to increase feasibility, efficiency and quality. Some standardizing also will be central to enhancing privacy protections, particularly employing privacy-enhancing technologies (e.g., secure multi-party computing). While a data facility may serve as the implementation vehicle for some of the standards discussed herein, all would require the participation of agencies beyond the facility.

This memorandum focuses on actions designed to affect multiple Federal agencies and it is therefore amenable to government-wide standard setting. These actions would sometimes affect states and other entities as well. The memorandum generally excludes actions that only one specific agency would take (e.g., a Food and Nutrition Service regulation affecting state program office data standardization). To the extent that the Commission heard such recommendations, they will be addressed in subsequent memoranda.

Stakeholders would like to see standardized: ***(1) Data inventory/metadata; (2) Datasets; (3) Data access agreements; (4) Technical data "exchange;" (5) Application, "qualification" and approval processes; and (6) Data product release protections.***

Data Inventory/metadata. Having "data about data," or information about what is in a dataset, is vital to improving accessibility and linkage. Standardized metadata also is essential to employing privacy preserving technologies such as secure multi-party computing as the means for linking datasets. Multiple stakeholders emphasized the importance of the Federal government creating and maintaining a list of government-held datasets, along with basic information about mechanisms for

accessing, as the most basic form of metadata. Stakeholders described metadata as a critical component of assessing fitness for use and data quality, and the Commission heard numerous recommendations to create robust administrative dataset documentation, ideally via standardized metadata for key datasets. Contributors called for an explicit role for stakeholders to ensure that standards are useful for evidence building. Results from the CEP Federal Office Survey (CEP survey) found that all 13 Principal Statistical Agencies collect and create metadata that includes variable definitions, units of measurement and response ranges. Also, the CEP survey showed that overall, the higher percentage of budget an office has for statistical, evaluation, research and analysis purposes, the more likely the office is to create metadata or technical documentation.

Datasets. Beyond metadata, multiple stakeholders stated that there is a need to standardize the underlying dataset elements. Comparability of concepts and definitions improves usefulness in two ways: by allowing data from disparate sources, such as states, to be used in tandem with each other and by allowing results originating from different datasets to be more comparable. Datasets can be partially or fully standardized, from:

a. *Variables used for file linking*. Common examples are name, SSN, EIN, date of birth or address. High quality linkage keys make linkages feasible and valid. A Federal working group concluded that in the absence of a universal business identifier, agencies should use multiple variables that, as a set, come close.
b. *Key variables*. Comparable definitions of the most used research variables are important for utility and validity. The Commission heard two cautions: standardization requires upfront funding and works at cross-purposes with the richness of each dataset incorporating nuances relevant to a particular context. Some sought standardization of Federal program reporting metrics as a way to reduce burden on states and to create incentives to standardize (e.g. the definition of "postsecondary recognized credential").
c. *All variables*. Full standardization facilitates easier use and comparability for policy and research. For example, local law enforcement agencies can participate in the National Incident Based Reporting System (of detailed crime incidents) only by fully adhering to the standards set out by the FBI.

Data access agreements. Formal data access agreements (e.g., MOUs) between two or more parties can take years to develop. Perhaps the most commonly identified data access barrier cited by stakeholders was the non-standardized and lengthy processes by which access is negotiated and memorialized via an agreement. In the CEP survey, one-quarter (23/83) of responding offices said that "Lack of staff, policies, and procedures to establish data sharing agreements" in their own office made it hard for them to get access to the specific data set they needed. Creating a uniform agreement that covers both Federal data and relevant data at the state level would reduce redundancy and save time.

Technical data "exchange". Standardizing protocols can reduce burden while enhancing security. A UN report on accessing administrative data states, "There are a number of international standards for data and metadata transmission....It is therefore important to agree which standards are to be used." Some commenters noted the time and effort associated with determining the technical data exchange standards to which a project will adhere, indicating that given lack of expertise in many agencies, especially at the state level, a standardized process would reduce cost and risk.

Application, "qualification" and approval processes. Researchers apply for access to data and agencies "qualify" researchers and approve their projects based largely on individual agency-defined processes. It is cumbersome and time consuming for the researchers and duplicates effort across agencies. A standardized set of processes, implemented by agencies, a data facility, or a combination, could reduce burden, time delays and redundancy. Stakeholders indicated that some datasets only seem to be available to researchers with "an in" and that even where application processes are public, they are lengthy and burdensome, in some cases making them too lengthy to be practicable for a graduate student's timetable, for example. Some also commented that being "qualified" to use one agency's data should be more transferable to another agency.

Data product release. Statistical agencies, and increasingly other agencies, release summary statistics and datasets publicly, especially in light of "open data" initiatives. In addition, agencies approve researcher "output" using agency-specific procedures and rules. A few stakeholders discussed the benefits of a tiered-access system of data release with formal statistical disclosure limitation (SDL) processes. The CEP survey found that of the 32 offices who responded yes, they provide researcher access to data, all Principal Statistical Agencies (13/13), 50% of evaluation offices (3/6) and 39% of other offices have a process for reviewing output to prevent disclosure. Offices that conduct disclosure review do not always have a process for documenting the results. Only 62% of Principal Statistical Agencies (8/13) have such a process. None of the 3 evaluation offices that conduct disclosure review have such a process, and only 3 of 5 other offices that conduct disclosure review have one. Given risk of re-identification, either from an agency's own multiple releases or from the "mosaic effect" of many entities releasing related data, each agency needs a consistent process to help it protect data against re-identification. These standards could include methods to measure and quantify risk, guidelines for using tiered access systems and statistical disclosure limitation techniques, including privacy budgets.

Standardization requires two policy-setting actions: (1) creating an actionable requirement for agencies to adhere to a standard and (2) creating the standard. Setting, then implementing, a standard requires an up-front investment. This initial investment has the potential for a long term payoff by allowing Federal and State agencies, especially those with limited infrastructure of their own, to utilize an already created 'how to' for data access.

OMB has relevant policy infrastructure in place to inform the Commission's thinking:

- *M-13-13, Open Data Policy—Managing Information as an Asset*, requires Federal agencies to create data inventories and provide updates to OMB at periodic intervals. According to OMB, the inventory will be built out over time, with the ultimate goal of including all agency datasets, to the extent practicable. Today it is of uneven quality and completeness, especially with regard to datasets that agencies have determined they cannot publicly release. It requires a very limited set of minimum metadata.
- OMB is authorized to set information and statistical standards that all agencies must follow, such as the Standard Occupational Classification and the standard for collecting and reporting on race and ethnicity. The race and ethnicity standard applies to even non-statistical (administrative) data collections, thereby flowing the requirement "downstream" to grantees and others reporting to the Federal government. These standards do not have the force of law as a regulation does. However, because they similarly affect both agencies and the larger public, OMB voluntarily develops them using interagency groups and the public notice and comment process required for regulations under the Administrative Procedures

Act (APA). Though it has not happened, an Administration could suddenly and unilaterally issue, rescind or change a standard without following APA procedures.

- In addition to setting policy designed to facilitate use of administrative data for evidence building, *M-14-06, Guidance for Providing and Using Administrative Data for Statistical Purposes*, provides an optional "model agreement for the provision of administrative records for statistical purposes." Its optional status combined with limited resources for government-wide implementation limit the uptake of such a tool.
- OMB's Federal Committee on Statistical Methodology has a standing subcommittee, the Confidentiality and Data Access Committee (CDAC), author of Statistical Working Paper (WP) 22, long considered the "Bible" of statistical disclosure limitation methodology. Despite incremental updates, CDAC's products do not fully incorporate concepts the Commission has explored of quantifying disclosure risk, such as via k-anonymity, or managing risk of complementary disclosure across multiple releases, such as via a privacy budget. While widely cited and used, WP 22 is voluntary and small agencies or those new to data product release under Open Data often have no familiarity or capacity to adhere to it.

Some individual agencies have tackled program-specific standard setting and enforcement in recent years. For example, recently the Centers for Medicaid and Medicare Services (CMS) developed extensive standards to address long-standing quality and timeliness concerns about the Medicaid data provided by states to CMS. These types of standards have also benefited human service data in states that developed integrated systems, though such an approach is not sufficiently cross-agency, which is the scope of this paper.

PRELIMINARY RECOMMENDATIONS

► In order to unleash the potential of the government's already-collected data to build evidence, OMB should leverage its existing ***policy and standard-setting*** authority and

Within six months, direct agencies to:

1. Ensure they are fully complying with OMB's 2013 policy (M-13-13) to make public complete ***inventories*** of datasets, especially of those datasets not publicly available
2. Identify ***existing*** dataset ***documentation*** that could usefully be made public

Within one year, **set standards** for

3. Improved ***minimum metadata*** of inventoried datasets to help the public learn how to access them
4. A template that all agencies use when entering into ***interagency agreements*** to enable authorized data access and use
5. A single, streamlined ***process*** through which ***researchers*** external to the government may ***apply, qualify, and are approved*** to access government data not publicly available

Within 18 months, **set initial standards** for

6. ***Comprehensive metadata for [20] high priority data sets*** identified by the Commission as having the most immediate potential to boost evidence-based policymaking. The standard would be sufficient to allow two things: assessing quality and implementing privacy-enhancing data matching technologies.
7. Risk-based and quantifiable confidentiality ***methods*** agencies must employ ***to protect against re-identification*** in released data products.

Within two years

8. Direct agencies to **implement** the three new standards: minimum metadata; interagency data access and use agreements; and researcher application, qualification and approval.

Within three years

9. Direct agencies to **complete initial implementation** of comprehensive metadata and confidentiality standard

► Recommend that NIST, in close collaboration with OMB, identify the most appropriate ***data exchange standard*** for use across Federal agencies and Federal grantees, in support of evidence building and also consistent with or useful for operational or administrative data exchanges.

These recommendations build on existing law, policy and practice while addressing widely cited barriers to accessing data for evidence building. OMB is the primary actor because of its central role in government, existing legal authority and experience in setting information and statistical policy that applies to all agencies, reserving one type of technical standard for NIST. OMB should include key stakeholders throughout the process, both within and external to the government. The placement at OMB and the inclusion of stakeholders enable more seamless connection to the implementation process that would follow release of each standard by ensuring the surfacing of costs, identification of resources, setting of realistic timetables, and overall buy-in to the standards.

The recommendations build on an existing OMB requirement in M-13-13 to inventory and create minimum metadata for all Federal datasets, and emphasizes the need to improve the "as is." The recommendations recognize the cost and burden, yet criticality, of creating comprehensive standardized metadata by focusing only on a few high priority datasets to start. The recommendations exclude standardization of data elements or datasets (the second item in the background) because it would be expensive and infeasible without a sharp focus on key datasets or variables that are relevant across many Federal programs and agencies. Stakeholders did not identify specific data elements that could be standardized by executive action alone.

Items 4 and 5 differentiate processes between agencies and those between a researcher and an agency. Here, all agencies engaged in seeking and providing access and use of data would be required to use the standardized agreement template, whereas the OMB's current standardized template is voluntary, and generally not used. The Commission may wish to offer OMB a more definitive recommendation on the content of a "qualification" standard given that the Commission is to "evaluate… which types of researchers, officials, and institutions should have access to data and what the qualifications of the researchers, officials, and institutions should be…." If so, that additional guidance could be stated as a separate recommendation or, if very brief, incorporated in the text above, when ready.

Standards for comprehensive metadata and confidentiality (6 and 7) are "initial" standards in order to balance the complexity of the task with the urgency of the need. OMB and agencies undoubtedly would want to revisit and improve upon the initial standards over time.

Item 5 addresses concerns about complementary disclosure (re-identification) risk posed by multiple releases including actions of entities beyond the releasing agency. It creates a consistent methodology across agencies, recognizing that few, if any, currently quantify risk as part of the release procedures. This recommendation is chiefly about the need for a standard and who would lead the effort. Given the extensive Commission focus on data facility design, including the role of agreements that reduce the risk of re-identification, it is likely that recommendations on that topic would serve to refine or compliment this one.

The executive branch could advance each of these standards with or without adopting the Commission's recommendations on a data facility, though the facility itself would be a primary means of implementing some of them. In particular, the facility would be, by design, the entity administering agreement processes and templates and facilitating application for access by researchers. It also would be the point of release for research and potentially a limited set of "production" statistical results created as a part of its ongoing analytical capability.

SUMMARY OF COMMISSIONER CALL

On the March 27 call, Commissioners favored quick, incremental progress over a comprehensive longer-term approach. Nancy recognizes that her office would be responsible for implementing this recommendation and, with clarifications (incorporated), supports it. Commissioners recognized that implementing the recommendations requires new or redirected resources.

LIKELY REACTION TO RECOMMENDATION:

- *Congressional:* No adverse reaction anticipated.
- *Executive:* All agencies will incur largely modest up-front costs, even if just the cost of meeting an earlier requirement adequately. Statistical, research, evaluation and analysis agencies stand to gain from these requirements, so are more likely to be supportive than program offices, which must play a central role in implementation but may not see direct benefits to program operations. Some may genuinely resist a requirement to change existing processes.
- *State and Local:* Some adverse reaction is anticipated to the extent that any requirement affects them, particularly if it is not explicitly funded. Like Federal agencies, opportunities to participate in the process and possible efficiency gains should somewhat assuage them.

COST IMPLICATIONS

This set of tasks, along with implementation-related ones we will consider later, requires modest additional staffing at OMB. Agencies will likely be expected to absorb the new requirements within existing budgets so cost will take the form of elongated implementation schedules or other activities receiving lower priority. Participation in standard setting costs to states and other grantees and the public will be minimal and implementation costs borne as part of participating in a Federal program. The Commission will address implementation and other costs more completely in a future recommendation memorandum on funding issues.

ALTERNATIVE OPTIONS

- *Standards Are Voluntary for Grantees, including States.* The standards could be focused on Federally-held data and optional for states or other grantees. By definition, having only some states participate limits a standard's intended benefits, and voluntary status likely lengthens the timetable for participation. States holds many of the priority datasets, so the Commission would effectively be indefinitely deferring improved access to those datasets.

- *Prescribe Additional Details as Inputs to Standard-setting.* The Commission could weigh in on key elements of any given standard. For example, it could specify which elements should comprise a minimum metadata standard or could prescribe a specific role for the data facility in overseeing or carrying out the data release methodology on behalf of agencies. To weigh in with specific standards content, Commission staff would need to do additional research, which may not be feasible given the remaining areas of recommendation already identified for further development.

FOLLOW-UP TO RECOMMENDATION MEMO #4

Access to Data: Federal Policies and Standards

Lead Commissioner: Potok
Lead Staff: Martinez and Stefanik

Summary of discussion: During the April 3 meeting, Commissioners discussed establishing standards across the Federal government on issues identified as problematic to stakeholders. Commissioners discussed whether each item presented should be explicitly listed in recommendations by the Commission, with what priority and timing, and how the list squared with existing OMB priorities and authority. The items were: the current Federal-wide data inventory; metadata for key data files to enable data quality assessments and use of privacy enhancing technologies; data access agreements; technical data exchange; application, qualification and approval process; and data product release protections.

Next steps: Commissioners generally agreed to recommend that OMB and/or NIST lead a process by which each of the items should be standardized. They asked staff to sharpen the language, reorganize the items less around timing and sequencing and more around priority, with enhanced "top 20" data source metadata near the top. Staff should include the ideas that standards should be living documents, that agencies should use a "yes, unless" approach to making publicly available useful documentation, and that part of the processes that should be streamlined beyond researcher application and approval is disclosure review. Commissioners also indicated interest in saying more about potential standards content, especially about qualified researchers and data product release (e.g., differential privacy). That content will be developed in *Minimizing Risk* (Memo#10). They also asked for a little more information about the relationship between OMB and NIST and how they differentiate their roles to help nuance the discussion.

* * * * *

UPDATED RECOMMENDATIONS

The recommended standards must reflect the need to keep up with the complex, ever changing methodological, technological and privacy environments. Structuring them as living documents that can be updated to include emerging understandings and techniques will allow them to be issued more quickly and to remain relevant.

This list of recommendations will continue to be refined as the memo process goes on in order for the standards to be consistent with other Commission recommendations.

❖ **Finding:** The Commission finds that setting standards for increasing data access, use, and quality are necessary to facilitate the continued generation of rigorous evidence about Federal programs and policies.

► **Recommendation 4.1:** The Commission recommends that the President direct OMB to fully leverage its existing *policy and standard-setting* authority in order to unleash the potential of the Government's current collection of data to build evidence.

►**4.1.1:** *For "priority" datasets, require the creation of comprehensive metadata* for those datasets identified by the Commission as having the most immediate potential to boost evidence-based policymaking. The standard would be sufficient to allow assessing quality and implementing privacy-enhancing data matching technologies.

►**4.1.2:** *Adopt a "yes, unless" approach to*

a. Ensure agencies are fully complying with M-13-13 to complete public *inventories* of datasets, with an emphasis on those datasets not publicly available currently
b. Identify *existing* dataset *documentation* that could usefully be made public
c. Require improved *minimum metadata* of inventoried datasets to help the public learn how to access the datasets

►**4.1.3:** *Create government wide standards for*

a. A "common form" *interagency agreement template* to enable authorized data access and use
b. A single, streamlined *process* through which *researchers* external to the government may *apply, qualify, and are approved* to access sensitive government microdata not publicly available
c. Risk-based and quantifiable confidentiality *methods to protect against re-identification* in released data products (superseded by 10.2)

►**4.1.3:** *Consult with NIST on*

a. The most appropriate **data exchange standard** for use across Federal Departments and Federal grantees in support of evidence building, consistent with or useful for operational or administrative data exchanges.

These recommendations need to be implemented in a timely manner to ensure each is appropriately prioritized within the Federal government. Prioritizing #1, this standard should be set and implemented within in a year. For the remaining standards, OMB will provide a window for implementation, ranging from 6 to 24 months, depending on required resources.

OMB/NIST Relationship

The National Institute of Standards and Technology (NIST) is a part of the Department of Commerce. NIST does not itself set standards, the institute helps agencies and organizations develop, maintain, and retain standards that they want to set. NIST acts as a facilitator between the standard setter and the stakeholders the standard would affect. For the purposes of OMB's standard setting process on data exchange technologies, OMB would work with NIST to develop this standard, using NIST as a facilitator and to set up working groups of experts in the field of data exchange technology.

DRAFT

RECOMMENDATION MEMO #5

Increasing Access to State-Collected Administrative Data

Staff Lead: Anne Fletcher

ISSUE: What approach should the Commission recommend to increase access to state-collected administrative data about Federally-funded programs for statistical, evaluation, research, and policy analysis purposes?

Many high-value administrative data associated with Federally-funded programs are collected and maintained at the state level. The Commission will need to consider what, if any, recommendations it would like to make related to increasing the availability of state administrative data for statistical, research, evaluation, and policy analysis (SERA) purposes, including as part of recommendations about a data facility.

This memo provides background on current state data quality and access issues, and introduces the first two decision points where the Commission may wish to make recommendations: 1) recommending a data sharing *requirement* versus incentivizing increased state administrative data access, and 2) identifying the appropriate mechanism by which to make access to state administrative data more readily available for SERA purposes. Note that this memo assumes the continued presence of CARRA, or a more robust data facility, as may be recommended by the Commission.

* * * * *

BACKGROUND AND ANALYSIS

A large array of Federal programs are administered at the state level, in collaboration with states or directly by states on behalf of the Federal government. Common examples of priority programs that operate in this model include Medicaid, the Supplemental Nutrition Assistance Program (SNAP), unemployment insurance (UI), and the Temporary Assistance for Needy Families (TANF) program. These programs have a broad variety of funding and administrative structures that include the Federal and state government as partners, and often include counties and grantees as partners, as well.[1] The administrative and financial management systems established for the administration of these programs are a valuable source of microdata for use in SERA activities, but there are

[1] The Federal government funds the full cost of SNAP benefits and reimburses the state for approximately 50 percent of their administrative costs. Medicaid is jointly funded by the Federal and state government, where the Federal government pays states for a specified percentage of program expenditures calculated as the Federal Medical Assistance Percentage (FMAP). TANF operates as a block grant to states, and the program may be state or county administered, depending on the structure of the state.

numerous, well-documented barriers to accessing and using these data.[2,3,4,5,6] Not unlike the barriers faced directly by Federal agencies, frequent barriers at the state-level include: 1) the utility or quality of state administrative data sources for the purpose of research and evaluation, 2) lack of capacity within states to transform the administrative data to make it suitable for use in SERA purposes, 3) state legal barriers that either expressly prohibit or tightly restrict data sharing for the purposes of SERA activities (including state interpretations of Federal law), and 4) administrative and/or procedural variations for accessing data across different state agencies, or even among different programs administered by the *same* state agency, leading to protracted and often duplicative administrative requirements. These barriers are compounded for projects in which multiple administrative datasets are intended to be linked, as the path to access each individual administrative dataset can be unique.

What state administrative data sources are we talking about?

For the purposes of this memo, the state administrative data sources prioritized by the Commission include:

- Unemployment Insurance (UI) Wage & Payment Data
- Medicaid
- Children's Health Insurance Program (CHIP)
- Vital Records
- Supplemental Nutrition Assistance Program (SNAP)
- Head Start
- Child Welfare
- Temporary Assistance to Needy Families (TANF)

These programs each include their own unique histories, interactions between the Federal government and states, and funding mechanisms. Each program also operates under a unique statutory requirement for submitting program data to their Federal sponsor—ranging from the Medicaid program, for which states are required to submit microdata on all beneficiaries to CMS, to SNAP, under which states are required to submit samples or aggregate statistics only. Adding to the complexity of reaching a single decision about the options for increasing access to state data described below, each of these programs prioritized by the Commission provides mixed levels of support for IT systems and program administration to the states. For example, while Medicaid

---

[2] OMB. (2016). Barriers to Using Administrative Data for Evidence-Building. Washington, D.C.: Office of Management and Budget (OMB), Executive Office of the President.

[3] HHS/ACF/OPRE. (2016). Using Administrative Data in Social Policy Research, OPRE Report #2016-62. Washington, D.C.

[4] Northwestern University/University of Chicago Joint Center for Poverty Research. (1997). Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development.

[5] HHS/ACF/OPRE. (2015). Cheaper, Faster, Better: Are State Administrative Data the Answer? OPRE Report 2015-09. Washington, D.C.

[6] Maxwell, Kelly. (2017). Issues in Accessing and Using Administrative Data. HHS/ACF/OPRE Report 2017-24. Washington, D.C.

mandates data collection from states, the Federal government spends $2-3 billion per year to support state IT systems to enable the data transfers.

Data quality challenges

Recognizing the limitations of state-held administrative data about Federally-funded programs represents an important element of a conceptual cost-benefit analysis. One specific challenge in using state administrative data for SERA purposes is that these data were collected for program administration, not SERA purposes. Thus, in order to use administrative data for SERA activities, existing data must be transformed, potentially including the development of data documentation, which can be time intensive and require a deep knowledge of the administrative data collected.[7] Importantly, while data within a single state or program may be analyzed with relative ease, when data from multiple sites or states are linked for analysis, substantial cleaning and consistency edits must be made for comparability. Several publications explicate the challenges inherent in assessing the quality and utility of administrative data, and provide procedures for transforming the data for SERA purposes.[8],[9],[10]

In January, Robert Goerge (Chapin Hall) stressed to the Commission the value of state data *in situ* over data submitted *by* states to federal agencies for reporting purposes, as the data are often transformed, de-identified, or sampled when they are submitted by the states, and thus are less useful for analyses of groups across states. He noted that raw state administrative data can provide more contextual information about how programs are implemented. When asked about his reaction to the notion of directing states to submit complete administrative datasets to Federal agencies with clear reporting requirements, Mr. Goerge responded that this would reduce burden on states who wish to share their data, but this approach also creates more work for Federal agency staff who prepare the data for analysis. This perspective is not universal, however. David Mancuso from Washington State expressed concerns that less structured reporting requirements would not necessarily be less burdensome for states. The burden is only one aspect of the limitation, however. Leaders from the Workforce Data Quality Campaign informed Commission staff of some interest in receiving cross-state analyses of the data accessed by the Federal government, though such promises have been made historically that were never fully satisfied.

---

[7] Rothbard, A. (2015). Quality Issues in the Use of Administrative Data Records. In J. Fantuzzo and D. Culhane, (Eds.), Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government, pp. 72-103, Palgrave MacMillan.

[8] Goerge, R.M. and Lee, B.J. (2001). Chapter 7: Matching and Cleaning Administrative Data. In Studies of Welfare Populations: Data Collection and Research Issue, pp. 197-219, National Research Council.

[9] Iwig, W., Berning, M., Marck, P., and Prell, M. (2013). Data Quality Assessment Tool for Administrative Data.

[10] Hatry, H. (2015). "Using Agency Records" in Handbook of Practical Program Evaluation. Eds. K. Newcomer, H. Hatry, and J. Wholey. Fourth Edition. John Wiley & Sons. See pp. 325-342.

Challenges in making microdata available across different government entities
Another set of challenges for states relates to their willingness to make microdata data accessible to Federal agencies, or even to other agencies within their own state. The Federal government often provides direct assistance to states, either in the form of funding or technical assistance, to help build the capacity of their information technology infrastructure and/or staff capacity. These funds can frequently be used to improve the quality of and access to administrative data made available for analysis. The PostsecData Collaborative noted in comments that the lack of capacity and limited funding at the state and local levels has created challenges to collecting, processing, and using these data; the implication is that sustainable funding would help incentivize states to pursue these activities. However, Maria Cancian noted that even with sufficient technology and data systems, rules and policies that allow data sharing, and empowered governance that supports sharing, state agencies will still need cultures that recognize the benefits of sharing data with other entities. In September, Michael Basil identified the challenges in doing so within the state of Illinois, and highlighted his work to create an enterprise MOU to help address some of the cultural and legal interpretation issues that sometimes arise.

Expert testimony provided to the Commission reflects the importance of the Federal government engaging states as partners when seeking to access and make use of their administrative data as one strategy for encouraging increased data access for Federal agencies. For example, Mr. Goerge stressed that when states are not obligated to share their data, developing a partnership with state data providers is necessary to increase access to data. Successful partnerships may include participation of agency staff and providing states with an opportunity to receive study results before public release to comment or react to the findings. Ms. Cancian echoed a similar point in her presentation to the Commission, indicating that a key objective of data sharing is improving services and measuring performance. She noted that if the Federal government wants to use state data, they have to help build a system that is useful to the people who own and create those data so they have the opportunity to improve their own performance. The range of comments related to approaches to building capacity highlights the tension between either of two approaches: 1) investing in resources to support states in increasing their own internal capacity to collect, curate, and analyze their own administrative data, and 2) investing in resources that are designed solely to increase access to administrative data managed by states for the purposes of Federally-funded SERA activities. Consideration of this specific issue will be addressed in a subsequent memo.

Accessing state administrative data for the purposes of SERA activities
The challenges to accessing state administrative data are well documented and include the barriers described earlier in this memo, as well as state legal barriers that either expressly prohibit or tightly restrict data access for SERA purposes, confusion regarding the Federal statutes and rules regarding whether enabling data access is allowable, and administrative complexities throughout the data access process. Accessing state administrative data can be fraught, whether it occurs across programs operated by the same state agency (such as TANF and child welfare programs operating under the same human services agency), across state agencies (such as a State housing finance agency sharing data with the State Medicaid agency), or between states and Federal agencies. Several entities have

developed how-to guides[11], or documented the challenges to accessing administrative data for SERA purposes.[12]

Strategies that have been implemented to address identified barriers to increased data access

Federal agencies have implemented numerous strategies with the dual goal of improving the quality of State administrative data and cultivating the willingness of the state to enable access to administrative data by Federal agencies for SERA purposes. Experts have noted to the Commission that there is still a fair amount of confusion at the state level regarding what data states can make accessible, under what conditions, and with whom. Federal agencies themselves are often not clear, and clarifying the rules around data access and broadly sharing these rules with the states was acknowledged as a critical step. Both the Food and Nutrition Service (FNS) and the Administration for Children and Families (ACF) recently issued memos to state program directors clarifying what activities are permissible under current law in regards to making TANF and SNAP microdata available to CARRA.

Another approach involves providing resources and technical assistance to states, which enables states to modernize their information systems, serving the dual purpose of enabling states to improve their capacity for performance measurement and increasing their capacity to make high quality data available to the Federal government for SERA purposes. Examples of these kinds of investments include grants administered by the U.S. Department of Education to support the development of State Longitudinal Data Systems to track education outcomes of children from K-12, and investments made by CMS to support states in developing more efficient Medicaid eligibility and enrollment systems as well as Medicaid Management Information System (MMIS) claims systems.

A third approach that some Federal agencies have implemented amounts to a quid pro quo. One example is the Vital Statistics Cooperative Program, which is the relationship between the National Center for Health Statistics (NCHS) and the 57 jurisdictions that collect vital records. In exchange for data collected in a uniform manner, NCHS offers the states funding, training, and technical assistance. A non-financial arrangement includes the data analysis offered by CARRA to state programs that are willing to make their microdata accessible to CARRA. The benefit of the service offered by CARRA was illustrated by witness Erin Ulric, WIC Program Director of the state of Colorado, who described how the agency uses the data provided by CARRA to track program performance and identify populations for targeted outreach.

---

[11] Feeney, L., Bauman, J., Chabrier, J., Mehra, G., and Woodford, M. 2015. Using Administrative Data for Randomized Evaluations. J-PAL North America.

[12] Doshi, Jalpa A.; Hendrick, Franklin B.; Graff, Jennifer S.; and Stuart, Bruce C. (2016) "Data, Data Everywhere, But Access Remains a Big Issue for Researchers: A Review of Access Policies for Publicly-Funded Patient-level Health Care Data in the United States," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 4: Iss. 2, Article 8.

RECOMMENDATION OPTIONS

Decision Point 1: What approach should the Commission recommend to increase access to state-collected administrative microdata for SERA purposes about Federally-funded programs? Increasing access to administrative data collected by states for Federally-funded programs can essentially be accomplished in one of two ways: 1) requiring states to make their microdata in priority datasets accessible, or 2) incentivizing states to make their microdata accessible. Both options are used for some datasets and programs currently, and the benefits and limitations of each are discussed further below.

Option 1: Data Sharing Requirement. Require states to make administrative microdata available for "priority" Federally-funded programs, for SERA purposes.

Option Overview. This option would require states to provide access to microdata to either a data facility or appropriate Federal program agency, for data collected by states in programs that are fully- or partially-funded by the Federal government. Implementation of such a requirement could include (1) targeted statutory modifications to direct states to make data available for priority datasets, or (2) imposition of additional requirements through *existing authority,* such as directives in apportionments from OMB, regulatory provisions, or cooperative agreements, conditioning the provision of program funds on access to microdata. Either approach could be strengthened if combined with a prevailing statute or *notwithstanding* provision in statute that allows access to and use of administrative data for statistical purposes (see CEP Recommendation Memo #3).

Benefits of a Data Sharing Requirement. The most obvious benefit of a data sharing requirement would be the resulting availability of universe administrative microdata for the subset of programs identified as having a high value for SERA purposes. While states have historically objected to the one-size-fits all approach to Federal requirements related to program implementation, given differing demographic and geographic constraints for key programs, a requirement for data sharing may be comparatively simple using existing and evolving IT infrastructures. In many of the programs listed among the Commission's priorities, states already receive substantial funding to support IT systems and administrative overhead (e.g., SNAP, Medicaid, TANF), suggesting that a baseline infrastructure exists to enable a microdata sharing mandate to be implemented at a low or negligible cost in most, if not all, states for the priority programs. In fact, many states already submit administrative microdata, whether the universe of program participants or a sample, to their Federal funding agency either in response to a statutory or programmatic requirement. Some states also voluntarily provide administrative microdata to CARRA, so the cost and burden of a mandate would be imposed on a lower universe of states when compared to the status quo (e.g., half of states already share complete TANF microdata with ACF).

State Perceptions of a Data Sharing Requirement. Regardless of how such a data sharing requirement might be imposed, states would likely view this approach as a mandate. Mandates have long been imposed on states by the Federal government, often termed "unfunded mandates" when states are

required to take a new, specific action perceived as important to the Federal government, but without increases in grants-in-aid or other direct funding provided to states. Thus, even programs that are partially Federally-funded can include "unfunded mandates." Mandates have traditionally been opposed by many states due to perceived lack of flexibility in implementing programs, including those authorized by the Federal government and those that are partially funded by a combination of state and Federal funds (i.e., with matching requirements). In the 1970s and 1980s, concerns among states specifically increased due to a belief that insufficient funding was provided by the Federal government to support new, complex, and evolving requirements, including costly mandates imposed in major welfare and environmental reforms during that period.[13] In response, Congress and President Clinton enacted the bipartisan Unfunded Mandates Reform Act of 1995 (UMRA), which explicitly sought to recognize that mandates can impose substantial burden on states. The legislation required prospective analyses of mandates prior to issuing major regulations and or considering bills in Congress, but did not technically limit the actual use of the policy mechanism.[14]

Philosophically, some consider unfunded mandates as an inefficient directive imposed by the Federal government on subnational governments.[15] Many state agencies would likely not be supportive of a data sharing requirement, even if such a requirement was accompanied with support to comply with the requirement. The National Association for State Workforce Agencies (NASWA) argued that "states should not be coerced to provide data through national or regional clearinghouses by way of conditional grants or other means; rather, the governance structure, technology framework, quality of and ease of access to data, funding environment, and [technical assistance] environment should provide the necessary incentives for governmental partners to participate."

In practice, many Federal programs combine mandates with grants-in-aid for program implementation, an aspect of "cooperative Federalism." For example, while compulsory state Medicaid programs also mandate the transmission of data to the Federal government,[16] the programs are funded through a mixture of Federal and state expenditures, with the Federal government funding at least half of the Medicaid expenditures up to a ceiling of 87 percent.[17] With regard to funding for Medicaid data, the Federal government funds that vast majority of the costs associated with state IT systems to support data transfers.

If a data sharing requirement were to be implemented, the provision of funding would lessen the ill

---

[13] Atwood, T. (1994). Home Rule: How States are Fighting Unfunded Federal Mandates. Washington, D.C.: The Heritage Foundation.
[14] Specifically UMRA requires legislation with unfunded mandates to be identified by the Congressional Budget Office, and for Federal agencies to consider such mandates in issuing new regulations.
[15] Additional background on unfunded mandates available in Dilger, R.J. and R.S. Beth. (2016). Unfunded Mandates Reform Act: History, Impact, and Issues. Washington, D.C.: Congressional Research Service.
[16] See 42 U.S.C. 1396b
[17] HHS. (2017). Medicaid Financing and Reimbursement. Washington, D.C. https://www.medicaid.gov/medicaid/financing-and-reimbursement/

will associated with the requirement, though will not eliminate it altogether. The very direction to provide data to the Federal government for research and evaluation purposes across a wide range of programs will likely be met with resistance by some, if not the majority of, states. Previously staff from the National Governors Association (NGA), for example, told CEP staff that their members generally opposed both funded and non-funded mandates. Because of this perspective, issuing a requirement may create the unintended consequence of delaying other evidence-based policymaking practices in states that view the Federal government's requirements as overreach, shunning further efforts to expand evidence-based policymaking in those states.

Further, the creation of a data sharing requirement for priority datasets would not be homogenous in implementation in instances where state funds also support data needs. Existing literature on cooperative Federalism suggests heterogeneous effects across states and policy domains for how states alter behaviors for use of their own funding in response to additional Federal funds – some states increase (crowd-in) while others displace existing resources (crowd-out). This suggests states that already allocate funds to support some data sharing with CARRA may shift state funds in the presence of a funded requirement, rather than entirely boosting funding.

Limitations of a Data Sharing Requirement. A data sharing requirement may be perceived as more onerous than implementation would actually necessitate and, therefore, may not have the force of action anticipated. Practically speaking, for most programs the threat of rescinding funding as a consequence of not sharing data is unlikely to result in a loss of Federal funding due to political consequences and negative public perception that would ensue. Instead, the inclusion of a data sharing requirement enables the Federal government to approach the issue with states with a heavier hand, who have an interest in complying with applicable laws to avoid adverse consequences in other areas of program implementation. Given the challenges in implementing such a requirement, the potential political costs to doing so can be substantial, including breeding distrust between Federal and state agencies, and potentially creating perverse incentives to address data quality or other standards that established for consistency and standardization (see CEP Recommendation Memo #4). This effect may be compounded in programs like the Longitudinal Employer-Household Dynamics (LEHD) program, which currently operates successfully in a purely voluntary and cooperative basis between the Federal government and states.
Implementation of a Data Sharing Requirement. Mechanically, a data sharing requirement may be executed through legislation, regulation, or even in cooperative agreements negotiated with states periodically. Thus, the requirement itself may originate from either the Congress or the Executive Branch (or both). To the extent the Commission wishes to pursue a data sharing requirement, it may be preferable to articulate a desire to do so without delineating a single specific strategy, allowing political negotiations to identify trade-offs and potential support mechanisms in areas beyond the Commission's scope.

Implementation of a data sharing requirement with new statutory language establishing clear authority to do so would be the simplest implementation of a mandate. In the absence of explicit statutory authority, some state Attorneys General may challenge the legality of such a requirement

imposed through regulation or other means, imposing additional costs on the Federal government and delays in full implementation of a data sharing requirement. If the Commission chooses to recommend a data sharing requirement for states to make state-collected administrative microdata available for priority Federally-funded programs, companion recommendations related to implementation should be considered as part of this recommendation. In comments, NASWA noted that their members report that resource limitations related to data access, funding, IT, and/or staff capacity impede or have even stalled research and evaluation activities in a large number of state workforce agencies. In fact, in commenting on proposed WIOA regulations, the National Governors Association joined NASWA in suggesting that states would need dedicated funding and Federal support to meet the evidence-building requirements of WIOA. Suggestions for modifying the existing data infrastructure in states included the provision of more predictable, coordinated and scalable funding support for the development of state longitudinal administrative data systems, and the development of electronic tools that deliver information in customer-helpful formats.

Option 2: Incrementalism. Incrementally improve from the status quo by *incentivizing* states to voluntarily make state administrative data available, except where requirements already exist in current law.

Option Overview. An alternative to imposing a data sharing requirement that would require states to make state administrative microdata available for priority Federally-funded programs is to encourage states to voluntarily make their data available, either by offering incentives or by articulating potential benefits to states. This option would build on existing voluntary relationships to encourage states to more quickly and readily make their data available for intended SERA purposes, without imposing new requirements to do so.

Existing Voluntary Relationships. Currently, many state agencies voluntarily submit some administrative data to CARRA, and a few states also provide access to microdata directly to their Federal funding agency. Over the past several years, CARRA used Title 13 authority[18] to conduct outreach to state agencies in order to access state-collected administrative data resulting from the operation of numerous programs.[19] As demonstrated by CARRA's current inventory, the Census Bureau's efforts have been moderately successful, though some obvious gaps remain in many Federally-funded programs in their inventory.

One aspect of CARRA's outreach involved Federal agencies working collaboratively with CARRA to issue statements encouraging states to share microdata with the Census Bureau for research and evaluation purposes. Several programs subsequently issued memoranda to states encouraging states to share data by clarifying rules or issuing guidance around the legality of such data sharing. For

[18] 13 U.S.C. § 6 - Information from other Federal departments and agencies; acquisition of reports from other governmental and private sources.

[19] See full administrative data inventory of CARRA as of 3/27/2017 here: https://www2.census.gov/about/linkage/data-file-inventory.pdf

example, both ACF and FNS sent memos to states and program administrators of the TANF and SNAP/WIC programs, respectively, encouraging states to share their microdata data with CARRA. CARRA, in return, provides technical assistance for sharing and offers participating states a modest amount of analytic support, enabling them to receive a set of analyses that can help them better understand their beneficiaries and trends in service for use in program operations.

Targeted incentives can be established or enhanced through either funding mechanisms or analytic support services to encourage voluntary participation in data sharing initiatives to a specified intermediary. For example, the Environmental Protection Agency (EPA) has similarly provided grants to states for more than a decade to encourage participation in the Exchange Network, the system through which EPA analyzes microdata from monitoring sites to make them publicly available. ACF provided Chapin Hall with pilot funding to explore an infrastructure for combining TANF data across multiple states and providing analytic services in exchange.

Implications of the Incremental Approach. The primary benefit of an incrementally improving, voluntary reporting structure is that states that desire services or support for research and evaluation can participate. Presumably these states self-select into a group of those most interested in pursuing evidence-based policymaking activities or perceive the infrastructure already exists to enable data sharing. Over time, as more states participate and perceive benefits of participation, an incremental, voluntary approach will likely benefit from policy diffusion as additional states opt to participate in the future.

However, exclusively voluntary reporting of state-level microdata inevitably means that universe files for Federal programs implemented in partnership with states will not become readily accessible, at least in the near term. Analytically, the absence of access to complete records means conducting national-scale analyses with incomplete cross-state data. This will impose some internal and external validity limitations – though we might suggest that gaining access to an increasingly greater share of the state-level data mitigates this limitation over time.

Implementation. Should the Commission adopt a recommendation to further incentivize and encourage states to voluntarily make their data available for SERA purposes, rather than recommend a data sharing requirement, except where currently required under law, the Commission may choose to recommend an enhanced set of incentivizes for states, such as increased funding through grants or additional administrative flexibilities for programs; the Commission could outline these concepts generally, leaving specific incentives to program offices. But, importantly, the range of strategies that have already been developed to address the ability and willingness of the states to share data with the Federal government could be expanded in support of this option. Notably, the magnitude and style of incentives will dictate the speed with which diffusion across states occurs, though enhanced incentives likely impose additional direct Federal costs.

Decision Point 2: Where within the Federal government should the states be directed to submit state-collected administrative data about Federally-funded programs? Regardless of whether the Commission recommends a data sharing requirement or a voluntary, incentive-based approach to increasing access to state-collected administrative data, there remains the question of *where* states should be directed to make their administrative data accessible to.

Option 1: States should be directed to make their administrative data accessible to the Federal agency that funds the program. Currently, several high priority state administrative data sources are already providing administrative data to their Federal funding agency in some form. For example, Medicaid data is required to be submitted to CMS, and states have the option to submit CHIP microdata to CMS as of 2010.[20] Vital records are shared with the National Center for Health Statistics (NCHS) as per a cooperative agreement that NCHS establishes annually with the 57 jurisdictions that collect vital records. Several agencies, however, are actively encouraging states to allow CARRA to access their administrative data, rather than the funding agency itself, including HHS/ACF (TANF data) and USDA/FNS (SNAP/WIC data). Program agencies may not have the capacity, nor the desire, to establish the necessary infrastructure for ingesting, securing, curating, and managing access to microdata, so requiring agencies to develop these capacities would require significant effort and capacity building. In addition, states may be less likely to comply with instructions to make microdata accessible to their Federal funding agency due to concerns that the data will be (or could be) used for monitoring and enforcement purposes, as opposed to, or in addition to, being used for SERA purposes.

Option 2: States should be directed to make their administrative data accessible to a Federal data facility. An alternative approach to directing states to make their data available to the Federal agencies that fund their programs would be to direct them to make their administrative data accessible to a data facility. As noted under Option 1, several programs already encourage and direct states to make their administrative data accessible to CARRA. The benefit of the central data facility is two-fold. For the agency that lacks the capacity to house and manage the administrative data generated by their programs, CARRA provides a valuable service. States may be less hesitant to make their data accessible to a data facility, which would have the clear purpose of collecting data exclusively for SERA purposes, thus reducing the state's concern that their data might be used for monitoring and enforcement purposes. A second benefit of the central data facility is the availability of a secure environment that could support the linkage of multiple administrative data sources for the purposes of more robust SERA efforts. However, several complex data sources, particularly the health data sources, have long-standing data submission arrangements and procedures that could be disrupted should the Commission recommend that states solely make their administrative data available to a data facility. Under the status quo, the health data sources are made available to researchers, both within and outside of the Federal government via some mechanism.[21] Therefore,

[20] Camillo, C. A. (2012). CHIP Data in the Medicaid Statistical Information System (MSIS): Availability and Uses. Medicaid Policy Brief 12. Mathematica Policy Research.

[21] Medicaid data is available via CMS *and* CARRA, and vital records are available via NCHS.

it is not clear that requiring state agencies that already have an established submission regime that includes submitting data to their funding agency to instead make their data accessible exclusively to a data facility would be a more efficient or desirable approach to states.

Option 3: A flexible hybrid approach, in which administrative data sources are each assessed individually to determine whether it would be more appropriate to direct states to make their data accessible to the Federal funding agency *or* to a data facility. The discussion under options one and two lays out various pros and cons to adopting a wholesale approach in which states would be directed to make their administrative data accessible to one entity (the Federal agency that runs the program) or another (the data facility). Option 3 proposes a more flexible hybrid approach in which each high priority administrative dataset is reviewed independently to determine whether making the administrative data accessible to the program agency versus the data facility would be the most reasonable approach. An advantage of the hybrid approach is that it would build upon data exchange procedures that are already established, particularly those that are required in statute such as the transmission of Medicaid data directly to CMS. Adopting a hybrid approach would not preclude the Commission from advancing recommendations that strengthen the relationship between Federal agencies that have access to certain administrative datasets and either the data facility and/or the external research community.

COST IMPLICATIONS

This section is TBD.

FOLLOW-UP TO RECOMMENDATION MEMO #5

Increasing Access to State-Collected Administrative Data

Lead Commissioner: Meyer
Lead Staff: Fletcher and Hart

Summary of discussion: Many high-value administrative data associated with Federally-funded programs are collected and maintained at the state level. At the April CEP meeting, the discussion about recommendations to increase access to state-collected administrative data focused on whether the Commission should recommend that states be *required* to make administrative data about Federally-funded programs available to the Federal government for statistical, evaluation, policy research, and policy analysis (SERA) purposes, or if the recommendation should instead encourage *incentives* for states to voluntarily make their data available. Most commissioners leaned toward adopting a "shall" approach versus a "may" approach. However, many expressed concerns that a requirement could be ill-received by states and may generate a negative reaction that expands beyond this specific recommendation, potentially impacting the reception of the entire report. The Commission also discussed the question of *to what entity* states would need to make their data accessible to, either the Federal funding agency or a data facility, and the Commission agreed to adopting a flexible hybrid approach, in which administrative data sources are each assessed individually to determine whether it would be more appropriate to direct states to make their data accessible to the Federal funding agency *or* to a data facility.

Next steps: Commissioners generally agreed to recommend encouraging incentives to increase states' willingness and capacity to make their administrative data available to the Federal government (the "may" option), while simultaneously identifying a subset of high-value programs for which to recommend a requirement for states to make their administrative data about Federally-funded programs available to the Federal government (the "shall" option). In order to further refine the proposed recommendations, this memo seeks to do three things. First, the memo seeks to provide sufficient background information on the current data collection, reporting requirements and accessibility of state-collected administrative microdata across a set of high-value data sources in order to enable the Commission to decide which, if any, specific programs should be included in the subset for which the Commission would recommend instituting a requirement for data access. Second, the memo describes the kinds of supports and incentives that the Commission might recommend be implemented to encourage either compliance with a requirement for states to make data available to the Federal government, or to *incentivize* the voluntary action of making data available to the Federal government in cases where a *requirement* is not established. Finally, the memo will propose a possible set of recommendations related to smoothing and expanding access to state-collected administrative data in cases where the data may already be held by a Federal agency, but simply is not readily accessible for SERA purposes.

REVISED BACKGROUND and ANALYSIS

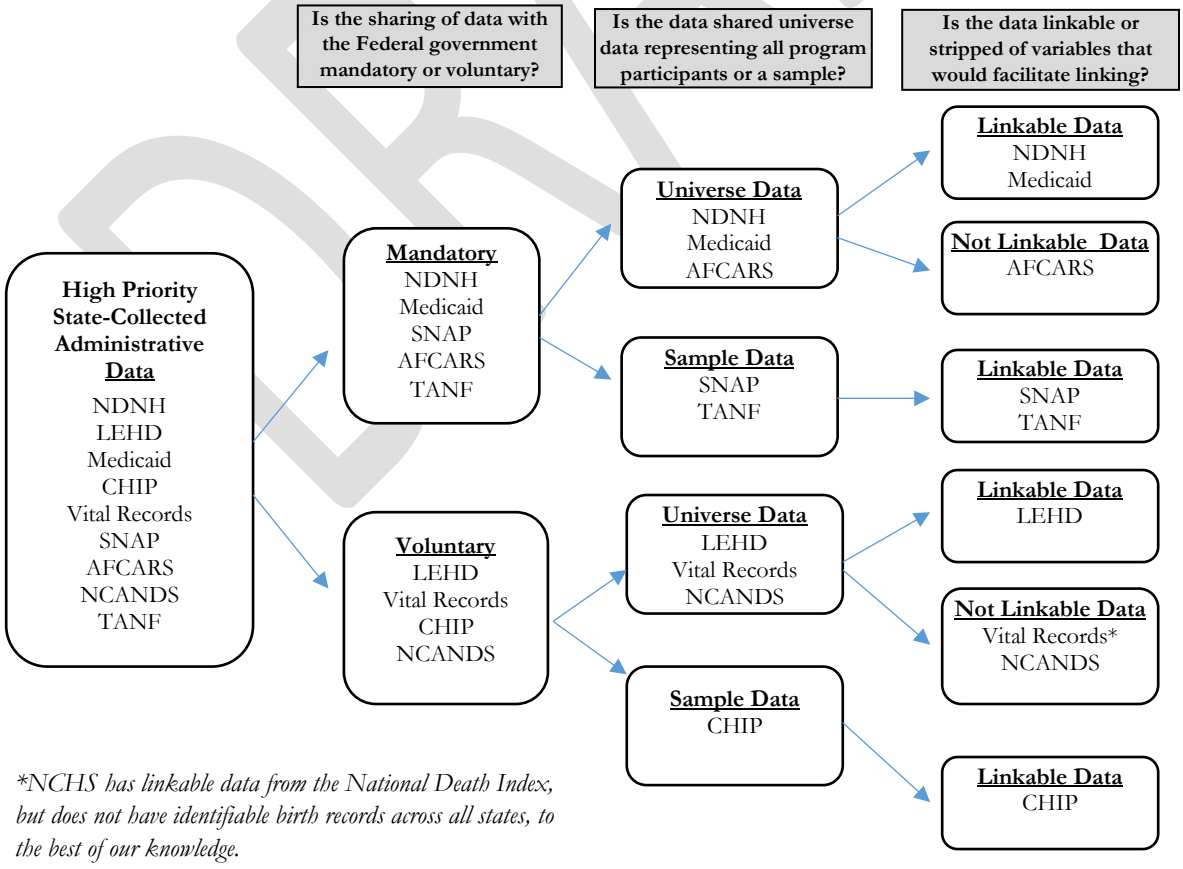
Because each program operates under a unique statutory authority and typically has both statutory and regulatory limitations and requirements regarding what data states may or must collect, the extent to which the Federal Government can require submission of the data from the state, and use

DRAFT

limitations for the data that is collected, each data source must be reviewed independently to assess the feasibility and practicality of requiring states to make their administrative data more accessible to for statistical, evaluation, policy research, and policy analysis (SERA) purposes. Table 1, which includes the set of high priority state-collected administrative data under consideration within this memo, has been expanded to now include information on the extent to which the data submitted to the Federal Government is linkable (does it contain identifiable enabling linkage to other data sources), a brief description of the incentives currently being offered to states by the Federal Government, for those programs in which the submission of universe microdata is voluntary, and a brief description of the various pathways via which a researcher might access these data sources currently.

For purposes of clarification, the administrative data of interest, and what is being discussed throughout this memo, relates to universe (meaning that it includes all program participants or beneficiaries, not just a sample) microdata that has sufficient personal identifiers to enable linking to other data sources. Figure 1 offers a way of thinking about the current landscape of data availability at the Federal level of a set of high priority state-collected administrative data sources by illustrating where the data sources diverge based on three important distinctions: 1) whether the exchange of data from the state to the Federal Government is mandatory or voluntary, 2) the extent to which the data being shared is the universe of program participants or beneficiaries, or just a sample, and 3) where the data shared contains sufficient personal identifiers to enable linking to other data sources.

Figure 1. High Priority State-Collected Administrative Data- What Does the Federal Government Currently Have?



**NCHS has linkable data from the National Death Index, but does not have identifiable birth records across all states, to the best of our knowledge.*

UPDATED RECOMMENDATIONS

A significant focus of Commission recommendations overall relate to supporting the establishment of a culture of evidence within Federal agencies, including a series of recommendations designed to strengthen and institutionalize the statistical, evaluation and policy research functions within Federal agencies with appropriate support. Overall, the Commission finds that state-collected administrative data is of high value for SERA purposes, and that there are numerous steps that could be taken by the Federal government to both *increase the volume* of linkable, universe administrative microdata that is accessible to the Federal Government, and to *broaden the availability* of the data to qualified researchers for the purpose of evidence-building.

Recommendation #1

Across the set of state-collected administrative data sources under review, Commissioners prioritized access to UI/wage and payment data. This is consistent with the volume of comments received by the Commission via hearings, testimonies, and public comment that also referenced the high value of increasing access to UI/wage records for SERA purposes. Thus, recommendation #1 offers options for increasing access to UI/wage records for the SERA purposes.

❖ **Finding:** Quarterly wage data is a valuable state-managed administrative data source for the purpose of research and evaluation. Accessing quarterly wage data directly from states can be challenging and time consuming, and in many cases not even possible, and some individuals may not be represented in the state-level data, including residents who work in a different state (e.g., federal employees, ex-service members, and postal workers). For a more complete picture of wage recipients, the access to quarterly wage data from a centralized source which allows access to multiple state records and/or a broader set of employees is critical. Currently, the Federal government already collects multi-state unemployment insurance and wage record data through two sources-- the National Directory of New Hires (NDNH), which is a mandatory collection required by statute, and the Local Employment Dynamics (LED) Partnership, which is a voluntary federal-state partnership. While both sources have existing authority to allow access to researchers for SERA purposes, this authority is very narrowly crafted, which makes access to the data challenging.

► **Optional Recommendation #1A**: The Commission recommends that the Congress and the President enact the necessary statutory changes that would expand access to NDNH employment and earnings data and LEHD unemployment and wage record data to "eligible researchers."[1]

► **Optional Recommendation #1B**: The Commission recommends that the Congress and the President enact the necessary statutory changes that would provide the U.S. Department of Labor (DOL) authority to require states to participate in a national system of UI/wage record data. Such statutory authority would additionally include SERA purposes as an allowable use of the data collected.

[1] See the *Protecting Privacy by Increasing Data Confidentiality* Memo #10 for the meaning of eligible researcher

National Directory of New Hires (NDNH)

The Federal Office of Child Support Enforcement (OCSE) operates the National Directory of New Hires (NDNH), a database established for the purposes of assisting state child support agencies in locating parents and enforcing child support orders.[2] The NDNH database consists of three separate files: 1) the new hire file, which contains information on all newly hired employees as reported by employers to each State Directory of New Hires (SDNH), 2) the quarterly wage (QW) file, which contains quarterly wage information on individual employees from state workforce agency (SWA) and Federal agency records, and 3) the unemployment insurance (UI) file, which contains unemployment insurance information on individuals who received or applied for unemployment benefits, as reported by SWAs. In the statute establishing NDNH[3], Congress authorized specific state and Federal agencies to receive information from the NDNH for authorized purposes. Researchers with the support of a state or Federal agency are also eligible to receive NDNH data, without personal identifiers, "to conduct research found by the Secretary of Health and Human Services (HHS) to be likely to contribute to achieving the purposes of part A (TANF program) or part D (child support program) of the Social Security Act."[4]

Because of the limited, yet somewhat undefined parameters for qualifying for access to the NDNH data as a researcher, some have noted that access to NDNH data can be uneven[5]. In addition to a very narrow pair of eligible research purposes, if and when a researcher is granted approval to use the NDNH data for a research project, there are some restrictions to working with the data.[6] One challenge is the statutory data deletion requirements, in which NDNH cannot retain data beyond eight quarters. A second challenge relates to the output received from NDNH under an approved match agreement, in that the output file returned to the researcher consists of de-identified records.

Despite the challenges posed by working with NDNH data, numerous researchers and many individuals who provided comments to the Commission,[7] including the membership of the Interagency Council on Evaluation Policy (ICEP), noted the potential value of this data source for SERA purposes. Most recently, the FY 2017 President's Budget request for the Administration for Children and Families (ACF), as the owner of the NDNH database, included a set of NDNH access proposals that would include expanding access to NDNH data for statistical and evaluation purposes, stating "this proposal grants NDNH access to specified federal statistical agencies and units (including evaluation offices), and their designees for statistical, research, evaluation, and performance measurement purposes associated with assessing positive labor market outcomes."[8]

---

[2] https://www.acf.hhs.gov/sites/default/files/programs/css/a_guide_to_the_national_directory_of_new_hires.pdf
[3] https://www.ssa.gov/OP_Home/ssact/title04/0453.htm
[4] 42 U.S.C. §653(j)(5)
[5] Unpublished memo from Abt Associates to the U.S. Department of Housing and Urban Development
[6] http://www.urban.org/sites/default/files/publication/25971/412688-Investigating-Alternative-Sources-of-Quarterly-Wage-Data-An-Overview-of-the-NDNH-LEHD-WRIS-and-ADARE.PDF
[7] Organizations and individuals who provided comment to the Commission supporting expanded access to NDNH include: Institute for Higher Education Policy (IHEP), New America, MDRC, Stephen A. Wandner, National Association of State Workforce Agencies, Workforce Data Quality Campaign and the PostSec Data Collaborative.
[8] https://www.acf.hhs.gov/sites/default/files/olab/final_cj_2017_print.pdf (pp. 303-307)

Local Employment Dynamics (LED) Partnership

Under the Local Employment Dynamics (LED) Partnership, which is a voluntary federal-state partnership started in 1999, states agree to share Unemployment Insurance earnings data and the Quarterly Census of Employment and Wages (QCEW) data with the Census Bureau. This data supports the Longitudinal Employer-Household Dynamics (LEHD) program, which combines administrative data, plus additional administrative data, and data from censuses and surveys to create a series of statistical products based on linked employer-household data.

Since 2000, the Census Bureau has entered into voluntary agreements with individual states, the District of Columbia, and eligible U.S. Territories (heretofore referred to collectively as "states"). At present, 54 states[9] have an agreement in place with the Census Bureau, which commits the entity to providing historical and ongoing administrative records on workers and employers to the Census Bureau. In return, the Census Bureau produces a longitudinal data infrastructure from which new data about the dynamics of local employment and the locations of jobs and workers can be produced.

Researchers seeking access to LEHD data, including the UI wage record microdata, must access the data through an FSRDC, and thus must comply with the requirements for conducting research at an FSRDC, specifically, the research project must aid in the mission of the Census Bureau. The data use agreement that states enter into with the Census Bureau includes an addendum agreement that offers the option for the state to allow their data to be used in external research projects. If the research project will include the release of state or sub-state-level specific results (as opposed to results from a group of states), the state will have the opportunity to review and approve the external research project. If the intent of the external researcher is to release pooled results only, states can choose to waive their opportunity to review each proposed project. If this option is not selected, even external projects that will release pooled results only would need to be reviewed and approved by the relevant states. Currently, only 12 states have selected the option that allows for blanket permission for the use of its confidential data for an external research project.

Individuals who provided comment to the Commission noted that these valuable data remain underutilized because of limitations to access, and expressed support for increasing access to the LEHD data.[10] Since the LEHD program is not authorized in statute, the Census Bureau relies on the voluntary participation of states in order to ensure that the data needed is available. These data use agreements with each state typically are in effect for a period of five years, requiring a continual renewal cycle in order to maintain full participation. Importantly, full participation is not guaranteed, as evidenced by the State of Wyoming exiting the program in 2015 upon the expiration of their agreement. The statutory authority governing the required disclosures under the Federal-state unemployment Compensation Program (20 CFR § 603.6) includes a list of Federal Departments to

---

[9] Among the 54 executed agreements, 49 of the states are included, with Wyoming as the only state that does not currently have an active data use agreement with the Census Bureau.

[10] Organizations and individuals who provided comment to the Commission supporting expanded access to LEHD include: Workforce Data Quality Campaign, PostSec Data Collaborative, JPAL, Donna Ginther, University of Kansas, and New America.

whom unemployment compensation data must be shared. As one option, this list could be expanded to allow the LEHD program to be included among the required disclosure purposes, thus ensuring complete coverage.

National System of UI/Wage Record Data

An alternative approach for making UI wage record data more accessible for SERA purposes would be the creation of a national system, into which states could be required to submit the same UI wage record data that they currently are mandated to provide to NDNH, and which the vast majority of states (all but Wyoming) currently make available to the Census Bureau for the purposes of the LEHD program. The establishment of such a system would require statutory changes to the Federal law that governs the UI program. Labor Economist Stephen Wandner submitted comments in response to the Commission's RFC in support of this approach, noting "In 1935, when the UI program was created as part of the Social Security Act, it may have made sense to leave interstate wage record exchange and data use for performance and research purposes to the states. It certainly does not make sense now, given voluntary state participation in the system and lack of support for using the data for research and evaluation purposes." There could be some resistance to this approach, however, based upon the public comments submitted by the National Association of State Workforce Agencies (NASWA), which commented "In general, state workforce agencies are not supportive of a single Federal clearinghouse, and support a more federated system with communications pathways at the national level to facilitate data sharing." Commissioner Abraham further noted that during a recent discussion she had with the NASWA Labor Market Information Committees, there was considerable sentiment that feeding data into a centralized system could mean that the data became divorced from the people who really understand them.

There could be a strong efficiency argument for such an approach, however. States have consistently expressed interested in being able to access interstate wage record and UI data for SERA purposes, and currently they have uneven access to such data, despite one mandatory reporting requirement (for NDNH) and several voluntary reporting efforts (including QCEW, LEHD, and the Wage Record Interchange System (WRIS)) that they may be part of. Creating a single system to which states would report UI/wage record data, would reduce the burden on states to report the same data to multiple entities, and would reduce the burden on Federal Agencies that spend a significant amount of resources negotiating agreements and MOUs with the states in favor of a single overarching arrangement. A single system would also streamline access to this highly valuable data source for SERA purposes, while potentially enhancing the security and privacy of the data through development of standardized procedures for data submission and access.

Recommendation #2

During the April CEP meeting, Commissioners expressed interest in the possibility of identifying a subset of high-value programs for which to recommend a requirement for states to make their administrative data available to the Federal Government. Recommendation #2 proposes two programs in which a statutory change requiring the submission of *universe*, linkable microdata may be most feasible given the current statutory requirement to submit *sample*, linkable microdata.

❖ **Finding:** The Federal government already has the statutory authority to collect a *sample* of linkable administrative microdata from states on beneficiaries/participants of both the TANF and

DRAFT

the SNAP programs. Efforts by CARRA to enter into data sharing agreements with states to access these data have been advancing slowly, with only 13 states sharing SNAP data, and eight states sharing TANF data. While not insignificant, the challenges of expanding the existing statutory authority to require the submission of universe data could be lessened significantly with the provision of supports to enable states to comply with a new requirement.

► **Recommendation #2:** The Commission recommends that the Congress and the President enact the necessary statutory changes that would require the submission by states of universe linkable administrative microdata in both the TANF and SNAP programs for SERA purposes, including any statutory changes required to address current use restrictions on both TANF and SNAP data.

Temporary Assistance to Needy Families (TANF)

The statutory language authorizing the TANF program[11] provides basic reporting requirements for states, including a requirement to submit a quarterly report containing disaggregated case record information on families receiving assistance. While the statutory language provides a set of required data elements to be included in the quarterly report, the set of required data elements are lacking any variables that would support linking to other datasets. The TANF final rule[12], however, expands upon the set of required data elements for the disaggregated case record information on families receiving assistance to include both SSN and DOB. The statutory language allows states to submit either universe data on all program beneficiaries, or to submit records on a "sample of families selected through the use of scientifically acceptable sampling methods approved by the Secretary." The statutory language further directs states to submit their TANF data report to the Secretary. As of April 2017, of the 54 states and territories reporting data to ACF, 31 entities submitted universe data, and the remainder of the states and territories submit sample data on a quarterly basis.

Currently, access to TANF microdata can only be achieved by approaching each state/territory individually to seek permission to access their data for a particular project. While ACF receives administrative microdata from states and territories on TANF recipients, there is not currently a mechanism by which ACF can make this data available to researchers. In 2013, ACF awarded a cooperative agreement to Chapin Hall to establish the Family Self-Sufficiency Data Center, where they had hoped to have Chapin Hall serve as a central location for TANF administrative data collected from the states, as well as other data sources relevant to measuring family self-sufficiency, but this effort has encountered significant difficulty in acquiring data and only a few states currently participate. In 2016, CARRA began contacting State Human Service Agencies requesting that states begin sharing their data with the Census Bureau in order to improve sampling procedures and sampling adjustment for various Census Bureau surveys, and ACF signaled their support for this effort by issuing a memo to Regional Program Manager encouraging data sharing. In addition to the agreements put in place between CARRA and the states themselves, ACF also provides to CARRA the TANF data submitted quarterly to the program office, as well. It is unclear if there are different

[11] https://www.ssa.gov/OP_Home/ssact/title04/0400.htm
[12] https://www.acf.hhs.gov/sites/default/files/ofa/1999_tanf_full_preamble_and_rule.pdf

use restrictions applied to the TANF data submitted by different sources, but currently, access to TANF data vis-à-vis CARRA is not facilitated.

Current law allows for states to submit sample data to comply with the reporting requirement, noting "A State may comply…by submitting disaggregated case record information on a sample of families selected through the use of scientifically acceptable sampling methods approved by the Secretary." Repealing the ability of states to submit sample data and requiring states to submit universe data would require a change in the existing statute. Furthermore, the statute requires quarterly reporting to be submitted directly to the HHS Secretary. If the statute were modified to require universe reporting, and reporting would continue to be directed to the program office, a mechanism to establish data access for SERA purposes would need to be established. In addition, a mechanism for increasing access to the universe microdata would need to be established, either via ACF or through CARRA or the proposed data facility.

Supplemental Nutrition Assistance Program (SNAP)

The statutory language reauthorizing the SNAP program, the Food and Nutrition Act of 2008[13], requires as a condition of eligibility that states must collect social security numbers for all SNAP applicants/beneficiaries. The statutory language states further that "the Secretary and State agencies shall have access to the information regarding individual supplemental nutrition assistance program applicants and participants who receive benefits…..but only to the extent that the Secretary and the Commissioner of Social Security determine necessary for purposes of determining or auditing a household's eligibility to receive assistance or the amount thereof under the supplemental nutrition assistance program, or verifying information related thereto." The language also includes a statutory use restriction for data collected by states under the SNAP program, allowing FNS to access and utilize the state SNAP data strictly for administrative or enforcement purposes.

FNS has a range of reporting requirements that states must comply with, primarily related to providing the agency with the ability to reimburse the states for program and administrative costs, auditing financial data to calculate error rates (for both overpayments and underpayments), and aggregate data on the characteristics of program participants. The majority of SNAP data submitted by states to the Federal government is in aggregate form, with the exception of the data collected and stored in the Quality Control System. States are required to conduct quality control (QC) reviews of their SNAP cases, and the Quality Control System maintained by FNS houses a sample of cases from each state. Each month, State agencies select an independent sample of participating SNAP households to submit to the Quality Control System. The annual required State samples depend upon the size of a State's caseload and generally range from approximately 300 to 1,200 reviews.[14] The administrative microdata submitted for this purpose is fairly extensive.[15]

Instituting a requirement for states to submit identifiable universe microdata would require a statutory change that would both require the submission of universe data rather than a sample, at some regular interval. In addition, the use restriction that exists in statue would also need to be

[13] https://www.fns.usda.gov/sites/default/files/snap/Food-And-Nutrition-Act-2008.pdf
[14] https://fns-prod.azureedge.net/sites/default/files/snap/2015-State-Activity-Report.pdf
[15] https://www.fns.usda.gov/sites/default/files/380-1-schedule.pdf

expanded to allow the data to also be used for SERA purposes. Finally, similar to the TANF program, the statute would need to be further changed to allow submission of the administrative data directly to the data facility, or a mechanism to establish data access for SERA purposes for qualified researchers would need to be established within USDA/FNS.

❖ **Finding:** Universe microdata is already available (somewhere in the Federal Government) but needs to be more accessible for SERA purposes.

► **Recommendation #3**: The Commission recommends that the Congress and the President direct Federal agencies that fund jointly administered Federal-state programs, and/or who have access to state-collected administrative data, to conduct a thorough review of the relevant statutes and regulations governing these collections to identify where the barriers exist to expanding access to administrative data for qualified researchers for SERA purposes.

Even when linkable universe microdata is being collected or procured from states by a Federal agency, researchers often face barriers to accessing the data for SERA purposes in a timely or efficient manner.[16]

Vital Records

National Center for Health Statistics (NCHS) is the federal agency legislatively mandated to produce national health statistics based on a cooperative, decentralized system in which data from more than six million vital-event records are collected each year by all states and U.S. territories and transmitted to NCHS for processing and dissemination[17]. NCHS has legislative authority and is mandated under 42 U.S.C. § 242k, Section 306(h) of the Public Health Service Act to collect vital statistics annually: "There shall be an annual collection of data from the records of births, deaths, marriages, and divorces in registration areas. The data shall be obtained only from and restricted to such records of the States and municipalities which the Secretary, in his discretion, determines possess records affording satisfactory data in necessary detail and form." Currently this data collection is limited to data from birth and death records (including fetal deaths), as NCHS discontinued the collection of individual-record marriage and divorce reports after 1995.

Data are collected through what is known as the Vital Statistics Cooperative Program. These data are provided through contracts between NCHS and vital registration systems operated in the 57 jurisdictions legally responsible for the registration of vital events – births, deaths, marriages, divorces, and fetal deaths. The contracts support the cost of training and technical assistance to help standardize data quality.

On November 20, 2007, the Division of Vital Statistics (DVS) and National Center for Health Statistics (NCHS) released a new policy on the release of and access to vital statistics microdata for

[16] Cheaper, Faster, Better: Are Administrative Data the Answer? The Mother and Infant Home Visiting Program Evaluation-Strong Start Second Annual Report. 2015. OPRE Report 2015-09.
[17] National Research Council. (2009). Vital Statistics: Summary of a Workshop. Michael J. Siri and Daniel L. Cork, rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

DRAFT

births, deaths, fetal deaths, linked birth/infant death, and matched multiple births. Effective with the 2005 data year, NCHS revised its microdata release and access policy to comply with state requirements, laws, and policies. *(From the National Vital Statistics System (NVSS) website: given changes in state laws and policies on confidentiality with respect to the re-release of vital registration data, NCHS has revised its micro-data release and access policy to comply with state requirements effective with the 2005 data year.)* This DVS revised policy reflects the dual goals to make data available as widely as possible while being responsive to concerns about confidentiality.

The current agreement with the states on the re-release of restricted data requires a review of all such data requests by the National Association for Public Health Statistics and Information Systems (NAPHSIS), which represents state vital registrars. The review by NAPHSIS is conducted prior to the NCHS review and includes both federal and non-federal requests for restricted data files. Researchers in Federal agencies, as well as their on-site or off-site contractors can submit project proposals that request exact dates of vital events. If needed, the file with exact date can also include geographic detail.

Non-federal researchers (including federal grantees) can gain controlled access to exact dates of vital event files only through the NCHS Research Data Center after NCHS approval.

Adoption and Foster Care Analysis and Reporting System (AFCARS)/ National Data Archive on Child Abuse and Neglect (NCANDS)

HHS manages several data systems that capture information about families engaged in the child welfare system, including two that appear on the list of high priority state-collected administrative data under consideration within this memo. The Adoption and Foster Care Analysis and Reporting System (AFCARS) is a statutorily required national data collection system[18] that provides comprehensive demographic and case-specific information on all children who are in foster care and adopted with title IV-E (child welfare) agency involvement (e.g. public funding). AFCARS also includes information on foster and adoptive parents. Child welfare agencies are required to report data to AFCARS two times per year.

The National Data Archive on Child Abuse and Neglect (NCANDS) Child File Dataset consists of investigations or assessments of alleged child maltreatment. In 1998, The Child Abuse Prevention and Treatment Act (CAPTA) was amended to direct the Secretary of HHS to establish a national data collection and analysis program, which would make available state child abuse and neglect reporting information.[19] HHS responded by establishing NCANDS as a voluntary national reporting system.[20] CAPTA requires states that receive CAPTA state grant funds to report annually—"to the maximum extent practicable," an in FFY 2015, 52 states (including the District of Columbia and Puerto Rico) submitted data to the NCANDS Child File.

---

[18] Section 479 of the Social Security Act

[19] Child Abuse Prevention, Adoption and Family Services Act of 1988, 42 U.S.C. §5101 et seq.; 42 U.S.C. 5116 et seq, (1988)

[20] U.S. Department of Health & Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. (2016). Child maltreatment 2014. Available from http://www.acf.hhs.gov/programs/cb/research-data-technology/statistics-research/child-maltreatment

Both AFCARS data and the NCANDS Child File Dataset are accessible at the National Data Archive on Child Abuse and Neglect (NDACAN)[21], however, researchers interested in accessing identifiable AFCARS or NCANDS microdata must access the data from the state child welfare agencies. While the AFCARS data system and the NCANDS Child File Dataset include case-level data, all identifiers that would enable linking are stripped from the dataset before it is transmitted to HHS/ACF and made available to researchers.

❖ **Finding:** Across the existing jointly administered Federal-state programs in which the submission of universe administrative microdata is statutorily required, the Federal government provides a significant amount of support to enable states to comply with reporting requirement. programs under which they are already required to submit universe administrative microdata to the Federal government.

❖ **Finding:** The provision of funding for IT infrastructure and technical assistance can assist states in building the necessary data collection systems and reporting structures to be able to comply with a new requirement for increased data collection and transmission, and serve the dual purpose of addressing issues of data integrity.

► **Recommendation #4:** The Commission recommends that the Congress and the President direct Federal agencies to explore and adopt the kinds of incentives that could be offered to support either compliance with a requirement for states to make data available to the Federal government, or to incentivize the voluntary action of making data available to the Federal government in cases where a requirement is not established.

Should the Commission choose to put forward a set of recommendations that would seek to *increase the volume* of linkable, universe administrative microdata that is accessible to the Federal Government, and *broaden the availability* of the data to qualified researchers for the purpose of evidence-building, the Commission may also choose to partner these with a recommendation designed to encourage either compliance with a requirement for states to make data available to the Federal Government, or to incentivize the voluntary action of making data available to the Federal government in cases where a requirement is not established.

Based on the number of efforts underway related to increasing state-level capacity for evidence-based policymaking[22], [23], there are strong signals that states themselves have an interest in increasing their own capacity. There is also some evidence that incentives provided by the Federal Government "stimulate the diffusion of policies through states," though "because of the interaction between states' desires and Federal demands, it is often difficult to untangle the federal incentive from the state's request for the Federal Government to help it do something it wants and needs to do."[24]

One approach for providing supports to states involves providing resources and technical assistance, which enables states to modernize their information systems, thereby serving the dual purpose of

[21] http://www.ndacan.cornell.edu/index.cfm
[22] https://www.aisp.upenn.edu/
[23] http://www.pewtrusts.org/en/research-and-analysis/reports/2017/01/how-states-engage-in-evidence-based-policymaking
[24] Welch, Susan and Thompson, Kay. 1980. The Impact of Federal Incentives on State Policy Innovation. American Journal of Political Science, Vol. 24, No. 4, pp. 715-729.

enabling states to improve their capacity for performance measurement and increasing their capacity to make high quality data available to the Federal Government for SERA purposes. Investment by the Federal Government at the outset of a new program or requirement can provide the impetus needed for states to begin the process of acquiring or updated new IT systems that better enable more robust and/or secure data collection and transmission activities. One such effort identified as highly successful by numerous commenters[25] to the Commission is the State Longitudinal Data System (SLDS) Program.[26] The SLDS grant program is administered by the National Center for Education Statistics (NCES), and provides grants to states to develop state-level longitudinal data systems designed to "enhance the ability of States to efficiently and accurately manage, analyze, and use education data, including individual student records." As noted by the Data Quality Campaign in their public comments to the Commission, "Many states had made some progress in building education data systems by 2009, but the infusion of federal dollars from the SLDS Grant Program was critical to securing state policymakers' interest and helped them move their systems from an emerging tool to a robust source of information. States responded to this "seed funding" by increasing their own investments to ensure long-term sustainability. While just a handful of states were funding their systems in 2009 when the bulk of the federal grant funds were distributed, 41 states were funding their data systems by 2014."

An additional benefit of assisting states financially in improving data collection processes is the ability for states to invest resources into improvements in data integrity, in turn improving the quality of the data that is made available to Federal agencies for SERA purposes. In response to the Commission's Request for Comment, Casey Family Programs noted: "Ensuring the quality, comparability, and accessibility of each contributing system's data is a necessary first step in any use of cross-system data. A high proportion of missing data or errors in data entry can undermine attempts to accurately determine whether a particular program is effective. State and local program administrators may require fiscal and technological resources, as well as technical assistance, to address issues of data quality."[27]

Recommending that Federal agencies that fund jointly administered Federal-state programs to consider the various tools that might be available to assist states in improving their data quality and to increase access to valuable state-collected data for SERA activities should also be viewed as a necessary step in establishing a culture of evidence within Federal agencies.

DRAFT

[25] Organizations and individuals who provided comment to the Commission describing the positive impact of the SLDS grant program include: Data Quality Campaign, National Association of State Workforce Agencies, Ohio Department of Jobs and Family Services, Washington State Employment Security Department, and RTI

[26] https://nces.ed.gov/programs/slds/about_SLDS.asp

[27] Public Comment submitted by Casey Family Programs

Table 1: High Priority State-Collected Administrative Data: Subset From the Top 20/Priority Data Sources Identified by the Commission/8

Data Source	Federal Funding Agency	How do Federal funding agencies currently receive administrative microdata from states or grantees?/1		Does Census have microdata? If yes, for how many states?/2	What percentage of the program is funded by the Federal government?	Is the data submitted linkable or not linkable?	If data is being shared voluntarily, what is the incentive being offered by the receiving agency?	How accessible is the data currently to outside researchers (either Federal agencies or non-Federal qualified researchers)?
		Mandatory/ Voluntary	What is available?		(None, limited, most, full)			
Unemployment Insurance/Wage & Payment Data								
National Directory of New Hires (NDNH)	HHS/ACF	Mandatory (42 U.S.C. §653)	Microdata from all states	No	most /4	Linkable, but output is deidentified, prohibiting further linking	N/A	Researchers seeking access to any of the three datasets comprising the NDNH (new hires, quarterly wages, and UI), must apply to HHS/ACF/OCSE and complete an application that satisfies the requirement that the project to be undertaken will likely contribute to achieving the purposes of Part A or Part D of the Social Security Act. (Social Security Act 453 (j)(5))
Longitudinal Employer–Household Dynamics (LEHD)	Census	Voluntary	MOUs with 49 states	Yes, 49 states	limited	Linkable	States voluntarily share their UI wage data with CES for the purposes of the LEHD under a voluntary data use agreement. No money changes hands, but states benefit from the products developed under LEHD.	Within the data use agreements, states can choose whether or not to allow CES to release their earnings data to qualified researchers or not. Currently, 16 states allow Census to release of the data; for the remainder, the researcher would need to approach the state for permission.
Quarterly Census of Employment & Wages (QCEW)	DOL/BLS	Voluntary	Cooperative agreements with 53 states/territories	No	most	Not linkable, as the record level is an establishment, not an individual	States are provided funds under a cooperative agreement to provide their data to BLS for use in the QCEW.	States produce establishment-level data built from individual wage records, and then submit the establishment-level data to BLS under the QCEW program. Individual-level wage records are not transmitted by the states to BLS through this program.
Health Data								
Medicaid	HHS/CMS	Mandatory (Section 4735 of the Balanced Budget Act of 1997; Section 6504 of the Affordable Care Act)	Microdata from all states on all beneficiaries	Yes, all states	most	Linkable	N/A	Researchers interested in accessing microdata on Medicaid beneficiaries can access data via CMS directly or may choose to approach the relevant state Medicaid agencies. To date, researchers have not sought access to Medicaid data from CARRA.
Children's Health Insurance Program (CHIP)	HHS/CMS	Voluntary	Microdata from 8+ states/7	Yes, 8+ states	most	Linkable	No incentive offered.	Researchers interested in accessing microdata on CHIP beneficiaries must access the data through the state Medicaid agency, unless the state is one of the 8+ states that administer an M-CHIP program. Data from these states can be accessed via CMS directly.
Vital Records	HHS/NCHS	Voluntary	Contracts with 57 jursidictions	No	limited/5	Linkable	Vital records are provided through contracts between NCHS and vital registration systems operated in the 57 jurisdictions that compile vital records. The contracts support the cost of training and technical assistance to help standardize data quality.	Vital records data can be accessed through NCHS or by seeking to enter into a data sharing agreement with the relevant state vial records agency.
Human Services Data								
Supplemental Nutrition Assistance Program (SNAP)	USDA/FNS	Mandatory (7 U.S.C. §2025)	A sample is drawn monthly from all states	13 states (CO, HI, ID, IL, IN, KY, MD, NJ, NY, OR, TN, TX, VA)	most /6	Linkable	CARRA matches program data submitted to Census data and returns data tables and a visual aid back to states	Researchers interested in accessing microdata on SNAP beneficiaries can access data for 13 states through CARRA, or may choose to approach the relevant state agency that operates the SNAP program. Currently, FNS does not have a mechanism to share microdata with researchers for SERA purposes.
Head Start	HHS/ACF	--- /3	N/A	No	most	N/A	N/A	N/A
Adoption and Foster Care Analysis and Reporting System (AFCARS)	HHS/ACF	Mandatory (42 U.S.C. §679)	Microdata from all agencies, every 6 months	No	most	Not linkable, case-level data is stripped of identifiers prior to being submitted	N/A	Researchers interested in accessing identifiable AFCARS microdata must access the data from the state child welfare agencies. While the AFCARS data system includes case-level data, all identifiers that would enable linking are stripped from the dataset before it is transmitted to HHS/ACF.
National Data Archive on Child Abuse and Neglect (NCANDS) Child File Dataset	HHS/ACF	Voluntary	Microdata from all agencies, annually	No	most	Not linkable, case-level data is stripped of identifiers prior to being submitted	ACF provides IT support and extensive technical assistance to states to support the states' voluntary compliance.	Researchers interested in accessing identifiable NCANDS microdata must access the data from the state child welfare agencies. While the NCANDS data system includes case-level data, all identifiers that would enable linking are stripped from the dataset before it is transmitted to HHS/ACF.
Temporary Assistance to Needy Families (TANF)	HHS/ACF	Mandatory (42 U.S.C. §611)	Roughly 50% of states submit sample data, and the remainder submit universe microdata	8 states (ID, IN, KY, MD, NJ, NY, TN, WI)	most	Linkable	CARRA matches program data submitted to Census data and returns data tables back to states.	Researchers interested in accessing microdata on TANF beneficiaries can access data for 8 states through CARRA, or may choose to approach the relevant state agency that operates the TANF program. While ACF receives universe data for 31 states, ACF does not have a mechanism to share TANF microdata with researchers for SERA purposes.

1/ Note that for data submissions that are categorized as voluntary, the actual holdings and availability of data may vary over time

2/ See full administrative data inventory of CARRA as of 3/27/2017 here: https://www2.census.gov/about/linkage/data-file-inventory.pdf

3/ Head Start data prohibited from collection by the Federal government

4/ NDNH is funded from the Child Support program, for which the system is mostly Federally funded; the system includes UI data, which is primarily state funded

5/ NCHS provides funding to jurisdictions to support data systems and standardization, as well as data sharing with NCHS

6/ SNAP benefits are fully Federally funded; administrative costs are split between the states and Federal government

7/ CHIP programs that operate as a Medidaid expansion (a choices selected by 8 states, 5 territories and DC, as of 2015) report their CHIP enrollment data in full along with their Medicaid data in MSIS

8/ Currently this table includes only data sources in which the Federal government funds some or all of the program under which the data is collected. The Commission may wish to consider strategies to access state-collected data for programs of high value, but where the Federal government is not making a financial contribution, such as the Workers' Compensation program.

RECOMMENDATION MEMO #6
UPDATED AFTER APRIL 24 CALL

Bans on Data Collection and Research
Lead Commissioner: Shea
Lead Staff: Hitt and Stefanik

ISSUE: In order to improve access for evaluation and evidence generation, to what extent, if any, should the Commission address the several statutory bans on data collection and research? If so, to what extent can setting a high bar for the use of a ban provide a solution?

Several statutes explicitly prohibit the assembly of datasets, or in some cases the use of data for evidence-building. As part of the Commission's charge to seek the "optimal arrangement," the Commission must determine what, if any, recommendations it will make with respect to bans on the assembly or use of data. This memo is one element of the response to *Sec. 4(a)(2)* of the CEP statute.

Relates to Legal (Memo #3) and Minimizing Risk (Memo #10).

* * * * *

BACKGROUND AND ANALYSIS

Certain policy domains include specific direction in law to *not* assemble datasets or conduct analysis on specific issues. Regardless of their merit, such prohibitions inhibit evidence-building about public policy and Federal programs. For decades, different iterations of evidence-prohibiting provisions have been included by either the Congress or the Executive Branch in directives, policies, and statutes. Many of these prohibitions on collecting or developing evidence relate to important policy areas of national significance, and often involve programs of substantial public investment. This practice is inherently in conflict with the Commission's charge to identify the "optimal arrangement" for evidence building and routine integration into program administration. Within the broad array of factors that influence policymaking, bans may be appropriate tools for achieving certain goals or setting priorities, but in the context of evidence building a ban on data collection or use cannot be reconciled with a goal of increased reliance on evidence to inform policymaking.

Through its discovery phase, several examples of data collection bans were raised to the Commission, and there are likely other such prohibitions in current U.S. law, which were not identified during the Commission's fact-finding phase. This paper explores several examples brought before the Commission in comments or testimony, but is not intended to be an exhaustive catalogue of all limitations. Of the bans raised, no other issued received as much focus as the student unit record ban.

Student Unit Record (SUR) Ban. The SUR ban prohibits any new Federal databases that include personally identifiable information on postsecondary students.[1] The statutory SUR ban includes provisions that explicitly limit the Federal government from collecting or maintaining data about students, with limited exception:

[1] 20 U.S.C. 1015c

> **(a) Prohibition.** Except as described in subsection (b), nothing in this chapter shall be construed to authorize the development, implementation, or maintenance of a Federal database of personally identifiable information on individuals receiving assistance under this chapter, attending institutions receiving assistance under this chapter, or otherwise involved in any studies or other collections of data under this chapter, including a student unit record system, an education bar code system, or any other system that tracks individual students over time.
>
> **(b) Exception.** The provisions of subsection (a) shall not apply to a system (or a successor system) that-
>
> (1) is necessary for the operation of programs authorized by subchapter II, IV, or VII; and [2]
>
> (2) was in use by the Secretary, directly or through a contractor, as of the day before August 14, 2008.
>
> **(c) State databases.** Nothing in this chapter shall prohibit a state or a consortium of States from developing, implementing, or maintaining State-developed databases that track individuals over time, including student unit record systems that contain information related to enrollment, attendance, graduation and retention rates, student financial assistance, and graduate employment outcomes.

The SUR ban was enacted in 2008 as a direct response to concerns about a plan from the National Center for Education Statistics (NCES) in the Department of Education to design a modern replacement for the cumbersome Integrated Postsecondary Education System (IPEDS), enabling evaluation of trends across institutions. Proponents of the ban publicly suggested that it would ensure protection of student privacy, but some higher-education organizations also lobbied for the ban.[3] Privacy was, in this case, a more convenient argument for supporting the ban than acknowledging that this system, if created, could be used to hold the higher education sector more accountable. Representative Virginia Foxx (R-N.C.) successfully amended the Higher Education Act reauthorization in 2008 to include the ban, while Representative Heath Shuler (D-N.C.) successfully amended it to exclude statewide systems. Ironically, the ban may well create greater privacy risk because the student level data already exist, but are now maintained across a variety of likely less-well protected state and private sector systems.

While the SUR ban amended the Higher Education Act, its implementation applies across both postsecondary education and Pre-K–12. Combined with similar bans covering data on participants in Head Start and programs authorized by the Workforce Innovation and Opportunity Act (WIOA), much of the government's investment in education programs—from young children through adulthood—are largely not subjected to the type of rigorous, national-scale evaluation that could be possible by leveraging administrative data. These programs collectively account for tens of billions in annual Federal spending.[4]

---

[2] Note exceptions apply to subchapters II -- Teacher Quality Enhancement, IV – Student Assistance (e.g., Pell Grants), and VII – Graduate and Postsecondary Improvement Programs. Non-excluded subchapters include: III – Institutional Aid, V – Developing Institutions, VI – International Education Programs, and VII and IX which include miscellaneous other programs.

[3] McCann, C. and A. Laitinen. (2014). College Blackout: How the Higher Education Lobby Fought to Keep Students in the Dark. New America.

[4] For example, in 2016 the Federal government obligated $9 billion for Head Start and $177 billion in postsecondary education and training activities, though a portion of the latter is covered by the exemptions referenced in Footnote #2 (e.g., $28 billion for Pell Grants). See OMB. (2016). President's FY 2017 Budget. Washington D.C.: Office of Management and Budget.

Concerns regarding privacy and confidentiality are often at the heart of the publicly articulated motivations behind support for data bans, and specific concerns about the capability to secure information are common objections to various databases. As noted above, the SUR ban, while framed on privacy concerns, was motivated in part by fears of increased accountability in part of the higher education sector.

A ban on collecting data, creating a dataset, or analyzing data is a blunt instrument. A ban entirely prioritizes privacy, or at least the perception of privacy, over any value or insight such information might provide. Likewise, bans on the collection and use of data for specific areas of research exclusively prioritize assumptions about outcomes over conclusions based on sound research and evaluation. In both cases, such actions are antithetical to the goals of evidence-based policymaking, which at its core encourages objectivity. That said, existing bans are the result of the culmination of public opinion, real and perceived risk of harm, and explicit choices of policy-makers.

The statutory SUR provision, when parsed, explicitly prohibits the creation of a "federal database of personally identifiable information" of individuals directly or indirectly receiving assistance, or otherwise involved in studies authorized under this act (Higher Education Act). The SUR ban does not prohibit data from existing or being retained, but does effectively prohibit those data from being assembled into a database held by the Federal government, thus rendering the value for statistical, evaluation, research or analysis (SERA) purposes largely moot.

Through various inputs, the Commission heard more about this topic than any other topic in scope. Via the Request for Comments, 65 percent of comments received were attributed to concerns about student records, with the majority of those comments in opposition to overturning the SUR ban or some form of an equivalent outcome. Many of the opposing comments submitted to the Commission were anonymous, and few included detailed arguments for retaining the ban. In most cases, it was indeterminable as to if the commenter was opposed to data collection on Pre-K–12 or postsecondary students or both, or if such a delineation was considered. However, some comments were substantive arguments, both in support and opposition to the ban.[5]

Notable contentions in favor of a continued SUR ban came from United States Parents Involved in Education (USPIE) who wrote, "We oppose the collection of this information without specific written consent of parents. We oppose the organization of this data into easily sharable formats. And we oppose efforts to make the data easily accessible. We are asking that this policy be dropped from consideration."[6] Similarly, comments from a jointly signed letter from the American Civil Liberties Union (ACLU), Parent Coalition for Student Privacy, and other signatories offered "we strongly oppose any proposal that would lead to the creation of a central federal clearinghouse or linked data sets containing personally indefinable[sic] information (PII) of all students."[7] While USPIE and the ACLU et al. comments did not address whether their concerns were applicable to both Pre-K-12 and postsecondary students, or just the latter, Commission staff infer, from the context of their submissions, a primary interest in Pre-K–12. Further, both sets of comments stressed concerns regarding the ability to actually safeguard such information, for example, USPIE continued, "the inability to adequately protect computerized data cannot be argued." During the

[5] Additional information also provided in the "Summary of Input from Public Hearings" and "Summary of Federal Register Notice Request for Comments" provided during the March 2017 Commission Meeting.
[6] USPIE RFC Comments
[7] PCSP, ACLU, et al. RFC Comments

Commission hearings, five witnesses testified against any repeal of the ban, citing potential re-identification and misuse of data other than for the original purpose, as notable concerns. One witness expressly rejected the idea that government must collect data for the universe of program beneficiaries to support research.[8]

Arguing to overturn the ban, many commenters focused on why the need for postsecondary unit-level information is important. Stephen Crawford of the George Washington Public Policy Institute offered, "Lifting that ban would enable EXISTING data systems to communicate with one another and thereby answer critical questions that can't be answered now, especially questions about the OUTCOMES of higher education programs in which the federal government/taxpayers invest billions of dollars."[9] The American Education Research Association (AERA) stated, "Simply put, students and families, policy makers, and the general public have a deep and legitimate need for high quality information about postsecondary education to inform everything from life changing individual choices to public policies affecting a $500 billion sector of the economy."[10] Likewise, various comments and a jointly signed letter from a coalition of 18 professional organizations detail the group's views of the high value of such data, along with specific limitations and protections.[11] The coalition suggests, "the student level data system as envisioned by our associations and organizations would be limited to postsecondary students, the vast majority of which are not minors. This system would be about college access, completion, debt, and workforce outcomes for adult postsecondary students. It would not be a K-12 national database on children and would not have sensitive information on health, religion or discipline." They state further, "the lifting of the ban would not negate protections in the Family Educational Rights and Privacy Act (FERPA)."

A careful read of the many comments and statements offered to the Commission on the SUR ban provides a window into the challenge of policy development for systems containing sensitive information about individuals, with an equally serious need for informed decision making around programs in which the Federal government invests considerable funds. Concerns regarding non-zero risk of cyber breaches are legitimate, yet likely no more concerning than for other policy domains where databases already exist or when there is an identified and legitimate need for informed decision-making. Government expenditures are also not the only investments to consider for general welfare as, for many American families, education expenses are second only to housing.

WIOA Database Ban. Another ban brought to the attention of the Commission is one embedded in the Workforce Innovation and Opportunities Act (WIOA) that prohibits developing a national database of program participants, inclusive of PII. This ban is modeled on the SUR ban statutory provision.[12]

---

[8] See hearing testimony from C. Cresswell.
[9] Crawford RFC Comments
[10] AERA RFC Comments
[11] Signatories include Achieving the Dream, Advance CTE, Association for Career and Technical Education, Association of Public Data Users, Association of Public and Land-grant Universities, Complete College America, the Council for Community and Economic Research, Ed Trust, Georgetown University Center on Education and the Workforce, George Washington Institute of Public Policy, The Institute for College Access and Success, Institute for Higher Education Policy, NASPA – Student Affairs Administrators in Higher Education, National Skills Coalition, New America Education Policy Program, Veterans Education Success, Workforce Data Quality Campaign, and Young Invincibles.
[12] 29 U.S.C. 3341

> "Prohibition on Development of National Database.-- (1) In general.--Nothing in this Act (including the amendments made by this Act) shall be construed to permit the development of a national database of personally identifiable information on individuals receiving services under title I or under the amendments made by title IV."

While the Commission did not receive extensive comments in favor of preserving the WIOA ban, many of the same objections and concerns related to student records would be equally relevant. Likewise, the motivations behind creating the WIOA ban are similar to that of the SUR ban. One group, the Workforce Data Quality Campaign advocated for overturning both the WIOA database ban and the SUR ban, in order to develop "a nationwide, inclusive data set to show how people are moving through a variety of education pathways."[13]

The population cohort receiving assistance under WIOA overlaps with those receiving assistance under various other education programs over time, essentially compounding the inability to gain insight in to the effectiveness of this collection of programs.

Similar examples of bans on data or databases include data regarding participants in the Head Start program, and a recent proposal prohibiting a database to support analysis about disparities in affordable housing. Many other examples of bans or restrictions on research populate the U.S. Code. Some of the more well-known bans affect research on gun-related crimes and injuries. For example, the Dickey Amendment, a rider to the 1995 HHS appropriations bill prohibits the Centers for Disease Control (CDC) from funding research on guns as a cause of injury. Similarly, the Tiahrt Amendment in 2014 prohibited the Bureau of Alcohol, Tobacco, and Firearms (ATF) from releasing any of its data on gun sales and traces except to law enforcement officials, effectively constraining potential national research on gun violence.

Privacy is at the core of many arguments behind bans on data collection, assembly, or use; yet there is little said as to why these specific datasets are held to a higher level of sensitivity than other government data, such as health data or tax data. Even the standards for data access in the Violence Against Women Act (VAWA), while creating an exceedingly high bar given the context of that program, do not constitute a ban equal to those that inhibit analysis of gun violence or postsecondary data.

In the Commission's draft vision statement: "We envision a future in which rigorous evidence is created efficiently as a routine part of government operations and used in policy making." Is there a scenario that adequately balances or mitigates privacy concerns leading to bans, with the Commission's draft vision?

SUMMARY OF COMMISSIONER CALL

During the April 24th call, Commissioners Rice, Hoynes, Shea, Meyer, Glied, Troske, and Haskins discussed an earlier version of this memorandum. Participating Commissioners generally agreed that bans on the collection or use of data are in opposition to the Commission's mandate, and recommended including a strong statement declaring this. Commissioners discussed that any mention of overturning bans in the Commission's final report, specifically the SUR ban, could limit the attention of other recommendations. However, participating commissioners recognized that the

[13] WDQC RFC Comments

commission was created, in part, to unlock sources of data for evidence building; we should proceed with recommendations that give Congress options for addressing current and future bans. This memo is updated to reflect the discussion, and a suggestion to include a sunset provision.

The discussion of the recommendation now includes three implementation options for Congress to consider.

DRAFT RECOMMENDATION

❖ **Finding:** Given that bans on data collection and use inhibit the generation of evidence, which is antithesis to the Commission's charge, the Commission finds that the use of bans should have limited use, if any at all.

► **Recommendation 6.1:** The Commission recommends that the Congress and the President should review any current or future bans that explicitly prohibit the collection and use of data for evidence building.

The Commission finds that bans are not optimal for the arrangement of information to support evidence building or to promote responsible and secure sharing of the data about Federal programs and policies. There are at least three distinct strategies through which Congress could address existing and future bans:

- **Option 1:** The Congress could overturn any ban directly affecting the collection and use of data for evidence-building purposes.
- **Option 2:** The Congress could include a sun-setting clause that states bans shall cease to be effective after 4 years, which allows stakeholders to re-state their case, and allows for reauthorization by Congress if they choose. This sun-setting clause would also apply to any new ban created in the future.
- **Option 3**: The Congress could establish criteria for evaluating existing and proposed bans on data collection or use.

While Congress can naturally help establish priorities for data collection and use—and the creation of bans may be perceived as one such mechanisms for setting priorities in a complex policy context—absent articulated criteria the bans are inconsistent with the goals of evidence-based policymaking. Thus, Congress should find bans that fail to meet the identified criteria to be antithetical to its interests and the interests of the American public. For example, criterion could establish a designation of a "Major Program," similar to that of an economically significant or major regulation, if projected annual expenditures exceed $100 million. In this example, data enabling evaluation of a Major Program should not be banned for explicitly statistical purposes.

Privacy expert Marc Rotenberg suggested to the Commission that "data can be used both for informed policy-making and for profiling, segmentation and discrimination." In other words, there is both a value and a harm associated with use of any data. There is also risk of harm associated with the investment of billions in public funds without adequate study of effectiveness and accountability. Thus, some means to weigh the competing interests would be of value, and has precedent in government. Regulatory actions measured to have a social cost exceeding $100 million are subjected to increased review and evaluation. A similar construct for government programs with annual costs

exceeding $100 million is equally appropriate, and would help achieve the Commission's vision routine and efficient production of evidence.

Current disparate treatment of different data types in statute lacks a strong rationale, however, as the Commission has discussed with regard to data tags, some types of information may merit different protections. However, there are far more surgical policy solutions that can enable comprehensive privacy protections without shielding an entire sector from review. A reasonable person could conclude that the motivations behind bans on data collection and use are less about privacy than about an intent to limit accountability by certain stakeholders. Given that, only in exceedingly rare cases should Congress allow programs receiving substantial public investment to be excluded from those subjected to rigorous policy research and evaluation.

LIKELY REACTION TO RECOMMENDATION:

- ***Congressional:*** As the originator of many of the current statutory bans, Congress is not likely to favor acting on a recommendation to overturn its past actions.
- ***Executive:*** Executive agencies currently impacted by bans on data assembly, while acknowledging the potential resource implications, will benefit from the ability to investigate and improve their program effectiveness.
- ***State and Local:*** State and local governments will have a mixed reaction, as Federal bans do not affect their data activities, but limit their ability to investigate program effectiveness across state lines.
- ***Other:*** Key stakeholders in the non-public higher-education community, as well as organizations such as the NRA whose policy perspective benefit from limited research will oppose this action vigorously.

COST IMPLICATIONS

If Congress were to act on the recommendation, short-term costs would likely be minimal because the recommendation does not require immediate action on the part of Departments, though allows it. Longer-term costs would fall to the Departments that will have access to relevant data. These Departments may see an increased demand for resources in the form of IT costs to maintain their IT systems and related administrative costs. On a state level, these data already exist, but states may not have a level of current capacity to securely share data with the Federal government, therefore additional resources may be needed for development and maintenance.

ALTERNATIVE OPTIONS

- ***Leave the Bans.*** The Commission finds that bans on data collection and use are justifiable in some cases and recommends no further action.
- ***Strong Statement Only.*** The Commission does not recommend any specific actions, but does include a strong statement in the report regarding the conflict between bans on data and evidence-informed policy.

RECOMMENDATION MEMO #7
UPDATED AFTER APRIL 27 CALL

Enhancing Collaboration in the Federal Evidence Ecosystem
Lead Commissioner: Abraham
Lead Staff: Hart and Martinez

ISSUE: The current institutional infrastructure and level of coordination among critical actors are not sufficient to meet the anticipated future demand for evidence. How can the infrastructure and coordination be improved to better achieve the Commission vision?

The Commission recognizes efficient and effective programs can benefit the public, and inefficient and ineffective programs may be harmful to members of the public. Thus, decision makers need more and better evidence to meet the demand articulated by both the Executive and Legislative Branches, as demonstrated by the creation of the CEP. This memorandum proposes recommendations designed to realize the Commission's ambitious vision of a highly collaborative and appropriately coordinated Federal Evidence Ecosystem, where evidence is efficiently and rigorously generated, and available to be used to inform policy decisions. The Federal government currently operates components of the envisioned evidence ecosystem, including a technical evidence-building community, but these components often operate in topical or organizational siloes that limit the overall effectiveness of the whole. Enhanced collaboration within the evidence-building community could foster better coordination and more effective interaction among units that perform statistical, evaluation, policy research, and policy analysis (SERA) functions, agencies that administer programs, OMB, and the proposed data facility. By strengthening the operations and enhancing coordination across the evidence-building community, the community will be able to meet the increased demand within the larger evidence ecosystem to inform policy decisions.

This memo presents a set of overarching recommendations to enhance the operations and coordination for key actors within the Federal evidence-building community and the Congress. Recommendations tailored to specific components of the evidence-building community are discussed in subsequent memos for Commissioners that will be presented during the May and June Commission meetings:

- ***Evaluation and Policy Research*** **(May Meeting):** Recommendations for the institutionalization of and incentives for the evaluation and policy research functions within Federal Departments (Memo #8),
- ***Statistical System*** **(May Meeting):** Recommendations for changes to facilitate robust coordination of the Federal statistical infrastructure that is foundational to evidence-building (Memo #9),
- ***Program Design*** **(June Meeting):** Recommendations to better integrate evidence building into individual program designs, for both existing and new programs (Memo #13), and
- ***Partnerships*** **(June Meeting):** Recommendations that recognize the evidence ecosystem encompasses non-governmental communities who can and do partner with government Departments to meet the demand for evidence (Memo #12).

Relates to Facility (#2), Evaluation (#8), Statistics (#9), Partnerships (#12), Program Design (#13), Administrative (#15), and Funding (#16).

* * * * *

BACKGROUND AND ANALYSIS

I. Why might the Commission want to adopt the Federal Evidence Ecosystem description?

Because the operation of the Federal government is complex and the variety of actors involved in both producing and using different types of evidence will inevitably vary over time and across policy issues, delineating a cohesive framework for envisioning how the apparatus could function is a critical first step to defining the optimal arrangement. The ecosystem image enables a plain-language explanation of how supply and demand for evidence relate and interact, and can help bridge the reality of how the many actors relate today, and how they could within a system that was fully functioning to meet increased demand for evidence.

Importantly, the ecosystem is an approach for clearly relaying the vision and principles of the Commission: "We envision a future in which rigorous evidence is created efficiently as a routine part of government operations and used in policymaking." The framing is a direct response to the questions of "optimal arrangement…to facilitate program evaluation, continuous improvement, policy-relevant research, and cost-benefit analyses…" *(Sec. 4(a)(1))*, appropriate modifications to data infrastructure (*Sec. 4(a)(2))*, incorporating outcome measurement and institutionalizing evaluation in program designs *(Sec. 4(a)(3)*), and how clearinghouse data and results of research can be used to improve programs (*Sec. 4(b)(2)(I))*. The ecosystem recognizes that across these areas the President and Congress directed the Commission to consider that there are interconnections and inextricable relationships for satisfying better approaches to evidence-based policymaking. But, the term "ecosystem" itself recognizes that the system is constantly changing based on activity within the system and as well as exogenous factors.

More broadly, the ecosystem framing also addresses concerns raised by stakeholders early in the Commission's fact-finding, appealing that the Commission speak expansively about the role and need for evidence generation. For example, in testimony during the November 4 meeting, Adam Gamoran appealed: "be the *evidence* commission, not just the *data* commission."[1] By emphasizing the many communities and relationships among actors, each community within the ecosystem can hold others accountable for the continued support and operation of the ecosystem by identifying gaps or unwilling participants that inhibit intended activities. For example, if a program agency plans to change an existing dataset that is important for evidence-building, a statistical agency can assert the importance of participating in the change process. Or if an evaluation office would find valuable access to a dataset that a statistical agency plans to acquire, the evaluation office can assert the importance of jointly approaching the program office to try to address both sets of needs met together.

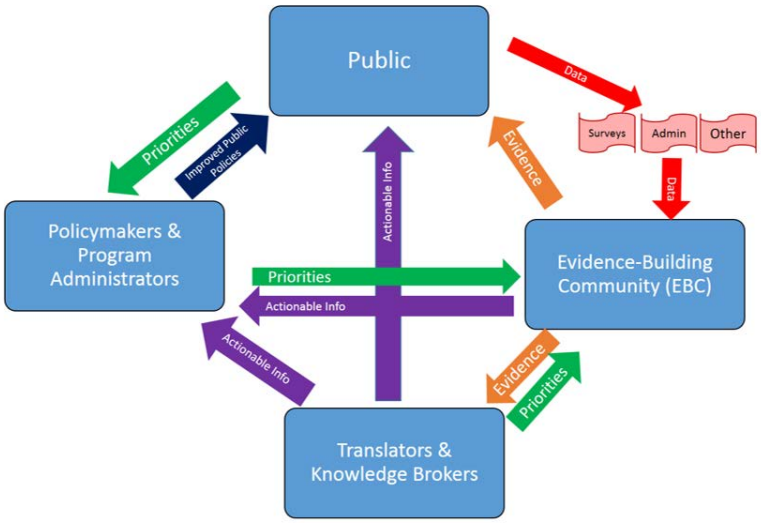
II. What is the Federal Evidence Ecosystem?

The *Federal Evidence Ecosystem* is an organizing construct that recognizes the complex network and interconnections present in generating and using credible evidence to inform policymaking in the Federal government. To the extent the Commission is interested in generating and using evidence more efficiently and rigorously as outlined in the draft Commission Vision Statement, the ecosystem offers a way to conceptualize important relationships between actors in the system and the available infrastructure. Notably, the ecosystem framework illustrates interactions between those who supply

[1] Nov. 4 Testimony of Adam Gamoran to CEP.

or generate evidence, and those who use evidence. All components of the ecosystem exist today, but the interactions are constrained by topical or organizational siloes that limit the overall effectiveness of the whole. In this paper and related memos we include recommendations that aim to reduce the tensions among actors seeking common goals, enabling more evidence generation and use by eliminating unnecessary barriers to building and maintaining capacity.

Figure 1. The Federal Evidence Ecosystem



As described in more detail below, within a fully functioning ecosystem, data flow from originating sources to data curators, users and analysts, who transform data into relevant and meaningful evidence, based on the priorities set by the public and policymakers. Evidence generated by the *Evidence-Building Community (EBC)* is then transmitted back to policymakers and the public, often through intermediary translators or knowledge brokers, a function sometimes played by members of the EBC. The evidence, once available and translated for appropriate audiences as actionable information, can then be used to hold programs accountable by policymakers or the public and ensure continuous learning and improvement within programs.

The Public. Individuals, businesses, and other organizations in the *public* provide critical data to support evidence-building activities, either directly to SERA agencies in the Evidence-Building Community (i.e., via surveys), or increasingly through an intermediary government program (i.e., administrative data). The public also plays an important role in helping policymakers and program administrators determine priorities for policy and research, including evaluation questions. In turn, the public receives information from the evidence generated to inform future priority setting, and benefits from improved policy-making and public programs as a result of their participation. (See discussion in *Transparency* (Memo #11)).

Evidence-Building Community (EBC). The Evidence-Building Community is comprised of SERA agencies and their external partners, such as university researchers, who are collectively responsible for producing evidence.[2] The EBC receives two key inputs: data transmitted about the public (directly from individuals or indirectly through programs) and priorities articulated by the public via policymakers and program administrators, which help define research and evaluation priorities for the EBC. In a fully functioning EBC, the community includes technical experts, researchers, statisticians, and evaluators who convert data into evidence, protect confidentiality and ensure that data use within the community is limited to SERA purposes (see further discussion in *Minimizing Risk* (Memo #10)). *Principal Statistical Agencies* (PSAs) form a critical hub for continuous data collection and curation that supports this community (Memo #9), and the proposed *data facility* and its network (Memo #2) would serve as a foundational component of the community by serving

[2] Note that many SERA offices are identified in OMB's annual *Statistical Programs of the United States*.

as an access point for administrative data and linking of those data to other relevant sources. Currently, PSAs protect data under the uniform protections of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA).[3] The community also includes evaluation and policy research actors (Memo #8) and research partnerships (Memo #12). Collectively, inputs or priorities to the EBC are used to shape the analysis conducted, and data are converted into evidence by actors within the EBC. Evidence generated by the EBC may be relayed to the public directly, to policymakers and program administrators, or to translators or knowledge brokers.

Policymakers and Program Administrators. Policymakers and program administrators are the intended audience of much evidence produced about government programs. These actors include Members of Congress, congressional staff, Executive Branch oversight agencies (e.g., OMB), Cabinet secretaries, and program administrators. Collectively these actors use evidence to either modify programs and policies for continuous improvements or for accountability. Depending on the individual role, the types of information needed for the different actors in this group will vary. As the primary users of the evidence, policymakers and administrators ultimately determine what is relevant and, therefore, play an important role in setting priorities for the EBC at the outset based on their own expertise, as well as input directly from the public. These actors are specifically addressed in recommendations within this memo as well as in *Program Design* (Memo #13).

Once evidence is generated, policymakers and administrators may receive actionable information directly from the EBC, to the extent evidence includes such information (e.g., recommendations). More common today, evidence produced by the EBC is relayed to policymakers and administrators through actors who serve as translators of a body of evidence, also known as knowledge brokers. Once administrators or policymakers possess the actionable information, they must decide how to use the information to improve decisions, policies, or programs. These actions are appropriately situated within the confines of the larger policy sphere, in which evidence generated by the EBC is realistically only one among many inputs.

Translators and Knowledge Brokers. Actors internal and external to government (and the EBC) who distill highly technical evidence generated by the EBC serve a critical role in relaying and translating what is known from the range of available evidence to address the questions raised by policymakers and program administrators, as well as to inform the public about the effectiveness of programs and policies. Within Federal agencies, this important function is often the role of policy analysis, and serves as a mechanism for quick turnaround responses when the real-world policymaking needs cannot wait for perfect information. Agencies may participate in translation for an external audience as well, by operating clearinghouses for final reports or publishing non-technical research summaries, with the aim of making the results of research and evaluation efforts available to decision makers or practitioners for quick guidance on program decisions. Critically, knowledge brokers are also essential for identifying gaps in current evidence that can be used to better inform decision makers and administrators for future actions or help the EBC set priorities for future research and evaluation. For example, if conducting a systematic review or meta-analysis the brokers identify an area where no evidence was available about a specific policy, that knowledge is an input to the priorities for the EBC.

This memo focuses primarily on challenges and recommendations related to actors within the Federal evidence-building community and policymakers with a role in interacting across the ecosystem, such as Congress; other memos will specifically address the public.

---

[3] Plus three OMB "recognized units." OMB is also authorized to designate other "units" as warranted. This component of the EBC could be scaled to include more units with a recognized need for confidentiality protections.

III. What is wrong with the current operation of the evidence-building community within the ecosystem?

The relationships among actors in the ecosystem described in Figure 1 contribute to identifying where this ideal-type breaks down in reality. This memo proposes recommendations designed to improve the collaboration and coordination of the Federal Evidence Ecosystem due to four issues: (1) unclear or inconsistent *institutionalization* of the SERA functions, including for missions, roles, and responsibilities within departments, (2) limited *cross-government collaboration* and coordination of SERA activities, (3) *prioritization* of evidence-building activities among senior leaders, and (4) inconsistent *signals about priorities* from the Legislative Branch. Other memos further discuss challenges around appropriately tailored *incentives* to support evidence production and use, and limited investment in *partnerships*.

Issue #1: Unclear or inconsistent institutionalization of SERA functions in Departments. Institutionalization involves both ensuring the existence of each of the SERA functions within Departments, as well as their coordination and collaboration. Many Federal Departments lack systematic coordination across components of the evidence-building community. Some departments lack critical SERA functions altogether; for example, Federal human services programs do not have a PSA to work with, though other parts of the Department have worked to address some of the limitations this poses. Other Departments have access to all (or many) of the components operating, but lack internal coordination. However, some Departments have successfully managed to navigate the complex arrangement of activities within their own contexts, developing internal groups and policies to ensure SERA activities are coordinated throughout the organization (e.g., HHS Data Council), generating considerable success stories.

The inconsistent institutionalization of SERA functions in agencies derives from two core issues: (1) missions and direction for programs do not always clarify the role of SERA activities in the evidence-building community or even with the larger operation of a Department, and (2) roles and responsibilities for supporting SERA activities within Departments may not be clear. In both cases, institutionalization of the SERA functions serves to enhance the evidence-building community overall, and the respective functions within Departments.

In terms of program missions, the role for program administrators asking for new evidence about program operations or outcomes is often not clearly articulated or consistently expected. For example, program agencies collecting administrative data may not see the provision of data as a priority for their office, instead prioritizing direct service delivery. This limited awareness of the vital role administrative data plays in generating evidence, limited resources for maintaining data quality and access, and the potential for data to be used "against" the program (i.e., threatened terminations if a program does not "work"), suggest complex motivational barriers in agencies and programs.

Similarly, statistical agencies may view legal missions narrowly around collecting data and disseminating statistics, without a specified role in analysis or technical support to other components, instead prioritizing resources or expressing concerns about maintaining objectivity and independence from policy activities. In both cases, the missions of the programs introduce gaps in the evidence-building community that some other component must recognize and fill. Ideally, to the extent PSAs exist in Departments, those offices can provide a core infrastructure to support continuous evidence generation, including by working with program offices and the proposed data facility to make administrative data available for SERA purposes. In turn, evaluation and policy

research offices, for example, can help programs develop and understand priorities for future evidence building.

Institutionally, roles and responsibilities for SERA activities are often dispersed across Departments, with resources and expertise to conduct the activities similarly dispersed, uncoordinated, or at worst, non-existent. Many Departments do not have identified components responsible for each of the SERA functions, in part or in combination, though many have pieces. This lack of coordination of resources means expertise may be inefficiently allocated among priorities or needs, limiting the volume and quality of SERA activities. Programs that want to pursue evaluations may even prioritize the activities themselves using in-house staff, but may not have the capability to develop the most technically sound or robust studies possible when experience is not otherwise available.

Despite the gaps in many Departments, some agencies have established internal structures and mechanisms to support the range of SERA activities. For example, the Department of Education's co-location of statistical, evaluation and research agencies within the Institute of Education Sciences could help facilitate collaboration and coordination of SERA activities within the Department. To the extent roles and responsibilities are made clear, senior leaders within Departments can reinforce the importance of routine and continuous evidence building activities and program administrators can be better connected to those within their Departments who carry out SERA functions.

Issue #2: Limited Cross-Government Collaboration and Coordination. Within the Federal government, coordination across and within agencies for evidence-building activities is varied and inconsistent. The Office of Management and Budget (OMB) and policy councils within the Executive Office of the President (EOP) are charged with facilitating interagency initiatives, but few efforts have successfully addressed the limitations for collaboration across Departments and SERA functions to routinely generate rigorous evidence to inform policymaking. As the actions of one Department's policies and programs are increasingly recognized to interact with those implemented by another, this lack of coordination is a growing concern.[4] The Government Accountability Office (GAO), for example, has long criticized the Executive Branch's lack of coordination for services provided to individuals with disabilities across vocational rehabilitation in the Department of Education, Disability Insurance in the Social Security Administration, the Administration for Community Living in the Department of Health and Human Services, the Office of Disability Employment Policy in the Department of Labor, and the Department of Veterans Affairs.[5]

One major reason for this limitation in cohesively implementing evidence-building activities is the siloed nature of research, statistics, information management, information technology, evaluation, performance measurement, and policy analysis functions across government. Historically, each of these functions evolved through an array of Federal laws and policies that established siloes. While each function has a distinct purpose, obvious synergies are often overlooked due to lack of communication between units assigned with implementing various activities. For example, performance measurement is conducted under the authority of the Government Performance and Results Act (GPRA) and the GPRA Modernization Act of 2010, which outline specific responsibilities for performance managers without explicit direction to coordinate with the other functions.

---

[4] For example, see NAS. (2016). Advancing the Power of Economic Evidence to Inform Investments in Children, Youth, and Families. Washington, D.C.: National Academies Press.

[5] GAO. (2008). Federal Disability Programs: Coordination Could Facilitate Better Data Collection to Assess the Status of People with Disabilities. http://www.gao.gov/new.items/d08872t.pdf

Importantly, OMB serves as a central coordinator across many government processes, but itself spreads core evidence-building responsibilities across multiple offices, which can result in confusion or inconsistent guidance for agencies (see Figure 3a). Within OMB today, statistical policy is the responsibility of the Chief Statistician and the Statistical and Science Policy Branch (SSP) of the Office of Information and Regulatory Affairs (OIRA); information policy is split between SSP and the Information Policy branch in OIRA; privacy policy is set by the Privacy Branch in OIRA; evaluation advocacy and policy is split between OIRA and the Evidence Team within the Economic Policy Division; information technology responsibilities, including for statistical purposes, are allocated to the Office of E-Government and Information Technology; performance measurement and management implementation under GPRA is coordinated by the Performance and Personnel Management (PPM) Division; and other Resource Management Offices (RMO) coordinate aspects of priority policies and resource allocations, including as actors who both facilitate supply and encourage the demand for evidence and evaluation. Each of these offices reports to a different political appointee, some of whom are Senate confirmed, and each can issue formal or informal guidance or directives to agencies that encourage, complicate or muddle activities in the evidence-building community. Yet, each has also identified "swim lanes" to operate within, though the boundaries are sometimes overlapping and unclear to staff in agencies (for example, an RMO may approve funding for an evaluation advocated by the Evidence Team that requires an information collection request from an agency whose study design SSP will review for consistency with OMB information quality standards).

Making implementation at OMB more complicated, the roles of offices within OMB also evolved over time. The authorizing statute for OIRA, the Paperwork Reduction Act (PRA), expressly provides direction that the OIRA Administrator facilitate government-wide "information management" policies.[6] The Chief Statistician position was established in OIRA to support the evidence-building community, though that role has been constrained by limited resources and, as a consequence, has largely focused on the PSAs and statistical policy, ultimately providing little coordination for smaller statistical programs, evaluation, policy research and policy analysis activities. Further, the importance of information policy within OMB for facilitating the activities of the evidence-building community should be no less important than budgetary, regulatory, or other functions of the organization. The E-government Act of 2002 further fractured implementation responsibilities at OMB by establishing a separate office charged with all information technology policy, often with overlapping or conflicting responsibilities to OIRA's information policy function.

In OMB's current organizational structure, agencies typically need to contact multiple offices within OMB—each reporting to different political appointees—before gaining approval to conduct new evidence-building activities.[7] This inherently places the burden of coordination on individual Departments, rather than OMB. Perhaps more problematic, many agency organizational structures have developed to mirror the organizational arrangement developed by OMB, adding new offices as additional responsibilities were assigned rather than exploring how to integrate the functions, when appropriate. This arrangement effectively ensures evidence-building activities maintain their siloes in OMB and in agencies throughout government.

Issue #3: Prioritization of Evidence-Building Activities. Even with the right organizational structure and sufficient staff capacity, Federal Departments may still not engage meaningfully in the ecosystem without the right set of incentives and expectations, given other important competing

[6] See Paperwork Reduction Act, 44 U.S.C. 3503

[7] Note that evidence initiatives in recent years have operated from multiple parts of OMB, often without sufficient resources, which limits the overall effectiveness and scope of the activities.

priorities. The recognition of senior leaders' role in support SERA activities can help encourage and facilitate the appropriate level of the activities throughout Departments. Currently there are few measures for holding senior leaders in Departments accountable for consistently supporting SERA functions. For the ecosystem to fully function, senior leaders must recognize the important role they play in the evidence-building ecosystem, be encouraged to participate in the SERA functions within the ecosystem, and be held accountable for their role in supporting the generation and use of evidence.

Issue #4: Inconsistent signals about priorities from the Legislative Branch. The previous two issues largely emphasized challenges that affect the Executive Branch, but the Legislative Branch also plays a critical role in establishing demand for evidence and as a user of evidence in the Federal Evidence Ecosystem. An upcoming report from the Bipartisan Policy Center (BPC) observes that "the fragmented nature of the legislative process, the clash of partisan interests, and the labyrinth of legislative procedures and practices make it more difficult to establish a coordinated structure for using research and evaluations in a more systematic fashion in the legislative process."[8] The same process can present challenges for individuals seeking to interpret priorities and signals from Congress about how to build better evidence to meet congressional needs.

SUMMARY OF APRIL 28 COMMISSIONER CALL

Co-Chair Haskins, and Commissioners Groves, Hoynes, Rice, Troske, and Wallin participated in a discussion with Commission staff on April 28, 2017, about an earlier draft of this memo. Additional feedback was received in writing from Commissioners Glied and Hahn, and from Commissioner Shea during a subsequent phone discussion. Through the various discussions, several revisions were made to this memo to clarify the Evidence Ecosystem involves participation with non-Federal actors, but is predominantly about evidence for programs funded or directed by the Federal government. Commissioners largely agreed that the ecosystem description and outline of the challenges were appropriate.

Modest revisions to the findings and recommendations were incorporated to clarify language and indicate priority status of certain recommendations. Recommendation 7.1 was reframed to emphasize the OMB reorganization is not intended as a massive expansion of OMB, rather a consolidation of fragmented parts of the organization. Recommendation 7.6 was similarly restructured to suggest some congressional infrastructure, without prescribing what that might look like more formally. A recommendation regarding GAO was shifted to the alternative option section, and a suggestion for evidence ombudsmen was also added as an alternative.

DRAFT RECOMMENDATIONS

Today's implementation of SERA activities in the Federal government falls short of the vision for what is possible in a coordinated, collaborative, and adaptable evidence system. This section outlines options that can substantially improve the status quo for the system as a whole and fill identified gaps to encourage actors within the ecosystem to work toward a set of common goals. Other memos for the May and June Commission meetings will address recommendations for specific components of the system.

---

[8] BPC. (2017). Congress and Evidence-Based Policymaking: Creating a 21st Century Legislature. Confidential Draft. Washington, D.C.: Bipartisan Policy Center, p. 13.

Recommendations Affecting the Executive Branch

The Commission proposes three Executive Branch recommendations, the first addresses OMB's organizational challenges, the second addresses challenges within individual Departments, and a third addresses interactions across Departments and the SERA functions.

❖ **Finding:** The Commission finds that the continued fragmentation of evidence-building functions within OMB hampers the organization's ability to sufficiently prioritize evidence building policies and coordination for government, limiting the Executive Branch's ability to realize the vision in current law for information policy. Further, the Commission finds that as a critical linchpin in ensuring that the full range of the Commission's recommendations are adopted, the existing coordination within OMB for evidence building is insufficient for maximizing the Federal Government's ability to use of evidence in decision making.

❖ **Finding:** The Commission finds that the Congress and President have successfully signaled the importance of establishing a dedicated leader within the Executive Office of the President (EOP) to focus on a specific, cross-cutting policy and management issue. As evidence-based policymaking increases in prevalence throughout the Federal government, the Commission finds that a similar role will be essential for continuous emphasis on developing and using evidence.

► **Recommendation 7.1 (PRIORITY):** The Commission recommends that the Congress and the President provide OMB authority to reorganize, consolidate, and sufficiently resource the information policy functions within the agency under a single senior leader to maximize support for developing and using evidence to inform decision making.

As a first step in achieving better coordination, the Commission recommends that OMB be provided sufficient legal authority under the PRA to reallocate evidence-building functions. While providing the authority alone as written in Recommendation 7.1 is not a guarantee the agency would pursue a reorganization of existing staff to better meet the need for improving coordination, establishing the expectation is a critical first step that would likely be paired with more detailed direction from Congress.

Importantly, the Commission's recommendation does not articulate precisely how OMB could achieve the desired level of coordination under a single senior leader, which would be intentional to increase the likelihood of adoption, though the Commission would also likely need to socialize the approach within OMB. The Commission could, however, choose to articulate suggestions for accomplishing the recommendation's goal. For example, one possible approach is to establish a new Office of Information Policy (separate from an Office of Regulatory Affairs) in which staff from the performance, evidence, information, privacy, and statistical policy offices would be brought together under an Administrator of Information Policy.[9] The separation of regulatory and information policy ensures that policy officials are no longer forced to prioritize regulatory policy over information policy.

As the demand for evidence in the policymaking process continues to grow, the operational siloes within OMB will likely only become more constraining for the timely production of evidence across government. Recognizing OMB's role in supporting collaboration and coordination across the Executive Branch and the evidence-building community, the Commission recommends OMB

[9] Alternately, the office could be named "Office of Evidence-Based Policy."

receive sufficient legal authority to consolidate the most important evidence-building responsibilities within OMB under a single policy official for information management, to be executed as soon as possible (e.g., see Figure 3b).[10] This recommendation will enable OMB to better coordinate the government's information policy and SERA functions, including through existing interagency councils described under Recommendation 7.3. Such a consolidation will also shift the burden for identifying the appropriate elements of OMB with whom to coordinate from Departments to OMB itself, enabling more clear and consistent consultation and advice.

Figure 3a. OMB Offices Engaged in Coordinating Evidence-Building with Agencies (Current)

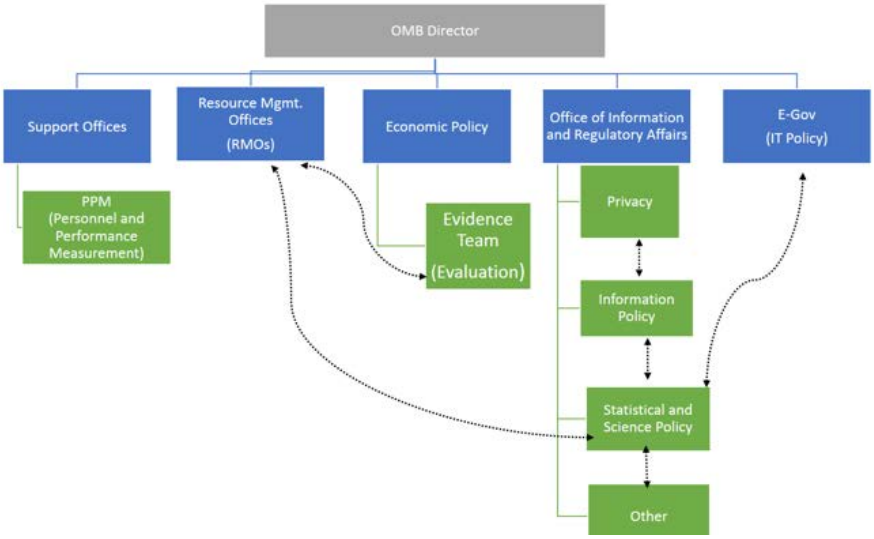


Figure 3b. Proposed Reorganization of OMB to Establish Office of Information Policy (Future)

In reorganizing and consolidating responsibilities, OMB may also seek to explore aspects of IT related to evidence building that could be located within this newly established office. Revisiting some of the IT responsibilities would continue the unfinished business initiated in 2002 when IT was removed from OIRA to the newly formed E-Gov office and "information policy" was essentially split between the two offices in a non-mutually-exclusive and arguably conflicting

[10] Specifically requires adjustments to 44 U.S.C. 3501 and 3503, in addition to general reorganization authority.

manner. Minimally the Administrator of an Office of Information Policy and the Administrator of E-Gov should consult on decisions affecting the IT infrastructure of the evidence-building community.

❖ **Finding:** The Commission finds that statistical, evaluation, policy research, and policy analysis (SERA) functions are each important and should be someone's responsibility within a well-functioning department. Further, it finds that coordination of SERA activities within agencies is too often under-coordinated and ill-calibrated to satisfy the Commission's vision for a more robust evidence ecosystem in the future.

► **Recommendation 7.2 (PRIORITY):** The Commission recommends that the President should direct Federal Departments to–

a) Ensure that each SERA function is present, regularly inventory units that perform SERA functions across the Department as part of the Strategic Planning process, and ensure adequate coordination of SERA activities; and
b) develop multi-year learning agendas in consultation with program and SERA offices, to be updated on a regular basis, containing both short- and long-term evidence-building priorities for each Department.

In some cases, the core infrastructure needed to routinely generate statistics, evaluation, policy research, or policy analysis does not even exist, or is not distributed throughout a Department in a useful manner. While interagency coordination can be facilitated through OMB and coordinating bodies, similar mechanisms must exist within each agency to coordinate and facilitate the maintenance of the basic SERA infrastructure, while addressing mission-specific needs and priorities of each Department. Recognizing SERA activities are most relevant when aligned with the structure and mission of an agency and require basic infrastructure and coordination, the Commission recommends that a senior manager in every major Department or agency regularly develop and revise a plan as part of strategic planning effort for establishing infrastructure for each SERA function and for appropriate intra-agency coordination.

The Commission recognizes the unique needs of agencies and does not recommend a one-size-fits-all approach. Rather, the Commission suggests a degree of flexibility within each agency in prioritizing how to best allocate SERA activities for accomplishing respective missions. However, such activities must be deliberate and recognized by senior managers as an important component of the decision-making apparatus. The recommendation might be implemented, for example, by requiring a senior manager to develop and periodically review the plan as part of the formal strategic planning requirement of GPRA Modernization, which could help ensure a higher degree of deliberate attention to the activity.

The establishment of a multi-year learning agenda is one approach to institutionalizing the role of SERA activities throughout a Federal Department, while recognizing knowledge gaps in existing evidence.[11] A learning agenda is a strategic research plan, which seeks to identify and prioritize critical research, evaluation, and policy questions related to Departmental programs and initiatives. The approach to developing a learning agenda may vary based on the size and complexity of the organization, but all learning agendas should be developed in collaboration with external

[11] USAID. (2017). Landscape Analysis of Learning Agendas: USAID/Washington and Beyond. Washington, D.C.: U.S. Agency for International Development, Bureau of Policy, Planning, and Learning. https://dec.usaid.gov/dec/GetDoc.axd?ctID=ODVhZjk4NWQtM2YyMi00YjRmLTkxNjktZTcxMjM2NDBmY2Uy&rID=Mzg4MDQz&pID=NTYw&attchmnt=VHJ1ZQ==&uSesDM=False&rIdx=NDkxODc5&rCFU=

stakeholders and take into account the ongoing work of the research and evaluation community across philanthropic, academic, and other institutions. Learning agendas should include both short-term and long-term priorities, and be viewed as flexible guidelines that can adjust to accommodate urgent needs which arise due to external events.

Smaller Departments may choose to develop a single learning agenda that spans the full reach of the Department's programs. An example would be the Research Roadmap developed by the Office of Policy Development and Research (PD&R) within HUD. [12] HUD's Research Roadmap was initially developed in response to a recommendation within a 2008 report of National Academy of Sciences[13] that found that in an era of limited financial and human resources, PD&R's research-agenda setting process had become "too insular" with "too much of a short term focus." As a result, the report concluded, PD&R was not "achieving its potential to contribute in a significant way to the ongoing internal and external discourses over the evolution of HUD programs and broader urban development policy." HUD refreshes the Research Roadmap every five years, engaging a broad stakeholder group both inside and outside of the Department, to ensure that the priorities captured reflect the needs of the Department and the broader field. Larger and more complex Departments may choose to develop program-specific learning agendas, which would then feed into a broader Department-wide learning agenda. An example of this approach would be DOL, which has 14 separate agencies that conduct statistical, evaluation, policy research, and policy analysis activities. Individual agencies within DOL may develop learning agendas, such as the Five Year Research, Demonstration, and Strategic Plan[14] developed by the Employment and Training Administration, with the Chief Evaluation Office within DOL given responsibility for coordinating the overarching Department's evaluation agenda as a whole to ensure efficiency and coverage.
Note: Memos on Evaluation and Policy Research (#8), Statistical System (#9), Partnerships (#12), and Program Design (#13) will also include recommendations to enhance the intra-agency SERA function to support the evidence-building community within the evidence ecosystem.

❖ **Finding:** The Commission finds that interagency coordination is necessary and important to the smooth operation of the many-faceted evidence building community.

► **Recommendation 7.3:** The Commission recommends that the Congress and the President appropriate and provide sufficient resources, giving authority to OMB for modernizing interagency information policy coordination, including to bolster existing interagency coordinating entities such as the Interagency Council on Statistical Policy (ICSP), the Interagency Council on Evaluation Policy (ICEP), the Privacy Council, and the Performance Improvement Council (PIC).

Similar to other commissions charged with exploring the Federal government's information policies and practices over the past half century, the Commission finds that interagency coordination is sorely needed in the Federal evidence ecosystem and that existing mechanisms to facilitate such coordination in an increasingly complex government must be appropriately resourced and supported to encourage cooperation across agencies. In fact, as the number of actors grows, the need increases, making coordination foundational to the government's ability to achieve the Commission's vision for evidence building and use. Recognizing the need for continuous interagency coordination in SERA activities and the long list of information and statistical policy standards the Commission is recommending that OMB develop and implement, the Commission recommends the continued

[12] HUD. (2017). HUD Research Roadmap: 2017 Update. Washington, D.C.: U.S. Department of Housing and Urban Development. https://www.huduser.gov/portal/pdf/ResearchRoadmap-2017Update.pdf
[13] https://www.nap.edu/catalog/12468/rebuilding-the-research-capacity-at-hud
[14] DOL. (2016). Five-Year Research Plan. Washington, D.C.: U.S. Department of Labor, Employment and Training Administration. https://www.doleta.gov/reports/fiveyear_researchplan.cfm

operation with sufficient resources of the OMB information policy activities, including leading interagency coordinating entities mentioned above. OMB's role as the overseer and facilitator of implementation for much of the evidence-building community's activity is key to the success of greater production of evidence across government. In order for many of the Commission's recommendations to be successful, OMB will need to be appropriately resourced to facilitate implementation of the recommendations in collaboration with SERA agencies.

For decades, the Interagency Council on Statistical Policy (ICSP) has existed to advise OMB on the needs of the statistical system—the evidence-building system of its day—and to bring together the heads of the statistical system in developing policies and practices for the efficient and objective operations of the statistical system. The PIC and ICEP came along more recently as the range of actors expanded.[15] Importantly, groups such as ICSP, ICEP, and the PIC provide valuable communities of practice within the Federal government to support knowledge generation, problem solving, best practice dissemination, and shared services, to the extent practicable. The creation of a data facility, with its proposed Executive Board (see *Data Facility,* Memo #2), could further exacerbate lack of clarity about roles; however, in conjunction with Recommendation 7.1, support for related interagency councils will help ensure roles and responsibilities are iteratively balanced and calibrated. Recommendation 7.1 is especially relevant for helping ensure the various coordinating councils are themselves coordinated and not offering conflicting guidance or sending their respective parts of the EBC in different directions.

The Commission specifically received public comments suggesting that the Federal government needs a coordinating operation for the evaluation function, to serve in tandem with other existing coordinating bodies to better support evaluation functions across agencies and to facilitate cross-agency outcome measurement and cross-cutting issues in evaluation.[16] Though not established in statute, such a group does exist today, as evaluation offices in some Departments and OMB created ICEP in 2015.

The Commission suggests that these types of interagency efforts be formally institutionalized in government, including their roles in providing advice to OMB and the new information policy component suggested above in Recommendation 7.1. Note that this is not envisioned to be a central coordination of specific evaluation projects, but rather a mechanism for reducing hurdles and barriers across government.

Recommendations Affecting Incentives in the Executive and Legislative Branches

Note: the following two recommendations were previously included in Institutionalizing and Incentivizing Evaluation (Memo #8) for the purposes of the call the earlier drafts.

❖ **Finding:** Individual incentives for leaders to produce and use evaluation and policy relevant research are lacking for Federal Departments, as are explicit accountability measures tied to the execution of either evaluation or policy research. The Congress and the President should hold Federal Departments directly accountable for pursuing evaluation and research activities by, and receive recognition when their efforts are laudable.

[15] PIC was established by Executive Order 13450 in 2007, and later codified in the GPRA Modernization Act of 2010. ICEP was established administratively in 2015.
[16] See Arnold Foundation VP Letter to the Commission

► **Recommendation 7.4:** The Commission recommends that the President and Congress direct Departments to incorporate measures into performance reviews of career senior leaders to be held accountable for producing evidence.

► **Recommendation 7.5:** The Commission recommends that the Congress, with the Senate confirmation process, and the President, through the Office of Presidential Personnel, seek affirmative responses from political appointees regarding their support for producing and using evidence to inform decisions prior to appointment or confirmation of individuals, consistent with the Commission's Guiding Principles.

For decades, literature in the fields of evaluation, performance monitoring, and public administration has noted the importance of leadership in supporting good management practices.[17] Former DOL Deputy Secretary Seth Harris reinforced this observation during the Commission's March meeting: individual incentives can motivate a culture that encourages the production and use of evidence for policymaking among leaders in Federal Departments. Recommendation 7.4 would require Departments to include an additional provision in performance appraisals for senior executives (defined as members of the Senior Executive Service), which could be achieved through a relatively minor adjustment to existing requirements for appraisals.[18] In order to implement such a legal change, OMB and OPM would need to collaborate and concur on guidance for Departments to implement the provision when appropriate.

Recommendation 7.5 is similar in intent and specifically addresses political appointees by raising the issue of evidence during the appointment and/or confirmation process. Positive affirmation has historically been a small, but meaningful, gesture to encourage new evaluations. For example, in 1990. EPA Associate Administrator Don Clay committed to a Senate panel that he would conduct a full national-scale review of the country's hazardous waste program, then proceeded to launch a major implementation study within weeks of confirmation. Clay's approach to evaluation has been referred to as leading to a pivotal shift in evaluation activities within part of EPA.[19]

Recommendations Affecting the Legislative Branch

❖ **Finding:** The Commission finds that well-coordinated oversight of the Executive Branch's evidence ecosystem by the Legislative Branch would well serve all stakeholders, and provide a clear coordination point to encourage and nurture the growth of evidence-building activities across the Executive Branch. Further, the Commission finds that the Legislative Branch similarly needs greater expertise to enhance Congress' use of evidence in the legislative process.

► **Recommendation 7.6 (PRIORITY):** The Commission recommends that the Congress establish a committee, or assign responsibility within existing committees to provide expertise to Members and congressional staff for drafting legislation that supports the Congress's evidence needs, and to provide coordinated oversight of the Federal offices within the evidence-building community.

Note: We envision further dialogue with Commission co-sponsors regarding this recommendation

---

[17] Cousins & Bourgeois, 2014; Labin et al., 2012; Taylor-Ritzler et al., 2013; Moynihan & Lavertu, 2012

[18] Such a proposal could be included in the list of criteria for SES appraisals established in 5 USC 4313.

[19] Hart, N. (2016). Evaluation at EPA: Determinants of the Environmental Protection Agency's Capacity to Supply Program Evaluation. Diss. Washington, D.C.: George Washington University.

In May 2017, the Bipartisan Policy Center (BPC) will release a white paper observing challenges in establishing evidence-based practices within the legislative process, concluding that Congress needs additional support structures.[20] Recommendation 7.6 builds on the BPC research and the Commission's recognition of dispersed expertise for evidence-building activities across Congress, inconsistent within the existing committee infrastructure. Improving the level of expertise among congressional staff would help ensure new legislation is technically appropriate for the needs of the evidence-building community, while also facilitating Congress's role as a user of evidence by establishing expertise that could help assess the quality of information received.

The recommendation acknowledges several potential paths to improving the overall expertise in Congress, especially among congressional staff who can sometimes draft prescriptive language that inadvertently limits flexibility in evidence building. One approach might be identifying appropriate expertise within each existing committee. Another approach might be establishing a new Joint Committee on Evidence, comprised of members of key legislative committees and supported by a staff comprised of experts in appropriate fields (e.g., economics, demographics, statistics, privacy and public administration). A joint committee could provide expertise and some level of oversight enable Congress's work to benefit effectively from a robust evidence ecosystem envisioned by the Commission.

LIKELY REACTIONS

- ***Congressional:*** Congress may not view decreased fragmentation of OMB as strategically valuable given the historic perception by Congress that OMB should have a limited role; however, Congress will recognize the important coordination role OMB can play within the Executive Branch. In terms of other recommendations, coordination with congressional committees would likely be key to success. Further work is needed to determine likely reactions to a recommendation about a potential joint committee, or increased expertise within existing committees.
- ***Executive:*** OMB may not willingly consolidate functions due to internal turf battles and perceived priorities for other functions within the institution, though on the whole the idea of better facilitating the ecosystem is likely to be viewed positively by OMB policy officials and staff. Similarly, across Departments, there will likely be some reluctance to engage in further intra-agency processes where they are not perceived as useful or in high demand. This may be mitigated by the inclusion of senior official commitments to pursue evidence-building through the appointment process and in performance appraisals. Finally, Departments may resist consolidation of activities within OMB unless there is a perceived benefit to agencies themselves, including through reduced burden in coordinating with fragmented components of OMB.
- ***State and Local:*** No negative reaction anticipated.
- ***Other:*** Researchers external to government may view the characterization of the "Federal Evidence Ecosystem" as limiting for their role; however, they are recognized as a critical component of the evidence-building ecosystem and can be incentivized to participate at a higher level based on access to the data facility and other recommendations included in *Research Partnerships* (Memo #12).

[20] BPC. (2017). Congress and Evidence-Based Policymaking: Creating a 21st Century Legislature. Draft. Washington, D.C.: Bipartisan Policy Center.

COST IMPLICATIONS

Overall, the recommendations included in this memo are envisioned to generate long-term efficiencies by better enabling research and evaluation about government programs in policies, which would likely achieve discretionary and mandatory savings over the long-term.
The vision of enabling OMB to better coordinate activities across government is not without tangible direct costs. Currently across the organization, 15-30 FTE are specifically devoted to the functions identified above, though some other related responsibilities are dispersed across other parts of the organization. Recommendation 7.1 envisions pulling those FTE together to better accommodate efficiencies across the operations, though fully enabling OMB to develop new standards (Memo #4), encourage data confidentiality efforts (Memo #10), and accommodate the interagency coordination envisioned in this memo will inevitably require further consideration of appropriate resource needs.

Intra-agency coordination is also not zero cost, but would be anticipated to be absorbed within existing resources at Departments. Similarly, every CFO Act Department already develops a strategic plan under GPRA, therefore the additional effort of including SERA coordination details is anticipated to be *de minimus* in terms of new direct costs. Recommendations 7.4 and 7.5 similarly are anticipated to be incorporated within existing Department processes and pose negligible direct costs.

For the legislative branch recommendation, the direct costs would be determined by how Congress approached the solution. If expertise was added within existing infrastructures, the additional direct cost for committees would likely be very low. However, if Congress elected to establish a new joint committee with supporting staff, the cost would largely be a function of the size of the staff. By comparison, the Joint Economic Committee receives a $4 million appropriation each year while the Joint Committee on Taxation appropriation is $10 million.[21]

ALTERNATIVE OPTIONS

- **Establish an Evidence ombudsmen positions in Congress and the Executive Branch.** Similar to the role of the Taxpayer Advocate in the Internal Revenue Service (IRS), well-positioned ombudsmen about evidence could help facilitate expertise within both the Legislative and Executive branches about SERA functions, including bill drafting. Ombudsmen could, for example, be attached to the Office of Information Policy established under Recommendation 7.1 or the committee recommendations in Congress under Recommendation 7.6. Such individuals would become obvious champions for SERA functions, though their level of influence may be minimal without specific authorities. Ombudsmen could also be used to address concerns among users of evidence in government regarding the credibility and quality of information, or as an independent authority to evaluate how well government is doing in implementing and achieving evidence-based policymaking. This option could be included in addition to other recommendations.

- **Direct GAO to Provide Oversight and Auditing Functions.** While much evidence-building will occur within individual topical domains, increasingly such work will span siloes across government. In order to facilitate oversight across department lines, the Commission could recommend that GAO play a role in overseeing the evidence-building community, in lieu of individual agency inspectors general. GAO could also be designated to conduct audits of the data facility, which would have an inherently interdisciplinary role.

[21] https://www.congress.gov/congressional-report/114th-congress/senate-report/258/1

Role of the Data Facility and Principal Statistical Agencies in the Federal Evidence Ecosystem

This paper expands on the Commission's discussion during the May meeting, and aims to outline a vision for how data infrastructure, access, and privacy protections can be optimized across the Federal Evidence Ecosystem by improving the coordination of certain functions for statistical activities. In particular, this paper highlights the role of the Data Facility and Principal Statistical Agencies (PSAs) in coordinating linkages for and access to confidential microdata within a highly protective legal environment, a secure technological infrastructure, and organizations that possess requisite expertise and experience.

* * * * *

In the Federal Government today, there are numerous components collecting, managing, and using data across all policy domains. Most of these diverse offices are specifically organized to administer programs, in some cases explicitly statistical programs; however, the vast majority of these offices collect and manage data for other than the purpose of producing evidence. The expertise, culture, and focus of these offices is the programs they administer. Unlike program administration, the production of evidence and facilitation of data access for evidence building to address today's complex questions invariably requires the linkage of data from across agencies – itself a practice necessitating specific expertise and careful practice. As confidential microdata are more frequently used for evidence building, encouraging more frequent and secure access to them can be beneficial, but uncoordinated access or an absence of standardization in how data are made available present clear potential for increased risk.

The provision of information is a core governmental function and governments will always have needs to fulfill and satisfy with some level of data releases, including reporting to Congress, public accountability, and production of microdata to support statistical activities. Increasing the use of evidence will only drive this need further, and while beneficial, must be balanced with the ability to protect the confidentiality of the data. However, the capacity of agencies to protect confidentiality, conduct disclosure review, and manage the risk of re-identification is uneven, particularly in agencies without specialized statistical operations. Of the offices that responded to the CEP Federal Office Survey and that collect confidential data, 38% provide external researchers with access to those data. Of those providing access, less than half (42%) conduct disclosure review of researcher output and less than half of those (38%) document the results of the disclosure review. One cannot reasonably recommend expanding access to microdata and additional linkages, without also improving the ability to measure privacy impacts in an ongoing way that consistently reevaluates the impact of ancillary information and new data releases. This suggests that success will require standardization of practice, by fewer actors with focused expertise.

Of the approximately 175 agencies that collect data, roughly 150 of them are program agencies whose work involves the day-to-day management of government programs. Nearly all of those program agencies perform at least a small amount of statistical activity as a secondary focus, which is to say they produce, handle, and examine data to create statistics or statistical analyses as one function. The remaining 25 include the 13 PSAs, most of which have existed for many decades, plus a few other mostly small and newer offices performing statistical or evaluation functions.

DRAFT

An Untenable Consolidated Statistical Agency

Past commissions and others who have studied statistical activities in the United States suggested varying degrees of centralization of the Federal government's statistical operations, similar to centralized statistical agencies in other developed countries. What would such an agency look like in the United States? A centralized statistical agency would require essentially combining the authority from all of the PSAs, allowing it to acquire data from all corners of the government, set rules for use and access, and employ many thousands of analysts, data scientists, information technology specialists, and administrators. As a reference, the 13 PSAs currently employ about 20,000 people, and its leaders would argue they are chronically understaffed.

There remain strong arguments for broad consolidation: as these agencies often have more in common with each other than they do their Departmental peers, and this Departmental heterogeneity is often the source of bureaucratic inefficacies (see Memos #7 and 9). But there is also real value in the dispersion of the statistical mission across government's many Departments. In particular, in the era of increased use of administrative data for statistical activities, proximity to program offices is of substantive value for both programs and PSAs. This diffusion also enables avoidance of the reality or perception of a central data repository, a central concern of the privacy community and a practical concern for data managers. Regardless of consolidation, there are still gaps in the system that must be addressed to improve data stewardship, the life cycle of administrative data, and effective access by government and non-governmental actors alike for evidence building. In other words, while much of the needed capacity may currently exist across a number of government components, there is still a need for new capacity and authority.

A Vision for Improved Coordination and an Optimal Arrangement

Between the existing 175 actors and the consolidation to a single actor, a middle ground may well best enable an "optimal arrangement." Building upon the foundation of the PSAs and recognized CIPSEA units, supplemented with a new Data Facility that will address existing gaps in privacy management, data access, data sciences R&D, and common services, the government can simultaneously achieve goals for improving secure access and enhancing privacy protections. By reducing the key actors responsible for responding to data access requests or combining datasets from as many as 175 to a number much closer to 14 and focusing the work of curating, stewarding, and protecting the government's data for evidence building to this smaller group of entities, we take a large step toward achieving the goals of consolidation envisioned for a single statistical agency. In essence, we are consolidating the roles of stewarding the government's data to a small group of agencies with a history of sound data management and protection. By stopping short of an actual consolidation to a single agency, we maintain an appropriate diffusion of authority, maintain proximity to programs and expertise, retain the disbursed stewardship of data, and balance both the program and statistical needs of the government. Additionally, we avoid the practical implications of such a substantial reorganization, which have historically been difficult to implement in the United States.

As discussed during the May Commission meeting, the Data Facility and the PSAs are crucial actors in the data infrastructure for combining data in the Federal Evidence Ecosystem, which, as noted, already includes hundreds of actors. Components that may eventually birth the Data Facility exist

DRAFT

today, but are not yet in an "optimal arrangement" to serve routine evidence building for Federal agencies or outside researchers. PSAs provide the foundational data and most of the existing infrastructure critical to evidence building, but lack consistent authority and capacity to acquire and integrate administrative data into production statistical datasets and derivative official statistics and analyses. Previous memos and discussions tended to focus on the independent roles of either the Data Facility or the PSAs; below we discuss how they work together within the ecosystem.

Status Quo

Data Facility. The closest construct for a Data Facility currently in government exists today within the Census Bureau infrastructure, specifically the Center for Administrative Records Research and Applications (CARRA) and the Center for Economic Studies (CES), including the group that manages the Federal Statistical Research Data Centers (FSRDCs).

The Census Bureau has a long history and culture of strong confidentiality protections and enjoys a high level of public trust. The Census Bureau's Title 13 authority protects its data and includes an authority to request administrative records. This authority, coupled with an exception in the Privacy Act (allowing agencies to provide data to the Census Bureau without obtaining consent) and a dedicated staff with expertise in negotiating MOUs, has provided Census moderate success in obtaining Federal and state administrative data. In addition, Title 26 authorizes Census to receive tax return information from the Treasury Department for the purpose of structuring of censuses and related statistical activities—although this broad authority is limited in practice as a result of Treasury Department regulations, interagency agreements, and project-specific approval requirements. The staff at CARRA have extensive expertise in secure data linkage and statistical methods for analyzing administrative data. The FSRDCs constitute an existing data enclave infrastructure that provides secure access to confidential data for outside researchers.

The Census Bureau's size means that it has more capacity than some of the other PSAs, but has also created an exhaustive bureaucracy that has been a source of frustration for Federal agencies and other entities trying to engage with the Census Bureau. Historic tensions between the Census Bureau and other agencies could also complicate a Census-based facility's ability to get their buy-in.

Principal Statistical Agencies. Congress provided specific authority to each of the Federal Government's 13 PSAs to perform an exclusively statistical mission and serve an enabling role for evidence generation in their topical domains. For example, Congress requires the Bureau of Transportation Statistics (BTS) to "collect, compile, analyze, and publish a comprehensive set of transportation statistics on the performance and impacts of the national transportation system, including statistics on [a number of specific topics]" and "to ensure that the statistics compiled…are designed to support transportation decisionmaking" by Federal, state, and local governments, transportation organizations, the private sector, and the public.[1]

PSAs have long served to fulfill the government's need for information; many are already combining survey and administrative data to produce descriptive statistics, statistical datasets, and analyses to inform policy and programs. For example, the Bureau of Labor Statistics (BLS) uses administrative data from the Employee Benefits Security Administration in the Department of Labor to refine their

[1] 49 U.S.C. § 6302

sampling frame for the National Compensation Survey and to code benefit details, thus reducing respondent burden. PSAs provide infrastructure within most Federal Departments to support day-to-day production and routine analysis of information for decision-making. Their history and culture is steeped in data protection, confidentiality, and stewardship.

Like Census, all other PSAs have implicit or explicit authority to acquire, protect, and use administrative data under their own statutes. For example, the Education Sciences Reform Act authorizes the National Center for Education Statistics (NCES) to use information collected "by other offices within the Institute and by other Federal departments, agencies, and instrumentalities." PSAs also benefit from some common legal authorities due to implementation by Congress and the President of key recommendations from past commissions. Most important was the enactment of CIPSEA as a common confidentiality law as recommended by the President's Reorganization Project for the Federal Statistical System.[2] CIPSEA provided an authority common to all PSAs to protect data, including administrative data, acquired for statistical purposes and enabled limited sharing among a subset of the PSAs.

All PSAs have the obligation to disseminate data for secondary analysis. They each have procedures for providing controlled access to confidential data and most have developed tiered access programs for wider distribution of aggregated and masked data. They have significant expertise in statistical disclosure limitation techniques and in disclosure review of statistical products. The workforces of the PSAs share a common emphasis on statistical and computer sciences, analytical disciplines, and data stewardship. They are pioneers within their respective disciplines.

Under the status quo, PSAs vary in their readiness to take on greater responsibility for evidence-building activities given legal, policy, and institutional placement differences. The current missions of the PSAs are focused on the rigorous and objective production of statistical datasets and statistics in each one's respective topical domain, as well as providing researcher access to their own datasets. Collectively these agencies comprise an expert workforce of more than 20,000.

The needs of a mature evidence ecosystem require information policy and collection coordination across government, a role that is largely underserved in the current environment. The PSAs, working in concert, are ideally situated to take on this mission within their Departments, as well as technical support for program offices in collecting, managing, and releasing data with strong privacy protections. Some departments do not have a PSA though some of them have an office with a statistical focus that may be able to serve as a point of contact for internal and external coordination. For departments with no such office, Federal and external researchers often have no obvious point of entry to learn about or access data resources for evidence building. This is a capacity gap that must be addressed.

All PSAs have implicit or explicit authority to *acquire* administrative data, but none has the authority to *require* it. The Commission discussed as part of the "Yes, unless" proposal in Memo #3 on Legal Frameworks a recommendation that Federal program agencies *shall* provide administrative data for statistical purposes. Such a recommendation addresses Federal agency and external researcher needs for access to administrative data for evidence building. The Commission's charge also includes

[2] Bonnen, et al. *Improving the Federal Statistical System: Report of the President's Reorganization Project for the Federal Statistical System*, reprinted in *American Statistician*, vol. 35, no. 4 (November 1981), pp. 184-196.

determining the optimal arrangement for access to survey and other statistical datasets. As PSAs increasingly create statistical datasets inextricably blended from administrative and survey data they collect and curate, the optimal arrangement should address Federal evidence-building community access to such statistical datasets.

One particular set of needs is for PSAs to be able to use one another's data to improve efficiency or quality of statistical series. For example, Census, BLS, NCES, NCSES and SOI all acquire data about universities in their various datasets. In a fully optimized state, they would benefit from each other's sampling frames, frame variables, and commonly-useful survey and administrative data content. As a further example, most PSAs conduct expensive survey screening operations to identify the specialized populations relevant to their statutory missions to be the principal source for statistics in a specific topical domain. The Census Bureau's American Community Survey is the largest household survey conducted by the Federal government, and is designed in part to permit follow-on surveys of specialized populations. When another agency sponsors such a follow-on survey, it cannot see the resulting confidential data because neither Title 13 nor CIPSEA allows sharing of such data from Census to the other agency at present. Given the centrality of such a dataset to a PSA's mission and day-to-day work, it is often untenable to operate under such practical constraints. One solution is to focus such activity within the new Data Facility. However, as PSAs increasingly rely on blended datasets for their core work, a larger and larger percentage of their staff will work with such data day-to-day. Requiring all PSAs to go through the Facility to share data would be akin to creating a centralized National Statistical Office. Short of this solution, which creates a single organizational bottleneck among competing statutory and operational requirements, the next alternative would be a common legal framework in which data can be shared among PSAs confidentially for exclusively statistical purposes with full transparency and safeguarding.

According to one BLS senior leader, barriers to greater use of administrative data for statistical production include "the potential for bias in outside alternative data, the potential for variables to be added to or dropped from external datasets, the potential for the meaning/definition of variables to change suddenly, and the need to fully evaluate the legal issues associated with informed consent for external alternative data where applicable."[3] Other PSAs cited these same barriers during the Commission's fact-finding stage. Office of Management and Budget (OMB) policy tells program agencies to consult in advance with their associated PSA before making changes to administrative data used for evidence building, but follow up has been uneven.

Other SERA Offices. Another dozen or so offices, such as HUD's Office of Policy Development and Research, perform statistical, evaluation or policy-related research activities, using survey and administrative data. These offices are organizationally distinct from program offices, yet are not recognized statistical agencies. Some could perform duties envisioned for PSAs, such as internal departmental coordination, and could be considered for recognition as a PSA in the future, particularly in departments that do not have a PSA.

Program Agencies. The existing Federal Evidence Ecosystem includes around 150 offices that run programs and produce administrative data, most of which conduct some statistical activities, often as a byproduct of the program work. Because administrative data are primarily collected for addressing

[3] Staff communication with Ken Robertson, BLS Assistant Commissioner for Industry Employment Statistics.

DRAFT

day-to-day program activities, such as reaching eligibility determinations or monitoring compliance, administrative data are not principally collected for statistical activities, though they can be invaluable for improving official statistics and for evaluating activities. An existing OMB policy (M-14-06) encourages program agencies to consult with PSAs and other SERA offices on changes to key administrative data collections, which can minimize adverse or unintended effects on those data when used for evidence building.

Future State: A Fully Mature System

Data Facility. In a fully mature Evidence Ecosystem, the Data Facility is a Center of Excellence for privacy-protected data access. Formed originally from CARRA and CES, the Facility has inherited the Census Bureau's mantel of data stewardship and public trust. Public transparency is paramount; the Facility provides public notice of approved projects for accountability about allowed, statistical-only uses of the Federal Government's data. The Facility has worked with external experts and the private sector to conduct research and development on privacy enhancing technologies. This investment has resulted in production-ready Secure Multi-party Computing (SMC) or a similar privacy protective process that enables the Facility to link data for analysis without having to physically acquire and store it. Because the data are rarely transmitted or stored, the Facility is hackproof. The Facility has made great strides in provable privacy methods such as *k*-anonymity and differential privacy to create datasets with minimal and quantifiable risk of re-identification. As a recognized PSA, the Facility enjoys strong privacy protection through an amended CIPSEA Subtitle A and data sharing authority through an amended CIPSEA Subtitle B. In collaboration with the other PSAs, the Facility gathers information on the use of the privacy budget across Departments in accordance with the amended law's re-identification risk management framework and reports it to OMB in compliance with established guidance.

Researchers have easy access to a complete inventory of data and metadata to use in evidence production. The Facility has substantial expertise in the linkage and analysis of confidential microdata; it has state-of-the-art query tools and data-masking processes for making linked data available in multiple tiers of access. The process to request access to linked confidential microdata is clear and consistent across datasets. External researchers undergo a qualification and training process that applies at the Facility and for accessing datasets held by the PSAs.

The Data Facility facilitates collaboration within the Federal Evidence Ecosystem. It has the authority to receive funds from other Federal agencies to fund operations and joint activities. Because of the importance of accurate metadata for SMC-like technologies, the PSAs and the Facility help program agencies develop and then maintain administrative metadata. Some PSAs and other SERA agencies look to the Facility to develop and maintain survey and other statistical series metadata, rather than maintaining that capacity in house. Some PSAs, other SERA agencies, and program agencies use the Facility to manage researcher access to their own data rather than doing it themselves. Some other agencies pay the facility to conduct their disclosure review according to OMB standards. The Facility and the PSAs jointly fund projects of mutual benefit, such as the Joint Program in Survey Methodology at the University of Maryland.

Principal Statistical Agencies. In a fully mature Evidence Ecosystem, the PSAs are thriving producers of blended statistical and administrative datasets used to produce official statistics and

analyses and made available to the Data Facility, other Federal offices, and external researchers for evidence building. They are experts in their topical domain and the stewards and owners of their own data also accessed by the Facility as required for statistical uses, through SMC or a similar privacy-protected network protocol.

Each PSA plays a key role in the stewardship of data resources within its Department. They are involved in high-level Department discussions about changes to programs that affect administrative datasets used for evidence building. Program and other SERA offices within each Department look to the PSAs for technical assistance in meeting OMB standards for managing re-identification risk and conducting disclosure review.

Because of their in-depth knowledge of their own data resources, PSA staff serve as expert consultants to Data Facility staff. Each PSA routinely details staff to the Facility to provide support and to develop shared expertise throughout the ecosystem. Periodically, PSAs ask the Facility to work with them to research new statistical products using linked data. Because the Facility and PSAs have CIPSEA Subtitle B authority and responsibility to share data while notifying the public of intended statistical purposes, PSAs are able to access data needed to produce and maintain new statistical series as they go into production.

As the ecosystem matures, some PSAs or administrative units that currently support them become Centers for Excellence in other areas critical to the routine production and use of evidence for policymaking. They may develop capacity and special expertise in functions that require nuanced understanding of evidence-building purposes, including procurement, partnerships with external organizations, information collection review, and human resources. Other PSAs and the Facility look to these Centers for consultation and support.

Other SERA Offices. In a fully mature Evidence Ecosystem, additional offices and agencies have become recognized CIPSEA units by demonstrating an ability to the meet requirements such as functional separation between statistical and other functions. These agencies engage with the Data Facility in much the same way as the PSAs.. Whether they pursue such recognition or not, they like program agencies, operate under clear guidance about requirements for protecting the privacy of confidential data and the importance of managing the risk of re-identification. If they have the capacity themselves, they manage their own privacy-protected researcher access, conduct disclosure review, quantify use of the privacy budget, and report it to the Facility. If they do not have such capacity, they work with a PSA in their Department or with the Data Facility to perform those functions to standards.

Program Agencies. In the fully mature Evidence Ecosystem, the 150 Federal program agencies also have clear guidance about privacy requirements and either develop capacity to meet them or get them met by their PSA or the Facility.

Program offices also work with the PSAs and the Data Facility to develop, maintain, and update their data documentation—including the metadata so critical to SMC and similar technologies. They understand the importance of their data for evidence building, so they communicate with their PSA on plans to change or eliminate administrative data collections.

DRAFT

Other SERA offices that need access to confidential data for evidence building designate staff to apply and train as eligible researchers. Those researchers can access linked data through the Facility or through a "node," the PSA in their own Department. The ability of these offices to securely analyze linked data results in a proliferation of policy-relevant research and routine evaluations to measure implementation and outcomes of Federal programs.

Initial Stage of Transformation—From Status Quo to Ecosystem Orientation

The initial stage of the transition will require the re-orientation of the PSAs and the components that will become the Data Facility towards an optimal arrangement for the Evidence Ecosystem. This stage will establish the consistent legal and policy framework required for efficient evidence production.

Data Facility. At its inception, the Data Facility would require statutory authority to do the following: constitute itself as an entity within the Census Bureau, with authorization for the Secretary of Commerce to create it as an independent Bureau within the Department or an independent entity attached to the Census Bureau at an appropriate stage of maturity (Recommendation 2.1.3); operate as a PSA (Recommendation 2.1); require agency sharing of administrative and survey data for exclusively statistical purposes (Recommendation 3.1); enter into contracts, grants, and cooperative agreements (Memo 12); sponsor a Federally Funded Research and Development Center (FFRDC) (Recommendation 2.1.7); and receive and use interagency fund transfers to support Facility operations and joint activities (Recommendation 12.3.1).

The Facility would begin the process of becoming the "one-stop shop" for external researcher access to confidential data by working with OMB to establish a transparent process for project application and approval and working with Departments to develop an inventory of existing data available for evidence building. At this initial stage, the Facility would focus on acquiring datasets critical for evidence building that pose significant access challenges. Because evolving privacy-protecting technologies are still in development, the Facility acquires the minimum amount of microdata necessary to perform linkages for specific projects. It would transparently provide public notice about planned statistical uses of acquired administrative data and apply best practices for the protection of direct identifiers used in linking.

Principal Statistical Agencies and Other SERA Offices. In this initial stage, the PSAs orient their expertise and activities towards coordinating data resources within their Departments and begin to develop specialized capabilities as resources for the Facility and other PSAs. Each Department designates the head of a PSA, or a similarly senior-level SERA office in the absence of a PSA, as the Department's "Senior Agency Official for Data Policy" (Recommendation 9.2). PSAs and the designated SERA offices provide technical assistance and support to program agencies in their departments to develop and maintain metadata for administrative records; these metadata are crucial to the Facility's pursuit of new technologies such as SMC. PSAs share data with the Facility under its new authority. They continue to any ongoing data sharing arrangements with one another under existing (mostly one-off) data sharing agreements and provide researcher access to confidential data using current authorities and procedures. PSAs detail staff to help stand up the Facility and lend topical domain expertise; they help to fund Data Facility operations and joint projects. Some PSAs

provide funds to the Facility for services such as external researcher access to confidential data and disclosure review.

Program and Other SERA Agencies. In the initial stage of the transformation to a mature evidence ecosystem, program agencies and SERA offices support the Data Facility and PSAs in creating a comprehensive list of administrative and survey data available for evidence building. They collaborate to develop accurate documentation and metadata for their data resources and include PSA staff in discussions about potential changes to those resources.

Interim Stages of Transformation—From Ecosystem Orientation to Evidence Production

The next stage of transformation sees the Data Facility moving out of the Census Bureau into a separate Bureau within the Department of Commerce or an independent entity attached to the Census Bureau. At this stage, the Data Facility has established research partnerships with external experts and organizations, developed expertise in creating tiered access datasets from linked confidential data, and is in the process of piloting privacy protective technologies.. The Facility and PSAs all operate under updated OMB guidance based on amended CIPSEA requirements for managing the risk of re-identification from the release of results based on confidential data.

As the ecosystem matures, the PSAs become the stewards of data within their own departments used for statistical activities, freeing up Facility capacity to focus on privacy protection and external researcher access. The rotation of PSA staff in and out of the Data Facility on detail expands content and technical expertise in the broader evidence-building workforce.

Program and other SERA agencies look to their PSAs as expert consultants and resources for maintaining evidence uses of survey and administrative data. They become qualified researchers and connect with data from other departments through their PSA node. Those agencies with sufficient capacity to protect, release, and manage re-identification risk for confidential data continue to do so; others go through the Facility or their PSA.

* * * * *

DRAFT

RECOMMENDATION MEMO #8
UPDATED BASED ON MAY 1 COMMISSIONER CALL

Institutionalizing and Incentivizing the Evaluation and Policy Research Functions Across Government

Commissioner Leads: Glied and Haskins
Staff Leads: Hart and Fletcher

Building on the Federal Evidence Ecosystem (Memo #7), which describes a vision for how a coordinated Federal system for producing and using evidence could operate, this memorandum specifically addresses strategies for institutionalizing and providing incentives to support the evaluation and policy research functions in Federal Departments. *[Sec. 4(a)(1), 4(a)(3)] Relates to Ecosystem Overview (#7), Facility (#2), Program Design (#13), Administrative (#15), and Funding (#16).*

* * * * *

BACKGROUND AND ANALYSIS

The Commission was specifically charged with determining an optimal arrangement of factors to "facilitate program evaluation, continuous improvement, policy relevant research…" (Sec. 4(a)(2)) and to "make recommendations on how best to incorporate outcomes measurement, institutionalize…[evaluation]…into program design" (Sec. 4(a)(3)). While much of the Commission's emphasis has been on how to make data available for building evidence, there are critical components of infrastructure for users of data that the Commission may also choose to address.

Currently, while many agencies have robust performance and statistical (S) functions that provide a broad perspective on policy issues, few have well developed evaluation (E) and policy research (R) functions that can provide precise analyses of policies and programs.[1] Consequently, the generation of evaluation and research about Federal policies using available data is intermittent, occurring only when the right series of factors align for sufficient capacity to emerge. Because of the lack of institutionalization, access barriers to data may not be the most important challenge hindering the growth of evaluation and policy research. This memorandum proposes a set of recommendations designed to directly address the institutionalization of these functions as well as a set of incentives designed to create and support a growing role for evaluation and policy research in the future.

The evaluation and policy research functions are emphasized in this memo, separately from other aspects of the Federal evidence ecosystem, precisely because the functions are more nascent in government today than other areas of the evidence building community. Thus, the evaluation and policy research functions require different solutions for bolstering their development across Federal Departments. The functions lack many of the existing structures and institutions found in, for example, the Federal Statistical System (see Memo #9) that benefits from explicit statutory authorities, formal organizing structures, and a long history and established culture in government. While some exemplars for evaluation and policy research *do exist* in the Federal government, more work is needed to meet the vision of the Commission and the charge outlined by Congress for better supporting these activities in government.

[1] The American Evaluation Association's critique of Federal evaluation suggests "for the most part, these evaluations have been sporadic, applied inconsistently, and supported inadequately. Too often, the units formed to conduct evaluations are short lived and under-resourced. Training and capacity building for evaluation have been inconsistent across agencies and, in many cases, insufficient to achieve the needed evaluation capacity and sustain it over time." See AEA. (2013). A Roadmap for a More Effective Government. AEA Evaluation Policy Task Force. Washington, D.C.

Defining Evaluation and Policy Research

Evaluation and policy research are related activities, but each with distinct roles in the evidence-building community and for users in the evidence ecosystem. *Evaluation* involves systematic application of methods to assess programs and policies, typically with formal recommendations for changes to improve or modify programs. [2] Because this role is often determining merit or worth of activities with value judgements, evaluation is itself an inherently political function that can be criticized for subjectivity when stakeholders or interests object to the conclusions of an evaluation.

Policy research[3] is naturally related to evaluation, and often precedes formal evaluation, but offers a fundamentally different role in the evidence-building community. Both applied and basic research are defined by OMB as activities in pursuit of new knowledge, the former directed at general applications or theory and the latter focused on practical objectives.[4] Policy research can then be considered gaining knowledge about policies and programs. For example, research about SNAP participation rates and the level of benefit penetration in a region would be an example of research about a program.

Within the context of evaluation and research activities, there are a range of organizing structures and processes pursued by Federal Departments. In practice, formal Executive Branch evaluations are most often conducted for exploring modifications to existing law through demonstration projects (e.g., Social Security Disability Insurance), testing new interventions or program changes (e.g., welfare waivers under 42 U.S.C 1115), or process/implementation studies. Rarely are evaluations in the Federal government truly about whether a program "works," relative to an absence of the program altogether. For example, even during the Obama Administration's recent efforts to encourage more limited evaluation in the Federal government, Departments primarily began exploring applications of more rapid cycle evaluation projects. These activities frequently focused on relatively small behavioral "nudges" which involved A/B testing changes in sequencing of actions or text in letters, for example.[5] The Program Design Memo (#13) will specifically address recommendations for how evaluation and other evidence considerations can be included in both new and long-operating programs.

Relationship to Other Actors in the Evidence Ecosystem

The Federal Evidence Ecosystem (see Memo #7) asserts that all statistical, evaluation, policy research, and policy analysis (SERA) functions are important in helping agencies achieve their missions, and must be appropriately resourced and coordinated to achieve their full potential. Information collected routinely by the statistical system can serve as vital input to monitoring outputs or outcomes for evaluations, particularly longitudinally. Analysts and knowledge brokers or intermediaries serve an important role in helping to translate evaluation and research findings back to decision-makers and the public. Decision-makers and program administrators also play a

---

[2] AEA (2013). According to the American Evaluation Association (AEA), "evaluation involves assessing the strengths and weaknesses of individual programs, policies, and organizations to improve their effectiveness, efficiency, and worth…It uses systematic data collection and analysis to address questions about how well government programs and policies are working, whether they are achieving their objectives, and, no less importantly, why they are or are not effective."

[3] Used synonymously with "policy-relevant research."

[4] OMB Circular A-11, Section 84.

[5] OSTP. (2016). Social and Behavioral Sciences Team Annual Report. Washington, D.C.: Office of Science and Technology Policy (OSTP), Executive Office of the President. https://www.whitehouse.gov/sites/whitehouse.gov/files/images/2016%20Social%20and%20Behavioral%20Sciences%20Team%20Annual%20Report.pdf

necessary role within the ecosystem in identifying priorities for SERA activities.[6] Realistically, in a world where resources for government operations are zero-sum, programs must prioritize where information needs are greatest to best serve the need for accountability to the American public while accommodating continuous learning for program administrators.

Brief History of Federal Evaluation and Policy Research

While the policy research and analysis functions emerged in the Federal government following World War II, largely as an outgrowth of the Department of Defense's needs to improve cost-benefit analyses of major procurements, more formal program evaluations evolved as a reaction to the Great Society programs in the 1960s and 1970s. The former Department of Health, Education, and Welfare (HEW) pioneered many of the approaches and techniques for executing the function in the Federal government over the past half century. HEW established the Assistant Secretary for Planning and Evaluation (ASPE), which created a basic infrastructure and resource mechanisms to support evaluation practices, with integration across HEW. Over time, as HEW responsibilities were dispersed over four different cabinet-level Departments and agencies, much of the evaluation emphasis transferred into the new institutional arrangements.[7] Today, two of these four HEW components are often cited as exemplars for institutional support of evaluation: the Department of Health and Human Services (HHS) and the Department of Education.[8] The other two components carried some of the similar authorities and mechanisms that were provided to HEW into the new organizations.

As early as 1973, calls for establishing a formal evaluation system in the Federal government emerged. Notably, the Urban Institute's early recommendations about Federal evaluation policy included expansions of overall capacity and infrastructure, including by creating evaluation responsibilities within research offices.[9] The formal "system" envisioned by the Urban Institute never materialized and instead culminated in a disparate, decentralized, heterogeneous collection of evaluation units and interests across Federal Departments, which as noted above are not consistent in their status within Departments, expertise, or resourcing.

Within offices that have substantial responsibilities for evaluation and policy research today, the statutory authorities to pursue evaluation are varied, as are the funding sources and levels of technical expertise. For example, the National Center for Education Evaluation and Regional Assistance in Education was established explicitly by statute, whereas the Office of Planning, Research and Evaluation (OPRE) at the Administration for Children and Families (ACF) in HHS emerged as a reaction to a need to coordinate policy research and evaluation activities across human service programs.[10] Policy research activities, which are defined by some Departments to include evaluation, can be housed in areas of Departments such as each agency's central policy office or R&D office, or even within individual programs or bureaus. The Department of Housing and Urban Development's (HUD) Policy and Research Division (PD&R) operates under the general authority of a research office in statute, but conducts a range of demonstration and evaluation projects.[11] To the extent central evaluation and policy research functions are not present, Departments delegate the evaluation function essentially to individual program offices. In such cases, evaluation and policy

[6] See Patton, M.Q. 2008. Utilization-Focused Evaluation. Fourth edition. Los Angeles: Sage Publications, Inc.
[7] Functions transferred to from HEW to ED (1979), SSA (1994), and EPA (1970), with residual in HHS.
[8] GAO. (2011). Program Evaluation: Experienced Agencies Follow a Similar Model for Prioritizing Research. GAO-11-176. Washington, D.C.: General Accountability Office (GAO). http://www.gao.gov/assets/320/314607.pdf
[9] Wholey, J. et al. (1973). Federal Evaluation Policy: Analyzing the Effects of Public Programs. Washington, DC: The Urban Institute.
[10] CRS authority summary to CEP
[11] CRS report for CEP

research compete with program services for attention and resources, such as when optional set-aside funding authorities are present and left to the discretion of Departments or individual programs.

Current Barriers to Evaluation and Policy Research Functions Across Government

Despite decades of experience conducting or funding evaluations of Federal policies and programs, the function has not been institutionalized across the Federal government due to a complicated morass of bureaucratic and institutional constraints coupled with a historic lack of consistent demand for studies.[12] To the extent Congress and the President have articulated a demand for these types of activities today, including through the creation of the Commission, the recommendations here focus on the identified production barriers.

Challenge #1: Uneven capacity. The capacity to perform evaluation and policy research activities across Federal Departments is uneven, with some Departments having a strong centralized approach for conducting this function, some Departments having the capacity for the function but dispersed across units and poorly coordinated, and other Departments having no capacity to conduct evaluation or policy research at all.

To the extent capacity is uneven, other priorities within Departments may be more likely to frequently crowd-out a high-performing and continuous evaluation function within Departments. For example, requirements to develop annual performance measures are aimed at providing timely information for decisions in the budget process, and in most Departments these mandatory activities consume an array of resources. However, to the extent performance measurement (or monitoring) is considered a subset of the broader field of evaluation,[13] the focus on short-term activities has crowded out the ability to prioritize long-term projects that may not align with the immediate time horizons of policymakers and program administrators. The capacity must exist for both short- and long-term activities. Additionally, as Memo #7 on the Federal Evidence Ecosystem suggests, the fragmentation of offices within OMB limits high volume production of policy research and evaluation evenly across government. However, some agencies with authority have been able to establish robust evaluation and policy research portfolios, even with a fragmented OMB structure.

Challenge #2: Difficulty hiring appropriately skilled staff. Federal agencies report having difficulty hiring staff with the appropriate skill set to match the work requirements for the program evaluation function. Much of this limitation may be closely associated with the nascent field of evaluation professionally: evaluation itself has only been recognized as a profession over the last several decades and does not bear the long history evident in, for example, the field of statistics. As the field evolved and matured, the methods and approaches for developing useful and relevant evaluations that meet appropriate standards for rigor and validity have improved. But the skills for conducting evaluation and policy-relevant research have also evolved in recent decades and may not be reflected widely in the current Federal workforce.

For Departments that have a strong existing infrastructure in the area of program evaluation, the issue of capacity is more fine-grained. Federally-supported program evaluation is typically carried out in one of two ways—internally, in which Federal staff carry out the program evaluation effort themselves, or externally, in which the program evaluation is carried out by a firm or other external

[12] See (1) Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation* 29(4), 443-459 and (2) Russ-Eft, D. and H. Preskill. (2009). *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change*. Second Edition. New York: Basic Books.

[13] Newcomer, K., & Brass, C. (2015). Forging a Strategic and Comprehensive Approach to Evaluation Within Public and Nonprofit Organizations. *American Journal of Evaluation* 37(1): 80-99.

agency under contract to the Federal government. There is, therefore, a need for staff capacity in both the conduct of program evaluation, as well as the management of contracts that support program evaluation. Contract management in support of research and program evaluation efforts tends to be a more complicated subset of the contract management function with riskier contract types, higher dollar values, and/or more challenges to assessing successful performance, thus requiring staff with both substantive program evaluation expertise as well as contract management skills.

Challenge #3: Lack of incentives, or the presence of disincentives. Even with the right organizational structure and sufficient staff capacity, Federal Departments may still not engage in meaningful evaluation and policy research without the right set of incentives.

Perceptions among program managers for how evaluation or policy-relevant research may be used to a program's detriment are an oft-cited reason for aversion to evaluation.[14] During the George W. Bush Administration, some Departments perceived that the inclusion of evaluation in the Program Assessment Rating Tool (PART) was more likely to be used as a punishment for pursuing evaluation than as a benefit.[15] Similarly, within the context of reaching centralized budget decisions, evaluations that suggest program outcomes are not being fully met can be perceived as causing harm to a program's interests rather than necessarily being used as a tactic to improve programs. Combined with the recognition that most of the funding mechanisms for evaluation activities are permissive rather than directive (i.e. ability to use a portion of funds from a set-aside), program administrators perhaps rightfully but too frequently perceive evaluation as a contradiction to achieving their programmatic missions, informed by their own expertise. These negative perceptions limit the perceived utility of supporting evaluation and research activities.

In turn, the historical constraints on institutionalization of the functions suggest tremendous value for the Commission in making strong statements about the need and role for evaluation and policy-relevant research. Such a strong statement could help outline a definitive transition from the mere hope that Departments pursue these activities to an expectation that they will, and the establishment of supports that will facilitate successful implementation. Perhaps more importantly, recognizing the multiple needs for evaluation and policy research activities to both improve programs and be used for accountability recognizes that the tools can be used for program and service improvement that benefit the lives of members of the public. Thus, a call for a greater degree of institutionalization offers the opportunity to both signal the role of and expectation for these evaluation and policy research activities.

SUMMARY OF COMMISSIONER CALL ON MAY 1

Co-Chairs Abraham and Haskins, and Commissioners Glied, Hahn, and Shea participated in a discussion with Commission staff on May 1, 2017, about an earlier draft of this memo. Participating commissioners considered the background and framing of challenges appropriate. Recommendation #1 was slightly reframed below to make a stronger requirement. Recommendations below on incentives were modified only slightly, and several individual-level recommendations were moved to Memo #7 and applied to the entire Federal Evidence Ecosystem.

DRAFT RECOMMENDATIONS

[14] Various in ECB literature (Footnote 12), CEP staff discussions

[15] However, some agencies have also suggested that the question enabled programs to pursue evaluation contracts that otherwise would not have done so. GAO asserted a similar conclusion in GAO. (2005) Performance Budgeting: PART Focuses Attention on Program Performance, but More Can Be Done to Engage Congress. Washington, D.C.: GAO. http://www.gao.gov/products/GAO-06-28

As described in this memo, Federal evaluation capacity and activity is uneven across the Federal Government. The Commission received input from numerous Federal Departments that currently operate high-functioning evaluation and policy research units, documenting a series of barriers that prevent their organizations from maximizing their evaluation and policy research efforts. Many of these barriers relate to procurement and other bureaucratic hurdles that discourage evaluation, or a lack of coordination *across* evaluation and policy research offices within Departments. These barriers are more unique to Departments who already have a high-functioning evaluation and policy research unit, and recommendations directly addressing these barriers will be included in memos on Administrative Activities (#15) and the Evidence Ecosystem (#7).

The recommendations proposed in this memo are intended to advance the institutionalization of the evaluation and policy research function across *all* Federal Departments and to incentivize the implementation of these activities. To this end, the recommendations fall into three distinct categories: (1) ***requiring*** Departments to establish an evaluation and policy research unit if one does not already exist, and/or strengthen and centralize this function in Departments where the activities may be disconnected across the organization; (2) ***enhancing basic capacity*** to ensure that Departments are properly resourced with the appropriate staff to execute a meaningful portfolio of evaluation and policy research activities, and (3) ***providing incentives*** for institutionalizing and prioritizing the functions, through funding, administrative flexibility, or recognition.

Requirement Recommendations

❖ **Finding:** Programs may contribute to low demand for evaluation, or be reluctant partners in an evaluation, for fear that evaluation findings may reflect poorly on their program and be used to de-fund or dramatically alter the program in a way they perceive as negative. The Commission recognizes there are multiple purposes for evaluation activities, including both accountability and continuous learning, and finds that the need for continuous learning and improvement is tremendous.

❖ **Finding:** The capacity to perform evaluation and policy research activities across Federal Departments is uneven, with some Departments having a strong centralized approach for conducting this function, some Departments having the capacity for the function but dispersed across units and poorly coordinated, and other Departments having no capacity to conduct evaluation or policy research at all.

►**Recommendation #8.1 (PRIORITY):** The Commission recommends that Federal Departments [16] identify or establish a Chief Evaluation Office responsible with (1) establishing evaluation and research policies to apply Department-wide that encourage rigor, credibility, independence, and transparency, (2) coordinating and supporting technical expertise for the functions within the Department, and (3) identifying and setting priorities for Departmental program evaluation and policy research, with appropriate attention to the mission and context of each Department. In addition to the Chief Evaluation Offices, Departments should also be encouraged to establish or identify other centers for evaluation, as appropriate.

During the Commission's fact-finding phase, numerous stakeholders and Co-Chair Haskins suggested having someone in charge of the evaluation function in Departments, such as establishing a Chief Evaluation Officer. Establishing an expectation for a role within each Federal Department

[16] For the purposes of these recommendations, "Department" refers to the 24 Departments and agencies identified in the Chief Financial Officers (CFO) Act.

would help ensure the presence of a champion, or group of advocates, who can also coordinate technical aspects of evaluation throughout agencies, as well as developing and maintaining the policies for conducting the activities.[17] Given the decentralized and loosely coordinated operations of most Departments' evaluation and policy research activities, this notion of greater centralization and coordination is a valuable approach to strengthening these functions.

By establishing or identifying a single lead office within each Department, the role of evaluation and policy research will be necessarily elevated to receive the attention of senior leaders (e.g., DOL's Chief Evaluation Office). Beyond the lead office within each Department, some agencies may need flexibility that suits their own contexts. In such cases, consistent with comments provided by the American Evaluation Association, some Departments may appropriately need multiple central hubs to meet needs (e.g., distinguishing between farm and nutrition programs at USDA). This recommendation sets a floor for the expectation that each Department be responsible for implementing these functions, while encouraging some flexibility to determine how the evaluation and policy research functions may be most effectively integrated within existing structures, potentially building on existing infrastructure or institutional arrangements.

The value of a coordination function for evaluation and policy research activities is that gathering and cataloging expertise in conducting evaluations and managing evaluation contracts can provide a resource to individual programs that wish to pursue, or are directed to execute, evaluations in the future. The Chief Evaluation Office will also play a direct role in increasing the objectivity and independence of evaluation activities in every Department. Staffing the offices with a senior official also provides a natural champion for encouraging greater production of evaluation within Departments.

The Chief Evaluation Office, and other evaluation centers in a Department, can also play a role in conducting evaluability assessments (i.e., determining whether programs are at an appropriate level of development for conducting different types of evaluation) and help prioritize evaluation and research studies for the Department. To the extent a Department views statements about evaluation policy useful, such as those issued by USAID, the U.S. Department of Housing and Urban Development (HUD), DOL, and HHS/ACF over the past three years, the coordinating center(s) would also serve as a natural place to coordinate the development of such statements and policies. Further, the designated Chief Evaluation Officer would ensure that the Department's evaluation community had adequate representation on the Interagency Council on Evaluation Policy (ICEP) described in the ecosystem discussion (Memo #7).

Basic Capacity Recommendations

During the process of establishing Chief Evaluation Offices for coordinating and supporting Departmental program evaluation and policy research as described under Recommendation #1, Departments may identify gaps in staff capacity that could inhibit a high-quality expansion of these functions. Specifically in support of the program evaluation function, staff with the appropriate skill set must be recruited and provided a meaningful career ladder. Departments seeking to generate capacity in program evaluation would need staff expertise in both the conduct of program evaluation, as well as the management of contracts that support program evaluation. The following recommendations relate to enhancing staff capacity to ensure that the skill set of the staff match the skill set required for a more robust portfolio of program evaluation and policy research.

[17] Haskins, Arnold, Nightingale

❖ **Finding:** Federal Departments report difficulty hiring staff with the appropriate skill sets to match the work requirements for the program evaluation function.

► **Recommendation #8.2:** The Commission recommends that the President direct the Office of Personnel Management (OPM) to take immediate steps to support the growth and development of the program evaluation field, including the establishment of occupational structures to address critical needs.

A relatively small step the Commission could recommend is for the Office of Personnel Management (OPM) to collaborate with agencies that already have high functioning program evaluation offices to identify steps to support the growth and development of the program evaluation field through a set of human capital strategies. Currently lower volume evaluation units may struggle to hire sufficient staff with expertise to scale efforts within an agency. OPM could identify mechanisms to assist Chief Evaluation Offices to develop hiring strategies that enable the recruitment and retention of professionals with the unique skills associated with managing evaluation contracts or directly executing evaluation in the Federal government. Multiple evaluation offices acknowledged challenges in hiring sufficient expertise in the Commission's Survey of Federal Offices. Steps might include the establishment of a targeted occupational series that would allow Departments flexibility to hire based on specific criteria most relevant for the position and the function, and help smooth tensions between evaluation units and human resource officers.

This recommendation builds on and goes beyond the requirement under GPRA Modernization for OPM to identify skills and competencies needed for performance management, including "evaluating programs," and incorporating those competencies in *existing* position classifications.[18] For example, by carving out program evaluation as a unique occupational series, Departments would be better able to tailor their recruitment and staffing efforts to ensure sufficient human capital to initiate and oversee a robust portfolio of program evaluation.

❖ **Finding:** The existing performance measurement and management network established under GPRA and GPRA Modernization provides a productive infrastructure for expanding evidence-building activities across government, though too much emphasis has been placed on metrics that are measurable within the short-term.

► **Recommendation #8.3:** The Commission recommends that Federal Departments assess their existing capacity to staff SERA functions, including identified evaluation and research offices, and reallocate resources to the extent possible and practicable from short-term performance monitoring activities to longer-term evaluative and research functions to ensure an appropriate balance among the activities.

To the extent that gaps in capacity are assessed while establishing Chief Evaluation Offices for coordinating and supporting Departmental program evaluation and policy research, Departments may choose to reallocate staff or resources from more short-term performance management functions.

Incentive Recommendations

Even with the right organizational structure and sufficient staff capacity, Federal Departments may still not engage in meaningful evaluation and policy research without the right set of incentives. The

[18] See GPRA Modernization Act of 2010, Section 12

recommendations below include approaches to incentivize agencies with increased funding and administrative flexibility.[19]

❖ **Finding:** Offices that conduct or manage program evaluation and policy research consistently reference funding as a primary barrier to scaling up their efforts.[20]

► **Recommendation #8.4:** The Commission recommends that the Congress and the President establish and grant Federal Departments access to *Evidence Incentive Funds* to supplement the production of future research, evaluation, and related activities outlined in Departmental learning agendas (see Memo #7).

► **Recommendation #8.5:** The Commission recommends that the Congress, through the appropriations committees, and the President provide Departments in active pursuit of a learning agenda access to multi-year (or no-year) funding to pursue articulated priorities. *[Note this recommendation could be included in multiple places, but the mechanism itself is an incentive/flexibility]*

The literature on bureaucratic constraints recognizes that individuals within complex systems seek to gain power of "slack," or the resources available to line managers for which they have direct access and control.[21] Mechanisms that give managers resource flexibility may encourage greater uptake of the goals of the Commission, including the production of more and better evidence.

The creation of an Evidence Incentive Fund for each Department, where 10 percent of unobligated balances at the end of a fiscal year can be allocated for future evidence-generating activities, is one such mechanism for re-purposing funding productively for increasing evidence production. By establishing a supplemental funding source for some evidence generation similar to the use of "Innovation Funds" in recent years,[22] the Funds would also address perceived threats from program managers that evaluation and policy research activities reduce the availability of funds for direct services or program operations. Similar mechanisms exist today for other purposes. For example, certain HHS block grants can recapture and reallocate funds to states, extending the funding's period of availability. Similarly, the Social Security Administration's unobligated balances at the end of a fiscal year are transferred specifically for information technology improvements. *[Note: If the Commission agrees to pursue this recommendation, Commission staff will work with OMB to devise an estimate of funding available under this recommendation.]*

Multi-year funding is also a coveted resource for Departments because the constraints imposed on multi-year appropriations are relaxed and the burdens of the annual discretionary appropriations process are reduced. The Department of Labor, for example, receives two-year funding through its set-aside in the Department's budget with notification to appropriations committees about evaluations to be conducted.[23] The Department currently receives multi-year funding for SERA activities that can be used as an incentive for centers or units that are actively pursuing a learning agenda, and incidentally reduces a series of constraints raised for efficiently executing evaluation and research contracts that may themselves be for multi-year studies.[24]

---

[19] Note that individual incentives are now addressed in Memo #7.
[20] See the November 4th CEP meeting testimonies and *Proceedings of a Workshop: Principles and Practices for Federal Program Evaluation:* https://www.nap.edu/catalog/24716/principles-and-practices-for-federal-program-evaluation-proceedings-of-a
[21] Wilson, J.Q. (1991) Bureaucracy.
[22] For example the Social Innovation Fund
[23] Division H, Sec. 107 in Consolidated Appropriations Act, 2017 (Senate Amendment to H.R. 244).
[24] ICEP comments to CEP.

Finally, recognition of successful evaluation and policy research activities can be a powerful symbol of the influence of evidence on policy. While not included as an explicit recommendation, the Commission could suggest that periodic awards to successful units may help provide for a continued high profile status for evaluation and policy research activities, and motivate staff to continue to develop and use evidence to improve programs and policies. The President, for example, could direct OPM to establish a new annual award for Federal agencies, or establish a new category for evidence-building activities in the President's Quality Award to recognize Departments, units, or individuals for the completion *and use* of policy research and evaluation in improving programs and services for the public.

❖ **Finding:** The Commission finds that initiatives to reduce reporting burdens provide opportunities to encourage Federal Departments to shift emphasis toward activities that benefit generation and use of evidence for decision making. Flexibilities such as the waiver authority for program rules provided to HHS and SSA demonstrate promise in encouraging participation in research and evaluation initiatives; and similar mechanisms that provide administrative flexibilities could be expanded to encourage more activity among Federal Departments.[25]

► **Recommendation #8.6:** The Commission recommends that the Congress and the President develop mechanisms for providing increased flexibility in program rules, reporting, or other administrative requirements for Departments that pursue and accomplish objectives in multi-year learning agendas.

Opportunities and rewards to provide flexibility can be devised for Departments pursuing evidence-building activities directly, including evaluation and policy research. Recommendation #8.6 above suggests offering flexibilities related to evidence-building administrative requirements that might be adjusted or modified for units or Departments in active pursuit of their identified learning agendas.

For example, Congress could decide to forego an annual report on a specific topic if a Department was actively pursuing better or stronger evidence about the program, as identified in the learning agenda. Congress could also consider whether other, more routine reports from Executive Departments may be streamlined in the presence of routine evaluation information. Relaxing certain administrative requirements associated with performance reporting may also be a productive means to providing reasonable flexibilities, or relaxing penalties directed at agencies for failing to meet targets for performance measures established under GPRA Modernization. Another approach could be linking organizational incentives to ongoing regulatory or reporting burden reduction initiatives within the Executive Branch, which for this purpose could be expanded to more intentionally focus on the burden within agencies themselves.

LIKELY REACTIONS

- ***Congressional:*** Congress may most likely object to the idea of relaxing reporting requirements, though objections are anticipated to be low.
- ***Executive:*** Departments with high performing evaluation and policy research functions will likely welcome the additional attention and potential stature the suite of recommendations can offer their organizations, recognizing some of the recommendations are already implemented in some Departments. Other Departments without a history of centralized evaluation and policy research may view the requirements as burdensome at the outset, including establishing a Chief Evaluation Officer; this burden is intentionally minimized by recognizing Departments may designate an existing office for the role.

[25] Waiver authority refers to 42 USC 1115

- ***State and Local:*** No objections anticipated.
- ***Other:*** The professional evaluation community will welcome this entire series of recommendations and view the establishment of CEOs as a positive move in the professionalization of evaluation.

COST IMPLICATIONS

The new direct costs to Departments associated with these recommendations are anticipated to be minimal, though some Departments may seek new appropriations to better support a Chief Evaluation Office including for hiring individuals in the new occupational series. The funding flexibility recommendations may reduce the availability of funds available for annual discretionary cancellations by appropriations committees, but otherwise rely on existing appropriations mechanisms to better support evaluation and policy research. If targeted appropriately, relaxing reporting requirements as a flexibility may achieve net administrative savings over time.

ALTERNATIVE OPTIONS

- **Mandate a single Chief Evaluation Officer and coordinating council for CFO Act Departments.** Draft legislation is currently being floated among stakeholder groups to establish Chief Evaluation Officers uniformly in all CFO Act Departments, with a coordinating council. The draft legislation would effectively direct coordination within the performance and evaluation communities, but not with other aspects of the Federal Evidence Ecosystem as envisioned in CEP Memo #7. This recommendation would provide for a single recognized lead in each Department for evaluation and policy research, though would limit flexibility for Departments that have topically diverse missions.

- **Establish new independent Federal evaluation and research agency.** Similar to the General Accountability Office (GAO) in the Legislative Branch, the Commission could recommend a new entity be tasked with providing expertise on behalf of all Departments in the Executive Branch. While this might epitomize independence, such an entity would likely face substantial challenges coordinating with existing Departments and managing bureaucratic and institutional constraints. Such an office would likely need to be inefficiently large to cover the full range of Federal Departments, or lack appropriate expertise. For comparison, GAO's annual budget exceeds $550 million and the organization includes more than 3,000 personnel.

- **Mandate evaluation for major Federal programs that exceed a direct Federal cost of $x million.** A more heavy-handed approach for encouraging increased evaluation and policy research would be to establish mandates for the activities at a policy or program level, rather than expectations that the functions be present. This approach would be similar to that deployed in Canada, where the fiscal office requires periodic evaluations for continued funding of programs. However, as recognized in Canada when implementing the provision, national scale evaluations can have major limitations for both execution and interpretation for use in mandated funding decisions. The mere presence of evaluation activities rather than directed use may be a means to mitigate such concerns.

* * * * *

APPENDIX

Role of Legislative Directives in Supporting Evaluation and Policy Research

Congress can, and often does, play an important role in encouraging the production of specific evaluation projects with authority included in periodic program authorizations or in appropriations bills. Evaluation of innovative proposals for improving human services programs are frequently funded through Section 1115 waiver authority, which permits states to receive approval from HHS or SSA to waive certain program requirements for mandatory funded programs if states develop evaluation plans and produce evaluations for these demonstration projects.[26] States tend to support the use of this vehicle because the authority allows Governors to pursue policies they may prefer within certain programs, and as a trade-off the projects contribute to increasing the body of knowledge available for future policy changes.

In 2015, Congress also reauthorized SSA's demonstration authority for the Disability Insurance program.[27] Currently this program has legally uncapped mandatory authority to develop demonstration projects that can be used for future program reforms, expected within the next decade. The reauthorization included several detailed proposals specified in law that Congress asked SSA to evaluate. Congress included similar directives for the Centers for Medicare and Medicaid Innovation (CMMI) under the Affordable Care Act (ACA), anticipating the need for future knowledge to improve program implementation.[28]

In each of the examples cited above, Congress created an expectation for evaluation and research to support future decisions, whether through demonstration projects or more routine evaluations. However, the vast majority of evaluation currently conducted within the Executive Branch is not explicitly directed by Congress, though it may be permitted because of Congressional authorization to do so. According to the Department of Labor (DOL), for example, only 10 to 20 percent of the evaluation activities are explicitly directed by Congress, with the residual directed by the Department's discretion and an evaluation agenda that prioritizes activities based on annual feedback from bureaus within DOL.[29] In turn, DOL includes these plans in their formal Departmental Strategic Plan and submits notifications to Congress as part of the appropriations process.[30] Similarly, from an inventory of more than [80] evaluation and research projects conducted with support from the Environmental Protection Agency (EPA), few were explicitly directed by Congress.[31]

---

[26] See 42 U.S.C. 1315. Currently used for child welfare, child support, Medicaid, and certain SSA programs. TANF waivers were suggested during the Obama Administration, though pulled back due to congressional concerns about implementation. Of note, Gov. Kasich (OH) submitted a request in 2016 for a TANF waiver.
[27] Included in Bipartisan Budget Act of 2015 (P.L. 114-74, Sec. 821).
[28] other potential examples include home visiting, HUD's family options study, the P3 program
[29] Email correspondence, CEP staff and DOL.
[30] DOL receives a funding set-aside, subject to congressional approval. Submission of the plans is part of the formal approval process for spending the funding.
[31] Hart 2016

RECOMMENDATION MEMO #9

The Role of the Federal Statistical System within the Evidence Ecosystem

Lead Commissioner(s): Potok
Lead Staff: Martinez, Hitt, Boivin, Howell

ISSUE: The Commission's vision of an integrated evidence ecosystem implicitly updates both the historical scope and definition of the Federal Statistical System. What targeted improvements to the mission and roles of the Principal Statistical Agencies (PSAs) and other major statistical programs will best facilitate this vision?

The Commission's charge includes determining the optimal arrangement for which *administrative and survey data, and related statistical data series* may be integrated (Sec. 4 (a) (1)). Effective integration of administrative and survey data is a necessary precursor to achieving the Commission's vision of "a future in which rigorous evidence is created efficiently as a routine part of government operations and used in policy making."[1]

Federal statistical programs have been a cornerstone of the evidence base for many decades, producing fundamental information to inform public and private decisions on a wide range of topics including the economy, the population, the environment, agriculture, crime, education, energy, health, science, and transportation policy domains.

This memorandum specifically addresses strategies for enhancing the existing Federal Statistical System to support its current work producing important data series, and recommendations to enable its envisioned role as the bulk of the "Federal Statistical Data Network," along with the Data Facility discussed in Memo #2 and other future components. The memo describes the existing infrastructure and capacity of the Federal Statistical System and provides recommendations for leveraging *existing* resources to create the means for the "optimal arrangement" of data integration, while protecting and improving data.

Relates to Facility (#2), Ecosystem (#7), Ecosystem: Evaluation (#8), Ecosystem: Program Design (#13), and Increasing Confidentiality (#10).

* * * * *

BACKGROUND AND ANALYSIS

Federal statistical programs have been a cornerstone of the U.S. government's evidence base for many decades, producing and disseminating information to inform public and private decisions on a wide range of policy domains. *Enhancing Collaboration in the Federal Evidence Ecosystem* (Memo #7) describes how, in a well-functioning evidence ecosystem, Principal Statistical Agencies (PSAs) provide the core infrastructure to support continuous evidence generation, by working with program offices to protect and transform administrative data for statistical, evaluation, policy research, and policy analysis (SERA) purposes. As described by the Office of Management and Budget (OMB), "The Federal Statistical System engages in a wide variety of evidence-building functions. These functions include the collection, compilation, processing, analysis, and dissemination of data to create general purpose, policy- and program-specific, and research oriented statistics and datasets."[2]

---

[1] Vision from CEP Memo #1.

[2] OMB. (2017). *Statistical Programs of the United States for Fiscal Year 2017*, Washington, D.C.: Executive Office of the President, Office of Management and Budget. https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2017.pdf

Distinguishing Between Statistical and Nonstatistical Purposes

As a matter of law, data held by the Federal government are available and governed for one of two purposes, either nonstatistical or statistical. The first involves actions that affect individuals instead of groups, while the second involves analysis of groups, without identifying individuals.

Current law is clear in stating what activities are considered beyond the boundaries of a statistical purpose, defined in the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) as "**nonstatistical purpose**." Nonstatistical purpose under CIPSEA means "the use of data in identifiable form for any purpose that is not a statistical purpose, including any administrative, regulatory, law enforcement, adjudicatory, or other purpose that affects the rights, privileges, or benefits of a particular identifiable respondent…."[3] For example, a program administrator reviewing a beneficiary's profile to assess program eligibility is considered a nonstatistical use of data, and is governed by the program's authorizing statute.

In contrast, **statistical purpose** is the use of data for purposes unrelated to any individual or business, but for the creation of descriptive and inferential statistics. The Privacy Protection Study Commission of 1977 (PPSC) defined statistical purposes as "the developing and reporting of *aggregate or anonymous information not intended to be used, in whole or in part, for making a decision about an individual*…."[4] A full 25 years later, PPSC's definition was codified in CIPSEA. Under current law, CIPSEA similarly defines statistical purpose as "(A) means the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups; and (B) includes the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support the purposes described in subparagraph (A)." OMB has further clarified that this definition is interpreted to include program evaluation, performance measurement, and public health surveillance.[5] Thus, statistical purposes are inherently about analyzing groups, analysis which policymakers can use to inform decisions about multiple individuals, but not based on a single individual's information. Statistical purposes are the focus of the Commission's work.

Current State of the Federal Statistical System

Efficiently expanding the government's evidence-building capacity will necessitate capitalizing on the strengths of the Federal Statistical System (FSS), arguably the most developed and capable national statistical system in the world. Unlike the evaluation and policy research function (Memo #8), which is nascent in government today, the FSS features a long history and established processes, protocols, and institutions. The missions of agencies comprising the system are among the earliest and most fundamental in government.[6] The Federal Statistical System includes more than 100 agencies and offices with statistical missions, but the vast majority of the government's statistical expertise and capacity reside in a subset of that group, known as the 13 Principal Statistical Agencies (PSAs). PSAs are agencies for which statistical functions are their primary mission. As a matter of purpose and mission, the PSAs perform the daily work of data collection, production, and protection – as well as the facilitation of access by both governmental and external researchers.

---

[3] Pub. L. 107–347, title V, Sec. 502(5).
[4] Pub. L. 107–347, title V, Sec. 502(9).
[5] See Footnote 2.
[6] For example, Article 1, Section 2 of the U S Constitution mandates the decennial census, which for many years covered people, agriculture and business. As a further example, at its inception, the U.S. Department of Education's primary mission is what the National Center for Education Statistics (NCES) does today.

While highly developed, the U.S. statistical system is also highly decentralized, which is both an asset and a limitation. While the 13 PSAs are spread across 11 departments, they share relatively common characteristics in the mission, capacity, and infrastructure essential to core statistical functions. However, by being embedded within departments alongside their program office peers, they each also have topical emphasis, permitting a high degree of subject matter expertise about the policies and programs covered. Beyond their common capacities, and the benefit of their varying subject matter expertise, they also benefit from some common legal authorities, due to implementation by Congress and the President of several key recommendations from past commissions. Most importantly is the enactment of CIPSEA as a common confidentiality law, as recommended by the President's Reorganization Project for the Federal Statistical System.[7] CIPSEA contributed an essential element to the FSS infrastructure by providing an authority common to all PSAs to protect data acquired for statistical purposes, and enabled limited sharing among a subset of the PSAs.[8]

It is because of their common legal confidentiality protections, culture of data protection, and substantive and technical expertise, combined with their decentralized positioning conducive to government-wide coordination of data for statistical purposes, that PSAs are the building blocks of a modern system of evidence building. This has long been true with respect to survey and other data collected for explicitly statistical purposes. This role only becomes more critical as PSAs and other actors in the evidence-building community increasingly turn to administrative data. Administrative data, which by definition originate outside the statistical system, can be increasingly used by PSAs to improve data quality, reduce data collection and processing costs, and reduce respondent burden. Administrative data are often a core input for evidence-building efforts and by definition exist at agencies beyond the scope of the statistical system. But, for the promise of administrative data to be fully realized for evidence building, it requires both technical and legal transformation into statistical data. Accomplishing this, along with an increased capacity to serve both internal and external researchers, is the essential evolution of today's FSS to tomorrow's "Federal Statistical Data Network."

The Commission's Vision for a "Federal Statistical Data Network"

The building blocks of a clear and consistent framework of data protection and access for Federal evidence building is the *existing* capacity and infrastructure of the statistical system. However, today's FSS will require a few new authorities, intentional placement, and some new resources to address the limitations of decentralization and the gaps between today's reality and the Commission's vision. This is true with or without the addition of a Data Facility into the mix, the primary open question being the exact division of labor, but not the essential authorities, tasks, and need to coordinate it all.

With respect to the limitations of decentralization, past commissions who studied the U.S. statistical system raised concerns regarding decentralization and lack of coordination, and some of these issues persist today. In 1995, Norwood summarized findings of previous commissions noting that despite some coordination, "[t]he U.S. system has neither the advantages that come from centralization nor the efficiency that comes from strong coordination in decentralization. As presently organized, therefore, the country's statistical system will be hard pressed to meet the demands of a technologically advanced, increasingly internationalized world in which the demand for objective

---

[7] Bonnen, et al. *Improving the Federal Statistical System: Report of the President's Reorganization Project for the Federal Statistical System*, reprinted in *American Statistician*, vol. 35, no. 4 (November 1981), pp. 184-196.

[8] Another example is the creation of the Committee on National Statistics at the National Academies of Sciences, Engineering, and Medicine, though the original vision was to be primarily as an advisor to OMB's Chief Statistician rather than primarily to the PSAs.

data of high quality is steadily rising."[9] As a solution, these commissions have recommended some combination of a centralized statistical agency, a centralized statistical statute, centralized statistical coordination authority, and a centralized statistical data center.[10]

Consolidating several PSAs into a single statistical agency, however, would not address all current limitations, and would likely introduce new challenges. For example, as currently constituted, a consolidation of the 13 PSAs would result in an agency of 20,000 plus employees now removed from proximity to the programs and policies upon which they develop data. Moreover, removing this proximity seems even more problematic when considered in the context of the growing volume of administrative data, and the increased coordination necessary for a robust evidence ecosystem. An alternative to the physical centralization is the "virtual" centralization by harmonizing of certain legal authorities, roles, and institutional placement. For example, the PSAs provide existing infrastructure within most departments to support day-to-day production and routine provision of information for decision-making. With necessary technical independence to facilitate objective information, the PSAs remaining in their departments are well-situated to support evaluation and policy research activities that rely on data acquired and curated for statistical purposes, while not being linked to the conclusions and value statements asserted in evaluation reports or the policy judgements and claims derived from such information.

Another approach, complementary to the harmonization of authorities, and as outlined in Memo #7, is greater coordination across all SERA functions. This need is the basis for the recommendation in Memo #7 to enhance interagency and intra-agency coordination for SERA functions. Yet a modern and well-functioning evidence-building community requires more than basic statistical capacity and light coordination of the statistics function. The increasing awareness that both survey and administrative data possess value for other actors in the evidence-building community, when combined with a decentralized system of agencies acquiring and producing those data, requires a statistical system with both vertical and horizontal coordination. While Memo #7 envisions a reorganization of the "headquarters" of this system, here the focus is on the "field offices" that comprise the "Federal Statistical Data Network." This network builds upon and reinforces the existing culture, expertise, and legal frameworks of the PSAs, along with a new Data Facility to improve access, to support the entire evidence-building community both within government and beyond. The vision for the "Federal Statistical Data Network" features expanded statistical agency *missions*, clear operational *legal frameworks*, and routine interactions between appropriate senior-level *actors* across government.

Mission

The current missions of the PSAs are focused on the rigorous and objective production of data for statistical purposes in each one's respective topical domain. To realize the needs of a robust evidence ecosystem, these missions would need to be expanded to explicitly include coordination of information policy, information collection, data management, and facilitating access to information. In effect, OMB ("headquarters") would be sharing coordination responsibility with the 13 PSAs – who horizontally coordinate across their departments – and with the proposed Data Facility, whose emphasis is the facilitation of data access for researchers external to government. Appropriately expanding and tailoring the missions of the PSAs will provide the basis to ensure that priorities for evidence generation across government remain focused on continually improving the system and supporting the ecosystem over time as technologies advance, budgets change, and human capital

[9] Norwood, J. (1995). *Organizing to Count: Change in the Federal Statistical System.* Washington, D.C.: The Urban Institute Press.

[10] See Norwood (1995) and materials from the March 13 CEP meeting.

evolve. For example, the effective use of administrative data means that its potential statistical purpose must also be considered throughout its lifecycle. For those data to remain fully useful for statistical purposes – be that reducing the burdens of survey data or as direct inputs into an evaluation – changes or discontinuation of those data should be informed by these statistical stakeholders, and not just program agency needs. Empowering PSAs to provide leadership will ensure that happens and increase the value of administrative data for statistical purpose. Likewise, data management – a rapidly developing field – increasingly requires highly specialized and multidisciplinary expertise of the type that can efficiently be centered in data-centric PSAs, while also serving the broader needs of a home department. Missions that emphasize technical, data-focused roles for the PSAs in support of the larger evidence ecosystem also protect the statistical function from political interference, operating within safe legal structures with strict boundaries on allowable use and disclosure of information.

Legal Framework

CIPSEA, as currently constructed, offers a strong starting point for the needed common legal framework to protect and facilitate access to statistical data for statistical purposes. CIPSEA Subtitle A (discussed in detail in Memo #10) codifies an authority to protect data acquired for statistical purposes, as well as specific requirements for disclosing those data, and rather severe penalties for violating those protections. Subtitle A grants this protection authority to agencies recognized by OMB.[11] Subtitle B permits the sharing of data protected under the authority of Subtitle A, but currently only business data and only between three "designated statistical agencies" which are specified in the statue. In essence, CIPSEA creates a legally distinct "island," where recognized actors can protect data for statistical purposes (and currently, three of them can exchange their data with each other). This "CIPSEA Island," wherein access to data is tightly controlled, and the penalties for violating those protections exceedingly high, could be expanded to include all of the 13 PSAs, and apply to data beyond business data. Doing so would provide a common authority for them to protect data at a strong, common level, as well as provide each other access to those data to support the full range of statistical activities. Further, the proposed Data Facility should also hold CIPSEA authorities, for the purpose of facilitating appropriate access to data held on the "CIPSEA Island." Access to the "CIPSEA Island" is tightly controlled. Allowable uses of data within the island are explicitly specified, disclosure of data beyond the island is expressly limited, and violations are punishable. In exchange for the tight restrictions on the "CIPSEA Island," covered statistical system actors are granted certain authorities and privileges for use of microdata that do not apply to other entities.

Actors

There are five actors in this vision. The first is the "headquarters," the Office of Information Affairs at OMB (see Recommendation #7.1 in Memo #7), responsible for government-wide coordination of information policy, including policies and standards regarding administrative and statistical data collection, protection and use.

Coordination with each department is accomplished via the second actor, a senior level official designated to serve as the "Senior Agency Official for Data Policy," the coordination point across their departments for information policy, data management, and data access by non-CIPSEA entities. This is not intended as a new position, rather it is intended as a consolidation of responsibility for coordinating a given department's data interests in concert with other officials with related responsibilities. For the 11 departments that house a PSA, this would in all cases be the head

[11] CIPSEA authorizes OMB to identify PSAs and to recognize other statistical units for coverage by CIPSEA, provided they meet certain requirements including the ability to demonstrate functional separation.

of the PSA. In some instances, departments may need multiple senior leaders to work collaboratively where a department's data inventories and activities may be better served by combined expertise. The senior officials are critical for advising OMB on strategies for leveraging data for both administrative and statistical purposes.

The third actor is the PSAs, which while continuing to provide relevant, rigorous, and objective information about topics, programs, and policies within their departments, also serve as their department's "node" in the "Federal Statistical Data Network" and thus the conduit for access to data held within the "CIPSEA Island" by the Data Facility or other authorized users. As a "node" in the network, these entities will be well-positioned to identify data sources in other departments beneficial to their own department's evidence-building needs.

The proposed Data Facility, itself the new 14th PSA, is the fourth actor and serves as the primary coordination point for external researcher access. The Commission envisions a single entry point for access to CIPSEA-protected data for external researchers, not so much as a mandate, but rather an approach to address the many concerns expressed to the Commission about multiple and confusing processes. The Data Facility would be responsible for enabling this single entry point.

The fifth actor is the "Recognized CIPSEA Unit," an office or agency who meets the legal specifications necessary for a CIPSEA designation by OMB. These entities would, after meeting the stringent requirements, share the benefits of access to the data held within the CIPSEA Island. While few of these would exist immediately, this is an important element to ensure future capacity for the evidence ecosystem.

How will Federal staff access data for statistical purposes under the vision?

Much of this is dependent on the Commission's specific determinations in Memo #10, but generally, an internal researcher requiring access to data protected by CIPSEA would coordinate access with an appropriate PSA or the Data Facility.

How will external researchers access data for statistical purposes under the vision – and what is the role of the Data Facility?

If the external researcher's needs involve the use of CIPSEA-protected data, that activity will only be allowed on the "CIPSEA Island," facilitated either by a PSA or the Data Facility. As one of the Commission's key recommendations is to provide a consistent application and approval process for external researchers (Recommendation #4.1.3 in Memo #4), the preferred route would be for the external researcher to contact the Data Facility to apply for access to the needed data. Regardless of whether the researcher wants to access data that are already curated and linked by the Data Facility or access a dataset in a single PSA, the Facility provides a one-stop shop, or a single point of access to a standard application and approval process for external researchers. In other words, the Data Facility is the ferry the researcher uses to access data protected on the "CIPSEA Island."

Current Challenges to Achieving the "Federal Statistical Data Network"

While the vision for the "Federal Statistical Data Network" above clearly delineates how the PSAs could operate and interact within the broader evidence-building community, the reality is that today we are far from this optimized arrangement. The missions, legal framework, and actors are today unclear and only loosely coordinated because of two predominant challenges: misaligned authorities of PSAs and dispersion of the statistics function without strong coordination.

Challenge #1: Authorities of PSAs not fully aligned with goals of evidence building

PSAs vary in their readiness to take on greater responsibility for evidence-building activities given legal, policy, and institutional placement differences. The Commission can identify changes or expansions to existing authorities to ensure that the actors across the statistical system have the appropriate legal standing.

As a benchmark for consideration, the United Nations (UN) model statistical law contains four minimum tenets for a decentralized system, which the US has only partially covered.[12]

Legal authority to acquire data. CIPSEA Subtitle A already provides an authority for the PSAs to protect data for statistical purposes, and this requires further enhancement by ensuring a counterpart authority (Memo #3's "Yes, Unless" and Memo #6 "Bans") to ensure clarity for statistical agencies to acquire and program agencies to provide data.

Confidentiality/statistical purpose law. CIPSEA meets the UN criteria for a confidentiality law that limits use to exclusively statistical purposes. However, implementation is uneven because some statistical agencies do not regularly use CIPSEA. The National Center for Education Statistics (NCES) does not use it because its own authorizing statute contains a requirement that it make its information available for anti-terrorism efforts. Absent that provision being revoked, NCES and OMB have determined that it is not appropriate to invoke CIPSEA except where NCES is not in direct control of any directly identifying data. Other PSAs are concerned about their ability to uphold their CIPSEA pledge, which indicates that no one other than agency employees and agents can see or use the data. In some cases this is because the PSA has lost direct control of information technology (IT) systems during OMB-required IT consolidation. Also, some departmental CIOs – particularly emboldened since the recent passage of the Federal Information Technology Acquisition Reform Act (FITARA) – have been reluctant to comply with the PSA's requests to train and qualify IT staff as CIPSEA agents. Each of these poses challenges to the real functioning of the "CIPSEA Island," and require resolution.

Common data-sharing rules. Common data-sharing rules pertain in this case to a mandate for program agencies to share administrative data with statistical agencies as well as sharing of statistical data among statistical agencies. The Commission has discussed the former as part of the "Yes, unless" proposal under Memo #3. One additionally important implementation component of such a law is the requirement on program agencies to consult in advance with their associated PSA before making changes to such datasets. OMB policy tells agencies that they should be doing this, but follow up has been limited.[13]

Subtitle B of CIPSEA focuses on the "statistical efficiency" to be gained by sharing of statistical business data among three "designated statistical agencies" (the Bureau of the Census, the Bureau of Labor Statistics, and the Bureau of Economic Analysis), considered appropriate once Subtitle A of CIPSEA provided a common data protection framework. The Commission has heard recommendations to complete the unfinished portion envisioned as part of CIPSEA by

[12] See Annex 1 in United Nations. (2003). *The Handbook of Statistical Organization: The Operation and Organization of a Statistical Agency*. Third Edition. New York: United Nations, Department of Economic and Social Affairs. http://unstats.un.org/unsd/publication/SeriesF/SeriesF_88E.pdf

[13] Guidance for Providing and Using Administrative Data for Statistical Purposes, February 14, 2014, https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf

encouraging passage of a narrow amendment to the Federal tax code that would permit sharing of tax data, or data comingled with tax data, under Subtitle B as well. This is currently limited, even among the three designated agencies, because one – Census – "comingles" census and tax data together in its economic census and survey programs, and Subtitle B does not apply to tax data. Presumably, this proposal would be subsumed under the broader recommendation on tax data access that the Commission will take up soon.

The remaining inconsistency is among statistical agencies for other statistical data. If the Commission recommends that program agencies have authority to provide data for statistical purposes, implying that those data can be linked to a statistical dataset, then the question naturally arises whether two statistical datasets collected by two different CIPSEA agencies should be able to be shared and linked. This would be resolved by expanding the applicability of Subtitle B to the remaining 10 PSAs, and applying that authority to share to types of data beyond business data.

Statistical coordination law. The responsibility for statistical coordination is given in law to OMB. Memo #7 suggests an improved ability for OMB to provide coordination, in part to help departments follow suit by betting aligning disparate functions. OMB's statistical policy function has historically been resource-constrained and insufficient, and previous commissions tend to recommend more than one tact for increasing the ability of the system to coordinate itself. The vision of the "Federal Statistical Data Network" is one such idea. However, the ability of the PSAs to serve in the coordination role envisioned is currently uneven given differences in law and institutional arrangements within departments. The designation of the head of a PSA as the "Senior Agency Official for Data Policy," will partially resolve this, but in some cases other remediation must be considered depending on the placement of a PSA in its home department, as well as the scope of the department and the PSA itself.

In the absence of sufficient law in the United States, OMB has tried to fill in gaps with policy. OMB's Statistical Policy Directive #1, issued in 2014, addressed many of the elements of the UN companion policy to the model statistical law, *Fundamental Principles of Official Statistics*. Statistical Policy Directive #1 established four responsibilities for Federal statistical agencies and recognized statistical units: relevance, credibility, objectivity, and trust:[14]

1. *Relevance* requires "***communication across and within Departments*** with planning information collection and dissemination activities" and that statistical agencies ***collect program and policy-relevant information from "administrative records collected and maintained by the agency, or other government agencies.***"
2. *Credibility* requires that agencies "apply sound statistical methods to ensure statistical products are accurate."
3. *Objectivity* requires that agencies "***produce data that are impartial, clear, and complete and are readily perceived as such by the public***" and further requires that "***statistical units must function in an environment that is clearly separate and autonomous from the other administrative, regulatory, law enforcement, or policy-making activities within their respective Departments***."

[14] Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units, 79 Fed. Reg. 71609 (December 2, 2014) https://www.gpo.gov/fdsys/pkg/FR-2014-12-02/pdf/2014-28326.pdf

4. The *trust* of information providers requires agencies to ***ensure "confidentiality and exclusive statistical use" of their data***.

Directive #1 further notes that "the benefits to Federal statistical data users and the Nation of maintaining and enhancing the quality of official Federal statistics envisioned by this Directive become fully realized when Federal statistical agencies and recognized statistical units, *with enabling support and facilitation from their Departments*, achieve these mutually-reinforcing responsibilities concurrently."

As noted, technical independence can be a challenge for some PSAs. In particular, conflicts with departmental Chief Information Officers (CIOs) have been a growing source of friction in this regard. The Committee on National Statistics at the National Academies of Sciences, Engineering, and Medicine and the American Statistical Association leadership have both expressed concern about this and recommended that PSAs have sufficient legal authority implemented or restored.[15]

Several PSAs have protections in law that are seen as addressing some of these technical independence concerns. These include, for the agency head:

- Fixed term of office
- Appointment by the President, some with Senate confirmation
- A requirement for the appointer to notify Congress in advance of any termination
- Direct reporting to the Secretary or Deputy Secretary
- Specific authority to act with technical independence as described in Directive #1 (e.g., to publish statistics without review by departmental policy officials)
- Authority to have an advisory committee or other formal external consultation processes

Several commenters reminded the Commission of the need to reinforce the importance of these cornerstone agencies and the laws and policies that undergird them, rather than taking them as a given. In particular, the public's ability to trust them to handle their data must be maintained and, given the increased data sharing contemplated, even strengthened.

Challenge #2: Historic dispersion of the statistics function inhibits evidence building

As discussed, the presence of a PSA in a department provides the opportunity for a "lead agency" on critical data resource needs. However, some departments do not have a PSA. PSAs are present in 11 of the 18 departments and major agencies.[16] In a few of these departments, there are "candidates" that could potentially take a similar role, though OMB does not currently recognize these statistical units under CIPSEA and they may need other remediation to play the role effectively.

Those departments without a PSA currently have no obvious point of contact or internal coordination for information policy, data management, and statistical support for evidence building.

---

[15] It is worth noting that both the United Kingdom and Canada have taken action in the past couple of years to increase the independence of their statistical agencies, especially in light of increased authority and need for access to administrative data and including addressing concerns about IT encroachments. IT-focused recommendations will be addressed in forthcoming memo on administrative processes (Memo #15).

[16] Eighteen includes all Federal departments, plus the Environmental Protection Agency, the National Science Foundation and the Social Security Administration. It excludes a few CFO Act agencies such as the National Archives and Records Administration. The seven without a PSA are the Departments of Defense, Homeland Security, Housing, Interior, State, Transportation, and Veterans Affairs.

This also means that researchers, internal or external to departments, often have no obvious point of entry to learn about or access data resources for evidence building.

Also, there is some evidence to suggest that the capacity to look across the department to assess data needs and to meet them through survey or administrative data use has lessened over time as units have moved deeper down into the organizational chart of departments. Today, many statistical agencies are farther down the hierarchy in their departments than in the past. Only the Bureau of Labor Statistics (BLS) and the Energy Information Administration (EIA) report to their cabinet secretary; other agencies are one, two or three levels down the organization chart. This layering can affect not only an agencies' ability to argue effectively for their priorities, but also their ability to collaborate with other agencies and maintain independence from political influence.

Even if a department's needs do not necessitate a fully developed PSA, under the Commission's vision for an enhanced "Federal Statistical Data Network" it will be imperative that each department have some individual responsible for interagency and intra-agency coordination of information policy and data access. This is not only essential to an effective evidence-building ecosystem, but also a best practice given the increasing reliance on data and information. The current arrangement often creates conflict and poor compromise when the needs of information policy and statistical activities are not expressly aligned with the IT needs and policies for nonstatistical functions. This condition is only likely to worsen over time.

DRAFT FINDINGS AND RECOMMENDATIONS

The recommendations proposed to achieve the "Federal Statistical Data Network" involve providing sufficient authority, designating senior officials, enabling consultation across information policy activities, and other institutional mechanisms.

Enabling Data Sharing Within the Federal Statistical System

❖ **Finding**: The decentralization of the U.S. Federal Statistical System is an asset, for in most cases it achieves a desirable proximity of a PSA to the programs and policies about which the PSA develops data. The limitations of decentralization can be overcome when coordination of statistical activities is legally enabled and prioritized appropriately; in order to achieve the Commission's vision of a well-functioning Evidence Ecosystem, the PSAs must be properly positioned to support evidence building. Greater coordination and sharing of data resources across and within the decentralized system, including within departments, is essential.

❖ **Finding:** CIPSEA Subtitle B ("Statistical Efficiency") authorized the sharing of *statistical* business data among three PSAs in order to create statistical efficiency within the context of CIPSEA's data protection provisions. Congress indicated that this data sharing was an initial step that it could expand if successful, and 15 years of experience with CIPSEA have demonstrated success. This provision has enabled several new data sharing initiatives that have improved efficiency for all three agencies.[17]

[17] For a summary of several of the successful data sharing initiatives that were enabled by CIPSEA Subtitle B, see "Confidentiality of Data Collected by BLS for Statistical Purposes" at https://www.bls.gov/bls/cipsea-report.htm.

❖ **Finding:** Achieving a fully robust evidence ecosystem necessitates researchers and evaluators having access to data across organizational silos in our decentralized system, enabling not only efficient evidence building, but also replicability.
►**Recommendation 9.1 [PRIORITY]:** The Commission recommends that Congress and the President include under the current statistical protection and efficiency law the authority for the remaining Principal Statistical Agencies and the [Data Facility] to access each other's data when the alternative would be initiating a new data collection or being unable to conduct a study.

This recommendation addresses one of the major legal barriers identified for improving the statistical system's ability to support basic production of official statistics and the system's ability to support greater policy research and evaluation activities in the future. The recommendation would enable PSAs and recognized units, including the data facility, to rely on a common set of data for evidence building. It recommends a change to CIPSEA to expand the applicability of Subtitle B to all 13 PSAs and the proposed Data Facility, and further expands its applicability to all data held by these agencies and protected by Subtitle A or a PSA's organic statute. Finally, the recommendation empowers OMB to designate additional offices as recognized CIPSEA units as needed to support future capacity.

Opening access to Subtitle B of CIPSEA is likely among the more politically sensitive in the slate of Commission recommendations developed to date, because such a recommendation directly acknowledges data may be shared across a subsection of offices within the Federal government. However, this change is a natural outgrowth of modernizing the statistical system to operate in a coordinated manner in support of increased evidence building and is appropriate because access is limited to statistical purposes and within the confines of the tightly controlled CIPSEA uses and access parameters. The ability to share data under CIPSEA Subtitle B also becomes a key incentive to encourage agencies to increase data protection to become eligible for recognition under CIPSEA Subtitle A.

Senior Official Role and Collaboration within Departments

❖ **Finding**: Central to the success of a [Federal Statistical Data Network] is the coordination of information policy, data management, and statistical activities throughout government, especially in a decentralized system. This necessitates that each department formalize the authority to perform this function to the heads of PSAs.

►**Recommendation 9.2 [PRIORITY]:** The Commission recommends that Congress and the President require each department to designate the head of a Principal Statistical Agency, where one exists, as the department's "Senior Agency Official for Data Policy," or a similarly senior-level departmental official in the absence of a PSA, responsible for coordinating with OMB and other departmental officials to ensure appropriate attention to statistical needs and information policy within departments.

►**Recommendation 9.3:** The Commission recommends that Congress and the President amend the Paperwork Reduction Act (PRA) to require intra-departmental collaboration when developing new data collections or modifying existing collections, including with the "Senior Agency Official for Data Policy," to calibrate collected administrative data to best achieve administrative and statistical purposes.

Recommendation #9.2 establishes or designates the heads of the PSAs as their department's "Senior Agency Official for Data Policy" or a similarly appropriate official at departments that do not house a PSA. The recommendation encourages departments to integrate the increasing role data, new data collection, and ongoing data management play in the evidence ecosystem. This is not intended as a new position, rather it is a recognition that the needs of data users and data management issues need equal influence when departments are considering actions involving their CIOs, Chief Privacy Officers, Chief Performance Officers and the like. As a matter of bureaucratic principal, if an essential stakeholder is not present for a decision-making process, that set of views is often weighed lower. This is intended to ensure that the unique and growing needs of the statistical and evaluation community are equally represented among senior department decision makers. In order for PSAs to best meet the needs of the many actors who engage them from across the ecosystem – including Congress, evaluators, researchers, and external researchers – their heads need parity to existing cross-departmental officials such as Chief Financial Officers, Chief Information Officers (under which OMB policy indicates that privacy officers typically sit), etc. As the head of a PSA, the "Senior Agency Official for Data Policy" would also be well positioned to help facilitate access requests external to the department, including from researchers via the data facility, providing for a seamless implementation of the "one-stop shop" vision. While each department needs a "Senior Agency Official for Data Policy," the Congress can provide sufficient flexibility for the President to calibrate this recommendation to the specific needs and structures of departments and agencies and consider where more than one "Senior Agency Official for Data Policy," each with a specific portfolio, may be appropriate.

Recommendation #9.3 addresses a longstanding limitation within departments where individual programs modify data collections that PSAs rely on, or want to rely on, but without notifying or consulting them. Since these will often be datasets of interest to evaluators and researchers, this also provides a mechanism for their needs to be addressed. A relatively minor adjustment to the PRA could encourage greater departmental collaboration, mitigating the disconnects that are far too frequent in Federal data collection, which OMB is not well positioned to remediate under current resource constraints. Fundamentally, while the recommendation suggests a relatively minor bureaucratic process change in departments that do not currently facilitate such intra-departmental collaboration, the recommendation contributes to shifting the expectation within departments that one program's information collections can and often do affect others and the production of statistics within the statistical system.

Strengthening PSAs

❖ **Finding**: The integrity and objectivity of Federal statistics is foundational to a well-functioning democracy and an essential ingredient for the larger evidence ecosystem. Some, but not all, PSAs have authority and institutional arrangements in law that promote such integrity and objectivity.

►**Recommendation 9.4:** The Commission recommends that Congress and the President provide each PSA authority and institutional support for maintaining the integrity and objectivity of Federal statistics by providing:

- A clear authority and mission to advise the Secretary about information policy matters, acquire and protect administrative data under CIPSEA, provide the basic infrastructure in support of evidence building, and support the needs of other departmental evidence needs;
- Relief from any conflicting statutory requirements;

- A statutory requirement that the appointer notify Congress's committees of jurisdiction 60 days in advance of any intent to terminate the head of a PSA, including the rationale;
- Direct reporting to the Secretary; and
- Authority to establish consultative processes with experts and the public, including advisory committees

►**Recommendation 9.5:** The Commission recommends that Congress and the President enact legislation to codify elements of Statistical Policy Directive #1 and provide sufficient whistleblower protection for employees in the PSAs to report suspected violations.

Given the idiosyncrasies of history, each existing PSA has slightly different legal authority and institutional arrangements, not all of which are optimal for fully taking on the new responsibilities envisioned by The Commission. There are certain core practices that the Commission could recommend, recognized as best practices within the U.S. and international statistical community, such as authority to engage the public and professional community directly and regularly, rather than through an advisory function from a superior organizational layer with a broader set of responsibilities. The practices in Recommendation #9.4 include clearly stating a mission that supports the full range of evidence building, addressing for example concerns that today external researchers must identify a "benefit to the Census Bureau." This recommendation also recognizes that some PSA heads sit under layers in their departments such that they could never effectively coordinate and lead on data policy without a more corporate perch to be aware of major issues and priorities.

Recommendation #9.5 further strengthens statistical institutions by placing in law the expectations of technical independence outlined in OMB's Statistical Policy Directive #1.[18] The placement of these responsibilities explicitly in statute strengthens the government's commitment to the policies, makes the principles more difficult to change over time (currently through an OMB policy process), and enables violations of the responsibilities to be reported under whistleblower protections for Federal employees established in Title 5.[19] Placing the directive in law also provides members of the statistical system a stronger basis with which to defend encroachment of administrative processes on the needs of the statistical system, to the extent the processes may jeopardize the principles outlined in statute. This approach also essentially brings the United States in line with the UN's model for a national statistical law. Note that administrative processes will be further discussed in Memo #15 for the June meeting.

COST IMPLICATIONS

Given the existence of 13 PSAs and over 100 other agencies who routinely perform some statistical activities,[20] departments are not expected nor need to create new bureaus to serve the roles outlined in the recommendations, but rather to better leverage existing ones either within their own departments, or potentially via an agreement with another department's PSA. For the former, we

[18] See 79 FR 71609: "It is the responsibility of Federal statistical agencies and recognized statistical units to produce and disseminate relevant and timely information; conduct credible, accurate, and objective statistical activities; and protect the trust of information providers by ensuring confidentiality and exclusive statistical use of their responses as described below."

[19] See 5 USC 2302

[20] See OMB's Annual Report to Congress on Statistical Programs of the U.S. Government, 2017, https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2017.pdf.

envision that some agencies that have not previously done so would submit to OMB the documentation required to demonstrate an ability to meet requirements for being recognized by OMB as a CIPSEA unit. For the latter, we have in mind that a department might enlist the data facility or one of the larger or topically appropriate statistical agencies to serve in a facilitating role. Departments would still need to identify a well-situated point of contact that could liaise with the facility if they chose an external agency to provide these services. Given expanded responsibilities, some PSAs will require small staff increases, potentially from elsewhere in departments.

It is also reasonable to expect that after a culture of strong coordination across a department's data needs is in place, efficiencies will be found across many aspects, including duplicative collections of variables or overlapping frequency. Additionally, by evolving the role of the PSA in its home department, other elements of the department will likely recognize efficiency by taking advantage of the expertise in data management, statistical methodology, and statistical inquiry that are common across PSAs.

ALTERNATIVE OPTIONS

- **Alternative Recommendation 1:** The Commission recommends that the President direct OMB to update its suite of statistical policies and standards to ensure they provide fully adequate guidance to departments about the boundaries and support necessary for PSAs to carry out both their fundamental statistical responsibilities as well as their new responsibilities as coordinators of statistical and information policy activities.
- **Alternative Recommendation 2:** The Commission recommends that Congress and the President require that the heads of PSAs be Presidentially appointed and Senate confirmed positions.

RECOMMENDATION MEMO #10
UPDATED AFTER CALL ON APRIL 26

Protecting Privacy by Increasing Confidentiality
Lead Commissioners: Rice and Sweeney
Lead Staff: Boivin, Martinez, and Hawes

ISSUE: What legal and regulatory improvements can the Commission recommend that will provide clear and consistent authority to protect Federal data used for evidence building from unacceptable risk of re-identification? How can current and evolving technology improve data access while simultaneously increasing data confidentiality and security?

The Commission is charged with considering the privacy implications of increasing data linkage for evidence building. Privacy, in this context, has two major aspects: 1) ensuring public confidence that data are being protected and used only for legal and allowable purposes (addressed in *Protecting Privacy by Increasing Transparency* (Memo #11) and, 2) technical approaches to protecting data confidentiality and security. This memo addresses technical approaches to protecting privacy by increasing the level of data confidentiality and security. Recommendations include establishing a clear and consistent legal framework authorizing the use of individually-identifiable data for statistical purposes (as part of statistical, evaluation, policy research, and policy analysis SERA activities) and permitting release of such data only if the risk of re-identification is objectively and quantifiably low enough not to pose a defined standard of harm to individuals or businesses. This set of recommendations further proposes to take advantage of the potential of emerging technologies that both minimize risk of re-identification and allow for increased use of data for evidence building. *Relates to Transparency (Memo #11), Facility (Memo #2), Legal (Memo #3), Standards (Memo #4).*

* * * * *

BACKGROUND AND ANALYSIS

In addition to considering the "optimal arrangement" of data access while weighing the risk of release of personally-identifiable information (PII), the Commission must "make recommendations on how *data infrastructure, database security, and statistical protocols* should be modified to best fulfill the objectives identified in paragraph (1);" and, in the context of a clearinghouse, "evaluate…*how to protect information and ensure individual privacy and confidentiality.*"

The CEP statute refers to making data available to "qualified researchers" and indicates that, for the clearinghouse, the Commission should evaluate, "which types of researchers, officials, and institutions should have access to data and what the qualifications of the researchers, officials, and institutions should be." This memo addresses this requirement by discussing relevant concepts under current law and making recommendations regarding how we might define "eligible researchers."

Privacy has been central to the Commission's charge. Five Commissioners were appointed based on their privacy expertise, the September and February meetings were devoted to the topic, at the December and January meetings experts on data access in the U.S. and abroad discussed how privacy is addressed in their models, and much of the public input from hearings and the Request for Comments related to privacy concerns.

The Commission envisions "a future in which rigorous evidence is created efficiently as a routine part of government operations and used in policy making."[1] As such, one of the five principles adopted by the Commissioners is that "individual privacy and confidentiality must be respected in the generation and use of data and evidence." The recommendations included in this memo seek to advance this principle.

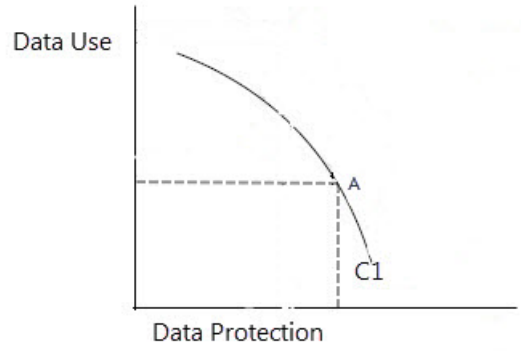
Definition of Confidentiality

Keeping information confidential means handling it "in a manner that would allow public identification of the respondent or would in any way be harmful to him is *prohibited* and that the data are *immune from legal process*."[2] This definition encompasses two important aspects of confidentiality protection: protection against public identification and protection against mandatory disclosure due to legal requirements.

Laws governing confidentiality protections for Federal data have historically required de-identification of individuals or businesses in datasets prior to release. Initially, this requirement meant stripping from the dataset direct identifiers such as name and address as well as some other indirectly identifying information. More recently, Federal statistical agencies have met this requirement by relying on statistical disclosure limitation (SDL) techniques that mask the data, as described in 2005 in Statistical Policy Working Paper #22, *Report on Statistical Disclosure Limitation Methodology*.[3] Working Paper #22 states that every agency or unit "that releases statistical data should be capable of selecting and applying suitable disclosure limitation procedures to all the data it releases." This reflects the prevailing view of the time that using SDL techniques on a release-by-release basis was sufficient to prevent re-identification.

Current understanding of re-identification acknowledges the cumulative risk from release of information about individuals and businesses both across datasets and across time. A more nuanced definition of re-identification from Sweeney et. al (forthcoming) is "the ability for an interested adversary to use reasonable effort to match details in the de-identified dataset to distinct persons sufficient to contact them."[4] The more information available to the interested adversary about distinct persons the greater the chance of re-identification. As shown in Figure 1, increased data use results in decreased confidentiality protection. Thus, efforts to reduce the information released in individual datasets by stripping individual identifiers and applying SDL techniques are necessary, but not sufficient, to prevent re-identification. Any paradigm seeking to protect individual identity must consider all other publicly available information about those individuals.

Figure 1.—The Data Use/Protection Frontier Curve



[1] Memo #1, *Evidence Principles*

[2] United States. President's Commission on Federal Statistics. *Federal Statistics: Report of the President's Commission.* 1971, p. 222.

[3] FCSM, December 2005. *Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology.* https://www.hhs.gov/sites/default/files/spwp22.pdf

[4] Sweeney, L., Yoo, J., Perovich, L., Boronow, K, Brown, P., and Brody, J. *Re-identification Risks in HIPAA Safe Harbor Data: a study of data from one environmental health study.* (forthcoming)

Statistical agencies have released public use microdata files and statistics from surveys and censuses for decades. However, the context has changed as the amount of public information about individuals has exploded in recent years. Just within the Federal Government, the Office of Management and Budget's (OMB) Open Data initiative yielded over 150,000 datasets accessible through a single website, many of which were administrative datasets that had never before been released to the public.[5] Many program agencies have little or no experience applying SDL techniques to data releases.

Furthermore, data collected and released by state and local governments, private companies, non-profit organizations, and researchers contribute to the cumulative amount of information available about individuals and businesses that could be used for re-identification. Sweeney et. al (forthcoming) demonstrated how data collected by researchers on air and dust samples from 50 homes in two communities in California could be combined with data released under the Safe Harbor provisions of the Health Insurance Portability and Accountability Act (HIPAA) to "uniquely and correctly identify [in one community] 8 of 40 (20 percent) by name and 9 of 40 (23 percent) by address."

The Commission's charge to identify the optimal arrangement by which to increase the creation and use of evidence in a manner that protects privacy acknowledges that data confidentiality is inextricably connected to data access. The most privacy protective approach is to generate evidence within a well-coordinated Federal evidence ecosystem in which strong laws, policies, and technologies work together to protect individual privacy in every single government release of data or information.

How Confidentiality is Defined in Law and Regulation

Federal agencies that collect and release information about individuals must consider the legal context in which they operate. Sensibly, many of the same Federal laws and regulations that apply to data access (Memo #3) also govern data confidentiality. On the other hand, agencies with primary responsibility for evidence generation often must apply legal requirements from multiple statutes. The Privacy Act of 1974 applies to all individually identifiable information collected by Federal agencies for any purpose. The Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) covers Principal Statistical Agencies (PSAs) and three other OMB-recognized statistical units. Sector, agency, or program-specific statutes vary in their requirements related to statistical and administrative data. As a result, a typical scenario for Federal agencies is figuring out how to apply up to three relevant statutes when providing access to confidential data, and more if the dataset is the result of linking data from across different legal domains.

The Privacy Act of 1974

The Privacy Act of 1974 is the common floor of privacy protection in the Federal Government. The Privacy Act established confidentiality of individually identifiable information as a norm by requiring Federal agencies to protect personal information that "could result in substantial harm, embarrassment, inconvenience, or unfairness to [the] individual…"

The Privacy Act governs all Federal data "retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual." It says that "no agency shall disclose any record…except…with the prior written consent of, the individual to whom the record pertains." The Privacy Act provides exceptions for, among other things, routine uses by the agency, the Census Bureau under Title 13, and "to a recipient who has provided the

[5] Data collected on https://www.data.gov/

agency with advance adequate written assurance that the record will be used solely as a statistical research or reporting record, and the record is to be transferred in *a form that is not individually identifiable*." The Privacy Act also allows for disclosure of records under the Freedom of Information Act (FOIA). FOIA itself provides for release of records unless a listed exception is met, one of which is "specifically exempted from disclosure by statute." In practice, this means that data governed by the Privacy Act and lacking additional statutory disclosure protection can be released under FOIA.[6] In part because of the FOIA exemption, the Privacy Act by itself does not provide sufficiently strong privacy protection for confidential data used for evidence building in the Federal evidence ecosystem.

The Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)

PSAs and three statistical units recognized by OMB are eligible to protect their data under CIPSEA Subtitle A.[7] Subtitle A is designed "to safeguard the confidentiality of individually identifiable information acquired under a pledge of confidentiality for statistical purposes by controlling access to, and uses made of, such information."

CIPSEA relies on a clear, dichotomous division between PII and de-identified data. Similar to Privacy Act provisions, CIPSEA data are either identifiable or de-identified. CIPSEA provides strong protections for identifiable data while allowing full disclosure of de-identified records.

According to the 2007 OMB guidance on CIPSEA implementation, the actions necessary to ensure CIPSEA protection include: informing respondents about confidentiality protections and use of the data; collecting and handling information to minimize risk of disclosure, including properly training employees; ensuring the information is only used for statistical purposes[8]; reviewing releases to prevent identifiable information from being reasonably inferred by either direct or indirect means; and supervising and controlling agents who have access to confidential information.

One way CIPSEA protects confidential data is by limiting its use to ***eligible researchers*** called "agents." CIPSEA defines an "agent" as an individual with a contract or agreement to either: 1) perform exclusively statistical activities under the control and guidance of an office or employee of a CIPSEA agency, or 2) design or maintain data systems; agents must agree in writing to comply with all provisions of law that affect information acquired by the agency. Under this definition, external researchers and agency contractors alike may qualify as agents. Agents must be trained on CIPSEA requirements. Statistical agencies are required to: 1) emphasize to their officers, employees, and agents the importance of protecting the confidentiality of information in cases where the identity of individual respondents can reasonably be inferred by either direct or indirect means, and 2) train officers, employees, and agents in their legal obligations to protect the confidentiality of individually-identifiable information and in the procedures that must be followed to provide access to such information.

CIPSEA defines ***statistical activities*** as "the collection, compilation, processing, or analysis of data for the purpose of describing or making estimates concerning the whole, or relevant groups or components within, the economy, society, or the natural environment." The key feature of this definition is the emphasis on describing or making estimates concerning *groups*. CIPSEA further

---

[6] Microdata released under FOIA are redacted, but usually by agency FOIA staff or legal counsel, rather than statistical disclosure experts.

[7] 44 USC § 3501, https://www.gpo.gov/fdsys/pkg/USCODE-2010-title44/pdf/USCODE-2010-title44-chap35-subchapI-sec3501.pdf

[8] Memo #9 (*Federal Statistical System*) discusses the CIPSEA distinction between "statistical purposes" and "nonstatistical purposes."

clarifies that officers, employees, or agents (including contractors) of the agency may use confidential data exclusively for statistical purposes and "may perform exclusively statistical activities, subject to the limitations and penalties described in this title." Statistical purpose explicitly excludes uses requiring the identification of individuals or businesses, such as program administration or compliance monitoring. While the CIPSEA definition encompasses the statistical activities conducted for evaluation and policy research, these functions are not explicitly named in the statute. As a result, some agencies may not interpret allowable uses to include these functions.

Data acquired under a CIPSEA pledge of confidentiality "shall not" be disclosed in identifiable form, for any use other than an exclusively statistical purpose, except with the informed consent of the respondent, and even then only when authorized by an agency head and not otherwise prohibited by law. CIPSEA establishes ***penalties*** for any officers, employees, or agents of an agency who willfully disclose protected data collected for statistical purposes to a person or agency not entitled to receive it, including a class E felony charge and imprisonment for not more than 5 years and/or fines of not more than $250,000.

As noted in the memo on the *Federal Statistical System* (Memo #9), OMB has the authority to recognize additional Federal statistical offices under CIPSEA. One of CIPSEA's strengths is that it requires an agency or unit to have clear functional separation between statistical and non-statistical activities. Therefore, the mission of the agency or unit is typically 100 percent statistical, allowing the mission and culture to reinforce the legal and OMB policy requirements.

It is important to remember, though, that CIPSEA can be applied to specific datasets and not others. Agencies must clearly identify under which, if any, confidentiality law they are protecting a given dataset. If under CIPSEA, they then must follow the required steps to promise and maintain confidentiality. Some PSAs have chosen not to protect some or all of their datasets under CIPSEA as discussed in the memo on the *Federal Statistical System* (Memo #9). We are aware of two reasons why. First, the National Center for Education Statistics does not promise CIPSEA confidentiality because of provisions in the Patriot Act of 2001 that resulted in amendments to their law to the effect that "the data may now be used with a judge's order for matters relevant to an offense concerning national or international terrorism."[9] Second, as noted below under the section on data security, some PSAs have ceded control of their information technology resources under OMB-directed "IT centralization" initiatives. These PSAs do not believe they can meet all of the CIPSEA implementation guidance requirements, especially the assurance that no one other than a sworn agent can access the data. Memos on the *Federal Evidence Enterprise* (Memo #7) and the *Federal Statistical System* (Memo #9) propose institutional changes at OMB and in Departments with PSAs to address this challenge.

Agency-, Sectoral-, and Dataset-Specific Statutes

As noted above, access to and confidentiality of most Federal datasets used for evidence building are governed by multiple statutes, including those specific to the agency, the sector, or the dataset itself. In 1977, the Privacy Protection Study Commission noted that "the common element in…agency statutes and regulations is the prohibition of the release of information that can be associated with or identified to a particular [individual or business]."[10]

---

[9] https://nces.ed.gov/statprog/conflaws.asp. NCES emphasizes that the data are still kept confidential because "even in the case of a judge's order for matters relevant to an offense concerning national or international terrorism, the Attorney General must protect the confidentiality of the data."

[10] *Personal Privacy in an Information Society: The Report of the Privacy Protection Study Commission* (1977), https://epic.org/privacy/ppsc1977report/c15.htm

Many agency-, sectoral-, or dataset-specific statutes either prohibit "disclosure" without defining the necessary protections to prevent it or rely on a dichotomous approach to confidentiality like that in CIPSEA that prevents re-identification by direct or in-direct means through reducing and masking the amount of information available for analysis. For example, the National Database of New Hires (NDNH) is governed by the Privacy Act, the amended Computer Matching and Privacy Protection Act of 1988, and the child support program's authorization in Title 42.[11] The program's regulations, *Safeguarding and disclosure of confidential information,*[12] prohibited disclosures including "any information relating to…an individual who can be identified by reference to one or more factors specific to him or her, including but not limited to the individual's Social Security number, residential and mailing addresses, employment information, and financial information." PSA's authorizing statutes include confidentiality provisions that apply in addition to the Privacy Act and CIPSEA. Many of those, such as the Census Bureau's Title 13, reflect the dichotomous identified/de-identified approach to confidentiality.

Other statutes and implementing regulations, like the Privacy Rule of HIPAA and the 2008 regulations for the Family Educational Rights and Privacy Act (FERPA),[13] attempt to establish more nuanced standards of acceptable re-identification risk in public releases of data. The HIPAA "Safe Harbor" de-identification standard provides a list of 18 identifiers that must be removed from protected health information in order for the resulting data to be deemed de-identified. Central to the Safe Harbor approach is an element of geographical aggregation, achievable in most cases by reporting only the first three digits of the zip code, so that there is an underlying population of at least 20,000 individuals in each reported geographic area. Safe Harbor, therefore, acknowledges that publicly released data carry a risk of re-identification and specifies an approximate quantity of acceptable risk.

Studies about re-identification risk under Safe Harbor have generally found a reduced risk[14] when matching on demographic fields such as year of birth, gender, and the first three digits of the zip code (people registers). A forthcoming study by Sweeney et. al matched to data collected under HIPAA using these fields and additional data on the characteristics of housing units (property registers).[15] Using both people and property registers increased the risk of re-identification as verified by the researchers. In one of the two communities studied (the one with the most heterogeneity), the study correctly identified 8 of 40 (20 percent) of the HIPAA records by name and 9 or 40 (23 percent) by address. The authors note that "our results are not necessarily the worst case for re-identifying the data" as additional HIPAA fields could have been used for matching and because "a re-identification expert may know more than our Attackers."

The standard in FERPA's 2008 regulations is less specific than HIPAA's Safe Harbor, but it also acknowledges the public release versus re-identification risk dilemma. Under the 2008 FERPA regulations, educational agencies and institutions are prohibited from releasing information from student education records if doing so would allow "a *reasonable* person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with *reasonable* certainty" (emphasis added). In contrast with the HIPAA Safe Harbor standard, FERPA does not provide an objective measurement of acceptable risk, opting instead for the much more

---

[11] 42 USC 653

[12] 45 CFR 303.21

[13] https://www.gpo.gov/fdsys/pkg/FR-2008-12-09/pdf/E8-28864.pdf

[14] Relative to non-"de-identified" data

[15] Sweeney, L., Yoo, J., Perovich, L., Boronow, K, Brown, P., and Brody, J. *Re-identification Risks in HIPAA Safe Harbor Data: a study of data from one environmental health study.* (forthcoming)

subjective "reasonable person" standard. Both approaches acknowledge that complete de-identification of data would be either impossible or impractical.

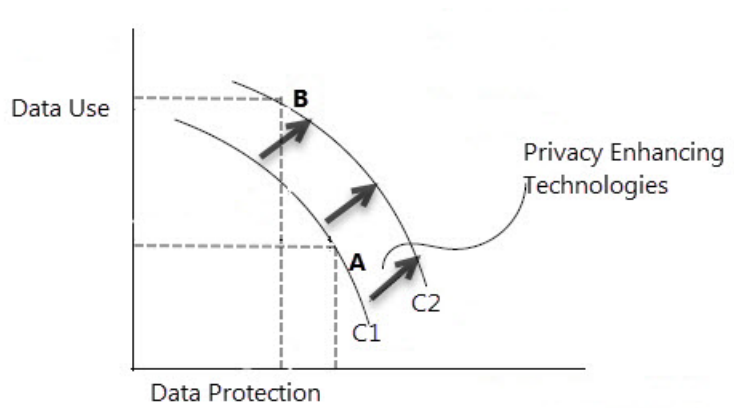
New and Evolving Technologies

New and evolving technologies have the potential to both minimize risk of de-identification and allow for increased use. At its February meeting, the Commission heard about new technologies that allow researchers to analyze data without accessing the underlying original identifiable microdata. Examples include secure multi-party computing and synthetic data with a verification server.[16]

Other new technologies, such as differential privacy and *k*-anonymity, make "provable privacy guarantees."[17] These formal protection models create alternative datasets that break the link with the original identifiable data but maintain their statistical properties. As a result, they actually shift the curve depicting the relationship between data protection and data use.

Figure 2 shows how privacy enhancing technologies can increase the amount of data use while improving confidentiality. Curve C1, from Figure 2 below, represents the relationship under current conditions. Curve C2 represents the increase in both data use and confidentiality protection that results from using privacy-enhancing technologies. Both curves show that as data confidentiality protections increase, the uses of data become more limited. For the same amount of confidentiality protection, outcome B provides greater capacity to use data. Correspondingly, for the same amount of data use, outcome A provides less confidentiality protection than a comparable point on C2.

Figure 2.—How Privacy Enhancing Technologies Shift the Use/Protection Frontier Curve



Technology is also evolving for data security, which considers data integrity and availability in addition to confidentiality. Preserving data integrity is particularly important for Federal evidence building. While few attacks in the current environment focus on data integrity, bad actors could try to change the data if they thought they could thereby change a policy decision. Furthermore, "hacks" on confidential data increase re-identification risk and may harm public trust as discussed in *Protecting Privacy by Increasing Transparency* (Memo #11). Data security in Federal Departments today is generally the responsibility of Chief Information Officers who implement standards promulgated by OMB in consultation with the National Institute for Standards and Technology (NIST).

As noted in the memo on the *Federal Statistical System* (Memo #9), technical independence to secure confidential data can be a challenge for some PSAs. In some cases, the need for PSAs to protect confidential data conflicts with the priorities of departmental Chief Information Officers (CIOs), many of whom, particularly since the recent passage of the Federal Information Technology Acquisition Reform Act (FITARA), have been reluctant to comply with the PSA's requests to train

[16] For more information see the briefing materials from the Commission's February meeting: https://community.max.gov/display/CrossAgencyExternal/February+24%2C+2017.
[17] Ibid., p. 52.

and qualify IT staff as CIPSEA agents. Sometimes the concern about independence for data security is because the statistical agency has lost direct control of information technology (IT) systems during OMB-required IT consolidation. Each of concerns poses challenges to data confidentiality for evidence building and requires resolution.

Data confidentiality and data security are both important, but distinct, parts of data protection in the Federal evidence ecosystem. Data security relates to the development and implementation of managerial, operational, and technical safeguards to protect the integrity of information systems.[18] Several recently publicized cybersecurity incidents (hacking) of Federal systems, which resulted in the release of millions of confidential records, has raised the visibility of vulnerabilities in the current level of data security of Federal systems. High standards for data security should be established so that agencies will strive to meet, and the technology community will have incentives to work toward, improved data security. Similar to the adoption of privacy protecting technologies and the establishment of procedures for quantifying and measuring the re-identification risk, procedures for strengthening data security should be flexibly designed, periodically reviewed, and refreshed as new challenges arise and new capabilities develop.

Monitoring and Mitigating Risk: "Near Miss" Reporting

In addition to establishing legal and technology-based approaches to protecting confidentiality, Federal Departments should consider developing systems to identify violations and re-identifications, determine their cause(s), mitigate those cause(s), and consider revising procedures to reduce the risk of future, similar incidents. "Near miss" reporting has become a common feature in some policy fields with the recognition that progressive attention to failures can be pivotal to the success of policy implementation. Matthew Syed summarizes that "if failure is a tragedy…learning from failure takes on a moral urgency."[19] Today, near miss reporting exists within the Federal government for cybersecurity breaches, aviation safety,[20] mineral drilling in sensitive areas,[21] occupational safety,[22] and other areas where the potential for harm is substantial. Third parties have also developed or considered similar reporting systems for bio labs, fire fighters,[23] and health care.[24] No consistent process exists for acknowledging similar feedback loops with regard to re-identification or confidentiality issues. In developing a feedback loop process, Departments should consider incentives for accurate and timely reporting. Auditing the feedback loop may provide meta-transparency, as discussed in *Protecting Privacy by Increasing Transparency* (Memo #11).

---

[18] Swanson, M. (2006). Guide for Developing Security Plans for Federal Information Systems (National Institute of Standards and Technology, Computer Security Division). http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-18r1.pdf

[19] Syed, M. (2015). *Black Box Thinking: The Surprising Truth About Success*. Hodder & Stoughton, p. 16.

[20] FAA Aviation Safety Information Analysis and Sharing (ASIAS). http://www.asias.faa.gov

[21] SafeOCS Outer Continental Shelf safety tracking. https://www.safeocs.gov/

[22] National Safety Council, *Near Miss Reporting Systems*. http://www.nsc.org/WorkplaceTrainingDocuments/Near-Miss-Reporting-Systems.pdf

[23] *A Framework for Collecting Emergency Responder/Roadside Worker Struck-by/Near-Miss Data*, 2013 March. Prepared by Gannett Fleming for American Association of State Highway and Transportation Officials (AASHTO).http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP20-07(321)_FR.pdf

[24] Crane, S. (2015). Reporting and Using Near-miss Events to Improve Patient Safety in Diverse Primary Care Practices: A Collaborative Approach to Learning from Our Mistakes. The Journal of the American Board of Family Medicine, 28(4), 452-460. http://www.jabfm.org/content/28/4/452.full

SUMMARY OF COMMISSIONER CALL

On April 26th, 2017, CEP staff discussed the first draft of this memo with interested Commissioners available during that time. Chair Abraham and Commissioners Meyer, Rice, Troske, and Wallin joined the call. Commissioner Groves later provided feedback by email and Commissioner Sweeney provided new, forthcoming research to help flesh out some of the memo's points.

Commissioners indicated that the memo should acknowledge explicitly the fallacy of the protective effects of de-identification of data in statutes like CIPSEA and discuss how new developments in technology have the potential to both exacerbate risk through improved linkages and to reduce risk through privacy enhancing technologies. Commissioners liked the principles embodied in the Five Safes framework but preferred to use it as basis for the elements that should be considered to protect confidentiality and security rather than an organizing framework itself. They suggested that the memo use the provisions of CIPSEA as a structure to address allowable uses and eligible researchers. Such a structure would also help Commissioners identify recommendations to Congress for changes in statute separately from recommendations to OMB about ways to use its existing authority to set standards for a risk management process that frames, assesses, responds to, and monitors re-identification risk and information security.

As a result of feedback from the Commissioner call and additional input from Commissioners Rice and Sweeney, this memo and its recommendations have been substantially revised, shifting away from primarily implementation strategies to legal and policy ones.

DRAFT RECOMMENDATIONS

❖**Finding:** The Commission's charge is to identify the optimal arrangement by which to increase the creation and use of evidence in a manner that considers privacy. The Commission finds that the most privacy protective approach is to generate evidence within a well-coordinated ecosystem in which strong laws, policies, and technologies work together to protect individual privacy in every single government release of data or information.

❖**Finding:** The laws governing confidentiality protections for Federal data—including the Privacy Act of 1974, the Census Bureau's Title 13, and the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)—have historically required de-identification of individuals or businesses in datasets prior to release. This has meant removing direct identifiers such as name and address and masking other indirectly identifying information. Current understanding of re-identification acknowledges the cumulative risk from the release of information about individuals and businesses across datasets and across time. The more information readily available about distinct persons or entities the greater the chance of re-identification. Therefore, the Commission finds that any paradigm seeking to reasonably protect privacy must seek to consider all other readily available information.

❖**Finding:** CIPSEA currently uses a de-identification approach to privacy protection. CIPSEA protects certain data against the risk of re-identification by establishing criteria for a variety of technological data privacy approaches and for allowable uses ("statistical purposes") and for eligible researchers ("agents"). The Commission finds that these established criteria define the strict conditions under which researchers may access and use sensitive data and, therefore, help to protect privacy.

►**Recommendation #10.1:** The Commission recommends that the Congress and the President enact a single standard specifying that data collected pursuant to a representation that it will remain confidential may only be publicly disclosed as de-identified data when the risk of re-identification is sufficiently low that it will not pose an unreasonable harm to named individuals or businesses.

►**Recommendation #10.2:** The Commission recommends that the Congress amend relevant laws governing data access and confidentiality to include legal standards defining allowable uses and eligible researchers comparable to those in CIPSEA.

Recommendation #10.1 encourages Congress to establish a single standard for determining re-identification risk, overriding current standards for disclosure across datasets governed by Federal law. One approach to implementing this recommendation could be changing key laws that the Commission has heard concerns about, such as the Privacy Act, CIPSEA, HIPAA, and FERPA. Another approach would be a "notwithstanding" provision that would supersede existing laws, including some with which the Commission is less familiar but may have similar shortcomings (e.g., those governing privacy in financial data). Recommendation #10.1 addresses the critique that some laws set a standard for public release of data that is demonstrably too low (see, for example, Sweeney's forthcoming paper on HIPAA re-identification). Since data releases outside the Federal Government also affect the overall risk of re-identification, enacting a new Federal statutory standard to address the risk of re-identification would bring current law in line with today's understanding of the importance of risk management. To the extent that a given law, such as HIPAA, governs activity outside of government, those practices would also be required to adapt to the new legal standard. Furthermore, recommending a standard that applies to the large inventory of Federal data that could be used for unauthorized re-identification is clearly within the Commission's charge and would provide increased confidentiality protection in and of itself.

Recommendation #10.2 encourages Congress to provide administrative data used for evidence building with the same data access provisions that currently help protect the confidentiality of data under CIPSEA. While the Commission could propose to extend CIPSEA coverage to key administrative datasets, this recommendation gives those datasets essential protections without requiring that units within program agencies themselves qualify as CIPSEA units, though in practice there may be a similar practical effect. Taken together, Recommendations #10.1 and #10.2 will enable program agencies to develop tiered access mechanisms for their datasets by defining allowable uses for eligible researchers to analyze more sensitive and identifiable data for evidence building and by developing query tools or risk-evaluated masked public use data for other purposes and actors. Taken together, these two recommendations offer a comprehensive approach to addressing re-identification risk across the board, not just for evidence building and not just under CIPSEA, since the activities more broadly have an inextricable effect on what happens within evidence building and agencies that conduct or facilitate it.

Rather than the "notwithstanding" approach (the "*gold option")*, the Commission could consider one of two less comprehensive alternatives.

1. *Silver option:* The Commission could apply Recommendations 10.1 and 10.2 to CIPSEA-protected data and a few key additional datasets. This alternative would bring data collected or protected under those laws into compliance with risk and data access standards. HIPAA should likely be considered a priority under this option because of new research showing the risk of re-identification is greater than previously believed. Commissioners could choose additional datasets from within the Top 20.

2. *Bronze option:* The Commission could apply Recommendation 10.1 to CIPSEA-protected data only. This alternative would fix the current outdated standard of de-identification in CIPSEA. Having a higher bar for disclosure based on re-identification risk limited to statistical data governed by CIPSEA would improve confidentiality protection for those data. However, the existence of Federal administrative data protected at lower standards could compromise even CIPSEA protections because the risk of re-identification for individuals increases with each public release.

Discussion Questions:

1. Is "unreasonable harm" the right standard for #10.1?
2. Should the standards in #10.1 and #10.2 apply to all Federal data (*gold option*); CIPSEA-protected data plus a few additional administrative datasets in the Commission's top 20 such as HIPAA-protected health data and FERPA-protected education data (*silver option*); or only to CIPSEA-protected data (*bronze option*)?

❖ **Finding:** The Commission finds that the protections enabled for statistical activities in CIPSEA provide essential basic parameters necessary for increasing data confidentiality and could be applied consistently to program evaluation and policy research.

► **Recommendation #10.3:** The Commission recommends that the Congress and the President enact legislation clarifying the applicability of CIPSEA protections for the statistical functions within program evaluation and policy research.

Historically, CIPSEA was intended to cast a broad net for providing protections across data collection for statistical purposes, which includes evaluation and policy research activities. However, initial guidance prioritized implementation in statistical agencies. Notwithstanding OMB's encouragement that statistical activities can include evaluation and research[25], many evaluation offices and other units within agencies currently collect data that could reasonably be protected by CIPSEA, yet the evaluation offices continue to perceive that the law is not relevant for their use.

A relatively modest amendment to Subtitle A of CIPSEA's definition of "statistical activities" would enable evaluation and policy research that conforms to the definition of "statistical purposes" to be eligible for CIPSEA protections.

❖**Finding:** The Commission finds that privacy protecting technologies and protocols that enable a more comprehensive assessment of the risk associated with any given data release can expand access to data while not adversely affecting, and potentially improving, privacy. Because technology is always evolving, the Commission does not recommend the adoption of any one specific approach, but rather acknowledges the critical need for leadership and partnership that will advance and accelerate the adoption of technical strategies to enhance data confidentiality.

❖**Finding:** Data confidentiality and data security are both important, but distinct, parts of data protection in the Federal evidence ecosystem. Preserving data integrity is particularly important for Federal evidence-building. Information technology requirements imposed on agencies in the

[25] *Guidance for Providing and Using Administrative Data for Statistical Purposes*, M-14-06, https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

evidence building community may conflict with those agencies' priority to maintain confidentiality.

►**Recommendation #10.4:** The Commission recommends that the President direct Federal agencies, such as the Census Bureau, to spearhead efforts to accelerate the adoption of privacy enhancing technologies as a way to make data access more secure and confidential.

►**Recommendation #10.5:** The Commission recommends that the President direct Federal agencies, such as the Census Bureau and the National Institute of Standards and Technology, to spearhead Federal efforts to develop procedures for quantifying and measuring the re-identification risk inherent in public releases of data.

►**Recommendation #10.6:** The Commission recommends that the President direct Federal agencies to regularly review security protocols for the transmission and storage of confidential data used for evidence building, consider needed variation in information technology needs of those agencies charged with holding and protecting sensitive data for evidence building, and ensure that such protocols adapt and evolve to address emerging threats.

Over the past several months, the Commission has explored a range of privacy protecting technologies, such as synthetic data and secure multi-party computation. These emerging technologies can help expand access to data while not adversely affecting privacy. Instead, these technologies potentially may help improve privacy. Numerous presenters over the past several months cautioned the Commission about seeking to identify a single "silver bullet," instead proposing that the application of numerous technologies would be the most effective approach to maximizing privacy protections. The privacy protecting technologies the Commission has explored are in various stages of development and few of the newer ones are ready to be deployed at scale at the current time. That said, several Federal agencies are currently investing in a portfolio of research that seeks to accelerate the development of these technologies, and are conducting a set of pilot efforts to apply them in practice. This kind of innovation is required in order to keep pace with changes in technology. The Federal Government can and should continue to advance a set of standards that requires agencies to continually assess the extent to which they are maximizing their capacity to make data access secure and confidential. This recommendation is also consistent with those included in the recently released NAS report *Innovations in Federal Statistics.*[26]

The first recommendation in the *Standards* memo (Memo #4) directs OMB to "create agency-wide standards for risk-based and quantifiable confidentiality *methods* agencies must employ *to protect against re-identification in released data products*." This recommendation grows from the acknowledgement that the current process for assessing the risk of re-identification within a particular data release is insufficient because each release is assessed in isolation. An updated standard, incorporating formal privacy protection models such as differential privacy and *k*-anonymity, would enable stronger confidentiality protections for individuals represented in the dataset. In order to establish these more rigorous standards, however, additional work must be done to develop the procedures necessary for quantifying and measuring the re-identification risk inherent in public releases of data.

Data management policy for evidence building is related to, but distinct from, information technology policy. The Commission's recommendations to establish specific and unique legal

[26] Innovations in Federal Statistics. (2017). National Academies Press. https://www.nap.edu/catalog/24652/innovations-in-federal-statistics-combining-data-sources-while-protecting-privacy

authorities and confidentiality requirements for the components of the Federal Statistical Data Network (see Memo #9) will invariably lead to the need to expressly calibrate information technology policy accordingly. Moreover, the complexities and challenges of information security, as well as the benefits of consolidated IT service provision, can likewise present challenges for PSAs when viewed as components of departments, rather than elements of a broader system.

Discussion Question: Is Recommendation #10.6 specific enough about data security improvements the Commission would like to see? What additional recommendations should be included?

❖ **Finding:** The Commission finds that feedback loops can successfully provide valuable information about unplanned events that do not result in tangible harm, but had the potential to do so, and can be a productive mechanism for preventing future harms in policy domains that expose individuals to risk. The strengthening of feedback loops in government can be expanded to establish a stronger system for assessing and mitigating potential vulnerabilities for re-identification with proactive monitoring, recognizing that risks continually evolve.

►**Recommendation #10.7:** The Commission recommends that the President direct Federal Departments to monitor, evaluate, and implement both remedies and penalties for procedural violations and unauthorized re-identification.

The Commission's principles articulate respect for individual privacy and confidentiality as an imperative. With this in mind, developing more robust processes and procedures within government to routinely monitor and improve data protection is a necessity. Recommendation #10.7 establishes the expectation that Federal Departments will engage in such monitoring and review to learn about, and mitigate, potential failures. OMB would likely play an important role in developing standardized processes for use by agencies to implement this recommendation, but the ultimate responsibility to do so would fall to Departments.

In addition, Recommendation #10.7 recognizes that Departments may need to consider a range of realistic remedies or penalties, short of the criminal and civil penalties specified in CIPSEA. For example, a Department could identify violations that merit limiting an individual's or institution's access to data for a defined period of time.

Discussion question: Does the Commission want to make a recommendation about the importance of auditing for protecting privacy? Does such a recommendation with the feedback loop or elsewhere?

LIKELY REACTION TO RECOMMENDATIONS:

- ***Congressional:*** Acknowledging the limitations of the current de-identification standard may concern legislators already skittish about the ability of the Federal Government to protect privacy. Taken in isolation, these findings and recommendations could be used to further restrict access in the name of privacy. Critiquing HIPAA could affect ongoing discussions about health care. Establishing a "notwithstanding" provision to establish risk measurement and management as the prevailing standard for protecting confidentiality is the most comprehensive solution to achieving a clear and consistent legal framework for data protection. But "notwithstanding" provisions, reasonably, have a high bar for passage and must be considered in light of legislative preferences related to specific datasets. In contrast, changing CIPSEA and other priority statutes involves opening and coordinating multiple amendments. However, because Congress explicitly asked the Commissioner to consider

how to protect privacy in evidence building, they may welcome the opportunity to set a consistent high bar of privacy protections.

- ***Executive:*** Federal agencies believe they are already appropriately applying de-identification and risk management standards pertaining to their data. They may see the establishment of a new management standard as an added burden. PSAs may welcome the "notwithstanding" provision because it would bring consistency and clarity to their authority to protect data. Furthermore, having a high and consistent confidentiality bar in statute would make it possible to allow CIPSEA Subtitle B sharing to be expanded across PSAs, as recommended in the memo on the *Federal Statistical System* (Memo #9). PSAs would also welcome the recommendation to consider separate IT standards for confidential data used for statistical purposes because of current concerns about Federal cybersecurity monitoring of transmitted data. Program agencies may have concerns about the implications of the "notwithstanding" provision for their own data use; likely these can be addressed in guidance. On the other hand, program agencies should welcome clarification about allowable uses and eligible researchers. Federal units performing evaluation and policy research functions will almost certainly welcome clarity about the inclusion of these activities in "statistical activities" under CIPSEA. Furthermore, agencies should welcome leadership by the Census Bureau and others in developing and piloting new technologies to improve confidentiality.

RECOMMENDATION MEMO #11
UPDATED AFTER MAY 8 CALL

Protecting Privacy by
Increasing Transparency About Evidence Generation and Benefits
Lead Commissioner: Groves
Lead Staff: Martinez, Boivin, and Hawes

ISSUE: What improvements to the transparency of Federal data collection and use can the Commission recommend to bolster public trust while achieving the goal of increased evidence building?

The Commission is charged with considering the privacy implications of increasing data linkage for evidence building. Privacy, in this context, has two major aspects: *Protecting Privacy by Increasing Confidentiality* (Memo #10) describes technical strategies to ensure data confidentiality and security. Second, this memo describes the use of communication strategies to meet the legal and ethical requirements for transparency, by assuring the public that their data are being protected, used only for legal and allowable purposes, and for the public good.

Transparency throughout the entire research process can help improve public trust and credibility of statistical, evaluation, policy research, and policy analysis (SERA) activities, while ensuring individuals whose data are used for those purposes have the opportunity to participate and know how their data will be used and for what benefit.

One vehicle for transparency is the use of notice and consent prior to data collection, typically required under several different Federal laws. This paper specifically addresses the limitations of current notice and consent processes for *secondary* statistical uses of administrative data and includes recommendations concerning the concept of increased transparency. It identifies recommendations for communicating to the public about the data confidentiality and security protections afforded by a strong legal and policy framework (discussed in Memo #10), improving communication around specific SERA project uses of data, anticipated benefits, and results with members of the public to demonstrate benefits. [Sec. 4(a)(1), 4(a)(3) & 4(b)(2)(K)] *Relates to Confidentiality (#10), Facility (#2), Standards (#4).*

* * * * *

BACKGROUND AND ANALYSIS

Privacy is a central element of the Commission's statutory charge. While identifying the "***optimal arrangement*** for which ***administrative data…,*** survey data, and related statistical data series may be ***integrated and made available*** to facilitate" evidence building, the Commission was directed to consider how "integration might lead to the intentional or unintentional access, breach, or release of personally-identifiable information or records" and make recommendations on data security. These topics are addressed in *Protecting Privacy by Increasing Confidentiality* (Memo #10). In addition, the Commission is to consider "how ***individuals whose data are used*** [in a clearinghouse] should be ***notified***…."

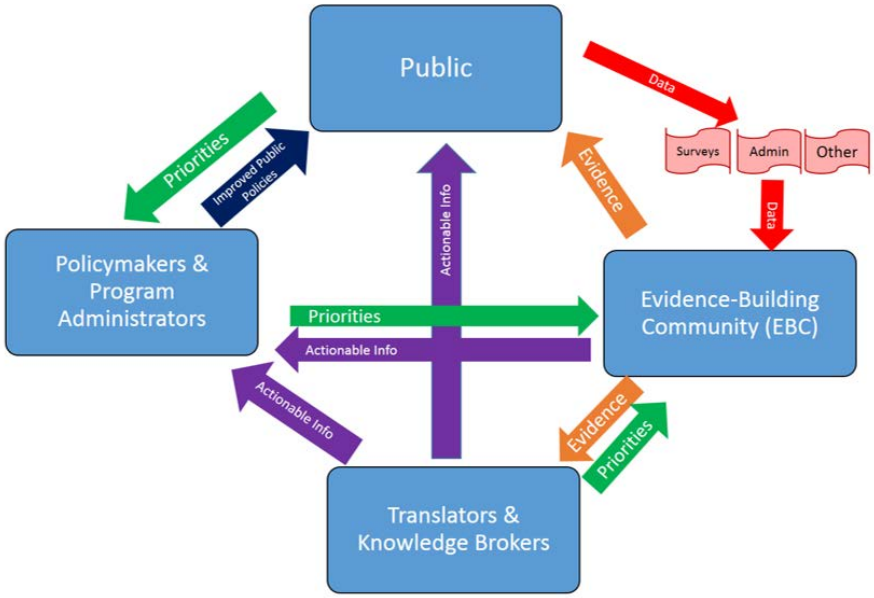
Notification in this context is about respect for personal autonomy and for guaranteeing effective notice, and serves both a practical/legal purpose and an ethical one. Providing effective notice seeks to accomplish three of the privacy objectives upon which the entire Federal privacy framework is

constructed: the existence of federal data systems containing personal information should not be a secret; individuals should know what information about them is being collected; and information about individuals should not be re-used for other purposes without their consent.[1] These principles, and others expressed in the 1973 Health, Education, and Welfare Report, became known as Fair Information Practice Principles (FIPPs). *Protecting Privacy by Increasing Confidentiality* (Memo #10) recommends legal and technical improvements that the Federal government could make to address privacy risk around re-identification and improper use of data. This memo provides recommendations on how to address the privacy principles of transparency, notice, and consent essential to maintaining public trust and credibility in the Federal evidence-building ecosystem.

Introduction

As depicted in Figure 1, data from the public are an essential input for evidence builders. Traditionally, most evidence building was based on primary data collection via surveys, and most of those were voluntary. Therefore, agencies for whom such collection is their mission have long understood the critical role of public cooperation, premised on the public trusting the information collector and its intended use and handling of the data.

Figure 1. The Federal Evidence Ecosystem



The United Kingdom had to confront the relationship between transparency and public trust in the wake of several data breaches in the 2000s, described by Tanvi Desai at the Commission's January meeting. A subsequent UK report on privacy and transparency stated that "the success of a transparency programme depends on public trust, which is more likely to be preserved if the public feels that those in charge of the programme respect its privacy concerns. Privacy is extremely important to transparency. The political legitimacy of a transparency programme will depend crucially on its ability to retain public confidence." The report further acknowledged the importance

[1] HEW. (1973). *Records, Computers, and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems.* U.S. Department of Health, Education & Welfare.

of auditing as a component of transparency, saying that, "Transparency about transparency – meta-transparency – is essential for preserving trust and confidence."[2]

The Congressional Research Service issued a report in 2012 on government transparency that addressed the importance of ensuring that the public understands government information:

> "Transparency may be defined as the disclosure of government information and its use by the public. Transparency, under this definition, requires a public that can access, understand, and use the information it receives from the federal government. Certain government actions and records are protected from public release for, among other reasons, national security and personal privacy reasons….Attempts to make the executive branch more transparent require Congress, the public, and the executive branch itself to ensure that appropriate information is released and that the information can be used and analyzed. Users of such released information, therefore, must have the time and the tools to interpret and understand the data in order to hold government accountable."[3]

Therefore, building evidence about the effectiveness of government policies and programs, and making those results available to the public, contributes to government transparency. Yet, because evidence building itself requires use of government data, transparency about the acquisition and use of those data is also important.

Former Census Deputy Director Tom Mesenbourg said that "the collection of accurate and complete information…is predicated on trust….any effort to maximize the quality of an information collection requires ensuring that the individuals from whom information is collected trust the statistical agency to protect the data's confidentiality, to ensure the respondents' privacy, and to use the information appropriately. Without that trust, data collection will be undermined before it even begins…"[4] The Commission heard similar comments through its public input process.[5]

As Figure 1 also shows, the output of evidence building is intended to inform a mix of policy makers and the public. If they do not trust the entity generating the evidence, then they may not trust the results. This lack of credibility means that the results may not be used, and worse, the mistrust may harm the larger evidence effort by creating disregard or suspicion more generally.

OMB issued Statistical Policy Directive One: *Fundamental Responsibilities of a Federal Statistical Agency* as a mechanism to assist principal statistical agencies (PSAs) in maintaining public confidence. It builds from both principles espoused by the United Nations and from the National Academies' long-

---

[2] O'Hara, Kieron. (2011). *Transparent Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office.* United Kingdom: School of Electronics and Computer Science, University of Southampton, Highfield, p .57. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/61279/transparency-and-privacy-review-annex-a.pdf

[3] Congressional Research Service. (2012). *Government Transparency and Secrecy: An Examination of Meaning and Its Use in the Executive Branch*. https://fas.org/sgp/crs/secrecy/R42817.pdf.

[4] Mesenbourg, (2009). How Statutory and Regulatory Changes May Create New and Unintended Challenges for Statistical Agencies. Draft.

[5] Levenstein, from Chicago Public Hearing Testimony.

standing *Principles and Practices of a Federal Statistical Agency*.[6] Of Directive One's four "fundamental responsibilities," two speak to the relationships described above:

- Responsibility 2--Conduct credible and accurate statistical activities.
- Responsibility 4—Protect the trust of information providers by ensuring the confidentiality and exclusively statistical use of their responses.[7]

In short, public trust is essential to the evidence building process.

The Congress and the President asked the Commission to consider how to increase evidence building through greater *secondary use* of administrative and statistical data the government already collects. All of its recommendations are designed to improve evidence building capacity and volume while also significantly increasing privacy. The Commission seeks to achieve this goal by removing barriers to secondary uses of existing data while providing enhanced confidentiality protections.

Evidence constructed through secondary analysis of administrative data is one step removed from the point of collection. Developing additional evidence by linking with other administrative and survey data moves the data even further down the line from the original collection. In this context, public trust and transparency becomes even more essential. Margaret Levenstein, in her public comments to the Commission, stated that the criticality of maintaining public trust "…is just as true when administrative data is re-purposed for statistics. Undermining this trust undermines statistical measurement as well as the effectiveness of the programs upon which the statistics are based."

She posited that "[t]he integration (linking) of administrative and/or survey data in a clearinghouse without question increases the risk of disclosure of entities within the data; however, the federal statistical community and the research data community have a long history and reputation for protecting confidentiality. This reputation must be maintained and protected.…"[8] Mesenbourg identifies two additional "distinct challenges to respondent privacy and confidentiality [from linkages]: they can be viewed as an excessive intrusion into an individual's personal life…and they can increase the degree to which individuals do not understand the collection and uses of their information."

The perception or excessive intrusion may occur because when individuals are asked to provide data about themselves (either through a survey or administrative form), they tend to discount the future privacy risks and uncertainties associated with secondary data uses and linkages and focus principally on the immediate context or transaction.[9] As *Protecting Privacy by Increasing Confidentiality* (Memo #10) demonstrates, integrating then releasing information about individuals from multiple sources increases the risk of unauthorized re-identification. Transparency about the use of data for *only*

---

[6] UN. (2003).*The Handbook of Statistical Organization: The Operation and Organization of a Statistical Agency*. Third Edition. New York: United Nations, Department of Economic and Social Affairs. & https://www.gpo.gov/fdsys/pkg/FR-2014-12-02/pdf/2014-28326.pdf

[7] Responsibility 1: Produce and disseminate relevant and timely information. Responsibility 3: Conduct objective statistical activities. https://www.gpo.gov/fdsys/pkg/FR-2014-12-02/pdf/2014-28326.pdf.

[8] Levenstein, from CEP Chicago Public Hearing.

[9] Acquisti, Alessandro. (2004). "Privacy in electronic commerce and the economics of immediate gratification." *Proceedings of the 5th ACM conference on Electronic commerce*. & Acquisti, Alessandro, and Jens Grossklags. (2003) "Losses, gains, and hyperbolic discounting: An experimental approach to information security attitudes and behavior." *2nd Annual Workshop on Economics and Information Security-WEIS*. Vol. 3. & Wilson, Dave, and Joseph S. Valacich. (2012). "Unpacking the privacy paradox: Irrational decision-making within the privacy calculus." & Norberg, Patricia A., Daniel R. Horne, and David A. Horne. "The privacy paradox: Personal information disclosure intentions versus behaviors." *Journal of Consumer Affairs* 41.1.

statistical purposes and the application of good data stewardship practice, backed by sufficient legal protections, will help members of the public who provide government data develop trust that the resulting information will never be used to affect them individually.

As Mesenbourg notes, greater temporal distance from the point of collection for secondary uses makes it even more difficult to communicate meaningfully what is happening with an individual's information. Concern about the privacy implications of data linkage is one of the main reasons that administrative agencies must follow strict review processes under the Computer Matching Act (CMA).[10] In that context, data linkages for the purpose of administrative program determinations may have a significantly higher privacy impact to the individual than the combined impact of the source datasets alone. Linkages for statistical purposes, however, are not subject to the CMA, precisely because Congress recognized that they are inherently different (i.e., by definition, not for the purpose of making "decisions concerning the rights, benefits, or privileges of specific individuals.") because they focus on describing the characteristics of groups.[11]

While the risk of harm to the individual is certainly lower for data integrated for statistical purposes than non-statistical ones, privacy considerations remain relevant due to the risk of lost public trust for the overall endeavor. As the Commission learned from Commissioner Ohm and others, the FIPPs are a nearly ubiquitous framework for addressing privacy concerns. The FIPPs suggest a three-part approach to building and maintaining public trust: apply strong data protection practices such as those outlined in *Protecting Privacy by Increasing Confidentiality* (Memo #10); be transparent about these data practices; and to be fully successful, be transparent about intended and actual data uses.

What do we know about public trust?

Public trust in the accuracy and validity of statistical data is strongly correlated with trust in the statistical agencies that produced them.[12] Recent well-publicized breaches of confidential data from the Office of Personnel Management have eroded public trust in the Federal Government's ability to keep data secure and resulted in comprehensive efforts across Departments to increase data security and restore the confidence of the American people. OPM established a cybersecurity resource center with information about the breach.[13] In 2014, the National Institute of Standards and Technology (NIST) created a cybersecurity risk framework[14] through collaboration between government and the private sector to address and manage cybersecurity risk.

The memo on the *Federal Statistical System* (Memo #9) describes how Principal Statistical Agencies (PSAs) have a long-standing reputation for maintaining data confidentiality and security. Regular consumers of federal statistics have higher levels of trust in the statistical system, both overall, and in comparison with other non-statistical agencies.[15] This finding suggests that familiarity with an agency's programs and awareness of its mission can improve the public's trust in the agency's

---

[10] The CMA amended the Privacy Act of 1974 in 1988.

[11] Computer Matching and Privacy Protection Act of 1988 (P.L. 100-503)

[12] Mitchell, Melissa, Jennifer Childs, and Morgan Earp. (2013). *Monitoring and detecting shocks that influence public trust towards the Federal Statistical System.* American Association of Public Opinion Research Conference Paper. (http://www.aapor.org/AAPOR_Main/media/AnnualMeetingProceedings/2013/Session_H-5-2-Mitchell.pdf

[13] https://www.opm.gov/cybersecurity/. A recommendation on data security is included in *Protecting Privacy by Increasing Confidentiality* (Memo #10).

[14] NIST. (2014). *Framework for Improving Critical Infrastructure Cybersecurity*. February 2014. https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf.

[15] *Ibid* and Hunter-Childs, Jennifer, Ryan King and Aleia Fobia. (2015). Confidence in Us Federal Statistical Agencies. *Survey Practice.* http://www.surveypractice.org/index.php/SurveyPractice/article/view/314/html_38

safeguarding of sensitive information.[16] Research by the Census Bureau shows the framing of data use for the public good matters; when respondents understand that using administrative data will save taxpayer dollars and help contribute to a quality census they are more likely to favor their use. The Census Bureau found that the public wants to know how that its information is protected and how it is being used.[17] During the December and January Commission meetings, the witnesses representing data access models reported procedural violations but no actual unauthorized re-identifications. Developing new ways to use existing data for evidence building requires data confidentiality protections including restriction of use to eligible researchers for statistical purposes. Outlining these protections in a carefully planned communications strategy can improve transparency and build public trust.

What do agencies do to provide transparency today?

At the September meeting, Marc Groman, then-Senior Advisor for Privacy at OMB and Chair of the Federal Privacy Council, focused his remarks on the centrality of transparency, and how the government implements transparency requirements in law.[18] He said:

> [As] the United States Government we have an obligation to be transparent to our citizens about what we collect and how we use it. We make representations about use and we have to honor those representations...[by] explain[ing] how the data collect[ed] will be used. And that is a core principle of the Privacy Act or other laws and of privacy generally.
>
> And so one of the ways we think about that is that - and there are varying degrees of effectiveness agency to agency - but agencies are required by law to draft and publish a Privacy Act System of Records Notice for data that is in a Privacy Act system.
>
> That disclosure explains what the information is that is being collected and the purposes and how that data will be used, shared and transferred. I do not believe that Joe Consumer reads those and, they're not intended for Joe Consumer.
>
> On the other hand, I can tell you that advocacy groups do read it; IGs do read it. And it provides actually a benefit, because it's about transparency. And there is an accountability feature to it, as well as Privacy Act statements on different kinds of documents.[19] And so I think that is incredibly important. And that we need to honor the representations in those policies that explain how the data will be used.

As Groman indicates, the Privacy Act requires agencies to ***notify*** the public via a Systems of Records Notice (SORN) prior to the creation or expansion of a system of records.[20] In addition,

---

[16] Tolbert, Caroline J., and Karen Mossberger. (2006). "The Effects of E-Government on Trust and Confidence in Government." *Public Administration Review.* pp. 354–369.

[17] Presentation by Jennifer Hunter-Childs to CEP staff, January 2017.

[18] Congress gives responsibility for issuing implementation guidance for the Privacy Act and the Paperwork Reduction Act (PRA) to OMB. The PRA requirement largely mirrors those of the Privacy Act, extending them to information collections about entities beyond persons.

[19] While there is little evidence that such notices are read by the average data subject, third party groups' review of these notices are acknowledged to play an important role in evaluating and critiquing federal agency privacy practices.

[20] Improvements in technology and advancements in data system design and operations render some aspects of the Privacy Act of 1974 outdated. It requires notification for federal "systems of records," defined as "a group of records under the control of any agency from which information <u>is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual</u>" (emphasis added). A key privacy

agencies must ***notify*** individuals of several things at the time that information is collected, including the purpose and planned uses, whether voluntary or mandatory, whether any confidentiality protection is promised, and the estimated burden. Unless eligible for one of 12 exceptions, the Privacy Act also requires that an agency obtain written ***consent*** at the time of information collection. One of the 12 exceptions is for agencies providing their data to the Census Bureau for purposes consistent with the Census Bureau's legal mission. Agencies also are required any collection of information from 10 or more individuals or entities to submit for two public comment periods and for approval by OMB. Some agencies, typically a PSA conducting primary data collection, go even further and engage key stakeholder groups directly prior to data collection. So the public potentially has numerous opportunities to learn about and comment on any agency information collection.

Sometimes agencies indicate that data will only be used for specific, narrow purposes, such as "only for processing a benefit application." Such promises generally precludes later use for any other purpose, without re-contacting and obtaining new written consent. Understandably, agencies view this process as onerous and undesirable in most cases. Other times, agencies anticipate future statistical or other uses.

How does notice and consent work today?

Along with providing notice of an intended set of uses of information, agencies often tell prospective program participants that completing a form or computer-assisted application process is voluntary but failure to do so may result in delays or not being eligible for benefits. In effect, there is notice and consent is for the full set of indicated uses as a bundle. For example, on the application for a social security number (SS-5), the Social Security Administration says

> "The information you furnish on this form is voluntary. However, failure to provide the requested information may prevent us from issuing you a Social Security number and card."[21]

It then goes on to say

> "We rarely use the information you supply for any purpose other than issuing a Social Security number and card. However, …we may also…disclose information…To facilitate statistical research…needed to assure the integrity of the social security system."[22]

This approach meets legal requirements for notice and consent in the Privacy Act. Some would argue that the true choice is limited by the bundling of multiple purposes into a single "yes/no" decision. Of all of the uses that may be bundled, most by definition would be non-statistical, meaning they have the potential to affect the rights and benefits of an individual, and therefore would pose a greater risk of privacy harm than a statistical use. While CEP staff have not examined specific wording in all cases, many of the administrative datasets in Commission's "top 20 data sources" bundle a list of anticipated uses of the data similarly to the SS-5.

---

limitation is that the law only applies to systems where use of direct identifiers is the primary means of accessing the data; it does not cover data systems that contain direct identifiers but where access to records is performed through contextual queries or analyses not relying on the input of direct identifiers. Thus, one can envision a federal data system containing extremely sensitive data on a large number of individuals for which no Privacy Act notices are required, merely because the system does not retrieve records by direct identifiers.

[21] https://www.ssa.gov/forms/ss-5.pdf

[22] Ibid.

Inherent in secondary use of data is a limitation in specifying full details before or during initial primary data collection. As demonstrated, some agencies do not mention "statistical purposes" at all. Others do so in general terms. Research indicates that organizations that provide too much information may unnecessarily alarm respondents who actually read longer notices, though most do not read them.[23] A more successful method of providing greater transparency in the notice and consent process, without overburdening the respondent or unnecessarily alarming them, is to engage in multi-layered messaging, where simple notices are provided at the point of collection, with a link to the agency webpage where additional, more detailed explanation is can be made available.

When an agency uses *its own* administrative data for statistical (or another) purpose, it generally does not need to gain consent. Such uses typically are generally an exception under the Privacy Act because they constitute "an intra-agency disclosure of data for statistical purposes" and therefore are not subject to its requirements for obtaining consent. When those data are provided to a second agency or integrated with data from a different program, however, the Privacy Act requirements become relevant. Memo #10 (*Enhancing Data Confidentiality and Security)* describes a legal and policy framework for protecting data integrated with those from other agencies for SERA purposes under the Confidential Information Protection and Statistical Efficiency Act of 2002.

At the time Congress wrote the Privacy Act and OMB issued related guidance, the vast majority of evidence building occurred via primary data collection so it is notable that even then the Privacy Act and the OMB guidance both demonstrate a comfort level with secondary use of administrative data.[24] In 2014, OMB issued guidance designed to explain how these requirements apply in light of the increasing use of integrated administrative data for evidence building. It says in part:

> In the context of providing administrative data for statistical purposes, both program agencies and statistical components facilitate [the FIPPs] by, among other things:
>
> 1. Respecting the public's time and effort by minimizing the number of times they are asked to provide the same or similar information.
>
> 2. Being transparent by providing adequate notice about the planned purpose and potential statistical uses of administrative data (such as in SORNs and Privacy Act statements).
>
> 3. …Identifiable information should be provided only if the need cannot be met by relying on non-identifiable information, and even then, only relevant subsets should be provided.
>
> 4. Protecting data….And once //the data are provided to the statistical agency or component, …[ensuring they] are not provided from the statistical agency or component back to the program agency for non-statistical purposes.
>
> 5. Implementing a set of policy and procedural safeguards, including …applying sufficient expertise in statistical disclosure avoidance in final products in order to maintain confidentiality, taking into account risks posed by external influences such as the mosaic effect.[25]

---

[24]For example, the 1975 Privacy Act guidance distinguishes statistical records from non-statistical ones and states that only non-statistical records need to be "obtained directly from the individual whenever practicable," which OMB stated in 2014 implies that statistical records may be obtained from already collected administrative data. See OMB M-16-06.
[25] https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf

It also says:

> While most administrative datasets contain identifiable information, when those data are about an individual, the Privacy Act is an essential consideration. The Privacy Act is designed to protect individual privacy by, for example, ensuring appropriate limits on the collection, use, maintenance, and dissemination of information about individuals maintained by an agency. The Privacy Act prohibits agencies from disclosing information in a "system of records," as defined in the statute, without the prior written consent of the individual to whom the information pertains. However, the statute provides a limited number of exceptions to this general rule.…
>
> 1. "[T]o those officers and employees of the agency which maintains the record who have a need for the record in the performance of their duties." In some cases, agencies may be able to engage in an intra-agency disclosure of data for statistical purposes if there is a valid "need to know."
> 2. For a "routine use," as defined in the Privacy Act. This exception may apply when the information will be used for a purpose which is compatible with and related to the purpose for which the information was collected.
> 3. "[T]o the Bureau of the Census for purposes of planning or carrying out a census or survey or related activity pursuant to the provisions of title 13.".…

Historically, the Common Rule has been the standard by which Institutional Review Boards (IRBs) protect individuals who provide personal information for research. Recent revisions to the Common Rule, which go into effect in 2018, provide an exception allowing limited IRB review of confidentiality protections for Federally-funded research projects "designed to study, evaluate, improve, or otherwise examine public benefit or service programs." The revised Common Rule acknowledges the inherent benefit of using administrative data for SERA purposes and provides transparency and protection to the public by ensuring that IRBs review the risk of disclosure from such uses.

Notice and consent are important principles in protecting individuals from misuses of their personally-identifiable information. The ways in which the government offers information about potential uses and asks for consent may not always prove meaningful to the individuals providing the data. An applicant for Social Security, for example, understands the need to provide personal information to receive benefits and is likely much less focused on provisions allowing the data to be used for other purposes. The Federal government's tools for giving notice and consent may have limitations; however, the underlying principles remain important. In addition, Federal law and policy recognize the inherently privacy protective nature of "statistical purposes" and sometimes provide exceptions for those activities, though not uniformly. Nonetheless, this memo focuses on improved transparency for evidence building because it is fair and ethical to tell people how their data are being used in a way they can digest and understand and an evidence-building ecosystem grounded in strong privacy protections and transparency principles is best positioned to garner the public's trust.

Articulating a clear and consistent system of transparency

Commissioners have expressed interest in exploring how to improve public trust through "radical transparency." The concept conveys going significantly beyond the status quo to incorporate some combination of new communication methods, providing more information that is also more meaningfully conveyed. A system of transparency should ensure "clarity and consistency" of practice across the evidence ecosystem, especially within the "Federal Statistical Data Network" as described in Memo #9. Adopting Commission recommendations to permit a much greater share of evidence building to take place under a clear and consistent legal, policy, and implementation framework

through the network, in and of itself, will help the government implement clarity and consistency in its transparency efforts. Yet there remain additional opportunities for transparency along the data life cycle. These could include:

- Before and at the point of collection – Agencies could develop and employ best practices by which to implement traditional notice and consent practices[26]
- After the point of collection – clarifying what datasets the government already has (e.g., by improving adherence to OMB's data inventory policy as recommended in the standards memo #4)
- At point of proposed reuse – announcing proposed or approved types of use of single and linked data resources or individual projects
- At point of research completion – demonstrating benefits to various audiences by
    - publishing all results in an accessible manner
    - creating/identifying and sharing useful results with targeted audiences (highlighting the public benefits)
    - expanding What Works Clearinghouse idea (practitioner oriented),
    - creating state profiles when possible (state government oriented)
    - reporting to Congress (relates to Evidence Ecosystem Memo #7 idea of a more organized approach to congressional oversight)
- Broadly across the lifecycle
    - launching and possibly maintaining a public dialogue that could include elements of a statistical literacy campaign,[27] information about the use of data for public good, and participation by communications experts both within and outside of government
    - maintaining a feedback loop to the public about improved practices informed by the reporting system for violations and breaches proposed in *Improving Privacy by Increasing Confidentiality* (Memo #10), ongoing qualitative and quantitative research on public knowledge/attitudes,[28] and the continued development of privacy enhancing technologies
    - incorporating public input into policy decisions about what constitutes risk of re-identification sufficiently low that it will not pose a "clear and articulable" harm to individuals or businesses and in establishing objective and decreasing levels of acceptable re-identification risk (from Memo #10).

SUMMARY OF COMMISSIONER CALL

On May 8, 2017, CEP staff discussed the first draft of this memo with interested Commissioners available during that time. Commissioners Groves, Hoynes, Meyer, Rice, Shea, and Troske joined the call. Commissioners suggested that the framing needed to discuss privacy risks from integrating data for evidence building in the context of strong data protections, which is much preferable to integrating data for evidence building without improved privacy protections. How transparent one needs to be about statistical uses is related to how transparent one needs to be about non-statistical uses that are by definition inherently more concerning. Commissioners want evidence building to be

[26] A working group of the Federal Committee on Statistical Methodology (FCSM) has explored the available research and informally identified improved practices around a layered communication strategy.
[27] Recommended by Enrico Giovani, former head of UN Statistical Division at the FCSM Statistical Policy Seminar keynote address in 2014 as a major prong of a public trust building campaign by national statistical offices.
[28] The FSS public opinion survey was designed as an ongoing measure of public trust. The head of Statistics New Zealand maintains an ongoing qualitative research program to gauge public attitudes.

a best practice and so called for positive framing of the net effects of stronger confidentiality protections in this context. They also suggested a greater emphasis on the benefits of standardization as an improvement. They said that the discussion of trust reflects the obvious concerns about publicized hacking incidents in government. Finally, they suggested not using the term "radical transparency." This memo and its recommendations were revised after the call.

FINDINGS AND RECOMMENDATIONS

❖**Finding:** The Commission finds that the current Federal model for meeting notice and consent requirements often offers limited choice since many uses are bundled together. If not included in the initial bundle, consent is hard to get after the fact, which can preclude secondary uses of administrative data for evidence building. The Privacy Act generally requires agencies to give notice of intended uses and obtain consent before it collects information about an individual. To avoid foreclosing the opportunity to build evidence about any particular policy or program, program agencies could include future uses for statistical purposes in their notices on a more routine basis.

►**Recommendation #11.1:** The Commission recommends the President promote transparency by directing Departments to indicate routinely that statistical purposes are permissible uses of all information collections, barring a compelling privacy or administrative reason.

►**Recommendation #11.2: [PRIORITY]** The Commission recommends that Congress further standardize the government's privacy protection framework by extending the exception to consent requirements for the Census Bureau's secondary uses of data to all statistical agencies.

These recommendations build on the long-standing recognition, including by Congress, that using data for statistical purposes, which focuses on the attributes of groups and never on the individual, is inherently safer than uses designed to affect the rights or privileges of individuals, especially when coupled with strong confidentiality protections. As such, Congress granted two relevant exceptions in the Privacy Act: the first from consent requirements for any data that agencies provide to the Census Bureau and the second from case-by-case scrutiny linkages of datasets when used for statistical purposes. Yet, these exceptions are applicable unevenly across the evidence building landscape, which has grown since the enactment of the Privacy Act.

Congress has determined that statistical linkages (typically of administrative data to survey data or to other administrative data) do not require the extra scrutiny other linkages within government must undergo and there is compelling need to increase the quality and volume of evidence about Federal policies and programs. Therefore, it is reasonable to require agencies to default to a "yes" for evidence building.

Recommendation #11.1 directs OMB to enable more consistent public notice and consent by a "yes, unless" approach to notifying individuals about the use of information for statistical purposes. OMB could direct agencies to incorporate into all notices that the information might be used for statistical purposes, unless they have a good reason not to do so. OMB could develop standard language and best practices such as "layered notices" that provide more detail for those interested.

These actions could improve comprehensibility of consent messages and the availability of administrative data for evidence building over time. It would require diligence by OMB to ensure that program agencies are considering inclusion of new evidence-generation-friendly language.

Congress determined that the provision of administrative data for an exclusively statistical purpose, with strong confidentiality protections, by a "qualified" entity, is inherently safe enough to except it from the Privacy Act's consent requirements when it created such an exception for disclosure of any Federal data provided to the Census Bureau. Arguably, the Census Bureau was the only "qualified" agency at that time, as CIPSEA was not yet in place. Since CIPSEA was designed specifically to give other statistical agencies a statutory framework similar to the Census Bureau, it is logical to argue that all CIPSEA agencies and recognized units should have a similar exception.

If Privacy Act exceptions were permitted to statistical agencies and recognized units under CIPSEA, this would create an incentive for other SERA units to partner with a CIPSEA unit or establish one within their own agency in order to qualify, a net improvement in data protection.

Recommendation #11.2 goes to the root of access concerns by extending the special exception to consent requirements given to the Census Bureau in the Privacy Act to statistical agencies and units operating under CIPSEA. The recommendation recognizes that current law is inconsistent and therefore adds to the broader Commission effort to create a clear and consistent legal framework.

If Privacy Act exceptions were permitted only for CIPSEA statistical agencies and recognized units, this also would create an incentive for other non-CIPSEA SERA units to partner with a CIPSEA unit or establish one within their own agency in order to qualify, a net improvement in privacy protections.

❖**Finding:** Federal law requires agency transparency by requiring multiple forms of notice, opportunities for public comment, and individual consent before information is collected. The Commission finds that some agencies go beyond minimum practice to engage stakeholders regularly before they collect information from the public. Yet, the current model often does not offer meaningful notice or opportunity for individual consent for secondary uses of information. While statistical uses of data are inherently pose a lower privacy risk, a primary purpose of notice and consent is to build trust in the government's collection and use of the public's information. Trust is a critical enabler of increased evidence building.

❖**Finding:** Personal data is shared widely in today's data rich networked society, but most of the data sharing is hidden and it is not possible for people who may be harmed from the data sharing to know. Similarly, it is difficult for the government to know whether current data protections are effective.

►**Recommendation #11.3:** The Commission recommends that Congress and the President promote public trust in evidence building by communicating clearly its value, use of state-of-the-art data protections, and why the public's information is a necessary ingredient. [*Placeholder until Commissioners choose from the options below.*]

►**Recommendation #11.4:** The Commission recommends that Congress amend relevant laws governing privacy to have providers of sensitive data on individuals publicly post a list of data recipients so that the public can know who receives personal data.

These recommendations communicate that the Commission values greater transparency of statistical uses of administrative and therefore wants the evidence-building ecosystem to meet a high standard.

The Commission could consider four options for addressing consent challenges by increased transparency:

Option 11.3.1: Incorporate notice and consent for statistical uses by default in routine agency processes

As discussed in 1-1 above, this option continues all of the public notice, public comment and individual notice and consent activities required under current law, and incorporates the "yes, unless" default language into program agency notices about secondary statistical uses of data. Notice continues to include information about intended uses and privacy protections.

Option 11.3.2: All of Option 11.3.2 plus a concerted effort to communicate

This option expands both the efforts to be transparent and the scope of what agencies are to be transparent about. Chiefly, OMB, with support from the statistical agencies and the facility, will undertake a significant public outreach effort around the development and implementation of the new data protection standards. OMB conducted such an extensive outreach campaign during its major revision of race and ethnicity standards in the 1990s. Statistical agencies have conducted similar outreach efforts when redesigning major surveys or helping OMB revise major classification standards. The goal would be to increase public comfort with the improved data protections put in place across government.

Option 11.3.3: All of Option 11.3.2 plus single web site notification of linking projects

Increased transparency is achieved by requiring either all agencies, all CIPSEA agencies, and/or the facility to post to a common location public notice of planned and actual record linkages as well as the products created as a result. Statistics Canada and the National Center for Health Statistics routinely publish information about linkage projects. Beyond communicating intended uses and protections promised, this approach creates an opportunity to comprehensively communicate public benefits from statistical uses of administrative data.

OMB has struggled to get all agencies to create and contribute to a single inventory of Federal datasets already in existence. Depending on the scope of agencies and activities include in this effort, it would be difficult to be comprehensive. Sub-setting to CIPSEA agencies is easier, yet these are the agencies with established track records and access to the strongest legal protections, so there is benefit to including a broader set. The effort could be further focused on the data facility's work for external researchers, as a routine byproduct of its application and approval processes.

Option 11.3.4: All of Option 11.3.3 plus a Data Privacy and Ethics Advisory Council containing community and advocacy members and a Chief Privacy Officer (CPO) position

This option adds to the transparency initiatives above by creating a council that provides citizens with a voice in the process, creating a forum for a sustained public dialogue, and by creating a new CPO. Both of these could be situated broadly across the full spectrum of SERA activities, those within the CIPSEA framework, or limited to external research conducted via the facility. The scope would likely drive the advisory council's placement and specific responsibilities.

LIKELY REACTION TO RECOMMENDATIONS:

- ***Congressional:*** Congress is planning to reauthorize the Paperwork Reduction Act (which includes CIPSEA) and would likely welcome Commission input on how to use that opportunity to enhance transparency and evidence building concurrently. Amending the Privacy Act, however, tends to be difficult, as many other proposed changes are often raised.
- ***Executive:*** Some agencies will find compelling the value of best practices to increase public trust through additional transparency measures in return for being afforded greater authority

for evidence building. Others will resist new requirements in an area that already feels onerous.

COST IMPLICATIONS

- Some of the added transparency measures could be expensive, such as a public awareness campaign. Others are "two for one" where transparency is a result of an action for a different or additional primary purpose (e.g., research on privacy enhancing techniques). For those the Commission is interested in recommending, we will include a more precise cost analysis in Memo 16.

ALTERNATIVE OPTIONS

Alternative Option #1: Extend Privacy Act exception to all SERA activities.

This option expands the consent exception beyond the CIPSEA framework to the full range of "safe" SERA projects as defined under the proposed OMB standard for managing risk of re-identification even if the data and agency are not associated with CIPSEA. To the extent that there are incentives for agencies to meet the requirements for recognition under CIPSEA, this option would remove such incentives by permitting them the same privilege without qualifying under CIPSEA.

RECOMMENDATION MEMO #12

Collaborating for Evidence Building

Commission Leads: Troske, Liebman
Staff Leads: Howell, Martinez

ISSUE: Enhanced collaboration inside and outside the government can support the generation and use of evidence. What recommendations can the Commission make to encourage more systematic use of partnerships at the Federal and state and local levels?

Actors outside of government play a critical role in supporting and extending the evidence-building community within government. This memorandum addresses the role of non-governmental researchers, philanthropic organizations, universities, and other partners in the broader evidence ecosystem. It includes recommendations that seek to enhance the ecosystem with continuous research on state-of-the-art privacy protections, data security, data science, and statistical approaches. It also includes recommendations to ensure that Federal agencies are able to engage with the Data Facility to achieve common goals. This memo relates to *Sec. 4(a)(2), 4(b)(2)(F),* and *4(b)(2)(J)* of the CEP statute.

Relates to Facility (#2), Standards (#4), Ecosystem (#7), Confidentiality (#10).

* * * * *

BACKGROUND AND ANALYSIS

During its fact-finding phase, the Commission heard about how partnerships can be invaluable for *promoting the use of evidence in policymaking*. The William T. Grant Foundation explained, "Partnerships between researchers and policymakers can improve the use of research evidence by guiding researchers to ask questions that respond to the needs of policymakers, building stronger practice-focused research networks or community-based participatory approaches, and creating a culture of learning in which administrators, policymakers and other government leaders include research evidence in their deliberations."[1] Researchers provide evidence on the effectiveness of policies and facilitate learning. In this way, partnerships between researchers and policymakers can help ensure that the feedback loop implicit in the evidence ecosystem framework (described in Memo #7) functions effectively.

Partnerships can also be an effective way to *expand government capacity* for the production of evidence. The Robert Wood Johnson Foundation encouraged the Commission to consider how "partnerships between researchers, service providers, and large institutions—both governmental and corporate—can improve the capacity to develop evidence."[2] Particularly in an era of shrinking budgets and ever-increasing demand for data, Federal, state, and local agencies can benefit from funding, expertise, and other resources from partner organizations such as universities, professional associations, and the private sector. These resources can enable expanded projects in terms of scope and scale. They can also be an important way to *leverage outside expertise* and *develop skills in the Federal workforce*. Margaret Levenstein (ICPSR) told the Commission, "Partnerships between state and local governments, federal governments, and academic institutions can provide the relevant training while

[1] William T. Grant Foundation RFC Comments
[2] Robert Wood Johnson Foundation RFC Comments

developing data resources of value to all parties for evidence-based program evaluation."[3] At the Federal, state, and local level, partnerships are used to *improve program administration and outcomes.*

Federal Government Tools for Engaging with External Actors

Federal agencies have a variety of tools to engage with non-governmental researchers, philanthropic organizations, universities, and commercial entities to achieve their missions. These include contracts, grants, agreements, and other arrangements such as temporary assignments of personnel. Developing and sustaining formal partnerships to leverage external expertise and to support evidence building requires administrative processes that ensure quality and accountability while building in flexibility.

Contracts require the external vendor to provide agreed upon deliverables (goods or services) paid for by government funds and produced according to the Federal Government's specifications. For the purposes of this memo, we will consider contracts to be out of scope because they represent a buyer-seller relationship between the Government and a third party rather than a partnership. Recommendations related to procurement are included in the Administration Memo (#15).

Grants and other agreements, on the other hand, give the external partner some level of control over the nature of the work. They are used "to carry out a public purpose of support or stimulation."[4] **Grants** enable a non-governmental entity to receive government funds to perform a specific service but, unlike contracts, they do not require a return of services to the government. This means that a grantee can provide services to others.

An agreement, such as a **cooperative agreement**, is an appropriate vehicle when the Federal agency anticipates a need for substantial interaction between the agency and the recipient to carry out the activity.[5] Under a cooperative agreement, both partners expend funds and share information and resources.

Joint project authority is similar to a cooperative agreement except that this authority does not allow for the transfer of resources. Rather, both sides must contribute equally to the project.

Cooperative research and development agreements (CRADAs) are a specific kind of cooperative agreement that allow the Federal Government to share the costs of research with its partners.[6] Both partners provide funds, personnel, facilities, equipment and expertise, but the Federal partner does not directly provide funds to the external partner.[7]

Federal agencies also have mechanisms for bringing on **temporary personnel** from colleges and universities and other institutions on a temporary basis, such as the Intergovernmental Personnel Act (IPA) Mobility Program. Assignments under the IPA program are intended to "facilitate cooperation between the Federal Government and the non-Federal entity" and to serve a "sound public purpose."[8] Some agencies have **fellowship programs** that operate similarly.

---

[3] Written testimony from Margaret Levenstein (ICPSR) from the January 5 CEP hearing
[4] 31 U.S.C. 6304-05
[5] Office of Management and Budget Circular A-110, "Uniform Administrative Requirements for Grants and Agreements With Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations," available at https://obamawhitehouse.archives.gov/omb/circulars_a110/#11
[6] 15 U.S.C. 3710a
[7] http://www.usgs.gov/tech-transfer/handbook/TT_Handbook-CRADA.doc
[8] https://www.opm.gov/policy-data-oversight/hiring-information/intergovernment-personnel-act/#url=Provisions

Additional Considerations for Interagency Collaboration

The ways that agencies can engage with one another are very similar to how they engage with non-governmental actors. However, the vehicle used is an **Interagency Agreement (IAA)**. Further, an agency's ability to receive funds from another agency in exchange for goods and services is generally governed by the Economy Act.[9] The Economy Act allows Departments and agencies to carry out their missions in a more efficient and effective manner by leveraging their comparative skills rather than operating individually in a stove-piped manner. It can also be used for multi-agency collaborations. For example, Regulations.gov is funded in this manner. However, there are several conditions that must be met. The amounts for the purchase must be available, there must be a bona fide need, the ordered goods or services cannot be provided by contract from a commercial enterprise as efficiently as could be by the Government, and the servicing agency must be able to provide or get by contract the ordered goods or services. Moreover, an Economy Act agreement must achieve full cost recovery. The Economy Act applies when more specific statutory authority does not exist.

Existing Impediments to Partnerships

The mechanisms described above provide the means for Federal agencies to establish formal partnerships with external actors and with one another. However, their use across government is inconsistent. For example, not all agencies have grant-making and cooperative agreement authority, which must be specifically authorized in an agency's statute. Other agencies have authorities that they do not take full advantage of, either because they lack the resources or expertise necessary to exercise the authority appropriately or because there is a lack of support within the organization for undertaking certain types of arrangements. Given the importance of these vehicles for enabling the government to expand its capacity, the Commission may wish to consider recommendations providing authority for the proposed Data Facility and other statistical, evaluation, policy research, and analysis (SERA) offices to use them.

Existing Partnerships

During its fact-finding phase, the Commission heard examples of existing partnership arrangements that have been effective in promoting the use of evidence in policymaking and in building evidence infrastructure. In the examples described below, the tool used to establish the partnership is shown in **bold**.

Promoting the use of evidence in policymaking

Effective partnerships signal the government's research priorities, leverage research expertise to hone research questions and methods, and ensure that the results will be useful for policymaking. As Adam Gamoran told the Commission, partnerships can "increase the likelihood that research evidence permeates the policymaking process."[10] For example, the Office of Policy Development and Research (PD&R) in the Department of Housing and Urban Development (HUD) has **noncompetitive cooperative agreement authority** to enter into research partnerships with universities, philanthropic organizations, and other Federal or state agencies. In the past, HUD has also used cooperative agreement authority to conduct significant research studies, including the original Jobs-Plus demonstration, which was a partnership between the Rockefeller Foundation, MDRC, and a number of other public agencies and private foundations. PD&R was granted *noncompetitive* cooperative agreement authority in 2014, which allows the agency to implement the

[9] 31 U.S.C. § 1535

[10] Written testimony from Adam Gamoran (William T. Grant Foundation) from the November CEP meeting

Research Partnership program,[11] through which PD&R can support high-quality and policy-relevant *unsolicited* research proposals. Since 2014, PD&R has received approximately $1 million in the Department's annual appropriation for this program. This authority allows PD&R to participate in research projects that align with the strategic goals of the Department and its "Research Roadmap," or learning agenda. Because these awards are cooperative agreements, they give PD&R the ability to "shape the research projects to make sure their design allows for the answering of policy and programmatic questions in ways that could inform new policy development efforts."[12] Since 2014, the Research Partnership program has funded 27 projects.

Noncompetitive cooperative agreement authority could be a valuable tool for agencies engaged in evidence building. However, with the considerable flexibility of noncompetitive cooperative agreement authority comes increased responsibility on the part of the agency to ensure that the funds are being used effectively. Projects that have been most successful (in terms of staying on budget and on time, adopting a sound research design, executing that design with fidelity, and producing findings that could be translated into policy recommendations) are those that have been managed by individuals who stay on top of their portfolio and have expertise commensurate with the difficulty of the project.[13] Transparency is also essential for maintaining credibility with oversight bodies (e.g., Congress). Therefore, it is clear that not every agency should be granted this authority; rather it should be reserved for agencies that have demonstrated capacity and expertise.

Federal research-practice partnerships—most common in education—have helped state and local agencies increase the use of evidence in policymaking. These partnerships can "build and harness research evidence, so that what is learned in one place need not be reinvented in another and the lessons accumulate."[14] In follow up to his hearing testimony, Matthew Stagner (Mathematica Policy Research) named the Researcher-Practitioner Partnerships in Education Research, a **grant program** funded by the Department of Education's Institute for Education Sciences (IES) as a model.[15] The IES program pairs research institutions and state and local education agencies to study high-priority education issues that have implications for student outcomes. According to IES, "Through this joint research, the education agency's capacity for taking part in research and using research results is expected to increase."[16] A unique aspect of the IES Researcher-Practitioner Partnerships is that they fund the development of a research plan. Between 2013 and 2016, the grant program funded 33 partnerships.

These collaborations between researchers, policymakers and practitioners can help make research more relevant and actionable. However, they can be "difficult to maintain because of differences in organizational cultures … ; a lack of trust and long-term commitment to the partnership; a lack of clearly defined roles; insufficient and unequal distribution of resources; and inadequate exercise of scientific rigor."[17] The Commission could consider ways to encourage these kinds of arrangements with the appropriate structure and oversight.

---

[11] https://www.huduser.gov/portal/oup/research_partnerships.html
[12] https://www.huduser.gov/portal/pdredge/pdr_edge_frm_asst_sec_050812.html
[13] From discussions with PD&R staff
[14] Tseng, V. "Evidence at the Crossroads Pt. 11: The Next Generation of Evidence-based Policy." Web blog post. *William T. Grant Foundation.* March 23, 2016, available at http://wtgrantfoundation.org/evidence-crossroads-pt-11-next-generation-evidence-based-policy
[15] Matthew Stagner Response to QFRs.
[16] https://ies.ed.gov/funding/ncer_rfas/partnerships.asp
[17] L. Palinkas, C. Short, and M. Wong. "Research-Practice Partnerships for Implementation of Evidence-Based Practices in Child Welfare and Child Mental Health." (August 2015) White Paper prepared or the William T. Grant Foundation, available at http://wtgrantfoundation.org/library/uploads/2015/10/Research-Practice-Policy_Partnerships.pdf

There are also examples of partnerships to promote the use of evidence in policymaking from outside the Federal Government. The Government Performance Lab (GPL) at the Harvard Kennedy School of Government uses a research model that involves providing technical assistance to state and local governments. Key to this approach is a **fellowship program** that places a full-time fellow in the government agency. Hands-on involvement allows GPL to "gain insights into the barriers that governments face and the solutions that can overcome these barriers."[18] GPL has three kinds of projects: Pay for Success social impact bond projects supporting governments exploring the use of Pay for Success contracting, results-driven contracting projects, and performance improvement projects to help governments improve the results they achieve with their core spending.[19] To date, GPL has worked with 38 jurisdictions spanning 22 states.

While technical assistance efforts like GPL's are typically focused on improving program development or implementation, the Commission could consider encouraging similar technical assistance projects for evidence building, particularly at the state and local level.

Expanding government capacity

Philanthropic organizations can accelerate evidence generation by providing funds that can be used to expand the size and scope of a project. For example, a foundation may support a larger evaluation than a Federal agency might otherwise be able to fund. The Laura and John Arnold Foundation (Arnold) and other philanthropic organizations have funded partnerships between researchers and state and local governments that aim to develop new policies and measure their success. Examples include GPL at Harvard, the Rhode Island Innovative Policy Lab (RIIPL) at Brown University, and the Abdul Latif Jameel Poverty Action Lab (J-PAL) at the Massachusetts Institute of Technology.

Partnerships with philanthropic organizations can be instrumental for establishing a new demonstration project or piloting an evaluation, but philanthropic organizations do not fund ongoing government activities. Partnering with philanthropic organizations also requires specific expertise to operate within the law and maintain objectivity. Some departments have designated units to coordinate the department's engagement with philanthropic organizations. For example, at HUD, the Office of International and Philanthropic Innovation develops collaborations with philanthropic organizations that align ideas, investments, and resources.[20] Not all Federal agencies have staff with such expertise.

The Federal Statistical Research Data Centers (FSRDCs) are examples of partnerships between Federal statistical agencies and academic institutions to increase capacity for research and evaluation. The partnerships are established through **joint project agreements** with the Census Bureau. Title 15, U.S. Code, Section 1525 authorizes the Department of Commerce to undertake joint projects with "nonprofit organizations, research organizations, or public organizations or agencies … on matters of mutual interest, the cost of which shall be apportioned equitably… ."[21] The agreement must include information about the mutual interest of the Census Bureau and the partner, the equitable apportionment of costs, and the nonprofit status of the partner.[22] The host institution provides a secure physical space for the research data center (RDC) and employs the Executive Director (typically a faculty member). The host institution also pays the salary and overhead costs

[18] https://govlab.hks.harvard.edu/gov-lab
[19] Ibid.
[20] http://c.ymcdn.com/sites/www.gcyf.org/resource/resmgr/event_materials/huds_office_for_internationa.pdf
[21] 15 U.S.C. 1525
[22] Establishing New Research Data Centers (RDCs), 63 Federal Register 14 (January 22, 1998).

for the RDC Administrator—a Census Bureau employee working on-site at the RDC—as well as a RDC network service fee.[23]

The RDC arrangement has been an effective way to expand capacity for researcher microdata access. Since 1998, the FSRDC network has grown from two pilot projects to 25 operating FSRDCs. The FSRDCs and their center directors are essential actors in the evidence ecosystem and will play an important role under the Commission's vision for a data facility.

Developing skills in the Federal workforce

One lasting partnership is between the Principal Statistical Agencies (PSAs) and the Committee on National Statistics (CNSTAT) at the National Academies of Sciences, Engineering, and Medicine. Established in 1972, CNSTAT has served as an advisor to the PSAs and has evaluated large statistical programs. The committee's core funding comes from a consortium of statistical agencies through a National Science Foundation (NSF) **grant**.

Statistical agencies have also collaborated with academic institutions and professional associations to develop skills in the Federal workforce. Since 1993, statistical agencies have partnered with the Joint Program in Survey Methodology (JPSM), a consortium of the University of Maryland, the University of Michigan and Westat. Sponsored in part by the Interagency Council on Statistical Policy and supported by the PSAs, the program's mission is to "educate the next generation of survey researchers, survey statisticians, and survey methodologists."[24] Students often intern at Federal statistical agencies, program graduates often go on to work at statistical agencies, and current staff of statistical agencies take courses to develop their skills in the area of survey methodology and research.

Although the JPSM program has been successful, the arrangement used to extend the partnership has faced challenges each time it is up for renewal. Originally funded through an NSF grant, the vehicle for partnering with JPSM was later changed to a **contract**. The Census Bureau has administered two multi-year contracts on behalf of the PSAs with the PSAs reimbursing Census through **IAAs**. However, this arrangement will not continue beyond 2017. The Census Bureau found the requirements imposed by each PSA's legal departments for writing and executing each year's IAAs made the process too unwieldy to manage. In addition, the Census Bureau conducted extensive market research and determined that there was no need to have a contractual agreement with the JPSM to fulfill the training needs of the PSAs. However, the experience showed that it would be efficient and beneficial to have a mechanism through which the PSAs could jointly enter into partnerships with outside entities to expand training and research opportunities and build capacity system-wide, and to collectively fund such efforts. In order to work, however, there would need to be standardized authority that would not be subject to revision by each agency's lawyers.

Leveraging outside expertise

The National Oceanic and Atmospheric Administration (NOAA) established the Big Data Partnership—a series of **CRADAs** with private sector firms—to help NOAA publish the large amounts of data it collects but cannot afford to make publically available. Major cloud services providers such as Amazon Web Services and IBM are working with NOAA to build out its data infrastructure at no cost to the government. The partners benefit from their investment by having

[23] Guidelines for the Development and Operation of a Federal Statistical Research Data Center (RDC) or RDC Branch (2016).
[24] https://jpsm.umd.edu/featuredcontent/256

access to data that can be used to build marketable products, such as improved weather prediction services.[25]

Collaborations with the commercial sector have great potential to spur technological advancements but also obvious disadvantages because the government has to give up some ownership over the outcomes in order to incentivize the partner to participate.

Improving program administration and outcomes

The Commission also heard about how Federal, state, and local government program agencies and other stakeholders are coordinating on service delivery to lead to better outcomes. Some of these efforts have adopted the label "collective impact," a term first coined by John Kania and Mark Kramer.[26] According to Kania and Kramer, **collective impact efforts** are unique because they involve "a centralized infrastructure … and a structured process that leads to a common agenda, shared measurement [of outcomes] … and mutually reinforcing activities among all participants."[27] Kania and Kramer cite the example of Strive, a non-profit that has brought together leaders in the Cincinnati area to improve education outcomes.

While the primary purpose of these kinds of efforts is programmatic, they can also be an effective way to build infrastructure for bringing administrative data together for research purposes. To the extent that program agencies are already trying to bring data together to improve service delivery, the Commission could encourage them to consider SERA uses of the data in building their data systems. In this way, these efforts can lead to the dual benefit of improving the effectiveness and efficiency of service delivery while simultaneously making more data available for research and evaluation. Palinkas, Short, and Wong (2015) describe a model for this type of partnership in which "the research agenda is informed by the need of the community partner (a state agency) to deliver high quality services to its clients, and by the researchers' desire to use the community as a 'natural laboratory' for developing, testing, and implementing evidence-based practices."[28]

SUMMARY OF COMMISSIONER CALL

Commissioners expressed concerns with the number and specificity of the draft recommendations and suggested that the core of the recommendation be distilled into one main recommendation with several action items underneath. Commissioners asked the staff to broaden the scope of the memo to acknowledge all the ways that government can engage with non-governmental actors, including contracts and to explore other vehicles that were not included in the memo (such as joint project or joint venture authority). These changes were incorporated into the section on "tools." Commissioners asked the staff to research tools for interagency collaboration. This will help to understand whether certain flexibilities should be recommended for the broader evidence-building community and which could be limited to the data facility. They also asked the staff to consider whether certain authorities should be granted to the proposed data facility as part of a "phase one" of implementation, where other agencies could be granted similar authorities in a later phase of implementation. These comments are reflected in this revised memo. In addition, the draft recommendations also take into account Commissioner comments on the full set of

---

[25] Center for Data Innovation RFC Comments

[26] J.Kania and M. Kramer. (2011). "Collective Impact." *Stanford Social Innovation Review,* available at https://ssir.org/articles/entry/collective_impact

[27] Ibid.

[28] L. Palinkas, C. Short, and M. Wong. "Research-Practice Partnerships for Implementation of Evidence-Based Practices in Child Welfare and Child Mental Health." (Augut 2015) White Paper prepared or the William T. Grant Foundation, available at http://wtgrantfoundation.org/library/uploads/2015/10/Research-Practice-Policy_Partnerships.pdf

recommendations that were distributed on June 8, including an interest in clarifying what existing authority Federal agencies have to transfer funds between one another.

[DRAFT] RECOMMENDATIONS[29]

Ensuring Federal Agencies Have the Tools to Establish Partnerships

❖ **Finding:** Partnerships are an essential tool for increasing evidence building, especially for small agencies with limited capacity. Tools like grant-making or cooperative agreement authority have been used to good effect at some agencies. Other agencies do not have such authority.

❖ **Finding:** As evidence-building agencies adapt to smaller budgets, they must still continue to develop their workforce and increasingly need to staff with expertise in emerging technologies and statistical methods. Some agencies have found ways to engage with academic institutions and other partners to develop their workforce including through assignments of temporary personnel, such as the Intergovernmental Personnel Act Mobility Program. However, agencies do not take full advantage of these programs.[30]

❖ **Finding:** State and local governments often lack the resources necessary to build and maintain their data systems. There are many organizations that are working at the state and local level to improve data quality, ensure data coverage, and set up data systems for both programmatic and evidence-building purposes.

❖ **Finding:** There are clear synergies between the Commission's charge to increase evidence production and partnerships that are being designed to increase the availability of data for programmatic purposes. While these partnerships are primarily focused on improving outcomes, they can also be an effective way to build infrastructure for bringing administrative data together for SERA purposes.

► **Recommendation 12.1:** The Commission recommends that Congress and the President ensure that Federal agencies in the evidence-building community have the necessary tools—such as grant-making authority and cooperative agreement authority—to encourage the use of partnerships to promote evidence building with foundations, researchers, and other potential partners to promote evidence building.

- *Action Item 12.1.1.* Congress and the President should provide Departments in active pursuit of a learning agenda the authority to enter into cooperative agreements to pursue articulated priorities.
- *Action Item 12.1.4.* Evidence-building agencies should explore ways to expand collaborations with academic institutions, professional associations, and researchers to develop the government's workforce's skillsets in privacy protections, data security, data science, and statistical approaches.
- *Action Item 12.1.5.* The President should encourage the increased use of programs that allow agencies to bring on temporary personnel from state and local governments and from academia to help develop their workforce and advance their learning agendas.

[29] For consistency with previous memos, we have labeled items in this memo as recommendations. However, Commissioners have expressed an interest in limiting the number of "top line" recommendations, so after Commissioners identify which of these they wish to pursue, we expect to identify most as "action items" or some other subordinate label within the broader set for discussion at the next meeting.

[30] https://archive.opm.gov/programs/ipa/index.asp

- *Action Item 12.1.6*. Federal agencies should explore ways to partner with philanthropic organizations and others to help improve data infrastructure at the state and local level in order to support evidence-building efforts, in particular efforts to make administrative data associated with Federally funded programs that are collected and maintained at the state level available for statistical purposes.

The President and Congress can take steps to encourage more systematic use of partnership arrangements to promote the production and use of evidence by providing the actors in the evidence-building community with flexibilities such as grant-making and cooperative agreement authority when appropriate. With the additional flexibility, it is reasonable to expect commensurate accountability. Action item #12.1.1 addresses this specifically by tying cooperative agreement authority to the development of a learning agenda. This is similar to Recommendation #8.5 (in the Evaluation memo), which would give departments that are in active pursuit of a learning agenda access to multi-year funding.

Commissioners on the May 31 call asked the staff to evaluate which flexibilities were needed by all members of the evidence-building community and which might be more narrowly provided to the data facility. The staff determined that a broad set of tools is desirable for all members of the evidence-building community to facilitate its work. An evaluation office, for example, could benefit from grant-making or cooperative agreement authority. For example, PD&R has both authorities and uses them for different evaluation projects. The emphasis of this recommendation is on providing agencies a variety of tools for their tool belt, while not recommending which specific authorities should be afforded to any specific agency. The one exception, is the data facility, which is addressed with the next set of recommendations.

Action items #12.1.4 and #12.1.5 suggest ways that SERA agencies can increase collaboration with academic institutions and other partners to build capacity. Action item #12.1.6 acknowledges that partnerships with philanthropic organizations, academic institutions, and others can provide additional resources at the state and local level to make administrative data available for research. This will be important given the Commission's recommendations related to mandating state reporting of certain datasets (see Memo #5).

Partnerships to Accelerate the Adoption of New Technologies [these recommendations would be included along with other recommendations related to the data facility]

❖ **Finding:** The Federal evidence-building community will benefit from the intellectual capital of academia and the private sector in order to develop and maintain state-of-the-art data access, transparency, and privacy protection technologies and methods. The decentralized system will benefit from a designated lead to coordinate such interaction.

► **Recommendation 12.1.2:** In establishing the [Data Facility], Congress should give the [facility] grant-making authority, cooperative agreement authority, and other flexibilities to engage with partner organizations to accelerate the development of new methods and technologies. *This could include the authority to sponsor a Federally Funded Research and Development Center (FFRDC) to research and develop increasingly privacy protective approaches. [italicized text derived from Recommendation 2.1.7.]*

► **Recommendation 12.1.3:** Other agencies in the Federal government should also receive sufficient authority in appropriations law to transfer funds to support activities conducted by the data facility that allow the evidence-building community to centralize certain privacy enhancing functions and research capabilities at the facility.

Under the Commission's vision, the Data Facility will lead on developing and adopting new technologies for privacy protection, data security, record linkage, and perhaps some other statistical methods. For the Data Facility to be successful, it will need to build partnerships with organizations outside of government. Recommendation #12.1.2 ensures that the facility will have the flexibility necessary to do so. This has been combined with a related recommendation in the Data Facility memo (Memo #2) that included the recommendation that Congress provide the facility with the authority to sponsor a Federally Funded Research and Development Center (FFRDC). (FFRDCs are established through a contract.) One could envision the Data Facility using a variety of tools to facilitate this work on R&D. In the vision for the evidence ecosystem, the Data Facility would perform essential centralized services to ensure clarity and consistency in procedures to ensure privacy and expand access to data. Recommendation #12.1.3 suggests that agencies have explicit authority to be able to transfer funds to the facility to fund these centralized functions.

RECOMMENDATION MEMO #13
Incorporating Evaluation and Policy Research in Program Design

Lead Commissioners: Hahn and Hoynes
Lead Staff: Hart and Fletcher

The Commission's charge specifically includes multiple references to supporting evaluation and policy research within program design. This memo documents the relationship between program design and evaluation and policy research, noting the importance of planning for future evaluative activities into the program design at the outset, where possible, but acknowledging the importance of also considering how to incorporate evaluative activities into ongoing programs. Program designs must facilitate and encourage the role of evidence for use in program operations and for accountability. The memo includes recommendations that address program activities to directly support continuous learning through a range of methods, including impact evaluation and cost-benefit analyses. The memo specifically addresses key elements to be considered during program design in order to meaningfully incorporate an approach for evaluation that will generate useful information for program managers and decision makers.

* * * * *

BACKGROUND AND ANALYSIS

During the course of the Commission's work, much emphasis has been placed on the need to better facilitate access to data and mechanisms for providing the *institutional* supports to increase the volume and quality of statistical, evaluation, policy research, and policy analysis (SERA) activities. This memo takes a step back and considers *how* a program might be best designed to support evaluation as a critical feature of program operations. Ensuring programs and policies are actually designed to readily provide for and encourage evaluation as an expectation—rather than merely a hope—can help achieve the statutory direction to the Commission to consider "how best to incorporate outcomes measurement, institutionalize randomized controlled trials, and rigorous impact analysis into program design" (Sec 4(a)(3)). The goal of this memo is to offer recommendations that address how evaluation can be built into program design as a necessary feature, enabling the evidence-building community to serve as partners in the design of policies, regulations, and public programs.

Encouraging Evaluation

Support for more and better evidence to support decision-making that improves government services and programs must begin with leadership. Congress, the President, and Departmental leadership can create demand for valid and reliable information and expectations for evaluation that will lead to the continuous production of evidence about programs and policies. In turn, program designs must facilitate and encourage the role of evidence for use in program operations. Program offices should routinely partner with central statistical, evaluation, and policy research units to align data collection activities with the needs for evidence-building. The Commission identified numerous examples where programs today *do* recognize the importance of developing rigorous evidence for program administration, though the Commission believes these should increasingly become the norm rather than characterized as exemplars.

A Portfolio of Evidence

A large body of public administration literature exists highlighting the varying information needs of different types of actors, including for programs at different stages of development. In short, different levels of evaluation, monitoring, and assessment information are needed at different stages of program implementation. Program administrators are faced with meeting these varying informational demands simultaneously, typically necessitating multiple types of evidence to address the range of needs and interests in both qualitative and quantitative forms. For example, consider a program in communities designed to support tutoring services to children with low test scores. A program manager may need to gather information about the communities in which the program is operating, the variation in service delivery approaches across grantees, the number and types of students being served, and any impact the program has on test scores. Gathering this information would require a variety of different approaches for collecting data.

During the Commission's fact-finding phase, we heard from multiple organizations specifically concerned about how to address the issues of program design in the Commission's recommendations. For example, Child Trends suggested:

> Research and evaluation may be used at multiple points in time to inform program design and implementation. Descriptive evaluations of participant outcomes and implementation can support high-quality programs through documenting activities and outcomes. Once a program is mature, impact evaluations may occur to determine if the program as a whole or individual pieces of the program are effective. Programs are all at different stages of implementation and need different types of research and evaluation support to guide evidence in their decision-making.[1]

This observation highlights the value of research and evaluation that takes place at various points throughout the lifecycle of a program, as well as the broad variety of research methods that may be employed to address different types of research questions. The OMB white paper *Using Administrative Data to Build Evidence,*[2] groups different types of evidence into two broad categories: 1) *foundational evidence*, which includes evidence such as aggregate indicators, descriptive statistics, trends and correlations, and 2) *policy-specific evidence*, which includes performance measures, process evaluation, impact evaluations (which include randomized controlled trials), and benefit-cost analysis. Taken together, all of these types of evidence make up a *portfolio of evidence*. A program that has a strong portfolio of evidence has a much more robust set of information by which to make continual improvements to the program when compared to a program that has limited information about the program operations and performance. As a Department seeks to develop a learning agenda to guide their future investments in statistical, evaluation, policy research and policy analysis activities (see Recommendation #8.2), a logical first step might be an assessment of the current portfolio of evidence that exists to support the implementation of the range of Departmental programs. A set of activities designed to fill the gaps identified during this assessment would be a strong start to the formation of a learning agenda.

---

[1] https://www.regulations.gov/contentStreamer?documentId=USBC-2016-0003-0227&attachmentNumber=1&contentType=pdf

[2] OMB. (2016). Using Administrative and Survey Data to Build Evidence. White Paper for the Commission on Evidence-Based Policymaking. Washington, D.C.: Office of Management and Budget. https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/using_administrative_and_survey_data_to_build_evidence_0.pdf

The Commission's challenge in addressing program design is not necessarily one of articulating the value of one methodological approach for generating evidence over another, but rather to identify how best to incorporate these different types of approaches within programs and policies to support continual learning. Incorporating evidence-building activities into an *existing* program, which may have been operating for many years will require a different approach than incorporating evidence-building activities into a *new* program, in which these activities can be a feature of the design of the program from the outset. The following two sections consider this challenge by acknowledging the differences and opportunities when designing new programs and policies versus modifications to longstanding or existing programs.

Designing New Programs and Policies

One challenge in developing recommendations related to the Commission's charge on incorporating research and evaluation in program design is that design often occurs *before* the evidence-building community is involved. Engaging the evidence-building community's experts as partners during the program design process) will help ensure that the program can be evaluated in a meaningful way, and ensures that the research questions will directly respond to the needs of program managers and policymakers. This early engagement may not happen due to limited evaluation capacity of a Department, a program office's concerns about how evaluation data will be used, or a lack of available or poorly structured data sources. For Departments that lack an institutionalized evaluation office, such a partnership may not currently be feasible. Recommendations from Memos #7 and #8 that strengthen the statistical, research and evaluation functions across government, and ensure their operation across all Departments will help address some of the existing gaps at an institutional level. Other recommendations that address facets of improving administrative data quality and metadata will help address potential data gaps for evaluating programs. But even with these improvements to the system as a whole, the Commission can also consider how to address core elements that are sometimes lacking in program design. Commenters to the Commission suggested these key elements often include:

1. Clearly stated goals, objectives, logic, outputs, desired outcomes (and system boundaries)
2. Recognition of role for different types of analyses based on program stage and development (e.g., prospective, pilots, implementation, retrospective, systematic reviews/meta-analysis) and incentives for evidence production
3. Sufficient legislative authority or flexibility to enable evaluation, including mechanisms to encourage innovation
4. Incorporation of data collection needs and requirements; and focus on data quality to ensure meaningful and useful information is gathered in order to assess program costs and outcomes. [3]

Because there is a vast array of program purposes, intended beneficiaries, operations, and desired outcomes, each program is therefore unique and faces distinct challenges in incorporating evaluative activities into the program design. Thus, there is no one-size-fits-all recommendation to be made in this arena to address all of the key elements described above.

The Federal government does have some existing guidance for agencies on program design, but this guidance is limited. For example, OMB's Cost Benefit Analysis and a series of Executive Orders on

[3] See LJAF comments, JPAL testimony, American Eval Association testimony, etc.

regulations offer existing guidance, including a requirement for *ex ante* analyses for regulatory activities of a certain scale or anticipated cost.[4] Some Federal programs also have requirements for *ex post* evaluation in their statutory authorizations, with applicability to grants or regulations. For example, grantees of the Corporation for National Community Service's Social Innovation Fund are required to participate in evaluations as a condition of the grants.[5] However, mandates to provide data or evaluations in program design are not necessarily sufficient if demand for evidence generation does not exist. For example, the Federal hazardous waste laws implemented by the Environmental Protection Agency (EPA) are required by law to be reviewed and evaluated every three years, though with few exceptions the hundreds of hazardous waste regulations have not been evaluated because there is little desire from industry or government to modify the rules.[6]

Given the difficulties in highlighting program design elements for evidence building, the Commission can consider recommendations that encourage these attributes be addressed at a program's outset. Based on an established pool of literature and comments received by the Commission, parties responsible for designing programs – whether Congress designs a program in legislation or Departments design a program through policies and regulations – should generally (1) clearly articulate the program or policy goals and desired outputs and outcomes, and engage routinely with stakeholders in the evidence ecosystem to ensure goals are commonly understood, (2) recognize that generation and dissemination of evidence about the program is a responsible component of program administration, (3) provide for flexibilities in implementation that allow for experimentation and analysis (e.g., phased or staggered implementation), (4) plan for and ensure that sufficient data of high quality are collected about program administration to support monitoring and evaluation of program activities, costs, outputs, and outcomes, and (5) incentivize continuous production of evidence to support decision-making.[7]

The regulatory context provides a unique opportunity to consider how evaluation of new programs and policies can be addressed at the outset. When designing significant new regulations, Federal Departments are required to complete a suite of *ex ante* analyses, which seek to calculate the anticipated impact or burden on the entity that will be impacted by the new regulation.[8] A valuable evaluative activity would be to assess whether the prospective assumptions about the regulation were appropriate. The information yielded from this type of analysis could be used to improve future efforts to complete similar *ex ante* analyses. Today, notwithstanding calls for decades to engage in these types of analyses, this is rarely done in any Federal Department.[9]

Modifying or Reauthorizing Existing Programs

Once programs are operating, modifying program design to incorporate evidence building is no less important, but it can be much more difficult to make changes midstream. The core elements listed in the previous section for new programs are also needed for existing ones, but may require inferring program goals and cobbling together information from other sources. While the existing data infrastructure may be useful in addressing key questions and may mitigate some new data collection needs, it may not be sufficient to address the questions that need to be asked to for continuous improvement or to assess program outcomes.

---

[4] e.g., EO 13563 encourages agencies to plan for retrospective analyses and make supporting data publicly available;
[5] https://www.nationalservice.gov/programs/social-innovation-fund/our-model
[6] Example from Hart 2016
[7] See Weiss (1998), Russ-Eft and Preskill (2009), Newcomer, Hatry, and Wholey (2016), etc.
[8] See EO 12866, EO 13563, and OMB Circular A-4
[9] E.g., Hahn, Dudley (2017)

Over the past 50 years, there are many successful examples of reauthorizations requiring evaluation frameworks for existing programs, including for many human services, education, and disability programs. However, failing to plan for evidence building at the outset limits the ability to develop counterfactuals to compare a program or policy to outcomes in its absence. Much of the interest in establishing the Commission was specifically aimed at recognizing these limitations and seeking guidance on addressing evidence building for the programs that *do* already exist.

Flexibilities, such as tiered evidence grantmaking and demonstration and innovation authorities, designed to enable continuous learning and innovation can be incorporated in long-running and new programs alike, while simultaneously incentivizing more evidence building. During the Commission's fact-finding phase, numerous commenters and staff research identified several models for the Commission to consider as promising approaches.

First, the "tiered evidence" model for grantmaking encourages programs to allocate funding toward approaches or policies that have developed a more robust body of evidence suggesting effectiveness, relative to other interventions.[10] The approach is intended to encourage grant recipients to increasingly pursue evidence-based approaches while also supporting efforts to test promising new approaches; both in an effort to continually build new evidence.[11] In 2016, the tiered evidence approach was used in four different Departments to stage increasing funding of evidence-based practices. Grant applicants for these Federal funds must provide evidence to support their requests, and a share of the total funding is allocated by the Federal agencies to support either the evidence-based practices (typically a bigger share of grant funds) or to encourage new evidence of other practices (typically a smaller share of grant funds). The tiered evidence model is not without critique, however, as some grantees and stakeholders observe the approach relies too narrowly on experimental research designs with inconsistent comparisons to demonstrate efficacy and sometimes relies on research with insufficient detail to justify external validity.[12] The other natural critique of the approach is that tiered evidence shifts funding perhaps unintentionally to support activities that have been evaluated with measurable outcomes sometimes in single studies, which may not necessarily be the "best" or most efficacious as determined by more robust systematic analyses or meta-analyses.

Second, demonstration and innovation authorities can encourage more evaluation within programs of timely and relevant interventions, evident through use of these approaches that underpin many evaluations conducted over the last 50 years. For example, in the late 1980s, a provision was added to the Social Security Act enabling -large-scale demonstration projects to explore improvements to antipoverty programs. The approach, known widely as "waiver authority" or Section 1115 projects, allows States to request exceptions to Federal rules for implementing a range of mandatory funded human services and health programs that result in no "extra cost" to government.[13] The approach encouraged evaluation of policies under Aid to Families with Dependent Children (AFDC), strategies for improving child support payments from noncustodial parents, improvements to the foster care system, and managed care plans under Medicaid, among many others. While many

---

[10] Gordon, R. and R. Haskins. (2015). A Bipartisan Moneyball Agenda. In *Moneyball for Government*. Second Edition. Eds. J. Nussle and P. Orszag. 114-145.
[11] GAO. (2016). https://www.gao.gov/products/GAO-16-818
[12] https://www.regulations.gov/contentStreamer?documentId=USBC-2016-0003-0009&attachmentNumber=1&contentType=pdf
[13] Guerron, J.M. and H. Rolston. Fighting for Reliable Evidence.

wavier projects included experimental designs, the approach has been used to support a wide range of studies and methods. The Supplemental Nutrition Assistance Program (SNAP) received a similar authority in the 2014 Farm Bill for employment and training pilots, aimed at helping program beneficiaries gain and retain employment to encourage self-sufficiency.[14]

Other programs like the Social Security Disability Insurance (SSDI) program feature similar authority for conducting demonstrations to explore national-scale program improvements rather than relying on state-by-state variation, and typically include implementation studies, impact studies, and cost-effectiveness studies. Importantly, these authorities are somewhat uniquely written for *mandatory* funded programs, and are of limited relevance for those funded through annual discretionary appropriations in the Federal government by virtue of its structure. But they were written in the authorizations for the programs after the programs were created, enabling continued evaluation and innovation for implementing approaches.

None of these identified approaches, however fully addresses the reality that continuous learning means that there is utility in re-evaluating a program and/or building a body of evidence around an intervention, policy or program model. In few instances will a single study, no matter how rigorous, be suitable to definitively "prove" that a program "works" in all contexts, rather building the body of evidence to help inform changing conditions ensures policymakers receive both timely and relevant information. For example, in 2014, HUD funded the Jobs Plus Pilot program, seeking to replicate the model tested under the Jobs-Plus Demonstration back in the 1990s and early 2000, which led to sustained growth in earned income among public housing residents. This current generation of the Jobs Plus program, however, differs from the Jobs Plus demonstration in some important programmatic ways, and also temporally, as the current program is being implemented almost twenty years after the initial demonstration in a very different employment market. Because of these important variations, HUD is supporting an evaluation of the Jobs Plus Pilot program, with the goal of documenting the programs established by the Jobs Plus Pilot Program grantees and laying the groundwork for future evaluative work that will seek to document the impact of the program.

A Role for Randomized Controlled Trials

During the public input phase in which the Commission specifically asked commenters about program design, the Commission heard some perspectives that prioritized certain types of evaluation over other types of analyses (specifically experimental studies or randomized control trials (RCTs) for impact evaluations). For example, the Heritage Foundation submitted comments encouraging the Commission to take a narrow view on defining evidence along this line: "The term 'evidence-based' should mean that experimental evaluations of a program model have found consistent statistically significant effects that meaningfully ameliorate a targeted social problem."[15] While not as directive as Heritage, other organizations encouraged greater use of low-cost RCTs, including comments from the Laura and John Arnold Foundation and J-PAL North America.

More frequently, however, the Commission heard concerns from individuals about over-emphasizing one type of evaluation design within programs over another, or prescribing research methods as part of program design before acknowledging what research or evaluation questions

[14] https://www.fns.usda.gov/2014-ET-Pilots

[15] Muhlhausen, D.B. (2017). Large-Scale, Multisite Experimental Evaluations Produce the Most Credible Evidence of Effectiveness of Federal Programs. https://www.regulations.gov/contentStreamer?documentId=USBC-2016-0003-0126&attachmentNumber=1&contentType=pdf

exist. For example, the Center for the Study of Social Policy (CSSP) suggested that "evaluation methodology should fit the purpose of the inquiry rather than a methodology driving the program design." The American Evaluation Association similarly suggested that the key is to first "identify the important evaluation questions that they need answered to effectively direct the future of Federal programs" enabling evaluators to then "identify which scientific methods are best suited to answer those questions." And the Annie E. Casey Foundation appealed for the "use of best evidence available rather than solely relying on RCTs."

As has been described in this memo, different research questions which may arise during different points in a program's lifecycle may be best addressed through a broad set of evaluative methods. Applying a "one size fits all" approach would greatly limit the range of information that might be gleaned about a program's operation or benefit. As such, the Commission could approach the issue with the increasingly frequent and more general framing around a "portfolio of evidence." This approach recognizes the necessity of a range of methods for building a portfolio of evidence relevant for different types of decision makers and information users, and recognizes the important contribution, though not an exclusive one, of RCTs. Echoing language from the Obama Administration, the Trump Administration's first Budget proposal even endorses the portfolio of evidence approach:

> Government agencies should use a range of evidence types and analytical and management tools to learn what works and what does not, for whom and under what circumstances, and how to improve results….A portfolio of evidence may include:
>
> - Impact evaluations, including randomized control trials and rigorous quasi-experimental designs, which can answer questions about a program's impact relative to a counterfactual—i.e. whether the outcome was achieved because of the program or due to some other factor.
> - Process or implementation evaluations that can answer questions about whether a program is implemented as designed and whether the program structure is sound.
> - Performance monitoring and measurement that can answer questions about program efficiency, outputs, and outcomes, but not about causal impact.
> - Statistics and other forms of research and analysis that can provide insight into trends, strategies, and underlying processes.
>
> There are multiple ways to assess policies and programs. The best approach or method depends on the specific information that is needed to answer key policy, programmatic, or operational questions, and on practical and methodological considerations. While many forms of evidence are complementary, some evidence that is useful for one purpose may not be useful for another. …Combining performance and evaluation information, and using the results of one to inform the design of the other, can be very powerful in understanding program performance and ensuring that a program is maximizing performance and impact on an ongoing basis.[16]

In short, the portfolio of evidence supports a range of evaluation and research methods to support information needs around not just "what works," but for whom, where, and why.

SUMMARY OF COMMISSIONER CALL

[16] OMB. 2017. FY 2018 Budget, Analytical Perspectives Volume. P. 55
https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/budget/fy2018/ap_6_evidence.pdf

A conference call to discuss this memo was held on June 14th. Participating Commissioners were generally comfortable with the framing of the memo, though there were several suggestions to add a paragraph up front documenting the value of evaluation and strongly recommending that all programs be evaluated to the extent practicable. These edits were incorporated by adding the new "encouraging evaluation" section at the front of the memo, and modifying one of the findings to strongly recommend program evaluation. Commissioners were also generally comfortable with the proposed recommendation and accompanying action steps, though several Commissioners made suggestions to streamline and simplify the language, and collapse two similar recommendations into a single recommendation. In addition, there was a suggestion to add an action item that would be designed to orchestrate better coordination between the statistical production activities undertaken by the Principal Statistical Agencies and program evaluation needs within Departments. These edits to the recommendations are reflected in this revised memo.

DRAFT RECOMMENDATIONS

❖ **Finding:** The Commission finds that the onus for ensuring the capabilities to produce evaluation collectively rest with the President, Congress, Departments, and program managers. Program designs must facilitate and encourage the role of evidence for use in program operations and for accountability.

❖ **Finding:** The Commission finds that expectations for evaluation established by the Congress and the President can promote continuous production of valid and reliable evidence about programs and policies, and strongly recommends that all programs be evaluated to the extent practicable.

❖ **Finding:** The Commission finds that the need for evidence about programs is context specific and should be tailored to produce the most relevant, valid, and reliable information needed for anticipated purposes of individual programs and policies. Further, the Commission finds that supporting a "portfolio of evidence" about programs, rather than recommending the application of a single methodological approach, is the best way to maximize evidence production.

► **Recommendation 13.1:** The Commission recommends the President and Congress provide sufficient and appropriate authority for Departments to design programs and policies that enable a portfolio of evidence to support continuous learning and accountability informational needs. Potential approaches to support design could include the following:

- *Action Item 13.1.1.* In establishing new programs in law or reauthorizing existing programs, the Congress and the President should strongly encourage pilot and demonstration projects to be evaluated, and provide flexibility to design policies that allow experimentation, including phased implementation options.
- *Action Item 13.1.2.* In providing direction to Departments to conduct evidence-building activities, Congress and the President should calibrate statutory language to ensure a portfolio of evidence can be developed to jointly achieve accountability and learning goals, specifically refraining from overly prescriptive methodological directives.
- *Action Item 13.1.3.* In establishing new programs in law or reauthorizing existing programs, the Congress and the President should consider enabling the use of set-aside authorities up to one percent to support the full suite of evidence-building activities, including data collection, data curation, policy-relevant research, and evaluation.

- *Action Item 13.1.4*. In appropriating programmatic funding to Departments, the Congress and the President should enable and encourage transfers across budget accounts that support multi-Departmental evidence-building needs.
- *Action Item 13.1.5*. Program offices should be directed to routinely work with central statistical, evaluation, and policy research units to align data collection activities with the needs for evidence-building, within available resources.
- *[In addition to other funding-related recommendations from Recommendations 7.2, 8.4 and 8.5, see Appendix below]*

ALTERNATIVE OPTIONS

- **Mandate evaluation for major Federal programs *or policies* that exceed a direct Federal cost of $x *million*.** A more heavy-handed approach for encouraging increased evaluation and policy research would be to establish mandates for the activities at a policy or program level, rather than expectations that the functions be present. This approach would be similar to that deployed in Canada, where the fiscal office requires periodic evaluations for continued funding of programs. However, as recognized in Canada when implementing the provision, national scale evaluations can have major limitations for both execution and interpretation for use in mandated funding decisions. The mere presence of evaluation activities rather than directed use may be a means to mitigate such concerns. Mandates also generate gaming of activities with unintended effects on the quality and relevance of produced evaluations.[17] (Note this was also included as an alternative in the Evaluation Memo (#8) but has not been discussed during a Commission meeting)
- ***Regulatory Review.*** The President should direct Departments, through OMB, to develop specific evaluation plans in proposed priority regulations reviewed under EOs 13563, 13579, and 13610 as a condition for approval, including capabilities to establish ethically appropriate counterfactuals to the extent permissible by law. [OR The Congress should exercise authority under the Congressional Review Act to discourage regulations that do not sufficiently plan for conducting retrospective review.]
- ***Biennial Evidence-Based Policy Report on Program Designs***. The Government Accountability Office should be directed to develop a biennial summary report to Congress indicating which existing programs have insufficient flexibilities or characteristics to pursue evidence-building activities, providing information as an input for congressional reauthorization discussions.
- ***Annual Evidence-Based Policy Report on State of Evidence.*** GAO should develop an annual report on the state of evidence-based policy in the government each year, including updates on the level of activities, a summary of findings, how findings are being used, and the benefits and costs.

[17] Dahler-Larson, P. (2012). The Evaluation Society. Stanford: Stanford University Press.

APPENDIX: Other Relevant Funding Recommendations Considered by the Commission

► **7.2 :** The Commission recommends that the President should encourage Federal Departments to–

a) Develop multi-year learning agendas in consultation with program and SERA offices, to be updated on a regular basis, containing both short- and long-term evidence-building priorities for each Department; and
b) **Ensure that each SERA function is appropriately resourced**, present, and coordinated with other SERA activities through a regular inventory of units that perform SERA functions across the Department as part of the Strategic Planning process.

► **8.4:** The Commission recommends that the Congress and the President establish and grant Federal Departments **access to *Evidence Incentive Funds*** to supplement the production of future research, evaluation, and related activities outlined in Departmental learning agendas.

► **8.5:** The Commission recommends that the Congress, through the appropriations committees, and the President provide Departments in active pursuit of a learning agenda **access to multi-year (or no-year) funding** to pursue articulated priorities.

RECOMMENDATION MEMO #15

Administration

Commission Leads: Wallin, Shea
Staff Leads: Fletcher, Howell

During the Commission's fact-finding phase, governmental and non-governmental researchers identified numerous barriers within the Federal government that hampered the efficient production of evidence. This memo includes recommendations that relate to two of the most commonly cited administrative barriers that were brought before the Commission, including: (1) requirements under the Paperwork Reduction Act (PRA) for the review and approval of Information Collection Requests of more than nine respondents prior to starting data collection, and (2) the procurement process for data collection, analysis, and evaluation contracts. Relates to Ecosystem (Memo #7), Program Evaluation (Memo #8), and Statistical System (Memo #9). [Sec. 4(a)(1)-(3)]

* * * * *

BACKGROUND AND ANALYSIS

Memo #7 presented the concept of the "evidence ecosystem," which depicts an ideal model for how data could be made available from originating sources to data curators, users and analysts, for the purposes of transforming data into relevant and meaningful evidence. Many of the recommendations being considered by the Commission are designed to move the current state of operations closer to this ideal model, in response to the Commission's charge and vision "to identify the optimal arrangement" that promotes "a future in which rigorous evidence is created efficiently as a routine part of government operations and used in policymaking." While some of the recommendations being considered by the Commission are foundational and designed to shore up the underpinnings of the data infrastructure within the evidence ecosystem, other recommendations—including those in this memo—are designed to address the administrative infrastructure and seek to address bureaucratic inefficiencies within the current evidence-building community.

During the Commission's fact-finding phase, both governmental and non-governmental entities identified numerous administrative barriers within the Federal government that currently hamper the efficient production of evidence, including issues related to human capital, agency capacity, procurement, funding, and the requirements under the Paperwork Reduction Act (PRA) pertaining to Information Collection Requests (ICR). In a follow-up to the Commission public meeting held in November 2016, the Interagency Council on Evaluation Policy (ICEP)[1] submitted a list of the top five barriers to increasing their evidence-building capacity highlighting bureaucratic barriers that "discourage evaluation, create inefficiencies and impose additional costs when conducting a Federal evaluation, particularly issues related to PRA, interagency agreements (IAAs) and procurement."[2] Recommendations related to human capital and agency capacity are addressed in memos #7 (Ecosystem) and #8 (Program Evaluation). This memo addresses the two distinct administrative challenges related to procurement and compliance with the PRA that were brought to the Commission's attention. Addressing these challenges would require modest effort yet would yield a substantial and immediate impact across the Federal evidence-building ecosystem.

---

[1] Signatories to the ICEP memo include representatives from DOL, HHS, HUD, USDA/FNS, DOJ, and NSF.
[2] Include the link to this reference

Administrative Functions Within the Evidence Ecosystem

Memo #7 describes the optimal arrangement of the evidence ecosystem, and notes that the current institutional infrastructure and level of coordination among critical actors within that ecosystem is not sufficient to bring this model to fruition. Inefficiencies within the administrative infrastructure that supports the evidence ecosystem contribute to the headwinds that prevent the realization of the envisioned optimal arrangement. Evidence-building requires support from many administrative functions, but these administrative functions are not always optimized for the particular needs of the evidence-building community. In some cases, there may be a gap within an administrative function that leaves a particular need unfilled, such as the challenges with hiring staff skilled in program evaluation (See Memo #8 on Evaluation and Policy Research), and in other cases, flexibilities exist within a particular administrative function, but are not evenly exercised across Departments, such as the case with procurement. Other administrative functions suffer from a lack of resources or organizational infrastructure that inhibit the full realization of the activity, such as the information management function envisioned under the Paperwork Reduction Act.

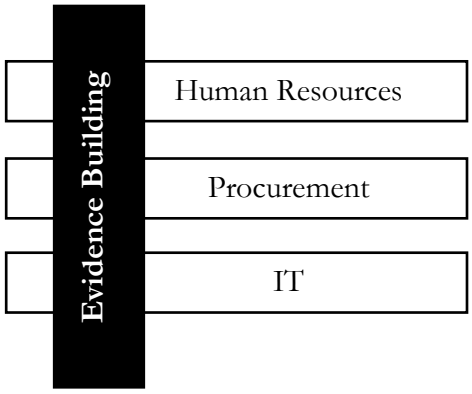


Figure 1.

The Paperwork Reduction Act and Information Collection Requests

The Paperwork Reduction Act (PRA)[3] was originally enacted in 1980, and was substantively amended in 1986 and again in 1995. The scope of the PRA is quite broad and was designed, among other purposes, 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."[4] The PRA notably established the Office of Information and Regulatory Affairs (OIRA) within the Office of Management and Budget (OMB) to "provide central agency leadership and oversight of government-wide efforts to reduce unnecessary paperwork burden and improve the management of information resources."[5] As required by statute, any collection of information, defined as "answers to identical questions posed to, or identical reporting or recordkeeping requirements imposed on, ten or more persons," to be undertaken by a Federal agency must be subjected to public comment and reviewed and approved by OIRA prior to the start of data collection.[6] In the final rule issued by OMB documenting the reporting and recordkeeping requirements under the PRA[7], a broad variety of information collection mechanisms are included under the definition of a "collection of information," ranging from surveys and questionnaires, to forms, to "any other techniques or technological methods used to monitor compliance with agency

[3] 44 U.S.C. § 3501-3520

[4] 44 U.S.C. § 3501

[5] Burrows, Vanessa K. and Copeland, Curtis W. *Paperwork Reduction Act (PRA): OMB and Agency Responsibilities and Burden Estimates.* 2009. Congressional Research Service (CRS) Report # R40636. Washington, DC.

[6] Note the threshold of 10 respondents was adopted for the PRA from the Federal Records Act of 1942.

[7] 5 CFR Part 1320. Reporting and Recordkeeping Requirements: Final Rule. Federal Register Volume 60, No. 167; pp.. 44978-44996

requirements."[8] These requirements are applicable for any collection that is "conducted or sponsored" by a Federal agency, and thus are applicable to any information collection conducted by nongovernmental third parties while under contract to, or in partnership with, the Federal government. As defined, the vast majority of data collection undertaken by the Federal government for statistical, evaluation, policy research and policy analysis (SERA) purposes is therefore covered under the information collection requirements dictated by the PRA. In addition, administrative data collections undertaken, sponsored, or required by Federal agencies are covered by the PRA.[9]

The PRA was intentionally written with broad language for information collection and quality, with wide implications across government. A 2005 Government Accountability Office (GAO) report documents that the legislation's drafters stated that the intention to revise the act in 1995 was "to make a more thorough and open agency paperwork clearance process to improve the quality of paperwork reviews and public confidence in government decision making."[10] This is captured in the purposes of the act, one of which reads "to improve the quality and use of Federal information to strengthen decisionmaking, accountability, and openness in Government and society."[11] The final rule developed by OMB detailing the reporting and recordkeeping requirements under the PRA hew closely to the statutory language and intent, echoing the threshold of "ten or more persons " and the requirement for two separate public comment periods. The PRA directs the OIRA Administrator to "oversee the use of information resources to improve the efficiency and effectiveness of governmental operations to serve agency missions, including burden reduction and service delivery to the public."[12] As noted in memo #7 (Ecosystem), the E-government Act of 2002 fractured implementation responsibilities for the PRA at OMB by establishing a separate office charged with all information technology policy, often with overlapping or conflicting responsibilities to OIRA's information policy function.

By law, there are three primary actors involved in processing an information collection for approval. First, the agency wishing to sponsor the collection prepares an Information Collection Request (ICR), which must meet the standards outlined in the PRA, including elements such as the need for, the plan for, and the public burden estimated for the collection. The Departmental CIO is responsible for certifying that the collection meets these standards *prior* to submitting the package to OIRA for review and approval.[13] The ICR, along with the proposed data collection instruments to be administered under the collection, must also be made available for public comment during this process. Figure 2. depicts the sequencing of the various steps involved in the ICR review and approval process.

---

[8] 5 CFR Part 1320. Reporting and Recordkeeping Requirements: Final Rule. Federal Register Volume 60, No. 167; pg. 44985.

[9] Secondary re-use of data is covered by the PRA, but is not subject to the PRA notice and review process requirements.

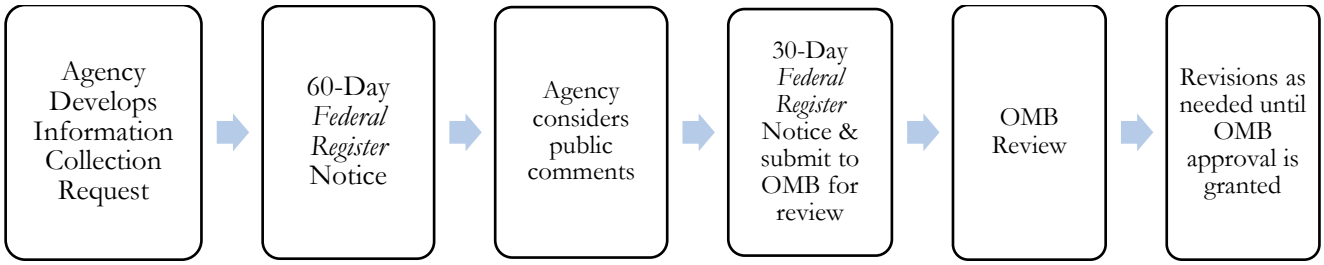
[10] U.S. Government Accountability Office, *Paperwork Reduction Act: New Approach May Be Needed to Reduce Government Burden on the Public*, GAO-05-424, May 20, 2005.

[11] 44 U.S.C. § 3501

[12] 44 U.S.C. § 3504

[13] The 1995 amendments to the PRA used the phrase "senior official," which was later changed to Chief Information Officer in the Clinger-Cohen Act (Pub. L. 104-106, Feb. 10, 1996; Pub. L. 104-208, Sept. 30, 1996).

Figure 2. ICR Review and Approval Process[14]



The steps outlined in Figure 2. must take place sequentially, and OMB is statutorily provided 60 days upon submission of the ICR package to OMB to provide notification to the sponsoring agency regarding approval. As such, the *minimum* amount of time required to process an ICR is four months, though in practice six to nine months is considered a more probable time frame as time must be budgeted for the review and incorporation of any public comment received, as well as internal Departmental clearances.[15]

Challenges With the Paperwork Reduction Act and Information Collection Requests

During the Commission's fact-finding phase, members of the research community both inside and outside of the Federal government identified the requirements of the PRA can be burdensome and time-consuming, causing, as noted by the Laura and John Arnold Foundation, "long delays and bottlenecks for agencies seeking approval of evaluations that have stymied efforts by agencies to increase the number of high-quality studies of important research questions."[16] Five of the ten evaluation offices that responded to the CEP Survey of Federal Offices (CEP Survey) noted that information collection requirements acted as a moderate or major barrier to their pursuit of SERA activities, and seven separate offices[17] identified compliance with the information collection requirements as the most critical barrier they face in conducting their work.

Notably, observations about the challenges associated with meeting the requirements for information collection under the PRA are not new. In 2009, OMB published a Request for Comments (RFC) in the *Federal Register* seeking comments specifically on "reducing current paperwork burdens, especially on small entities; increasing the practical utility of information collected by the Federal Government; ensuring accurate burden estimates; and preventing unintended adverse consequences." Over 80 public comments were submitted in response to the OMB RFC, and a number of the public comments addressed the adverse consequences resulting from the delays associated with the ICR review and approval process, such as the need to modify contracts to account for adjusted project schedules and/or missed opportunities to collect data within a specific window of time.

There are two primary areas of concern for those who identify the ICR review and approval process as a barrier to carrying out SERA activities with maximum efficiency. The first issue relates to the

---

[14] This figure depicting the PRA Clearance Process is frequently included in OMB Guidance materials, including M-11-07.

[15] Shapiro, S. (2013). The Paperwork Reduction Act: Benefits, Costs, and directions for reform. Government Information Quarterly 30: 204-210.

[16] eLetter submitted to the Commission by the Laura and John Arnold Foundation

[17] Seven Federal evaluation offices identified compliance with PRA rules regarding information collection as the most critical barrier to their work, including: USDA, NIH/OD, DOL, HHS/ACL, HHS/CDC, EPA, HUD/PD&R

broad net cast by the PRA in regards to what constitutes an information collection that requires OMB review and approval. The low threshold of respondents that triggers the ICR clearance process (ten or more) means most collections that the Federal government wishes to undertake require OMB review and approval--very few meaningful collections can be implemented with less than ten respondents. In addition, as noted by one administrative law expert, "the combination of a broad definition of "information" and a broad definition of "collection of information" leads to extremely pervasive coverage."[18] The broad definition of what constitutes an information collection means that almost every effort to collect standardized information from the public, which includes grantees who are spending Federal funds, triggers the ICR review and approval process.

The second issue of concern is the length of time required to draft the ICR, then complete the ICR review and approval process can be lengthy. This delay prior to starting data collection, which may be compounded by subsequent Institutional Review Board reviews for certain types of research, often means that a window of opportunity has passed for collecting critical data (such as baseline data from participants in a new program) and greatly inhibits the ability of agencies to gather real time information about program operations. The PRA requires two sequential public comment periods—a 60 day period followed by a 30-day period. Given the significant evolution of electronic communications since the most recent revision of the PRA in 1995, it is unclear if there would be any negative impact of reducing the initial public comment period to a shorter window of time.

It is further unclear that the introduction of the Departmental CIO into the ICR review and approval process has been a productive change within Federal Departments. A 2005 GAO audit of the extent to which Federal Departments have established effective processes found that CIOs often provide certifications for ICRs "despite often missing or partial support from the program offices sponsoring the collections. Further, although the law requires CIOs to provide support for certifications, agency files contained little evidence that CIO reviewers had made efforts to improve the support offered by program offices."[19]
As highlighted during Katherine O'Regan's testimony at the November 2016 CEP meeting, these challenges are most keenly felt by Federal agencies that do not operate their programs directly, but rather provide funding to local organizations to implement their programs. Dr. O'Regan noted that it is often critical to get new information from these organizations to address an un-foreseen issue, or even just perform the most basic oversight that is necessary for ensuring accountability. If the collection of such information cannot take place in a timely manner, agencies are "left to make policy decisions and program changes with very little information; often reacting to anecdote rather than a more complete picture."[20]

Considerations When Evaluating Possible Remedies to the Challenges Identified With the Paperwork Reduction Act and Information Collection Requests

Despite the challenges identified by researchers in complying with the PRA, the requirements of the ICR review and approval process flow from the larger set of purposes outlined in the PRA statute, including minimizing paperwork burdens on those who will be required to respond to Federal information collections. The intent of driving the vast majority of information collections through a

---

[18] Lubbers, Jeffrey S. 1997. *Paperwork Redux: The (Stronger) Paperwork Reduction Act of 1995*. Administrative Law Review, Volume 49, pp. 111-121.

[19] U.S. Government Accountability Office, *Paperwork Reduction Act: New Approach May Be Needed to Reduce Government Burden on the Public*, GAO-05-424, May 20, 2005.

[20] Testimony to CEP of Dr. Katherine O'Regan, Assistant Secretary for Policy Development & Research, HUD; November 4, 2016.

single approving entity supports the notion that information collections should be coordinated across the Federal government, provides an opportunity for a neutral assessment of the necessity and quality of a planned collection, and a concrete mechanism for implementing government-wide information policy. In addition, the public comment period built into the ICR provides for transparency, allowing members of the public to have advance opportunity to comment on the necessity, utility, and likely quality of the proposed collection, as well. As noted at the September CEP meeting by Marc Groman, then-Senior Advisor for Privacy at OMB and Chair of the Federal Privacy Council, "[As] the United States Government we have an obligation to be transparent to our citizens about what we collect and how we use it." He further noted that the efforts undertaken by the Federal government to achieve transparency, such as the publication of *Federal Register* Notices announcing a new information collections or the publication of Privacy Act System of Records Notices, provide an additional benefit of serving as a valuable accountability function to the public.

Flexibilities exist within the PRA statute which may currently be underutilized. For example, the PRA provides OMB with the ability to delegate authority to approve proposed collections of information as long as "the Director finds that a senior official…..is sufficiently independent of program responsibility to evaluate fairly whether proposed collections of information should be approved and have sufficient resources to carry out this responsibility effectively."[21] OIRA reviews between 3,000 and 5,000 information collection requests each year;[22] so exercising this flexibility could reduce the volume of reviews that must funnel through this one office, and create an opportunity to focus additional attention on other purposes of the PRA, such as coordinating information collections across Government and addressing issues of quality. Though, statutory adjustments could also provide OMB greater flexibility to prioritize key data collections for the highest level of scrutiny, rather than relying on a one-size-fits all process for collections that affect ten or more individuals that vary greatly in level of burden, quality, and utility.

Procurement

Procurement by the Federal government is defined at 48 CFR 2.101 as "acquiring by contract with appropriated funds supplies or services (including construction) by and for the use of the Federal Government through purchase or lease, whether the supplies or services are already in existence or must be created, developed, demonstrated, and evaluated." In fiscal year 2016, the Federal Government awarded just under a half billion dollars in contracts to outside entities to carry out work on behalf of the Federal government.[23] Federal acquisitions are governed by the Federal Acquisition Regulation (FAR), the primary regulation for use by all Federal Executive agencies in their acquisition of supplies and services with appropriated funds. Contracts are an important mechanism used by SERA agencies to carry out research and evaluation activities, but Federal agencies frequently identify procurement as a major barrier to the efficient production of evidence. In the CEP Survey, 62% of Principal Statistical Agencies and 50% of evaluation offices identified lack of ability to execute and manage contracts as a major or moderate barrier to their ability to use data for SERA purposes.

---

[21] 44 U.S.C. § 3507(i)(1)

[22] Burrows, Vanessa K. and Copeland, Curtis W. *Paperwork Reduction Act (PRA): OMB and Agency Responsibilities and Burden Estimates.* 2009. Congressional Research Service (CRS) Report # R40636. Washington, DC.

[23] See www.usaspending.gov for the total value of awarded contracts (as well as contracts, grants, loans and other assistance) made by the Federal government by fiscal year.

Challenges With Procurement

There are many procurement-related challenges when entering into a contract for the execution of SERA purposes, as the type of work sought under such contracts tends to be dynamic and is frequently novel. This memo will focus on three specific challenges for which concrete actions might be taken to improve the current state. The first challenge relates to the acquisition itself, specifically, how to structure the procurement properly at the outset to ensure that the work can be conducted in the most effective manner and result in the product desired by the government. The second challenge relates to the management of contracts that support SERA activities. As noted in memo #8 (Program Evaluation), contract management in support of research and program evaluation efforts tends to be a more complicated subset of the contract management function with riskier contract types and/or more challenges to assessing successful performance, thus requiring staff with both substantive expertise as well as contract management skills. The third challenge relates to the period of availability of funds that different agencies use to support evaluations. SERA activities are often considered "non-severable services," meaning that the work to be conducted constitutes a single undertaking for which the requesting agency receives no benefit until the entire project is completed. An example includes a research project where the requesting agency does not receive anything of consequential value until the entire project is completed and it receives a final report.

The FAR offers the Government a significant level of flexibility in structuring a procurement from the outset to maximize the efficiency of the contract and the likelihood of success. One of the earliest decisions an agency will face after determining that a contract is the appropriate mechanism for procuring a service is the selection of contract type. FAR Part 16 describes the different contract types, providing guidance and procedures for selecting the proper contract type for a specific acquisition. Contract types can be grouped into two broad categories: fixed-price contracts, in which the contractor has full responsibility for the performance, costs and resulting profit (or loss), and cost-reimbursement contracts, in which the contractor receives payments based on the work conducted, regardless of the quality of the product produced. There are many variations on contract type between these two poles. Because of the level of potential risk involved in contracts that are other-than-fixed-price, over the past decade, there has been increasing pressure on Federal agencies to default to a firm-fixed price contract, which requires the contractor to bear the full risk of the contract, as well as full responsibility for all costs, even if those costs exceed the value of the contract.[24] The FAR states, "A firm-fixed-price contract is suitable for acquiring commercial items or for acquiring other supplies or services on the basis of reasonably definite functional or detailed specifications when the contracting officer can establish fair and reasonable prices at the outset."[25] However, often times, research and evaluation activities may be more exploratory in nature and/or the field conditions might not be apparent in advance of data collection. As such, "reasonably definite functional or detailed specifications" may not be feasible in the early stages of a procurement.

The FAR recognizes these challenges, and thus includes Part 35, which was established to provide guidance on what are termed research and development (R&D) contracts. FAR Part 35 notes, "Unlike contracts for supplies and services, most R&D contracts are directed toward objectives for which the work or methods cannot be precisely described in advance."[26] Often times research

---

[24] See the March 4, 2009 Memorandum on Government Contracting accessible here: https://www.gpo.gov/fdsys/pkg/FR-2009-03-06/pdf/E9-4938.pdf
[25] Federal Acquisition Regulation, 48 C.F.R. § 3.502-1
[26] FAR Part 35

projects, particularly evaluations of large demonstration programs or novel research efforts, may include an extensive design phase during which various methodologies may be considered prior to determining the desired approach. The general purpose section of FAR Part 35 states, "The contracting process shall be used to encourage the best sources from the scientific and industrial community to become involved in the program and must provide an environment in which the work can be pursued with reasonable flexibility and minimum administrative burden."[27] FAR Part 35 acknowledges that R&D can be a more exploratory and/or fungible exercise based on field conditions, stating "because of the importance of technical considerations in R&D, the choice of contract type should be made after obtaining the recommendations of technical personnel. Although the Government ordinarily prefers fixed-price arrangements in contracting, this preference applies in R&D contracting only to the extent that goals, objectives, specifications, and cost estimates are sufficient to permit such a preference. The precision with which the goals, performance objectives, and specifications for the work can be defined will largely determine the type of contract employed. The contract type must be selected to fit the work required." Within R&D contracting, there is acknowledgement that riskier contract types, such as Time & Material/Labor Hour contracts may be more appropriate for accomplishing the goals of the procurement. As an example of the importance of fitting the contract type to the work required, a recent randomized controlled trial (RCT) conducted by HUD was originally anticipated to be conducted as a panel study, but after extensive consultation with experts, the Department determined that an RCT would be feasible and changed course. Because the Department had selected a more flexible contract type, this mid-course correction was adopted having to defaulting or cancel the contract.

The second challenge relates to the familiarity with and management of contracts that support SERA activities. Acquisition staff who may be unfamiliar with contracts that are not structured as fixed-price and with the flexibilities available under FAR Part 35, may seek to limit an agency's preference to utilize a riskier contract type in support of SERA activities. Similarly, the technical staff who will be responsible for overseeing the day-to-day management of the contract may be unfamiliar with the proper techniques for managing the performance of a riskier contract type. Differing goals between the acquisition professionals and technical staff may lead to difficulties in identifying the right contract type, and raise concerns about the capacity of the Government to properly manage the contract.

Finally, the third challenge relates to the period of availability of funds that different agencies use to support evaluations. SERA activities are often considered "non-severable services," which per Appropriations law must either be fully funded upfront, or structured as a multi-year contract. Because agencies often do not have sufficient funding to fully fund the entire cost of a study that is expected to stretch over several years, they frequently break these contracts into phases, ultimately resulting in a single study being supported by multiple complicated and overlapping contracts.

Considerations When Evaluating Possible Remedies to the Challenges Identified With Procurement

The Government's increasing focus on reducing risk in Federal contracting is rooted in evidence that suggests that Federal agencies may not be properly documenting their rationale for choosing the riskier contract types or demonstrating strong management of those riskier contracts within their portfolio.[28] In recent years, Federal agencies were directed to examine their use of non-competitive

[27] Ibid.
[28] GAO. 2009. Extent of Federal Spending under Cost-Reimbursement Contracts Unclear and Key Controls Not Always Used. GAO-09-921.

and cost-reimbursement contracts[29], and OMB Memorandum M-09-25 on *Improving Government Acquisitions* required agencies to reduce by at least ten percent "the combined share of dollars obligated through new contracts in FY 2010" that are awarded non-competitively or are either cost-reimbursement or Time & Materials/Labor Hours contracts.[30] These efforts, while well intentioned, may have the unintended effect of discouraging cost-reimbursement or riskier contract types, even when the work to be procured would be more appropriately conducted under such a contract type.

Procurement challenges across the evidence-building community are well-documented and are commonly shared across a broad variety of statistical agencies and evaluation offices, but there are some examples of procurement offices that are more familiar with R&D contract types. For example, the Program Support Center (PSC), which is an operating division within HHS, acts as a shared services division offering acquisition services to Federal agencies and has a depth of experience in procuring data collection, analysis, and evaluation contracts Any recommendations that seek to address the challenges that Federal agencies face regarding procurement should address the roles and responsibilities of both acquisition staff and technical staff. If there is a desire to more intentionally consider the full range of contract types in support of SERA activities, there should be complementary effort to ensure that technical staff are well-trained in the management in the riskier contract types.

SUMMARY OF COMMISSIONER CALL

A conference call to discuss this memo was held on June 7th. Commissioners discussed the rationale for including recommendations within this topic area, and the extent to which these topics were felt to be in scope. Overall, participating Commissioners were generally comfortable with the framing of the memo. Commissioners offered some refinements to the sharpen the specific action steps that are reflected in this revised memo, and a new action step was suggested to enhance agencies' abilities to coordinate their information collections. These comments were incorporated into this revised memo.

DRAFT RECOMMENDATIONS

Finding: The purposes embodied by the PRA are important for ensuring the necessity, utility, and quality of the information collections administered by Federal agencies. The PRA provides an underutilized statutory infrastructure for coordination across data collections undertaken, sponsored, or required by the Federal government.

Finding: The statutory language of the PRA intentionally casts a broad net to ensure that the majority of information collections undertaken by the Federal government are included in the required review and approval process. The capacity to review and approve important collections must be calibrated to optimize governments ability to achieve the goals of the PRA.

Finding: The current structure of the ICR review and approval process can be inefficient, leading to problematic delays in data collection and challenges to OMB's ability to maximize the coordination and transparency of Federal information collections.

[29] Presidential Memorandum on Government Contracting, Federal Register, Vol. 74, No. 43. March 6, 2009.
[30] OMB Memorandum M-09-25, *Improving Government Acquisitions,* July 29, 2009.

► **Recommendation 15.1:** The Commission recommends Congress and the President modernize the PRA, taking action to streamline the review and approval process for ICRs, while ensuring that information coordination and transparency remain critical features of PRA implementation. This could be done in any of the following ways, either individually or in combination:

- *Action Item 15.1.1.* Direct OMB to issue guidance to Federal agencies describing the flexibilities under the PRA, including the process for acquiring "delegated authority" to approve proposed collections of information.
- *Action Item 15.1.2.* Amend the PRA to shorten the initial public comment period from 60 days to 30 days in light of the widespread use of electronic communications that has taken place since the PRA was last amended.
- *Action Item 15.1.3.* Remove the Departmental CIO from the ICR review and approval process and instead assign the "Senior Agency Official for Data Policy" responsibility for certifying that the collection meets the standards outlined in the PRA.
- *Action Item 15.1.4.* Amend the PRA to provide to provide additional flexibilities to enable the Federal government to most effectively support evidence-building.
- *Action Item 15.1.5* Establish a searchable electronic inventory of all current and proposed information collections to better enable coordination and non-duplication of information collection activities.

The purposes outlined in the PRA are directly relevant to a well-functioning evidence ecosystem, but there are elements to the legal, policy and implementation framework that would benefit from modernization. The law contains a number of authorities and flexibilities under the PRA, but these flexibilities are rarely exercised, largely because neither agency CIO's nor OIRA are equipped to exercise them. In the absence of exercising those authorities and flexibilities, evidence building is hampered by inefficiencies, delays, and missed opportunities.

Recommendation #9.2 would require each department to designate the head of a Principal Statistical Agency within the department, or similarly senior-level departmental official as the Senior Agency Official for Data Policy. The certification of information collections follows as a natural part of the coordination role envisioned for the PSA heads and other designated Senior Agency Officials for Data Policy.

Finding: Federal agencies frequently identify procurement as a major barrier to the efficient production of evidence.

Finding: While there is good reason for the Federal government overall to maintain a strong preference for fixed-price contracting, the existence of flexibilities within the current regulations acknowledge that there are some types of services, specifically those included under research and development, which may be more appropriately procured with a more flexible, yet riskier, contract type.

► **Recommendation 15.2:** The Commission recommends OMB clarify the applicability of FAR Part 35 to contracts that support SERA activities, as appropriate, to support evidence-building. This could be done in any of the following ways, either individually or in combination:

- *Action 15.2.1* The Commission recommends OMB issue guidance to agencies that clarifies the applicability of procurements related to data collection, analysis, and evaluation

services under FAR Part 35, and encourages the use of FAR Part 35 when awarding contracts that support evidence-building, where appropriate.

- *Action Item 15.2.2*. The Commission recommends [insert actor here] develop trainings for both acquisition professionals and technical staff to address topics such as the applicability of FAR Part 35 to in support of select evidence-building activities, effective management techniques for oversight of service contracts of a riskier type, and the management of research and evaluation contracts more generally.
- *Action Item 15.2.3* The Commission recommends that Congress grant certain evaluation and statistical agencies certain flexibilities, such as multi-year funding, when pursuing contracts for evidence-building purposes.
- *Action Item 15.2.4* The Commission recommends the [insert actor here] establish a Govermentwide Acquisition Contract (GWAC) for use in issuing contracts for SERA activities and services.

The recommendations related to procurement are primarily related to strengthening implementation to make this administrative function more aligned with evidence-building purposes. Similar to the PRA, there are flexibilities that exist within the FAR that would benefit from clarification and expanded implementation.