Privacy and Confidentiality in the Use of Administrative and Survey Data[1]

I. Executive Summary

This white paper discusses data infrastructure, database security, and statistical protocols by reviewing some of the most relevant Federal laws and other agency- and sector-specific laws. It then describes the range of data access protocols used by various Federal agencies to provide researchers access to restricted data.

The Commission on Evidence-Based Policymaking ("the Commission") is charged with, among other things, examining how administrative data on Federal programs and survey data may be used to build evidence while also protecting individual privacy and confidentiality.[2] Federal evidence-building efforts provide insight about the Nation's population and economy and Federal policies and programs serving our national interests. The Federal Government may realize these benefits by extending the utility of previously collected data. However, this should be done in a way that protects the data, ensuring the trust of those individuals and businesses that provide them. Many Federal evidence-building activities rely on statutes, time-honored principles, and related practices to protect data appropriately. Several Federal statutes not only protect data but also establish and distinguish research and statistical activities.

As discussed in the white paper, *Overview of Federal Evidence-Building Efforts*, the U.S. government has a highly decentralized approach for building evidence.[3] As stated above, this paper will discuss data infrastructure, database security, and statistical protocols by first reviewing some of the most relevant Federal laws, including the Privacy Act of 1974 (Privacy Act)[4], the Confidential Information Protection and Statistical Efficiency Act (CIPSEA)[5], and the E-Government Act of 2002 (E-Government Act)[6], and mentioning other agency- and sector-specific laws. Following this review, the paper discusses how these laws provide a strong legal framework that guides the U.S. Census Bureau's Data Stewardship program. The Census Bureau is one of 13 principal statistical agencies whose primary missions include the collection, compilation, processing, analysis, and dissemination of statistical information.[7] The Census Bureau is highlighted because it is a leader in the technical work of bringing together data from multiple sources, protecting privacy and confidentiality, ensuring information security throughout the course of

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

[2] Evidence-Based Policymaking Commission Act of 2016, Pub. L. No. 114-140, § 4, 130 Stat. 317, 318. For the purposes of this paper, evidence-building includes the collection, compilation, processing, or analysis of data to better understand the characteristics, behavior, or needs of groups of individuals or communities. It excludes uses that affect the rights, benefits, or privileges of individuals but includes a wide range of analytic uses, where only aggregated and de-identified data are made public.

[3] The Federal evidence-building system includes principal statistical agencies, evaluation offices, and other evidence-builders.

[4] Privacy Act of 1974, 5 U.S.C. § 552a.

[5] Confidential Information Protection and Statistical Efficiency Act, Pub. L. No. 107-347, title V, 116 Stat. 2962.

[6] E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899.

[7] See CIPSEA, Pub. L. No. 107-347, § 502(8) and "Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units." https://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/cipsea/cipsea_statute.pdf.

the data lifecycle, and providing secure access to researchers inside and outside of government to conduct a broad array of policy- and program-relevant analyses.

The paper then considers a range of data access protocols used by different agencies to provide researcher access to restricted data. As an example of such access, it examines how the National Center for Education Statistics (NCES) within the Department of Education (ED) has implemented a data licensing program and how this program supports ED's mission by permitting researcher access to data while also meeting obligations to safeguard confidential information.[8] Some of these practices and processes may be particularly relevant to the Commission as it considers privacy, confidentiality, and data security in the context of how a Federal clearinghouse for program and survey data should be established. This discussion highlights some of the highest-capacity agencies and processes, and as noted in the white paper, *Barriers to Using Administrative Data for Evidence-Building*, there is inconsistency in agency capacity and infrastructure.

II. Background: Relevant U.S. Law

Much of the relevant law that governs Federal agencies' maintenance of information about individuals is based on the Fair Information Practice Principles (FIPPs).[9] While the precise expression of the FIPPs has varied over time and in different contexts, the FIPPs retain a consistent set of core principles that are broadly relevant to agencies' information management practices. In their application, the FIPPs inform the most basic information protections, including the importance that an individual know what information is collected, how it is used, and have the opportunity to correct inaccurate information. The FIPPs provide a framework for data handling and help to ensure that the data are reliable and kept secure.

The FIPPs are generally regarded as the basis for many Federal data protection statutes that provide the overarching framework for privacy protections, including the Privacy Act. The Privacy Act is especially relevant to the Commission's task, as it provides for both the basic protection of individuals' records[10] and includes provisions that pertain to the use of records for statistical purposes. The Privacy Act established a number of requirements that pertain to records that are maintained in a "system of records," as defined in the statute.[11] Among the statute's requirements is the requirement for agencies maintaining a system of records to publish a system of records notice that identifies the system of records, the categories of records maintained in the system, the categories of individuals about whom records are maintained, the routine uses to which the records will be subject (including the purpose of each use), and other details about the system.[12] In addition, the Privacy Act generally prohibits agencies from disclosing individuals' records without the prior written consent of the individual to whom the information pertains. There are a limited number of exceptions to this prohibition. As described in Office

[8] See 20 U.S.C. § 9573.

[9] The Fair Information Practice Principles are rooted in the United States Department of Health, Education and Welfare's seminal 1973 report, "Records, Computers and the Rights of Citizens." These principles are at the core of the Privacy Act of 1974 and are mirrored in the laws of many U.S. States, as well as in those of many foreign nations and international organizations. A number of private and not-for-profit organizations have also incorporated these principles into their privacy policies.

[10] As defined in the Privacy Act, the term "record" means "any item, collection, or grouping of information about an individual that is maintained by an agency, including, but not limited to, his education, financial transactions, medical history, and criminal or employment history and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph." 5 U.S.C. § 552a(a)(4).

[11] As defined in the Privacy Act, the term "system of records" means "a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual." *Ibid.* § 552a(a)(5).

[12] 5 U.S.C. § 552a(e)(4)(D).

of Management and Budget (OMB) memorandum M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes*, four of the exceptions may be relevant to efforts to provide data for statistical purposes.[13]

As M-14-06 notes, the Privacy Act provides an exception from the matching requirements[14] specified in the Act for "matches performed to produce aggregate statistical data without any personal identifiers" and "matches performed to support any research or statistical project, the specific data of which may not be used to make decisions concerning the rights, benefits, or privileges of specific individuals…."[15]

The Privacy Act defines "statistical record" as, with one exception, "a record in a system of records maintained for statistical research or reporting purposes only and not used in whole or in part in making any determination about an identifiable individual…."[16] OMB's 1975 *Privacy Act Implementation Guidelines and Responsibilities* further distinguishes statistical records from "virtually any other record."[17] That guidance states that only non-statistical records need to be "obtained directly from the individual whenever practicable,"[18] which implies that statistical records may be obtained indirectly from already-collected administrative data.[19]

Additionally relevant to the Commission's considerations are the requirements of the E-Government Act, with provisions covering information handling and "privacy impact assessments" (PIA), data security, and CIPSEA. One of the key privacy provisions of the E-Government Act is Section 208(b), which requires agencies to conduct PIAs. A PIA is "an analysis of how [personally-identifiable] information [(PII)] is handled: (i) to ensure handling conforms to applicable legal, regulatory, and policy requirements regarding privacy, (ii) to determine the risks and effects of collecting, maintaining and disseminating information in identifiable form in an electronic information system, and (iii) to examine and evaluate protections and alternative processes for handling information to mitigate potential privacy risks."[20] PIA refers to both the analysis undertaken and the formal document detailing the process and the outcome of the analysis. As a general matter, an agency must conduct a PIA, absent an applicable exception under Section 208(b), when the agency develops, procures, or uses information technology to create, collect, use, process, store, maintain, disseminate, disclose, or dispose of PII.[21] Each PIA is reviewed by the Chief Information Officer, or equivalent official, as determined by the head of the agency. Pursuant to OMB guidance, a PIA is not required where information relates to internal government operations, has been previously assessed under an evaluation similar to a PIA, or where privacy issues are

---

[13] Office of Management and Budget, M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes*, available at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

[14] 5 U.S.C. § 552a(o).

[15] *Ibid.* § 552a(a)(8)(B)(i)-(ii).

[16] *Ibid.* § 552a(a)(6).

[17] See "Privacy Implementation Act Guidelines and Responsibilities," 40 Fed. Reg. 28948, 28961 (July 9, 1975), *available at* https://www.whitehouse.gov/sites/default/files/omb/inforeg/implementation_guidelines.pdf.

[18] *Ibid.*

[19] 5 U.S.C. § 552a(b). For further discussion of how the Privacy Act applies to the use of administrative records for statistical purposes, see OMB M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes.*

[20] Office of Management and Budget, M-03-22, *OMB Guidance for Implementing the Provisions of the E-Government Act of 2002* (Sept. 26, 2003), *available at* https://www.whitehouse.gov/omb/memoranda_m03-22.

[21] *See* 44 U.S.C. § 3501 note. Section 208(b) of the E-Government Act requires agencies, absent an applicable exception under this section, to conduct a PIA before: (i) developing or procuring IT that collects, maintains, or disseminates information that is in an identifiable form; or (ii) initiating a new collection of information that – (I) will be collected, maintained, or disseminated using IT; and (II) includes any information in an identifiable form permitting the physical or online contacting of a specific individual, if identical questions have been posed to, or identical reporting requirements imposed on, 10 or more persons, other than agencies, instrumentalities, or employees of the Federal Government.

unchanged.[22] For example, a PIA is not required when all elements of a PIA are addressed in an interagency agreement permitting the merging of data for strictly statistical purposes and where the resulting data are protected from improper disclosure and use under Title V of the E-Government Act.[23]

The Federal Information Security Modernization Act of 2014 (FISMA) amended title III of the E-Government Act to require that "each Federal agency shall develop, document, and implement an agency-wide information security program to provide information security for the information and information systems that support the operations and assets of the agency, including those provided or managed by another agency, contractor, or other source…."[24]

Title V of the E-Government Act, CIPSEA, established uniform confidentiality protections for information acquired by agencies, including principal statistical agencies and recognized statistical units, under a pledge of confidentiality and for exclusively statistical purposes. CIPSEA requires that such information be used exclusively for statistical purposes and not be disclosed by an agency in identifiable form, for any use other than an exclusively statistical use, without informed consent.[25] CIPSEA defines a statistical purpose as "the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups" and explains that the definition includes "the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support [such purposes]."[26] The use limitations, data protections, and penalties of CIPSEA and other principal statistical agencies' laws are the basis for the strong framework that underpins not only partnership with administrative agencies and researchers to support evaluation and research, but also the dissemination of public data products.[27]

In addition to CIPSEA, there are a number of agency- and sector-specific statutes concerning privacy and data protection. The examples below incorporate use limitations, disclosure limitations, accountability and penalties, and other protections.

- The U.S. Tax Code provides strong statutory protections and penalties for the unauthorized disclosure of individual and business tax information and limits disclosure, including to specific agencies for certain purposes.[28]

[22] For more information, see OMB memorandum M-03-22, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*.

[23] *Ibid.*

[24] 44 U.S.C. § 3554(b).

[25] CIPSEA, §§ 512(a) and (b).

[26] *Ibid.* § 502(9).

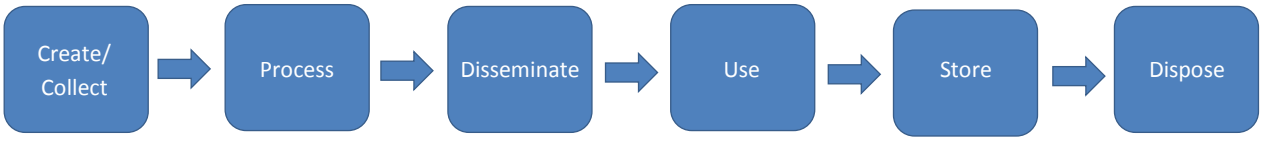
[27] Under CIPSEA, for example, knowing and willful disclosure by officers, employees, or certain agents of an agency of information that was acquired for exclusively statistical purposes "in any manner to a person or agency not entitled to receive it, shall be guilty of a class E felony and imprisoned for not more than 5 years, or fined not more than $250,000, or both." *Ibid.* § 513.

[28] 26 U.S.C. § 6103. Section 6103(h)(1) permits disclosure of Federal Tax Information (FTI) to employees of the Department of the Treasury whose official duties require inspection or disclosure of the information for tax administration purposes. Section 6103(b)(4) defines tax administration purposes to include "statistical gathering functions" under internal revenue laws or related statutes (or equivalent laws and statutes of a State), and tax conventions. Section 6103(j) provides for disclosure of FTI to specific agencies and components. Additionally, 6103(n) authorizes IRS to hire contractors to support the agency's mission. Section 6103(j) provides for disclosure of FTI to specific agencies.

- Title 42 of the U.S. Code protects the information states provide to the Department of Health and Human Services Federal Parent Locator System, which includes the National Directory of New Hires, permitting the disclosure to specific agencies for limited purposes.[29]
- The Family Educational Rights and Privacy Act (FERPA) protects the privacy of student education records, limiting both access and use.[30]
- The 1996 Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act (HITECH Act), addresses the use and disclosure of protected health information by specified "covered entities."[31]

Federal agencies are responsible for protecting the privacy of and honoring the pledges of confidentiality for the data they collect and maintain to support Federal programs and statistical activities over the course of the data lifecycle. Figure 1 shows the data lifecycle, or the stages through which data passes, typically characterized as creation or collection, processing, dissemination, use, storage, and disposition.[32] Agencies can and should consider what actions are appropriate through the data lifecycle, based on the relevant legal, technical, and policy frameworks.

Figure 1. The Data Lifecycle



III. Discussion: Overview of Data Protection and Privacy and Confidentiality Protections for the Data Lifecycle at the U.S. Census Bureau

This section provides an example of the legal, technical, and policy framework that Federal evidence-builders use to help ensure privacy and confidentiality in the acquisition and uses of administrative and survey data throughout its lifecycle. The section describes the administrative data lifecycle of the Census Bureau, one of the main repositories of Federal data. The Census Bureau's approach is based in its own law that provides an authority to acquire, utilize, and protect administrative data, which are utilized in a variety of programs and activities authorized under Title 13.[33] In utilizing administrative records, the Census Bureau complies with its law, as well as other applicable laws, and incorporates principles of data stewardship relevant to these laws, as well as the FIPPs, throughout the course of the administrative and survey data lifecycle, utilizing legal agreements, policies, procedures, and security controls to protect confidentiality and privacy.[34]

---

[29] 42 U.S.C. § 653(j).

[30] 20 U.S.C. § 1232(g)(b).

[31] *E.g.*, Pub. L. No. 111-5, § 13405, 123 Stat. 115, 264-68.

[32] See Office of Management and Budget, Circular A-130, *Management of Federal Information Resources*, *available at* https://www.whitehouse.gov/omb/circulars_a130.

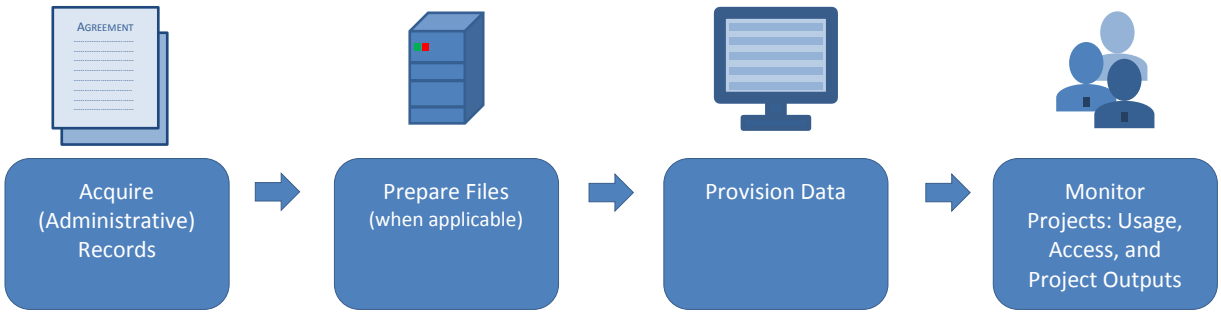
[33] 13 U.S.C. et seq.

[34] For more information about the Census Bureau's policies and procedures for safeguarding information, see, the "Data Protection," website (http://www.census.gov/about/policies/privacy/data_protection.html). For information about the Census Bureau's data linkage activities, including descriptions of the process for creating "Protected Identification Keys" (PIK), as well as research project reports, see the "Census Bureau Linkage Infrastructure" website (http://www.census.gov/about/adrm/data-linkage/about.html).

1. *Data Lifecycle at the Census Bureau*

Figure 2 outlines the Census Bureau framework for coordinating the specific requirements of Title 13, along with the requirements of other Federal privacy and data protection laws. At its most basic form, it encompasses several activities, from collection or acquisition of data through public dissemination. This framework is becoming increasingly significant because the acquisition and use of administrative records is becoming increasingly central to the Census Bureau's mission. As discussed in other Commission papers, administrative records are collected or maintained by Federal, state, tribal, or local government agencies or commercial entities to administer programs or to provide services. The Census Bureau uses administrative records to reduce inquiries on the public and produce and improve statistics on the American population and economy. As the Census Bureau and other Federal evidence-builders successfully integrate administrative records with surveys and other statistical collections to increase the utility of the underlying data through deeper, more detailed insights about groups and localities or to conduct program evaluations, a strong data stewardship program to protect confidentiality and privacy is essential.[35]

Figure 2. The Administrative Data Lifecycle at the Census Bureau



2. *Legal and Policy Foundations Undergird the Data Lifecycle*

At the Census Bureau, the integration of administrative records is facilitated and governed by the same set of laws and policies that protect the privacy and confidentiality of the Census Bureau's survey respondents. These laws and policies are directly relevant to each phase of the lifecycle. The Census Bureau's Data Stewardship Program helps ensure the proper use and protection of PII and other confidential data over the course of this lifecycle. [36]

The Census Bureau's collection authority, as well as its obligation to protect confidentiality, is governed by Title 13 of the United States Code.[37] 13 U.S.C. 9 provides strong protections for the information collected from individuals and businesses. 13 U.S.C. 6 provides the authority to acquire and use administrative records. This section authorizes and directs the Census Bureau to acquire and use records previously collected by other Federal agencies and state, tribal, or local governments, as well as private organizations, and to seek out this information instead of conducting direct inquiries.

---

[35] OMB, "Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units," 79 Fed. Reg. 231 (December 2, 2014). https://www.gpo.gov/fdsys/pkg/FR-2014-12-02/pdf/2014-28326.pdf.

[36] More information on the Census Bureau's Data Stewardship and Privacy programs is available at http://www.census.gov/about/policies/privacy/data_stewardship.html and http://www.census.gov/privacy/.

[37] 13 U.S.C. et seq.

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. One of the exceptions to the Privacy Act's written consent requirement permits agencies to disclose data to the Census Bureau to carry out censuses, surveys, and "related activity pursuant to the provisions of Title 13," and many administrative agencies cite this authority when transferring their data to the Census Bureau.[38] As administrative data are transferred to the Census Bureau for activities authorized under Title 13, the Census Bureau is obligated to protect their confidentiality just as it protects the information it gathers directly from individuals and businesses: these data can only be used for statistical purposes; no individual or business may be identified in a published report; and individual records may be accessed only by sworn officers or employees of the Census Bureau.

3. *Data Protections throughout the Data Lifecycle*

To help ensure adherence to the laws and other Federal policies that protect privacy and confidentiality, the Census Bureau has established a series of administrative and technical processes relevant to the use of administrative records over the course of the data lifecycle. These are made up of a well-established set of practices for acquiring administrative and commercial datasets, processing and removing PII[39] of data files, accessing data files by approved projects, and producing statistical output and other resources derived from census data and administrative records. These practices are reinforced in a required Data Stewardship training program that educates all employees and sworn agents on an annual basis.

Acquisition

The Census Bureau acquires administrative records by entering into data sharing agreements or joint statistical projects with other agencies, usually via an Interagency Agreement (IAA).[40] An IAA provides a vehicle to document both the legal authority for disclosing or providing data and the applicable data stewardship policies and practices that will protect data provided by the program agency for statistical purposes. For instance, Census Bureau IAAs typically indicate that a data file is being acquired under Title 13, Section 6 (or other applicable sections) and protected under Section 9 as well as the specific security and confidentiality controls and standards that will be applied at the Census Bureau. The other agency's authority to provide data, which may be another specific law, is also specified.

The IAAs describe such important topics as planned methods for file transfers, data retention, and any requirements for the providing agency to approve uses of the file. Files that the Census Bureau acquires are encrypted prior to delivery and then transmitted to the Census Bureau either via Secure File Transfer Protocol, dedicated Virtual Private Networks (VPN)[41], or on physical media. The IAA names specific agency contacts and data custodians, and these individuals may have reporting responsibilities through the life of the agreement. At the Census Bureau, all final signed agreements to acquire data are logged in a centralized interagency agreement repository system. All incoming data files are logged in a

---

[38] 5 U.S.C. § 552a(b)(4).

[39] Here we mean the process by which the Census Bureau removes PII and applies statistical disclosure limitation procedures prior to public release to ensure the confidentiality of the underlying survey and/or administrative data.

[40] For model agreements used by the Census Bureau and other agencies, see the Census Bureau's Interagency and Other Special Agreements page (https://www.census.gov/about/business-opportunities/resources/iosa.html) and Appendix B of OMB memorandum M-14-06 at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

[41] A VPN is a private network that uses a public network (usually the Internet) to connect remote sites or users together. The VPN uses encrypted, "virtual" connections routed through the Internet from an organization's private network to the remote site.

centralized Data Management System, which assigns responsibilities for tracking and reporting to a named Data Custodian.

Preparing Files and Provisioning Data[42]

Once files are acquired by the Census Bureau, they are initially only accessible by a small staff responsible for inventorying the contents of the file. This staff "curates" the data, conducting basic quality control (QC) checks to ensure the data are clean and can be processed, and removing explicit identifiers, such as name or Social Security Number (SSN).[43] This staff works in a secure physical environment and utilizes a highly restricted computing cluster that is behind the Census Bureau firewall and is only accessible to them. It is made up of sworn Census Bureau employees who must complete a moderate background investigation for Public Trust positions, which includes a personal interview and submission of a set of fingerprints to the Federal Bureau of Investigation.

First, this staff confirms that the received files are as described in the IAA, with no omissions and no additional variables included, since the Census Bureau is never permitted to receive more than what has been specified in an applicable agreement. If needed, an IAA could be modified to accommodate any additional variables. This staff also confirms that variables are not blank and that provided values and ranges are properly documented in either the agreement or layouts provided with the data.

Once files are accepted and basic QC has been performed, the staff replaces the explicit identifiers with a unique Protected Identification Key (PIK) that can be used to link the records to other databases held at the Census Bureau. This linkage is usually achieved by using a combination of variables including first and last name, SSN, date of birth, and place of residence. These variables are used to conduct a linkage between the incoming file and a "reference file" composed of records from censuses, surveys, and Federal-agency files. This reference file contains PII from the records the Census Bureau already holds and the PIK that corresponds to each record in the reference file. When a linkage can be made between the incoming file and the reference file, the PIK is appended to the incoming file. When this process is complete, the final "research version" of the file does not contain direct PII and instead contains "PIK-ed" records.

After files are accepted and PII has been removed, the "research version" of the microdata files are registered in the Census Bureau's centralized Data Management System. Each Data Management System entry indicates that a file is available for request by research projects and Census Bureau programs. A Data Management System record contains basic file metadata, file usage restrictions, and information about approvals required to use the file. These files are still confidential and protected by Title 13, though they no longer contain the PII variables from the original file. The replacement of PII with a PIK is one step that the Census Bureau takes to centralize linkage services and to limit the amount of PII that is available to downstream users within the agency, including researchers at local Federal Statistical Research Data Centers (FSRDCs, described in further detail below).

Monitoring Projects: Data Access, Usage, and Project Outputs

As the data lifecycle moves beyond file preparation and provisioning, key safeguards come into play regarding data access and data uses, which must be consistent both with a project's objective and any

[42] The Census Bureau acquires administrative data about businesses and individuals. This discussion focuses on the processes for files containing PII, although many of the steps are similar for business data.

[43] While out of scope for this paper, data curation is a critically important step of the process that is complementary with data protection.

underlying conditions set forth in an IAA. The final stage of the project is producing the outputs, which are reviewed to protect individual persons and establishment confidentiality.

Administrative records acquired by the Census Bureau become protected by the confidentiality protections of Title 13, just as survey data. Only Census Bureau employees or individuals with Special Sworn Status (SSS) may access such confidential data. SSS individuals are sworn to protect the data as Census Bureau employees are sworn. Title 13 permits the agency to swear in individuals to conduct work that specifically benefits a Census Bureau program,[44] and SSS individuals are subject to the same legal requirements and penalties as Census Bureau staff. In fact, SSS is the mechanism by which researchers are able to access data in the FSRDCs. Employees and SSS individuals must take annual Data Stewardship Training, Title 26 (Federal Tax Returns and Return Information) training, and any additional trainings required by the agreement associated with the particular administrative records being used.

To gain approval to conduct a project, a named Project Contact must write a proposal that indicates the project's methodology and objectives, anticipated output, benefit to the Census Bureau, and datasets required (including any that need to be acquired for the current project). The proposal must also include logistical details about where the work will take place, over what period of time, and the contact information for supporting staff who will participate.

Proposals to conduct research under Census Bureau authority in the FSRDCs are reviewed for their benefit to Census Bureau programs, scientific merit, feasibility, and potential risk of disclosure. The approvals process is managed by the Census Bureau's Policy Coordination Office and tracked within the Data Management System. Reviews are conducted by representatives of the Policy Coordination Office, Census Bureau subject matter experts, an Executive-level Division Chief, a local Census Bureau Information Owner who is responsible for knowing restrictions and terms of use, and (when indicated by an IAA) the agency that provided the administrative records being requested.

When a research project is approved, a data access team confirms that all researchers have completed required trainings, including any required by data supplying agencies, and then provides access within a secure computing environment. Users can access read-only versions of approved files, and they conduct all work within a project-level workspace that is shared by the approved researchers. All analysis of microdata takes place within this computing environment.

At the final stage, when researchers need to remove aggregated output, tables, or model coefficients from the FSRDC, they contact a Disclosure Avoidance Officer (DAO). The DAO follows protocols established by the Census Bureau's Disclosure Review Board in order to ensure that no individual information is revealed and that output is consistent with the original proposal. If the results pass DAO review, they are provided to the research team outside of the secure computing environment (usually via email). The research team can then produce reports, presentations, and other products outside of the secure environment. Depending on the terms of any relevant IAA, these products also may often go through a final review by the data-providing agency or Census Bureau subject matter expert to confirm that final products were consistent with the original proposed project.

Historically, the national network of FSRDCs has been managed by the Census Bureau to allow researchers secure access to restricted Census Bureau data. More recently, the FSRDCs have been expanded to also include restricted health data from the National Center for Health Statistics and the Agency for Healthcare Research and Quality. As the program has grown, the Census Bureau and the

[44] 13 U.S.C. § 23(c).

Interagency Council on Statistical Policy (ICSP)[45] recognized the potential benefits of a shared governance model and are collaborating to design this new model.[46]

IV. Discussion: Data Protections in the Context of Researcher Access

Federal evidence-building activities compile, analyze, and disseminate information in order to build evidence. This includes describing population characteristics and trends, producing and disseminating performance information, and facilitating and conducting research and evaluation. Significantly, this often means partnering with academic institutions, researchers, and other organizations.

Many agencies make public-use files available for evidence-building purposes, which do not contain PII. Generally, these files either contain information aggregated at a sufficient level to prevent re-identification of an individual or they contain synthetic individual-level data. Given the importance of personal identifiers for matching individual-level data for many evidence-building purposes, this paper focuses on the privacy and confidentiality considerations associated with restricted-use files.[47]

The principal statistical agencies and a select number of other Federal evidence-builders have established processes and practices that permit qualified researchers to access restricted data under highly controlled conditions that satisfy the dual concerns of data access and protecting the confidentiality of data about individuals and businesses. In most instances, these types of access are limited to qualified academic or other non-profit researchers, or qualified contractors acting on a Federal agency's behalf. In some cases, this access also includes Federal agency staff accessing data owned and maintained by another Federal agency. Agencies exercise varying levels of review of the scientific merit and final output from the projects. There are three primary models by which an agency conducting or supporting evidence-building activities provides controlled access to researchers outside of that agency:

- *Online data query systems* are analysis tools that allow the public to examine restricted-use data dynamically, creating tables, rates, and/or models. For example, the Bureau of Justice Statistics has created the Corrections Statistical Analysis Tool that allows the public to examine national prisoner statistics on inmates under the jurisdiction of Federal and state correctional authorities online.[48]
- *On-site access* allows eligible researchers the opportunity to gain access to restricted-use data for select research projects at the agency. Generally, interested researchers submit a project proposal that, if approved, allows them to conduct work with on-site microdata at little to no cost to them or their institution or organization. For example, the Bureau of Labor Statistics has an on-site visiting researcher program that allows approved researchers on-site access to confidential microdata.[49]

[45] The ICSP is made up of the heads of the 13 principal statistical agencies spanning nine cabinet departments and two other agencies, as well as representatives from other evidence-building programs. It provides advice and counsel to OMB on relevant statistical matters and is a primary vehicle for coordinating cross-cutting statistical work and information exchange about agency programs and activities.

[46] For more information on FSDRCs, see U.S. Census Bureau, http://www.census.gov/fsrdc.

[47] Note that as the amount of microdata available from governmental and non-governmental sources grows, and as the provision of data access through public- and restricted-use means are scaled up, the systems used to safeguard privacy may require increasingly technical and sophisticated tools and the training and resources needed to employ them.

[48] See Bureau of Justice Statistics. *Corrections Statistical Analysis Tool (CSAT) - Prisoners*. http://www.bjs.gov/index.cfm?ty=nps.

[49] See Bureau of Labor Statistics. *On-site Visiting Researcher Program: Access to Confidential Data Files at the Bureau of Labor Statistics.* http://www.bls.gov/bls/blsresda.htm.

- *Data enclaves* are secure environments in which qualified researchers may access restricted-use microdata for statistical purposes.
    - Data enclaves can be *physical* locations that are operated by an onsite Federal employee who administers a data laboratory and serves as a resource for visiting researchers. Physical enclaves often utilize a "client-server" environment, in which an agency hosts data that researchers securely access from an approved location through a "thin client" terminal.[50] For example, the FSRDC system administered by the Census Bureau and used by it and several other agencies is a physical enclave system. The Research Data Center for the Department of Health and Human Services, which is coordinated through the National Center for Health Statistics, is also a physical enclave.[51]
    - Data enclaves can also be *virtual*, existing outside of a Federally-run brick-and-mortar location but subject to similar controls. Virtual enclaves use either a "client-server" environment, a VPN, or sometimes allow qualified researchers to host data on computers at their home institutions under strict conditions. For example, the Centers for Medicare & Medicaid Services has created the Chronic Conditions Data Warehouse Virtual Research Data Center to provide qualified researchers with secure, online access to Medicare and Medicaid beneficiary, claims, and assessment data.[52] As described in the next section, NCES operates a virtual data enclave system through a data licensing structure, in which qualified researchers agree to a set of controls designed to create an environment similar to a physical data enclave.

V. Discussion: Department of Education's National Center for Education Statistics Restricted-Use License Program

Title 20 of the U.S. Code directs NCES, within ED's Institute of Education Sciences (IES), to collect, report, analyze, and disseminate statistical data related to education in the United States and in other nations.[53] It is this authority that enables NCES to conduct research in support of ED's broader mission to improve education in the United States.[54] The research is congressionally mandated and funded as part of ED's annual appropriations. NCES studies provide official national education statistics and underlying data used by policy-makers within ED, across other agencies, by international organizations, and by researchers.

Like other agencies involved in Federal evidence-building efforts, NCES provides a mechanism for researchers to access individual-level data when needed to answer their research questions. As described above, NCES uses a variation of the enclave model called a licensing program. [55] Licenses are contractual arrangements between an agency and researchers to make data available for statistical, research, or evaluation purposes to qualified researchers. Research and analysis that results from a licensing agreement must be consistent with the statistical, research, or evaluation purposes for which

[50] A thin client is a lightweight computer that is purpose-built for remote access into a server. It depends heavily on another computer (its server) to fulfill its computational roles. Thin client hardware normally requires only a keyboard, a monitor, and a network connection to access programs and resources. Dedicated thin clients usually do not have hard disk drives, CD-ROM drives, floppy disk drives, or expansion slots. This results in both lower cost and increased security.

[51] See Centers for Disease Control and Prevention. NCHS Research Data Center (RDC). http://www.cdc.gov/rdc/.

[52] See Chronic Conditions Date Warehouse. https://www.ccwdata.org/web/guest/home.

[53] 20 U.S.C. § 9543(a).

[54] *Ibid.*

[55] 20 U.S.C. § 9573(c) and (d).

the data were provided or are maintained, and the data must be used and protected in accordance with the terms and conditions of the license.

The NCES restricted-use data license program has been in place for over 25 years. The relevant system of records notices (SORNs) include notification to the public of ED's intent to allow qualified researchers access to restricted-use data to carry out specific research related to the purposes for which the information is collected.[56] In addition, the individually-identifiable information about the licensees are identified as an official system of records to allow NCES to maintain the information necessary for internal control and monitoring of restricted-use data licensees.[57] Licensing procedures and requirements have been updated since the program's inception in response to new laws and changes in existing laws. The current licensing process is bounded by and compliant with requirements in the Privacy Act, FISMA, the Education Sciences Reform Act of 2002 (ESRA) as amended,[58] and the E-Government Act (apart from FISMA).[59] Details about the program, including the online application system, are available on the NCES restricted-use data license program website at http://nces.ed.gov/statprog/instruct.asp.

NCES loans restricted-use data only to qualified organizations in the United States. Individual researchers must apply through an organization (e.g., a university, a research institution, or company). The process begins when an interested researcher submits an application for a data license and provides information about themselves and their organization; the proposed research project, including a description of the research objectives and the intended audience for the research, an explanation of why public-use data are not suitable, information on any proposed data linkages, and an indication of how long access to data will be needed; the proposed security plan, including physical location of the data, computer system information, security system information, file access management, and end of project procedures; and an agreement to a series of procedures and other data protections. NCES assists interested researchers through the application process by posting a manual on restricted-use data licensing procedures, an online application process, FAQs, and other information, including contact information to respond to individual questions.

A data license for restricted-use data is approved by NCES when important policy-relevant research questions can only be answered by using PII. In order for NCES to release data, the researcher and their organization must also certify that:

- Adequate safeguards are in place to protect against violations of law and policy and to prevent the unauthorized disclosure of confidential data,
- The researcher and their organization will abide by and comply with all of the terms of the restricted-use data license,
- The researcher and any authorized data users on the license will complete annual privacy and data security training, and
- The researcher and their organization agrees to follow the data destruction protocol.

The approvals process is managed by the Information Security unit within the NCES Statistical Standards and Data Confidentiality Staff and tracked within the electronic licensing system. Each data license application is reviewed by a team of information security staff, content and technical experts, the Chief Statistician, and the NCES Commissioner to assess the aforementioned criteria and to determine

[56] Federal Register Notices 18-13-01 and 18-13-03 can be retrieved at http://www2.ed.gov/notices/ed-pia.html#ies.

[57] Federal Register Notice 18-13-02 can be retrieved at http://www2.ed.gov/notices/ed-pia.html#ies.

[58] 20 U.S.C. § 3419.

[59] With the enactment of ESRA, the NCES confidentiality provisions that support the NCES restricted-use data licensing program were expanded to cover the statistics, research, and evaluation centers within the newly formed Institute of Education Sciences. As a result, the licensing program was expanded to include all IES centers.

whether the proposed project can be accomplished without PII and whether all requested data are necessary to complete the work. When proposed safeguards for privacy protection and protection against disclosure of confidential information are deemed insufficient, the NCES information security staff will recommend additional or alternative safeguards to the applicant's data security plan.

Following approval of the license application, security plan, and receipt of the signed license agreement and notarized affidavits of nondisclosure, NCES security staff send an encrypted CD-ROM to the Principal Project Officer via certified U.S. mail with a signature required for receipt. Upon receipt of the encrypted data, the licensee must contact the NCES security office for the password required to access the data. In the case that data security is breached for any reason, license holders are required to report the breach to NCES immediately. NCES has the right to make unannounced and unscheduled inspections of the licensee's facility to evaluate compliance with the terms of the license.

Licensees may publish the analysis and results derived from use of the licensed data, but the results may only be presented in summary form that does not disclose the identity of individuals either directly or by inference. NCES provides guidelines for how to avoid disclosure when publishing summary statistical information. To ensure compliance, NCES requires a licensee to submit for review in advance of publication all articles, reports, and statistical summary tables generated from licensed data. NCES' data stewardship practices in establishing data security standards, project and data use review processes, and output reviews are further demonstration of the Federal Government's commitment to protecting information and ensuring the utility of the underlying data.

VI. Summary of Themes

Federal agencies involved in building evidence are taking important steps to extend the utility of data they maintain, while helping to ensure fundamental protections for privacy and confidentiality are addressed throughout the data lifecycle. These protections are provided by a number of laws, such as the Privacy Act, E-Government Act, and CIPSEA, as well as agency- and sector-specific statutes. These laws and the requirements they represent are integral to data stewardship and assuring the trust of those individuals and businesses whose information the datasets contain. Many agencies, such as the Census Bureau and the other principal statistical agencies, have laid important groundwork in their approach to data stewardship, which is based on a strong set of laws and policies supporting statistical activities. They have also been working together to expand the FSRDC program, which is intended to provide a better mechanism for evidence-building agencies to work with researchers. The FSRDC program is an important opportunity to not only further extend the utility of Federal data but to also achieve greater consistency and safeguarding of the data made available to researchers. Finally, as the discussion of the various models by which agencies provide restricted access to data demonstrates, additional agencies build evidence and work with researchers in ways that both protect information and also extend the utility of the underlying data in support of agency missions.