

# Handbook For Administrative Data Projects

---

Policy Coordination Office

U.S. Census Bureau

Issued: June 2018

## Table of Contents

CHAPTER 1. Introduction.....	1
Purpose: handbook and policy companion .....	1
Definitions: administrative data & administrative data projects.....	1
Scope.....	2
Administrative data projects that are considered in-scope for this policy are those that: .....	2
Key roles and contacts in the administrative data project lifecycle.....	3
Chapter 2. Anatomy of an Administrative Data Project .....	7
Initial development of a project idea.....	7
Formal project proposal development and submission.....	7
Project proposal review.....	8
Agreement negotiation and acquisition of administrative data.....	9
Project proposal approval .....	9
Data access and project/agreement administration .....	10
Chapter 3. Roles and Responsibilities.....	12
Program areas and staff members .....	12
Governing bodies .....	13
Chapter 4. Internal Project Lifecycle & Project Contact Responsibilities .....	16
Project types.....	16
Project development.....	17
Project review and approval.....	18
Subprojects.....	20
Amendments .....	21
Project implementation and administration .....	21
Flow of events – internal research project proposal review and approval.....	23
Chapter 5. External Project Lifecycle & Principal Investigator Responsibilities .....	24
Project types and description of projects .....	24
Project development.....	24
Project review and approval.....	25
Project implementation and administration .....	26

Flow of events – external research project proposal review and approval .....	28
Chapter 6. Using Agreements to Bring in Administrative Data .....	29
General information .....	29
Negotiation.....	29
Using an agreement to acquire data .....	30
Data access and agreement administration .....	33
Chapter 7. Administrative Data and Confidentiality Laws.....	35
Background.....	35
General guidelines for determining confidentiality status .....	35
Specific guidelines for data sets obtained under an agreement for services .....	36
Making a determination .....	38
Resolution .....	38
Chapter 8. Security Protocol .....	39
Security Controls for Information Systems That Process Sensitive Data and Administratively Restricted Data .....	39
System Security Officers, System Owners and Authorizing Officials .....	39
Telecommunications .....	40
Auditing .....	40
Labeling of Data Displays .....	40
Electronic Storage.....	40

## APPENDICES

- Appendix A. Glossary
- Appendix B. Example of the instructions for conducting an Annual Review
- Appendix C. Checklist of Responsibilities for Project Contacts
- Appendix D. FSRDC Research Proposal Guidelines
- Appendix E. Template: Predominant Purpose Statement
- Appendix F. Examples: Predominant Purpose Statements

## FIGURES

- Figure 1. Key Roles & Contacts in the Administrative Data Projects Lifecycle
- Figure 2. Governing Bodies and Roles For Providing Guidance Regarding Administrative Data Use

## CHAPTER 1. Introduction

### Purpose: handbook and policy companion

The Census Bureau’s administrative data infrastructure supports a wide variety of our programs for which data from outside sources are essential. This handbook is a companion to the *Administrative Data Acquisition, Access, and Use Policy*, #DS001a, (issued October 2016).<sup>1</sup> This policy establishes a mandatory business process for all staff and special sworn status (SSS) associates of the Census Bureau who conduct projects that use administrative data. Staff and SSS associates should consider the contents of the handbook an extension of the policy; it sets forth procedures for access to and the acquisition, custodianship, and use of administrative data. The procedures include a project review process required for all project proposals involving administrative records.

This handbook instructs program areas on safeguarding confidential administrative data that the Census Bureau receives from outside sources. Together with the companion policy, these formalized processes and procedures specify authorized uses and users of the data and prevent unauthorized disclosures.

Compliance with these policies and procedures is especially important when the provider of the administrative data imposes restrictions on the data’s use and/or protection. However, these procedures are also relevant for the use of administrative data sets that are without external restrictions or protections because those projects must also be reviewed, approved, and tracked in the Census Bureau’s Electronic Project Tracking System for administrative datasets and projects.<sup>2</sup>

### Definitions: administrative data & administrative data projects

**Administrative data.** Administrative data refer to microdata records contained in files collected and maintained by administrative or program agencies and commercial entities. Government and commercial entities maintain these files for the purpose of administering programs and providing services. Administrative data are distinct from systems of information collected exclusively for statistical purposes, such as data from censuses and surveys that are produced under the authority of Title 13 of the United States Code. The Census Bureau primarily draws upon administrative data developed by federal agencies. However, we also use information from state, local, and tribal governments, as well as commercial entities. Data we acquire from commercial entities are often dubbed “third party” data. Administrative data may also be culled from public sources.

Prominent sources of administrative data used by the Census Bureau are the Internal Revenue Service (IRS), Social Security Administration (SSA), Centers for Medicare and Medicaid Services (CMS), United States Postal Service (USPS), Bureau of Labor Statistics (BLS), Housing and Urban Development (HUD), and state data that feeds into the Census Bureau’s Longitudinal Employer Household Dynamics (LEHD) program.

The Census Bureau uses several types of administrative data information, including information about people, households, businesses, and address data. Examples of these types of information include:

---

<sup>1</sup> Policies preceded by a “DS” indicate that the policy originated from the Data Stewardship Executive Policy (DSEP) Committee.

<sup>2</sup> Currently, the Data Management System (DMS) is the software solution the Census Bureau uses to track administrative data projects.

- People & households – personal identifiers such as social security numbers (SSNs) and income
- Businesses – identifying information that describes the operating, financial, ownership, and other characteristics of a business
- Address data – identifying information such as latitude/longitude points, telephone numbers, and address information

The Census Bureau also creates and uses “commingled” data, which is Census Bureau data that has been enhanced with administrative data. The administrative data typically used in commingled data is Federal Tax Information (FTI) that we receive from the Internal Revenue Service (IRS). For example, the Business Register is considered commingled data because it is the sampling frame of businesses maintained by Census that includes Federal Tax Returns and Return Information (i.e., FTI).

**Administrative data projects.** The Census Bureau tracks, manages, and reports on its administrative data activities both internally and externally at the “project” level. An administrative data project is a project that makes use of administrative data. An example of such a project is the “Demonstration of Administrative Records Improving Surveys.”

In general, a project is a set of activities that has a distinct mission and clear starting and ending points. Projects have a life cycle that proceeds from conception to planning, execution, and termination; however, some projects are extended or repeat indefinitely (e.g., surveys). There is generally a single point of contact for each project. A project is often part of a broader program and is a building block in the design and execution of program goals. Dependent relationships may exist between projects.

## Scope

Projects that use administrative data must adhere to this policy and the processes outlined in the *Administrative Data Projects Handbook*. The *Administrative Data Projects Handbook* covers all administrative data projects and datasets maintained by the Census Bureau, and the *Data Access and Transfer Policy* mandates that all datasets used for projects must be registered and tracked by the Census Bureau’s Electronic Project Tracking System. Tracking data sets in a centralized electronic tracking system, such as the DMS, ensures proper use and reporting and enhances the coordination of corporate information sharing across the Census Bureau.

When the Census Bureau collects information from individuals or businesses, that information is usually subject to Title 13 confidentiality restrictions. All Census Bureau employees and individuals with SSS take an oath to uphold the confidentiality of information that they access. However, administrative data we receive from external sources often comes with additional restrictions regarding access and use that go beyond the provisions of Title 13. In addition to Title 13 protections, these files are often protected by the data supplier’s own statutes or regulations, or by language in an agreement with the data supplier. The Census Bureau must meet all legal obligations imposed by the external providers of administrative data.

A significant example of data protected by Title 13 *and* a data supplier’s statute is all FTI, whether received from the IRS directly or from the SSA. Additional examples include the InfoUSA File or vital records from the National Center for Health Statistics.

Administrative data projects that are considered in-scope for this policy are those that:

- Plan to use administrative data files that are protected by Title 13, Sections 8 and 9 and/or carry constraints imposed by the data-provider; AND
- Plan to use administrative data files containing information about individuals or businesses; AND

- Involve internal, external reimbursable, or joint proposals; AND
- Involve research or production.

Furthermore, administrative data projects that are considered in-scope for this policy can include:

- New or novel activities OR previously-established, routine activities that involve significant modification to existing data (e.g., more detailed categories); OR
- Routine, vetted, ongoing production activities in support of regular, Census Bureau programs that use administrative data that carry provider-related constraints (e.g., the Business Register<sup>3</sup>).

### **Key roles and contacts in the administrative data project lifecycle**

Figure 1, on the next page, provides an overview of the Census Bureau's Administrative Projects Lifecycle, which includes the role of various program areas involved in, and responsible for, bringing administrative data projects on-line.

---

<sup>3</sup> The Business Register is comprised, in part, of FTI which carries IRS-related constraints on data use.

**Figure 1. Key Roles & Contacts in the Administrative Data Projects Lifecycle**

Area/Title	Project Development	Project Review & Approval	Project Implementation & Administration	Administrative Data Acquisition, Access & Agreement Administration	Policy & Guidance
Policy Coordination Office (DIR/PCO)					
Division Chief					Policy guidance
Administrative Records Coordinator	Advises and consults with program areas on approaching and contracting with providing agencies	Project Review Team member		Leads acquisition strategy meetings for data currently external to Census; guides acquisitions through the SharePoint Inter-agency and Other Special Agreement (IOSA) process	Policy interpretation/guidance
Center for Administrative Records Research and Applications (CARRA)					
Center Chief			De-identifies and links administrative data files, preparing the files for use in statistical research	Negotiates with major agencies for the acquisition of administrative data (along with ECON, CES, and GEO)	
Associate Director for Research and Methodology (ADRM) – Planning, Coordination, and Outreach (ADRM/PCO)					
Associate Director					
Research Project Coordinator & Project Review Coordinators		Responsible for coordinating reviews for internal and external projects			
Other Divisions/Areas/Committees					
Disclosure Avoidance Officers (DAO); one DAO is assigned for each Division	A DAO can bring your project proposal to the DRB for a feasibility discussion		Consulted for proper application of disclosure avoidance techniques to statistical results; bring results to the DRB for formal review and approval		Responsible for the application of proper disclosure avoidance techniques to statistical results and maintaining and disseminating standards of disclosure techniques and rules

<b>Area/Title</b>	<b>Project Development</b>	<b>Project Review &amp; Approval</b>	<b>Project Implementation &amp; Administration</b>	<b>Administrative Data Acquisition, Access &amp; Agreement Administration</b>	<b>Policy &amp; Guidance</b>
Information Owner (IO)	Answers questions about the datasets for which they are responsible	Project Review Team member; reviews and approves requests to access data		IO must consult the agreement(s) associated with the dataset(s) and decide whether to approve the use(s) of the dataset(s) for which they are responsible.	
Data Stewardship Executive Policy (DSEP) Committee		High- or moderate-risk projects may need to seek clearance/ approval from this committee			Policy-setting body on issues related to privacy, security, confidentiality, and administrative records.
Data Integration Policy Committee (DIPC)	Project Contacts present issues associated with research proposals to this group for feedback/ suggestions				Sub-committee of the DSEP, conducts policy and procedural analyses on issues relating to the use of administrative records and data linkage activities, vets questions about administrative data projects
Disclosure Review Board (DRB)	Project Contacts present their project proposal to this committee for feedback on feasibility and whether results could be released				Sub-committee of the DSEP, conducts a disclosure review for all statistical products for release or publication, vets feasibility questions regarding administrative data projects
Office of Information Security		Project Review Team member			Provides IT policy guidance

Chapter 2 provides a general description of the lifecycle of an administrative data project.

Chapter 3 defines the roles and responsibilities of both staff and the bodies that govern administrative data project activities, policies, and procedures.

Chapters 4 and 5 provide in-depth explanations of the project lifecycle; Chapter 4 describes the lifecycle of an *internal* administrative data project (generally carried out by staff at the Census Bureau's Headquarters). Chapter 4 also points readers to a Checklist for Project Contacts (see Appendix C) to assist them in complying with the policies, processes, and procedures described in this handbook. Chapter 5 describes the lifecycle for *external* projects (generally undertaken by a researcher affiliated with a Federal Statistical Research Data Center) and points readers to Proposal Guidelines (see Appendix D) that describes the research proposal process in detail for external researchers.

Chapter 6 describes the process of obtaining administrative data sets for use in a project, in the event that the Census Bureau does not already have access to the required data.

Chapter 7 describes the confidentiality laws that apply to administrative data.

Finally, Chapter 8 outlines the physical and information technology (IT) security controls that protect all the administrative data files at the Census Bureau.

## Chapter 2. Anatomy of an Administrative Data Project

In general, all projects that use administrative data undergo a rigorous vetting process and are required to follow the guidance set forth by the *Administrative Data Acquisition, Access, and Use Policy* and this handbook.

Research is undertaken on behalf of the Census Bureau both at Headquarters in Suitland, Maryland and at approximately twenty Federal Statistical Research Data Centers (FSRDCs) across the country. Administrative data projects may originate from researchers at any of these locations. Researchers are either Census Bureau employees, or they have qualified for special sworn status (SSS), meaning they have completed a background check, training, and have agreed to a code of conduct that subjects them to the same legal penalties for disclosing data protected by law as are Census Bureau employees.

Because administrative data project proposals originate from one of two of these particular research settings, the vetting process for each type is slightly different. These differences are outlined in detail in Chapters 4 and 5. However, each vetting process shares commonalities that are discussed below.

### Initial development of a project idea

In the initial phase of a project, research questions are developed by a Project Contact (i.e., the researcher) and the data types available at the Census Bureau are explored. At this step of the project, staff (and SSS individuals, as appropriate) are required to:

- Generate and begin documenting research questions and ideas for research that support the Census Bureau's mission.
- Determine whether administrative dataset(s) of interest is already held by Census, needs to be acquired, and can be used for such purpose(s).<sup>4</sup> Contact the Center for Administrative Records Research and Applications (CARRA), the Economic Area's International Trade Management Division's (ITMD) Data Management and Stewardship Staff (DMSS), the Geography Division, or the Policy Coordination Office's (PCO/DIR) Administrative Records Coordinator for more information or assistance in identifying appropriate administrative datasets for a particular project.
- If the data are not already in-house, then the Project Contact can begin informal discussions with the data provider. The process of negotiating to bring in additional data, and the formal agreements they entail, is discussed in detail in Chapter 6.

### Formal project proposal development and submission

At this step of the project, staff (and SSS individuals, as appropriate) are required to:

- Develop a new administrative project proposal that supports the Census Bureau's mission. In limited circumstances, work may be carried out under an existing, approved project, called a "sub project" (for more information, see Chapter 4, Sub Project section).

---

<sup>4</sup> Each administrative dataset has an Information Owner. Information Owners are required to safeguard datasets by restricting access to those staff members authorized to use them, which involves determining whether the proposed uses for the data are consistent with previously approved uses outlined in the agreement(s) and that the proposed uses are covered by a System of Records Notice (SORN) associated with the dataset. See more about the data stewardship requirements for Information Owners in the *Safeguarding and Managing Information Policy* (#DS007).

- Identify the administrative dataset(s) needed for the research by browsing the datasets registered in the Census Bureau’s Electronic Project Tracking System (i.e., the DMS).<sup>5</sup> Project proposals must identify and document the administrative datasets they wish to request and access.<sup>6</sup>
- Agree to use only demographic data that has received one-time pad or FIPS encrypted identifiers such as the Protected Identification Key (PIK). DSEP must approve any project requesting direct access to PII (i.e., names of individuals and SSNs).
- Agree to use only business data that has been identified numerically (e.g., EIN, Census File Number, Census Alpha, LBDnum, SEIN, Census geography, etc.).
- Begin informal discussions with the data provider(s) to acquire new administrative data (if administrative data are unavailable in the DMS catalog). However, formal acquisition processes must be followed to acquire the data. Researchers must work with the Policy Coordination Office (PCO/DIR) to carry out a formal agreement process. For more detail, see the *Agreement Negotiation and Administrative Data Acquisition* section below, or Chapter 6.
- Upload the finalized project proposal and all related documentation to the Census Bureau’s Electronic Project Tracking System to begin the review and approval process.<sup>7</sup> If the project plans to link datasets and is in-scope for the [Record Linkage Policy \(DS014\)](#), the documentation must include a [Record Linkage Checklist \(DS014\)](#).

## Project proposal review

At this step of the project, staff (and SSS individuals, as appropriate) are required to:

- Undergo a formal review and approval process for each administrative data project proposal.<sup>8</sup> The review ensures projects have scientific merit, a purpose that furthers the Census Bureau’s mission, sufficient oversight from the DSEP Committee, ample confidentiality and privacy controls, and are in compliance with Title 13, U.S.C. and existing data agreements. The review and approval process is tracked by the Census Bureau’s Electronic Project Tracking System.
- Present the proposal to the DSEP Committee, prior to project approval, if required. Proposals subjected to the approval of DSEP typically involve the linkage of one mandatory collection to another mandatory collection, or involve the use of names and/or social security numbers as part of the data linkage process.
- Be responsive at each stage of the review process (e.g., provide additional information, revise and resubmit to the Census Bureau’s Electronic Project Tracking System updated versions of the project proposal, etc.).

---

<sup>5</sup> The DMS is a software system that catalogs, documents, tracks, and controls access to all datasets maintained by the Census Bureau. Registration of datasets in the DMS became mandatory in January 2014.

<sup>6</sup> Once the data sets are identified, selected, and added to the project proposal documentation in the DMS. The DMS will alert the Information Owners (IOs) that the data sets for which they are responsible have been requested. The IOs will review and accept or deny the request.

<sup>7</sup> Currently, Census uses the DMS to electronically track the review/approval/vetting process for administrative data project proposals. This system is also used to archive all documents, decisions, approvals, and results related to these proposals and projects. The DMS Users Guide can be found online at <https://dms-web3.asd.census.gov:8085/datamanager/login?VM=false>.

<sup>8</sup> Greater detail is provided in Chapters 4 and 5 for the review processes specific to internal and external projects, respectively. The Census Bureau maintains minimum standards for project proposal reviews that must be undertaken for internal projects (see the *Research Proposal Review Process Policy*, #DS24 for details).

## Agreement negotiation and acquisition of administrative data

At this step of the project, staff and SSS individuals are required to:

- Contact the Administrative Records Coordinator (PCO/DIR) to arrange for the acquisition of new administrative data; this area of the Census Bureau oversees the review and approval of data agreements. Agreements are reviewed by the Census Bureau's Policy, Legal, and IT and OIS Security Offices. See the Census Bureau's *Policies & Procedures Manual* to determine the appropriate level of signatory approval (e.g., Director, Deputy Director, Assistant Director, etc.) required for various contract amounts:

[https://intranet.ecm.census.gov/apps/policyportal/Policy%20Repository/PPM\\_B1.pdf](https://intranet.ecm.census.gov/apps/policyportal/Policy%20Repository/PPM_B1.pdf)

- Contact the Administrative Records Coordinator (PCO/DIR) to arrange for the acquisition of new administrative data; this area can help coordinate a new data request with potentially long-standing agreements that are active in other parts of the Census Bureau.
- Use a standardized Census Bureau Inter-Agency or Other Special Agreement (IOSA) template when developing new agreements. These [templates](#) are available on PCO's/DIR website.
- Document the terms of the agreement by negotiating requirements for the use and exchange of the desired data with the data provider. Agreements must describe the authorized uses of the data, the legal authority under which the data are collected and protected, access restrictions, retention period, and disposition of data at the end of the retention period. Any and all uses of administrative data at the Census bureau must be consistent with the approved uses stated in the relevant agreement.
- Upload the agreement to the Interagency and Other Special Agreements (IOSA) Review and Approval System.<sup>9</sup> The Administrative Records Coordinator in the Policy Coordination Office (PCO/DIR) can provide assistance.
- Track the progress of the agreement in the IOSA Review and Approval System and be responsive to any additional inquiries resulting from the review and approval of the agreement.
- Register the administrative data, once it is received, in the Census Bureau's Electronic Project Tracking System (i.e., the DMS).

## Project proposal approval

At this step of the project, staff (and SSS individuals, as appropriate) are required to:

- Wait for the Project Review Coordinator to obtain approval to use administrative data for the proposed research from various stakeholders at the Census Bureau (i.e., Information Owners and reviewers). Approvals may also be required from other Federal agencies who supplied the data. For example, if a project uses FTI, researchers must obtain approval from the Census Bureau and the IRS before the FTI can be accessed and used.
- Wait until approval from all stakeholders is granted before beginning work, unless a particular phase of the project has been approved and partial work can begin that involves only the data for which approval has been granted (called a "Phased Approval").
- Undergo an additional formal review if the project was previously approved and new administrative data are sought, or if there are significant modifications to the project/analysis.

---

<sup>9</sup> The IOSA Review and Approval System is now in SharePoint.

## Data access and project/agreement administration

At this step of the project, staff (and SSS individuals, as appropriate) are required to:

- Access administrative data with the appropriate one-time pad or FIPS encrypted (demographic) or numeric (business) identifiers for approved project uses through designated, secure Census Bureau channels (e.g., through a Federal Statistical Research Data Center, Virtual Desktop Interface, or Census Bureau-controlled computers).
- Comply with the restrictions set forth by the providing agency while using the data.
- Abide by all data security and confidentiality requirements.
- Maintain a current list of Census Bureau staff and/or SSS individuals assigned to a particular administrative data project. This means updating the Census Bureau's Electronic Project Tracking System immediately, in real time, with an accurate accounting of staff still on the project (i.e., removing people who are no longer working on the project and adding new people who plan to work on the project).
- Take an oath to protect all data and undergo a background investigation.<sup>10</sup>
- Complete the mandatory, annual "Data Stewardship and IT Security Awareness" and "Title 26 Awareness" training by the June 30<sup>th</sup> deadline designated by the Policy Coordination Office.
- Conduct an annual review of each project and document this review in the Census Bureau's Electronic Project Tracking System.
- Maintain accurate server name and information for each project in the Census Bureau's Electronic Project Tracking System.
- Abide by disclosure avoidance procedures and submit any statistical products resulting from the research, prior to publication/release or sharing beyond those working on the project, to the Census Bureau's Disclosure Review Board (DRB), or another appropriate DRB (e.g., SSA's DRB). A DRB review is required for statistical products based in whole or in part on administrative data. Statistical products using administrative data might also be subject to additional disclosure review as required by the providing agency.
- Upload the memo documenting the request for disclosure avoidance review to the Census Bureau's Electronic Project Tracking System as part of the project's documentation. The Center for Disclosure Avoidance Research and the DRB will document approval/denial in the Census Bureau's Electronic Project Tracking System.
- Post non-confidential results (or a non-confidential summary of the results) from the research in the Census Bureau's Electronic Project Tracking System.
- Implement project close-out procedures by setting the project to "Completed" or "Terminated" in the Census Bureau's Electronic Project Tracking System.
- At the Project Contact's request, internal and external projects in "Completed" status may be reopened to allow the researchers to respond to external peer-review of the research. The Census Bureau recognizes that Title 13 benefits are not fully realized until the research has been reviewed and published. The integrity of research done under Census Bureau auspices depends upon the confidence of the scientific community in our adherence to the principle of peer review. Every

---

<sup>10</sup> This is directed at SSS individuals listed on any project; it is assumed that Census Bureau staff members have completed these steps as a condition of employment.

effort should be made to allow for timely response to peer review, consistent with the requirements of Title 13 and agreements with external data providers.

## Chapter 3. Roles and Responsibilities

### Program areas and staff members

A variety of staff is involved over the lifecycle of an administrative data project. This section briefly describes the roles and responsibilities of those involved. Chapter 4 provides a greater level of detail regarding a Project Contact's responsibilities.

**Administrative Records Coordination Branch, in the Policy Coordination Office (PCO/DIR)** – helps coordinate the agreement review and approval process between agencies or commercial entities to bring administrative data into the Census Bureau. The acquisition of any administrative data should start with or involve the Administrative Records Coordinator, and the data source (i.e., federal/state agency, vendor, etc.) will determine the acquisition path. The Administrative Records Coordinator also serves as a resource for researchers seeking to identify available administrative data sets and serves as a liaison between IT security and legal staff to determine how data stewardship and other policies should be applied to administrative data acquisition and use.

**Acquisitions (ACQ)** – facilitates and processes agreements to purchase and bring data into the Census Bureau. The acquisition of any administrative data should start with or involve the Administrative Records Coordinator, and the data source (i.e., federal/state agency, vendor, etc.) will determine the acquisition path.

**Center for Administrative Records Research and Applications (CARRA)** – owns the largest collection of administrative data about persons and households at the Census Bureau. CARRA anonymizes and links data sets so researchers at the Census Bureau and the FSRDCs can use them for approved projects. CARRA coordinates and carries out much of the research undertaken with administrative data.

**Center for Economic Studies (CES)** – is the initial recipient of much of the Census Bureau's administrative data from external sources (i.e., other government agencies and third party/commercial data suppliers). CES performs initial data quality checks before making it available to other areas of the Census Bureau or researchers at FSRDCs. CES also brings in a fair amount of business data, some on behalf of the Economic Directorate (e.g., FTI) and some that is purchased for the purpose of conducting research to develop new data products.

**Division/Center Chief** – must review and approve proposals for internal projects before it can be submitted for a formal review and approval by the Census Bureau and possibly other federal agencies.

**FSRDC Director/Principle Investigator (PI)** – oversee the operation of the FSRDC and is the individual responsible for approving or denying external project proposal requests. Directors/PIs work with researchers from outside the Census Bureau to develop and submit to the Census Bureau projects requesting administrative data use so that the research aligns with the mission of, and benefits, the agency.

**Project Contact (PC)** – is the individual responsible for documenting an internal project proposal in the Census Bureau’s Electronic Project Tracking System and shepherding an internal project from the formal review, implementation, and close out stages.

**Project Review Coordinator (PRC)** – The PRC is a role in the Census Bureau’s Electronic Project Tracking System. Project Review Coordinators coordinate the review and approval process for new internal and external administrative data projects and for amendments to administrative data projects.

**Project Review Team (PRT)** – Reviews and approves internal, external, reimbursable, joint, and ongoing Census Bureau administrative data project proposals, and also guides the Project Contact in next steps. The Project Review Team consists of:

- A Project Review Coordinator, from the Associate Director of Research and Methodology (ADRM), Planning, Coordination and Outreach (PCO/ADRM)
- The Administrative Records Coordinator (ARC), Policy Coordination Office (PCO/DIR)
- A member from the Office of Information Security (OIS), if needed
- The Information Owner (IO) of datasets requested for the project

**Research Project Coordinator (RPC)** – The RPC is a formal job title in the Research and Methodology Directorate. This job title applies to several individuals in ADRM's Planning, Coordination, and Outreach Office (PCO/ADRM). Research Project Coordinators coordinate the review process for all external Census Bureau projects, and serve as Project Review Coordinators (PRCs) for internal and external administrative data projects in the Census Bureau’s Electronic Project Tracking System.

## Governing bodies

Each governing body below supports the infrastructure of the Census Bureau’s administrative data use and research programs.

**The Associate Directors for Programs** meet regularly to consider a wide range of programmatic issues, including administrative data uses. Members of the Executive Staff may identify issues in this forum and refer them to the DSEP Committee.

**The Data Management Committee (DMC)** serves as the focal point for issues related to the management and the efficient and responsible use of Census Bureau information throughout its lifecycle. The DMC sets policy and governs management of the Data Management System (DMS)—the software system Census currently uses to electronically track and administer administrative data projects (i.e., the Census Bureau’s Electronic Project Tracking System).

**The Data Stewardship Executive Policy (DSEP) Committee** operates on behalf of the full Executive Staff, providing corporate guidance and setting Census Bureau administrative data policy. The DSEP is a decision-making body that sets policy by deliberating on issues that have been fully analyzed and vetted by one or more of the groups mentioned below. DSEP will review and approve/disapprove potentially sensitive projects, as well hear and consider appeals to decisions arrived at by the Project Review Team. Project Contacts or others staff with questions, ideas, or concerns about current administrative data policies should direct their comments to individual members of the Data Integration Policy Committee (DIPC).

**The Data Integration Policy Committee (DIPC)** is the operational-level administrative data issues and policy development group that supports the DSEP Committee. The DIPC develops administrative data policy and makes recommendations to the DSEP Committee, and it is the primary group that monitors and evaluates administrative data procedures and activities. The DIPC is responsible for keeping this handbook up-to-date and for considering any changes to related procedures. Members represent each directorate with a role in administrative data research or use. If the staff has questions about a proposed administrative data project, they should contact the Administrative Records Coordinator in the Policy Coordination Office (PCO/DIR). The Administrative Records Coordinator chairs the DIPC and will determine if the proposal should come before the committee.

**The Disclosure Review Board (DRB)** sets disclosure avoidance methodology and policy for statistical projects released by the Census Bureau. The DRB is consulted on the feasibility of administrative data project proposals and reviews summarized statistical products for the correct application of disclosure proofing methodologies.

**The Privacy, Policy, and Research Committee (PPRC)** support DSEP's effort in ensuring that the Census Bureau protects the privacy of its respondents and employees. The PPRC serves as the focal point for issue identification, research coordination, and policy development on issues related to privacy. Its role is to identify privacy-related issues requiring policy development, to conduct and coordinate policy research, to document the research, to analyze and make policy recommendations, to conduct appropriate stakeholder vetting, and to present its findings to DSEP.

Figure 2, on the next page, provides a summarized view of each participant's function in the process of providing procedural and policy guidance surrounding the use of administrative data.

**Figure 2. Governing Bodies and Roles for Providing Guidance Regarding Administrative Data Use**

<b>Roles</b>	<b>DSEP</b>	<b>DIPC</b>	<b>DRB</b>	<b>PPRC</b>	<b>DMC</b>	<b>Assoc Directors</b>	<b>Census Staff</b>
Identifies Issues	x	x	x	x	x	x	x
Conducts Analysis		x		x	x		
Calls for Additional Analysis	x						
Vets Proposals		x					
Assesses Project Viability	x	x	x			x	
Evaluates Process	x	x					
Makes Decisions & Approves Policy	x						

## Chapter 4. Internal Project Lifecycle & Project Contact Responsibilities

### Project types

A project is defined as a set of activities that have a distinct mission and clear starting and ending points. Projects have a lifecycle that proceeds from concept to planning, execution, and completion or termination. A project is often part of a broader program and is a building block in the design and execution of program goals. This chapter will focus on the lifecycle for internal projects, and Chapter 5 will focus on external projects. Below are descriptions of all the project types carried out at, or on behalf of, the Census Bureau:

- **Internal project** – is a project carried out in Census Bureau facilities by Census Bureau employees. In rare instances, an internal project may also involve individuals working on the project who are not Census Bureau employees, but who are individuals with Special Sworn Status. The majority of internal projects are routine, ongoing projects essential to the planning, execution and analysis of the Census Bureau’s core programs. Internal projects also develop new data products to measure the U.S. population and economy through data linkage and modeling.
- **Reimbursable project** – is a project that is requested and funded by, in whole or in part, another agency or organization. Reimbursable projects are usually conducted to provide a more complete picture of the topics covered by the Census Bureau’s censuses. Reimbursable projects must support the mission of the Census Bureau and be authorized by Title 13. As with all projects using FTI provided by the IRS, reimbursable projects using FTI may only be conducted under the authorization of Title 13.
- **External project** – is defined as statistical analyses project with benefits related to the Census Bureau’s mission but conducted by someone outside the Census Bureau, typically from other federal agencies or academic institutions. These projects are conducted through the Center for Economic Studies (CES) and its Federal Statistical Research Data Centers (FSRDCs).
- **Joint project** – are those projects the Census Bureau conducts in collaboration with other federal agencies to predominately accomplish work under Title 13 while also meeting the needs of another federal statistical agency. A team consisting of a representative from each contributing agency typically manages joint projects. Both agencies contribute data and other resources. Joint projects directly enhance the Census Bureau’s ability to conduct its censuses and surveys. Other federal agencies provide access to their own databases as part of these projects. Joint projects differ from reimbursable projects in that both employees of the Census Bureau and the other agency may have direct access to Title 13 data. Employees of the other agency must have Special Sworn Status.
- **Ongoing project** – are typically sets of discrete production activities that are periodically repeated within a program. The project cycles are defined at their inception as having specific starting and ending points, with the associated activities repeated periodically. Predefined production cycles could be quarterly, annually, biennially, quinquennially, decennially, or some other period. As long as the methodologies, types of data, and data linkage strategies used to create a product are unchanged over time, the project is not “new” despite the fact that it occurs more than one time. However, all ongoing projects have a not-to-exceed date of five years. Every five years, the project must be renewed. Additionally, if the project uses FTI, the IRS must approve the renewal of the project and a new five-year period is set. Annual reporting to IRS on the status of these projects is required as well. Examples of ongoing projects include production of the annual Intercensal Population Estimates and maintenance of the Business Register.

## Project development

The project development process includes all stages in developing a proposal from idea development to creating a formal proposal. This includes obtaining guidance from the appropriate internal Census Bureau staff to articulate the data needed for the project, with appropriate reasons for needing it.

A Project Contact (PC)—typically a researcher at the Census Bureau—in discussions with colleagues or management, has an idea for a project using administrative data. After thoroughly developing a formal project proposal, the PC submits the formal proposal to the Census Bureau’s Electronic Project Tracking System to begin the project review and approval process. The following aspects of the proposed project must be articulated and recorded:<sup>11</sup>

- List of datasets to be used in the project
- Start and end dates of the project
- Description of the project
- Methodology of the project
- Projected output of the project
- The server names where the project will be conducted
- Physical work location where those working on the project will access the data
- The employees and Special Sworn Status individuals that will work on the project
- Disclosure avoidance techniques that will be employed
- Title 13 Benefits and description of how the project will benefit the Census Bureau

The PC must determine if the datasets are available at the Census Bureau or if any or all of the datasets must be acquired. If any of the desired data is not available at the Census Bureau, the PC should contact the Administrative Records Coordinator (PCO/DIR) to determine if there are plans to acquire the data.<sup>12</sup> In order to acquire data, a PC must work with the PCO/DIR to prepare agreements with data providers for data that are not yet available at the Census Bureau.

Projects where data linkage occurs require a *Data Linkage Checklist* (#DS014) that must accompany the project through the approval process. This checklist is derived from the *Data Linkage Policy* (#DS014) and will assist the reviewers in determining the sensitivity of the proposed project and whether the proposal requires approval by DSEP.

Projects requesting administrative data from the IRS require a Predominant Purpose Statement (PPS) that the IRS will review, in addition to the formal project proposal. See Appendix E for the PPS template.

The Data Integration Policy Committee (DIPC) is available as a resource to PCs to assist in developing projects using administrative data. Contact the Administrative Records Coordinator (PCO/DIR) to request DIPC assistance or review of a project.

---

<sup>11</sup> The Census Bureau’s Electronic Project Tracking System is run by the DMS software solution. For the [DMS User’s Guide](#), click on link #2 in the *DMS Announcements* box.

<sup>12</sup> Many areas of the Census Bureau acquire administrative data for various purposes, e.g., CARRA, CES, and GEO.

## Project review and approval

The Project Review and Approval process includes all the steps that occur during the review of the project proposal to its subsequent approval or rejection; it involves consultation with all relevant stakeholders, which includes Census Bureau and any outside data providers, as required.

All projects that use administrative data must go through a formal review and approval process. Requesting to use some administrative data will require approval from outside agencies, including the IRS and the SSA. A project may begin using the data once the agency approves the use for the purpose stated in the project.

Once a PC has developed a project proposal, the project undergoes a subject matter review by colleagues. This review is not required, but may be helpful in ensuring the project is methodologically sound, feasible, and has scientific merit. Once these reviews are complete, the Division Chief of the requesting PC must approve the project in order for the project to move forward. At this time, the Division Chief may request changes to the project and, if so, the project goes back to the PC to make the requested changes. Once the Division Chief has fully reviewed the project for scientific merit and feasibility, and concludes it supports the Census Bureau's mission and objectives, then the project can move forward to the next step. If the Division Chief disapproves the project, however, it cannot continue through the process.<sup>13</sup>

Once the Division Chief approves the project, a Project Review Coordinator (PRC) will review the project. The PRC will review the project for completeness and to ensure that all required data elements are included, as well as additional documentation that is required for the project (e.g., a record linkage checklist, as needed, and a Predominant Purpose Statement, or PPS). If the PRC determines that changes are needed, the project will be returned to the PC to revise and resubmit through the approval process. This will continue until the PRC is satisfied that the project proposal is complete and will forward the project to the next step in the review process.

Project Review Team (PRT) review is the next step in the approval process. This team includes the following members:

- Administrative Records Coordinator
- Information Owner(s)
- Office of Information Security (OIS) (as needed)

This process begins when the Project Review Coordinator assigns the project for review to specific individuals and then sets the project's status to "Under Evaluation" (by Project Review Team). Each member of the team reviews the project independently for their particular responsibilities. Checklists have been developed for each reviewer to consider when reviewing the project. The Administrative Records Coordinator reviews the project for policy related criteria, including potential sensitivities, data linkage issues, data stewardship issues and other Policy concerns. The Information Owner(s) reviews the project to ensure that it can be carried out in accordance with the agreement covering the datasets and whether or not

---

<sup>13</sup> For Divisions that have no formal review process, minimal project review steps are outlined in the [Research Proposal Review Process Policy \(#DS24\)](#).

personal identifiers are requested and required. The OIS reviewer ensures that there are no IT security concerns.

If the project must go to DSEP, the Administrative Records Coordinator will work with the PC to prepare an issue paper that will go to DSEP for discussion. The Policy and Data Stewardship Branch in PCO will coordinate this effort.

The PRT may choose one of the following upon completion of their review:

- Revise and resubmit the project to the PC to make requested changes
- Approve
- Disapprove

If a member of the PRT requests changes, the project will go back to the PC to make the changes. If the PRT believes that the changes are substantive, they may request that the Division Chief review and approve the project again.

***External Agency Review and Approval:*** Some projects using external data may require approval from the source agency if required in the formal agreement with the agency. The Project Review Coordinator will coordinate this review by either submitting the proposal directly to the agency or working with the appropriate program area to obtain approval. These projects include those using data from the IRS or state agencies supplying social welfare program data (i.e., Supplemental Nutrition Assistance Program, or SNAP). In addition, some projects using SSA data may require approval from SSA. The following may be outcomes from an Other Agency Review:

- Source Agency Approval
- Recommendation to Revise & Resubmit
- Disapproval

The PRC will communicate with the PC and, if necessary, the Division Chief and PRT regarding any decisions or feedback provided by the source agency.

If the source agency approves the project then the project is ready to begin implementation, conditional on approvals by any other agencies. If the source agency determines that the proposal is incomplete, or requires additional information or clarification, then the proposal is sent back to the PC with a “Recommendation to Revise and Resubmit.” If the source agency disapproves the project, then the process ends for that dataset. Once all source agencies have approved access to data, the PRC sets the project to “Approved.”

***Phased Approvals:*** When a new project requests data from both internal and external sources (including data from one or more agencies that require their own independent review and approval process), phased approval allows researchers on the project to start working with a subset of approved datasets while the rest are still in the process of review by external agencies. Once the internal review process is complete, the Project Review Coordinator sets the status of the project to “Phased Approved.” At this point, the researchers may begin using all of the datasets on the project that are marked “Approved” by the Project Review Coordinator. This will include all of the Title 13 Census Bureau datasets, plus any datasets from external sources that do not require an independent review and approval process.

At this point, the Project Review Coordinator also initiates the review process for all externally-sourced data from agencies requiring independent review and approval. As these agencies approve the use of specific datasets the Project Review Coordinator marks these “Approved” in the Census Bureau’s Electronic Project Tracking System, at which point they can be accessed and used for the project. Once all datasets have been approved, the Project Review Coordinator sets the project status to “Approved.”

In the rare instance that an external agency does not approve the use of a dataset, the Project Contact must submit an amendment to remove the dataset from the dataset list and from the project description and methodology.

**Project Status:** During the course of the review and approval process, a project may go through a variety of statuses from inception of the project to approval.<sup>14</sup>

## Subprojects

In special circumstances, it may be possible to carry out additional work under an established project without having to create a new administrative data project. This additional work can be contained within a subproject. In these cases, the review and approval cycle can move more quickly than a completely new project. However, to be approved, subprojects must align with the parent projects’ previously established scope, methodology, servers, staff, etc. If these aspects of the parent project do not align with the proposed subproject, the PC must bring the parent project into alignment (as part of the amendment process, see next section) or a new project will need to be created and approved that includes the scope of this work. In either scenario, the proposed work must be documented in the Census Bureau’s Electronic Approval & Tracking System.

To document a subproject on an existing project, a subproject tab is available on each approved project that allows for the creation of subprojects. A job aid is available to assist in creating the subproject. All subprojects must be created by a project contact or alternate for the main project. Similar information is required to create a subproject as for the main project: a title, a subproject contact, subproject supporters, data, servers, methodology, and proposed outputs. Once the subproject has been created, it is submitted to the PRC who will review the subproject for scope and convene a meeting with the PRC team if necessary. A subproject can either be approved, rejected, or sent back to the PC/PCA for revision. Once a subproject is approved, work can begin.

Subprojects will be monitored by the PRC team, over time. If any future changes to the subproject renders the work out of scope under the existing parent project, the project team will be contacted. If the issue cannot be resolved, a new project must be started and the entire review and approval process must be executed from the beginning before the work will be allowed to continue.

Furthermore, if the statistical products resulting from a sub project's efforts fall outside the scope of its parent project, those products may not be cleared for release by the DRB.

---

<sup>14</sup> To explore various project status categories, see the [DMS User’s Guide](#), and click on link #2 in the DMS Announcements box.

## Amendments

During the course of a project, the PC may wish to make changes to the project. Some changes can be made to a project at any time without review and approval, such as changing the PC or Alternate PC, servers, work locations, and adding or removing employees or SSS individuals. All other changes to the project require review and approval, and these changes require an amendment. Amendments are reviewed and approved by the PRC, who may add the Division Chief or other reviewers. When datasets are added to a project, the Information Owner will automatically be included in the review process.

The PRC may request a meeting of the PC and PRT to discuss potential changes to a project. In turn, the PC may request a meeting of the PRC and PRT to discuss changes and pending approval, as well.<sup>15</sup>

If a PC is making significant changes to a project, a new project may be warranted. The types of changes that may make a new project necessary include:

- Significant change in scope
- Addition of datasets that may change the scope
- Guidance to revise the PPS for additional tax data added to a project

The PRC will help the PC determine if the amendments requested meet the threshold of new project creation.

## Project implementation and administration

The project implementation process starts when researchers begin accessing data according to the negotiated agreement, conducting the project, providing regular status reports on the project's progress, initiating changes to the agreements as appropriate, closing out the project, and appropriately disposing of the data or returning it to the external agency.

If the PC needs new data, or additional data while the project is underway, the PC submits an amendment to request the additional data.

The PC will submit regular status updates to the Administrative Records Coordinator (PCO/DIR) and relevant Information Owners as part of the Agreement Administration and Data Stewardship process. In addition, PCs must submit an annual progress report on the project to the PRC. Annual progress reports for projects using FTI are submitted to the IRS.

The PC will submit regular status updates to the relevant Information Owners on data access, use, and disposition. Once the project is finished, the PC will begin the closeout process.

If the PC decides that major changes to the project are necessary, the PC will revise the project, and resubmit it through the amendment process.

If the project must be terminated before completion, the PC will begin the closeout process.

**Annual Reviews:** In order to ensure that all projects are up to date, an annual review is required. PCO/DIR requires all PCs to review their projects and ensure that all fields are consistent with the work conducted in

---

<sup>15</sup> In essence, this arrangement can be used as an appeals process should an amendment be denied.

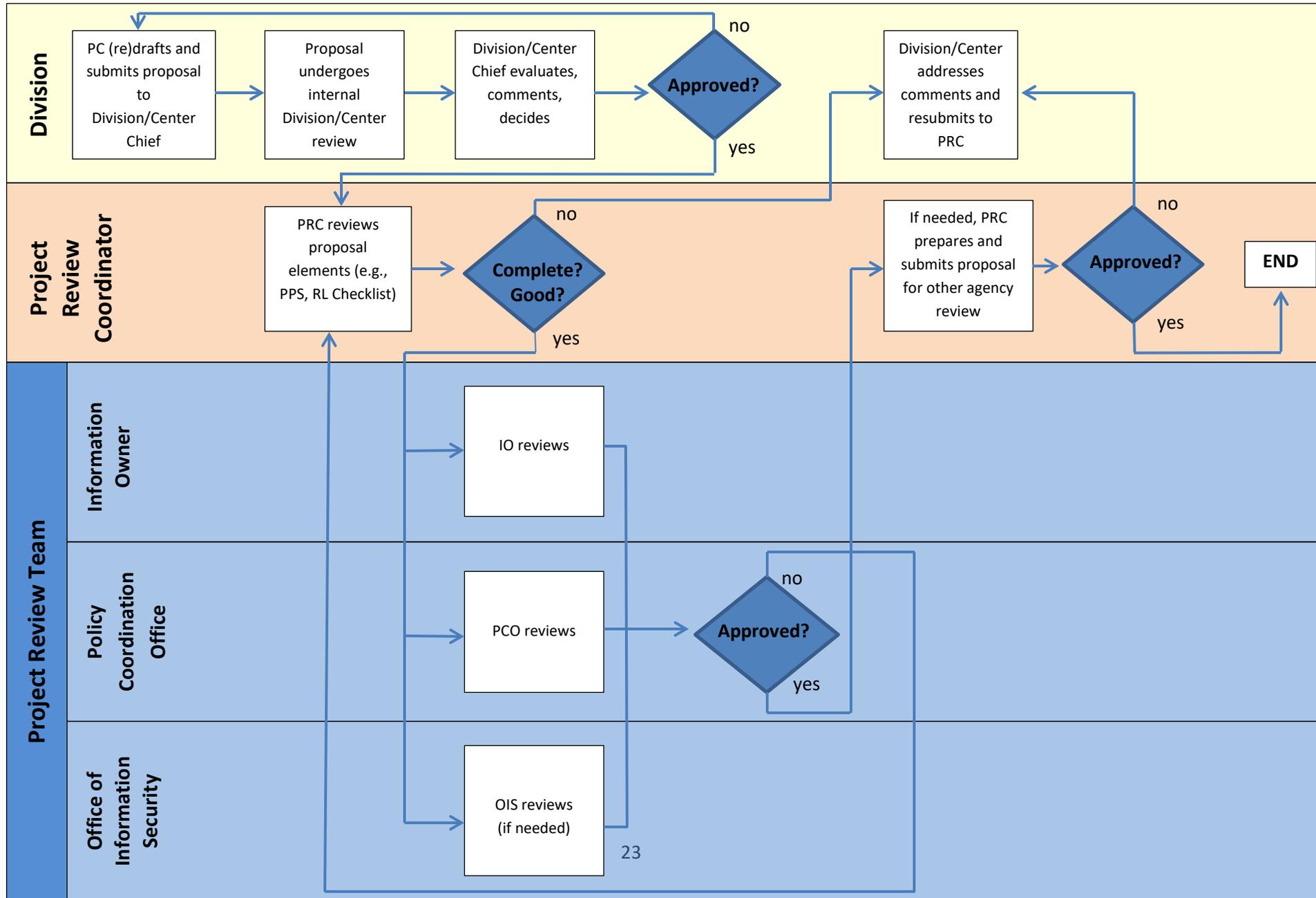
the project. It is important that projects are accurate for record keeping and reporting to outside agencies, including the IRS. When reviewing a project, the PC must be sure to update any fields that are not current, paying careful attention to the employees and SSS individuals listed on the project, servers used for the project, and the methodology. If a project is complete or no longer being worked on it should be closed out following closeout procedures. The Administrative Records Coordinator emails PCs annually with instructions on how to complete an annual review for their project(s). See Appendix B for an example of the instructions for conducting an annual review.

***Closeout (Data Disposition):*** When projects closeout—either because they are complete or because they were terminated prior to completion—PCs will prepare reports on project closeout and data disposition for the Information Owner and as part of the agreement administration and data stewardship process

The PC must ensure that all supporting documentation on a project is disclosure proofed.

See Appendix C for a Checklist of Responsibilities for the Project Contact.

## Flow of events – internal research project proposal review and approval



## Chapter 5. External Project Lifecycle & Principal Investigator Responsibilities

### Project types and description of projects

This chapter describes the project lifecycle and Principal Investigator (i.e., the FSRDC researcher with special sworn status) responsibilities for external projects, which are projects that utilize Census Bureau data and that are proposed and managed by persons with Special Sworn Status (SSS).

External project research takes place at FSRDCs. FSRDCs host data from a growing number of federal agencies, including the Census Bureau. Examples of data at FSRDCs supplied by the Census Bureau include the American Community Survey, the Survey of Income and Program Participation, and the Annual Survey of Manufactures, all of which are surveys collected under Title 13. Other data supplied by the Census Bureau include administrative records acquired by the Census Bureau and protected under Title 13, such as the SSA Numident File and CMS Medicare Enrollment Database.<sup>16</sup>

### Project development

Detailed instructions for developing an FSRDC proposal can be found on the Center for Economic Studies external website, here:

<http://www.census.gov/ces/FSRDCresearch/howtoapply.html>

Or, see Appendix D.

Researchers who wish to develop a proposal for external research at an FSRDC first contact the administrator at the primary center where the research will be conducted. The researcher discusses the proposed project with the administrator to determine whether the research meets the Census Bureau's requirements for external projects, which are that:

- The research is consistent with the Census Bureau's mission
- The research will provide benefits to Census Bureau programs under Title 13 of the U.S. Code
- The project is designed to produce model based statistical estimates, as opposed to tabular statistics
- The research output will not pose undue risk of disclosing information about survey respondents

Once the FSRDC administrator and the researcher have agreed that it is sensible to proceed, the researcher submits a completed Preliminary Proposal to the administrator, which contains a brief abstract of the research as well as information on data to be requested and proposed benefits to the Census Bureau. The administrator will review the summary, advise the researcher of any suggestions for refinement, and submit

---

<sup>16</sup> Not all projects at FSRDCs are Census Bureau external projects. Some research projects at FSRDCs utilize data supplied solely by agencies other than the Census Bureau. For example, both the National Center for Health Statistics and the Agency for Healthcare Research and Quality supply extracts of health data owned by their respective agencies. Such data are not protected under Title 13 and may be used for projects that do not have Title 13 benefits. These projects are out of scope for this chapter because they are governed by processes that were developed by the agencies supplying the data. In addition, while most projects carried out at FSRDCs are external projects, FSRDCs sometimes host internal Census Bureau projects as well. Internal projects are also out of scope for this chapter. Review, approval, and governance of internal Census Bureau projects taking place at FSRDCs follow procedures described in Chapter 4 of this Handbook.

it to the Center for Economic Studies via a Web-based project management application, thereby generating a project tracking number.

Once the Preliminary Proposal has been submitted the researcher continues to work with the FSRDC administrator to develop a full proposal package for submission. The final proposal consists of three separate documents: a project description of fifteen pages or less, a benefits statement describing how the research will provide benefits to the Census Bureau under Title 13 Chapter 5, and an abstract. Once the FSRDC administrator finds these documents to be of sufficient quality and completeness, they are submitted to the Center for Economic Studies for review.

## Project review and approval

Once a full proposal has been submitted, the CES project management application alerts Research Project Coordinators in the ADRM Planning, Coordination, and Outreach Office (PCO/ADRM) with an automated email. In consultation with the Assistant Center Chief for Research (ACCR/CES), a Research Project Coordinator assigns each proposal for review to:

- Two or more subject-matter experts, typically Census Bureau staff scientists with significant experience using the data requested in the proposal.
- The Administrative Records Coordinator in the Policy Coordination Office.
- The Disclosure Avoidance Coordinator in the Center for Economic Studies.
- Additional staff as required.

Reviews are submitted and uploaded to the project management system, typically within 21 days of assignment.

Based on all submitted reviews, a Research Project Coordinator drafts a review document that summarizes basic information about the project (duration, location, requested data, objectives, and Title 13 benefits) and synthesizes comments in the submitted reviews. The synthesis highlights concerns expressed by reviewers based on the following five evaluation criteria:

1. **Scientific merit.** This criterion refers to the project's likelihood of contributing to existing knowledge. The proposal should describe in some detail the nature and scientific basis of the research questions, the methodology (including models to be estimated, how model variables will be measured and any hypotheses to be tested), the data sets to be used (both Census and non-Census), and expected outcomes from the research. Given the proposal is seeking access to confidential micro-data, the technical aspects of data and methodology included in the proposal should be fairly detailed.
2. **Benefits to the Census Bureau data programs under Title 13, Chapter 5, U.S.C.** Proposals must demonstrate that the research is likely to provide one or more benefits to the Census Bureau under Title 13, Chapter 5 of the U.S. Code. If a project has as its *predominant purpose* one, or any combination, of the criteria specified in DS002 Articulating the Title 13 Benefits of Census Bureau Projects, it will be considered to have as its predominant purpose increasing the utility of Title 13, Chapter 5 data. The proposal should emphasize strongly the potential of the research to benefit Census Bureau programs. While it is unlikely that any project will provide benefits under every criterion, stronger projects propose benefits under more than one criterion.

3. **Feasibility.** The proposal must show that the research can be conducted successfully with the methodology and requested data and that it can be completed within the requested duration. In order to demonstrate this, researchers should discuss their familiarity and experience with the micro-data to be used in the project, and present a timeline for accomplishing major tasks in the project.
4. **Disclosure risk.** Output from all research projects must undergo and pass disclosure review. The proposal should describe the anticipated outputs from the project, whether there are any anticipated disclosure risks, and, if so, how they will be addressed by the researcher.
5. **Demonstrated need for confidential data.** The proposal should explain why publicly available data are not sufficient to meet the project's objectives. Specifically, it should address whether the project's objectives can be realized with the use of known publicly available data, and should make a strong case for access to confidential Census micro-data.

The Research Project Coordinator uploads the review summary to the project management system and notifies the FSRDC Director that the project is ready for a decision to approve or deny the request. The FSRDC Director records an approval or denial of the project, and the Research Project Coordinator communicates this decision to the researcher, in an email package containing the review synopsis, an explanation for the decision, and anonymized copies of the expert reviews. Should the FSRDC Director decline to approve a project, the researcher is invited to work with the FSRDC administrator on a new or revised proposal.

If the decision is to approve the proposal, the Research Project Coordinator must then determine whether separate approvals from federal agencies and other entities supplying data for the project are required. Examples of agencies requiring project-level approval for use of data accessed in FSRDCs include the Internal Revenue Service, the Bureau of Labor Statistics, and select state agencies supplying data to the LEHD program. Until these external approvals have been granted, the researcher may not start the work.

Once all relevant stakeholders have approved the project, a Research Project Coordinator notifies the FSRDC administrator who then notifies the researcher that the project has full approval and can start (assuming the researcher has already obtained SSS, below).

At this point, an ADRM Research Project Coordinator also creates a record for the project in the Census Bureau's Electronic Project Tracking System.

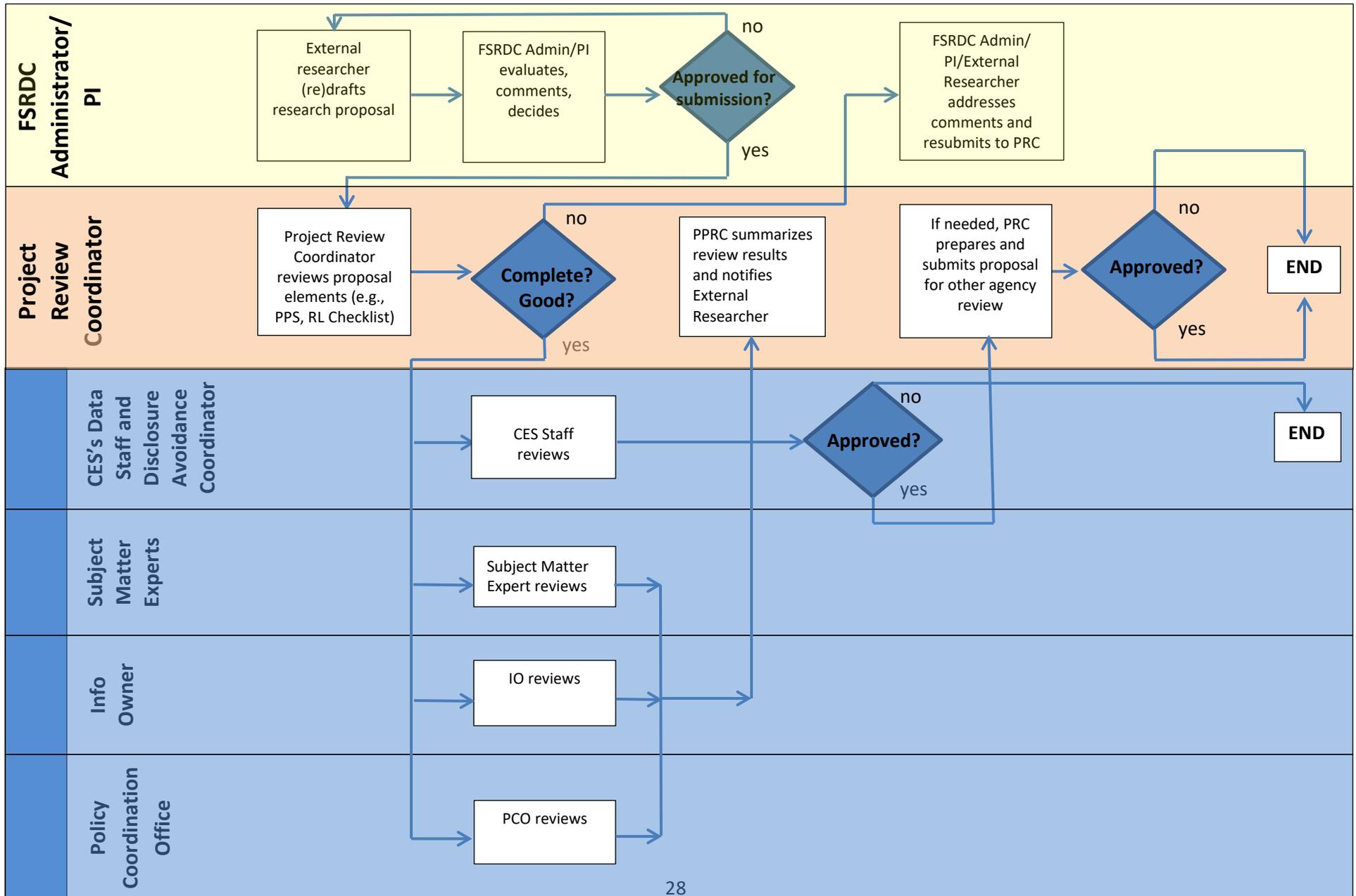
### **Project implementation and administration**

Any researcher who must access restricted data or undisclosed output during the course of the research must apply for and obtain Special Sworn Status with the Census Bureau before accessing data. In addition, all approved research projects are governed by a written agreement between the researcher(s) and the Census Bureau. The agreement stipulates the start and end dates for the project and responsibilities of both parties with respect to procedures and practices. Each researcher on the project must sign an agreement with the Census Bureau. If researchers are added to the project after the project starts, these researchers will also each sign an agreement. Separately, an agreement with the FSRDC partner institution may be required, as well as payment of fees to use the FSRDC.

The benefits statement contains a commitment by the principal investigator and includes as part of the findings that investigators are required to file at the end of the research project, a memorandum or report describing how the research project furthered a census, survey, or other statistical program authorized under Title 13, Chapter 5, including ways not anticipated at the outset of the research project. For projects using

Federal Tax Information, the Census Bureau submits this document as part of project close-out to the Internal Revenue Service Statistics of Income Division.

## Flow of events – external research project proposal review and approval



## Chapter 6. Using Agreements to Bring in Administrative Data

### General information

This chapter describes the process by which administrative data files, that support a particular project, are acquired by the Census Bureau. In most cases, an agreement is required which documents the terms of the contract. The agreement/contract must be mutually agreed upon and it must be documented in a written format; oral agreements and understandings are insufficient.

Data acquisition agreements may be for any period of time, but periods of time over five years should, as a general rule, be avoided (absent special circumstances). Circumstances may change for either agency over time, and it is wise to re-examine agreements regularly to ensure the data and associated work or research is still in the Census Bureau's best interest.<sup>17</sup> Often, lengthy agreements are subject to annual renewal conditions, meaning the agreement does not renew automatically without one party indicating they would like the agreement to continue for another year (this is usually accomplished via memo).

Negotiating an agreement can be a lengthy process due, in part, to the time required for agencies to agree on the agreement details. In addition, the agreement must travel through the Census Bureau's IOSA Review and Approval System (e.g., agreements are reviewed and approved by Policy, Legal, Budget, Finance, IT and OIS Security Offices, and sometimes Commerce), as well as the providing agency's review and approval system.

### Negotiation

The Census Bureau has a strong corporate interest in these source entities because they often provide administrative data that feed into a broad range of our programs. Some of those source entities Census receives data from include:

- Bureau of Economic Analysis (BEA)
- Bureau of Labor Statistics (BLS)
- Center for Medicare and Medicaid Services (CMS)
- Housing and Urban Development (HUD)
- Indian Health Service (HIS)
- Internal Revenue Service (IRS)
- Selective Service
- Social Security Administration (SSA)
- Supplemental Nutrition Assistance Program (SNAP)
- United States Postal Service (USPS)
- Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)

It is extremely important that the PC work directly with the Administrative Records Coordinator (PCO/DIR) to arrange for the acquisition of new administrative data from any of the source agencies above. This area of the Census Bureau oversees the review and approval of data agreements and will be aware of concurrent negotiations with the same source entity that may affect other programs.

---

<sup>17</sup> PCO/DIR's templates require that agreements with durations exceeding three years be reviewed after three years (or at three-year intervals).

If the source entity is not on the list above, the PC may reach out to begin *informal* negotiations.<sup>18</sup> Informal negotiations are exploratory in nature and will not involve obligating funds, nor should it be construed as a commitment on the part of the Census Bureau. Once the informal negotiations have concluded (e.g., are the appropriate data elements available, is it good quality data, does the source entity have the authority to share the data, is it affordable, how will data be delivered, timelines for delivery, methodology, potential products resulting from the research, etc.?), it is time to involve the Administrative Records Coordinator (PCO/DIR). The Administrative Records Coordinator (PCO/DIR) will help the PC navigate every stage of the agreement process.

The Administrative Records Coordinator (PCO/DIR) will assist the PC in coordinating with staff across the Census Bureau to arrange for the acquisition of new administrative data. The Economic Area acquires quite a bit of administrative data. For example, CES acquires data about people and households that contain personal identifiers, as well as business data that maintain federal economic data for the Business Register and other economic programs. While CARRA acquires data mostly about people and households. The Geography Division is responsible for acquiring address files to support the Master Address File (MAF).

The list below can be used to facilitate a successful negotiation strategy between the Census Bureau (i.e., PCO/DIR, ECON area, CES, GEO, and the Project Contact) and the providing agency:

- Work with the Administrative Records Coordinator (PCO/DIR) to identify stakeholders across the Census Bureau
- Set up a strategy meeting with all Census Bureau program areas that are involved
- Identify data files required
- Identify cost/deliverables from source agency
- Determine if costs can be justified
- Identify security issues
- Initiate formal contact with source agency
- Identify whether third party concurrence is needed (e.g., IRS approval)
- Identify who the Information Owner at the Census Bureau is or will be
- Identify and agree on a data transmission method

## Using an agreement to acquire data

Below are is a summary of the basic steps that PCs must follow to create and execute an agreement to acquire external data:

- The Census Bureau and providing agency should agree on a *working* draft of the agreement, revising the agreement as needed, until both parties are satisfied with the initial content.
- Submit the agreement to the IOSA Review and Approval process for Census Bureau review/approval.
- Submit the agreement to providing agency for review/approval.

---

<sup>18</sup> We speak in terms of the Project Contact (PI) here, instead of the Principle Investigator (PI), because PI's will typically bring in their own data sets that they wish to link to other Census Bureau data sets. PI's generally will not seek to acquire new administrative data outside of the Census Bureau's current holdings, unlike a researcher at the Census Bureau.

- Once feedback is received from all parties (i.e., the Census Bureau’s and the providing agency’s reviewers), Census Bureau staff will revise the agreement if needed and resubmit the final copy into the IOSA Review and Approval website.
- Secure approval signatures at the Census Bureau and the providing agency on the final version of the agreement.
- Upload the agreement, signed by both agencies/entities, to the IOSA Review and Approval website.
- Upload the agreement, signed by both agencies (along with any other documentation, e.g., approval by a third agency to use the data), into the Census Bureau’s Electronic Project Tracking System for administrative data projects.
- The providing agency transmits data to Census Bureau.
- The PC logs data set, and any relevant metadata, into the DMS (a mandatory policy).

To acquire external data, you will first need to decide which type of agreement to use. There are a few different legal authorities under which the Census Bureau acquires data and a variety of resulting agreements, including:

- **Payable or No Funds Exchanged Agreements** – Title 13, Section 6 of the Census Act permits Census to acquire data from other agencies “by purchase or otherwise.” Data acquired under this section must be acquired to conduct work under Title 13 and may also be protected by Title 13. The Census Bureau is authorized to compensate the supplying agency by payment of funds or returning an improved non-confidential data set.
- **Reimbursable Agreements** – Title 13, Section 8(b) of the Census Act, permits the Secretary to “make special statistical compilations and surveys, for departments, agencies, and establishments of the Federal Government, the government of the District of Columbia, the government of any possession or area..., or other public and private person and agencies, upon payment of the actual or estimated cost of such work.”<sup>19</sup> Reimbursable agreements under Title 13 are used for transactions in which the Census Bureau will receive funds. These types of agreements may permit the Census Bureau to maintain the resulting data set and use it for Title 13 purposes. The acquired data may or may not be protected by Title 13.
- **Joint Project agreements** – Authorized by Title 13, Section 8(b), which allows the Secretary to “engage in joint statistical projects, the purpose of which are otherwise authorized by law ...”. Joint project agreements must reflect mutuality of benefit and contribution. Joint project authority may be appropriate for agreements in which the work on the acquired data set will be collaborative. While the Census Bureau can pay out funds under a Joint Project Agreement, costs for joint projects must be “equitably apportioned,” meaning that the Census Bureau cannot alone be responsible for the majority of the project costs. Data acquired under Joint Project Agreements is acquired to perform work authorized by Title 13.

Use a standardized Census Bureau Inter-Agency Agreement (IAA) template when developing new agreements. Links to model agreements can be found on PCO’s/DIR Administrative Records Coordinator’s website; they contain the latest version of agreement language that has been vetted through our legal

---

<sup>19</sup> Other agencies might cite the Economy Act (31 U.S.C, Section 1535) to order statistical work from the Census Bureau. The act authorizes Reimbursable Agreements between federal agencies for the procurement of goods and services if it is in the best interest of the government. The procuring agency in an Economy Act Agreement must execute a D & F, a Determination and Finding, stating that the agency has determined that the ordered goods or services cannot be provided as conveniently or cheaply by a commercial enterprise and that the order is in the best interest of the government.

council. Templates for these agreements are available on PCO's/DIR website:  
<https://collab.ecm.census.gov/div/pco/intranet/Pages/Interagency-Agreements.aspx>

Using the appropriate agreement template for your project, document the terms of the agreement by negotiating requirements for the use and exchange of the desired data with the data provider. Agreements must describe the authorized uses of the data, the legal authority under which the data are collected and protected, access restrictions, retention period, and disposition of data at the end of the retention period. Any and all uses of administrative data at the Census bureau must be consistent with the approved uses stated in the relevant agreement.

Once the two agencies have initially agreed on the content and syntax of the agreement, upload the agreement to the IOSA Review and Approval website:

<https://share.census.gov/apps/iosa/SitePages/Submitted.aspx#>

The IOSA Review and Approval website is managed by the PCO's/DIR Administrative Records Coordination Branch. This branch uses the site to coordinate the review of all agreements and amended/modified agreements. Agreements are reviewed by the Census Bureau's Policy, Legal, and IT and OIS Security Offices. Our legal staff decide whether agreements need to be sent to the Department of Commerce for additional review and approval. The IOSA Review and Approval process can take anywhere from several weeks for a straight forward agreement, with no funds exchanged, to several months for more complex agreements with funding involved. On average, agreements typically approved at the Census Bureau in less than 30 days. The time it takes to get the agreement reviewed and approved at the data provider's agency, and any other agency whose approval is needed, is not factored into the 30-day timeline.

See the Census Bureau's *Policies & Procedures Manual* to determine the appropriate level of signatory approval (e.g., Director, Deputy Director, Assistant Director, etc.) required for various contract amounts:

[https://intranet.ecm.census.gov/apps/policyportal/Policy%20Repository/PPM\\_B1.pdf](https://intranet.ecm.census.gov/apps/policyportal/Policy%20Repository/PPM_B1.pdf)

PCs should track the progress of the agreement in the IOSA Review and Approval website and be responsive to any additional inquiries resulting from the review and approval of the agreement. The Administrative Records Coordinator in the Policy Coordination Office (PCO/DIR) can provide assistance.

Occasionally, a data acquisition agreement may need to be formally amended, or modified, when a material or substantial term changes. For example, if the amount of money paid changes, or if the expectations for confidentiality change, an amendment may be needed, and you should consult with PCO/DIR to facilitate this change. Examples of non-material changes include a change in the custodian or minor changes in delivery dates.

There is some special guidance and cautions related to agreements where data will be returned to the providing agency:

- **Confidentiality.** Issues of confidentiality and privacy are frequently raised in these kinds of agreements, particularly in instances where the proposed agreement involves the Census Bureau linking Title 13 data with the data supplied by the providing agency and then returning variants of the initial data set to the provider. These types of proposals require a higher degree of scrutiny; while this type of agreement may well pass legal and policy review, these proposals will require an extra level of sensitivity assessment and possibly a review by the DRB because there is the possibility that comingling issues will arise.
- **Other Agencies' Data.** Extra concern must be exercised when the proposed agreement would allow linkage with a data set that includes data acquired from a third agency, like the Business Register which contains FTI that Census receives from the IRS. The Census Bureau cannot

redisclose FTI data unless it is allowed by laws that protect that type of data and we have the IRS's consent.<sup>20</sup>

- **Security Issues.** Caution should be used when the proposed agreement may raise security concerns, such as remote access to the data or there is an off-site data access request lodged.
- **Data Use and Functional Separation between Statistical and Other Mandates or Activities.** The identity of the proposed data provider and what it intends to do with the returned data may also require a higher degree of scrutiny. A proposal to return data to a sister statistical agency is more likely to be approved than a proposal to return data to an agency with a regulator or enforcement mandate. Return of data to a commercial entity also raises issues. Similarly, the use that will be made of the returned data is important. The proposed use of returned data for anything other than statistical purposes raises issues of perception. Uses for law enforcement or commercial purposes, for example, are unlikely to be approved. For example, Census would not approve a data delivery under these circumstances that would give a private company a competitive advantage, give data to a law enforcement agency, or a direct marketing company.
- **Relationship to other Census Bureau Programs.** In general, extra caution must be used if the proposed agreement has the potential to adversely affect another Census Bureau program or project.
- **Disclosure Review Board (DRB).** Products generated from Title 13 data for public release, including release to a providing agency, must be cleared by the DRB.

If there is any doubt about how you should proceed with the details of an agreement, please contact the Administrative Records Coordinator in PCO/DIR.

## Data access and agreement administration

The formal administration of a data agreement begins immediately after the negotiation phase and signing of the agreement. Administration of an agreement occurs throughout the life of the agreement and project implementation and it ends when the agreement expires.

Responsibilities and deliverables are negotiated prior to signing the agreement, and include things such as custodianship of the data, period of agreement, security requirements, reporting requirements, quality assurance, data or service delivery schedules, and the disposition of the data at the conclusion of the data retention period. At the agreement administration step, all of these responsibilities and deliverables must be managed by the signatory of the agreement. If the PC happens to be the signatory, then they are the individual responsible for carrying out these terms.

The following guidance focuses on the requirements that must be fulfilled by the PCs.

The very first action taken to manage an agreement is for the Census Bureau to accept the data from the data provider. Once the Census Bureau has received the data, the PC is required to register the dataset into the DMS, along with any relevant metadata.

The Census Bureau's Electronic Project Tracking System (i.e., the DMS) archives all of the information and documentation associated with an administrative data project. Within this software system, the PC is required to record relevant project-related information such as:

- Project name
- Agreement number
- IOs, PC/PI are identified and recorded

---

<sup>20</sup> Furthermore, the current IRS Criteria Agreement does not permit disclosure under reimbursable agreements.

- Names of the datasets that the project will use
- Due date for various activities and deliverables
- Any predecessor requirements
- Names of staff and SSS individuals who are actively working on the project
- Status of the project
- Approvals for the project
- Any additional documents or comments regarding the project

PCs are required to do the following on a regular schedule for the life of the project (using the Census Bureau's Electronic Project Tracking System):

- Annually – conduct a project review (See Appendix B for an example of instructions).
- Annually – make certain everyone working on the project has received a certificate for having taken the Data Stewardship Awareness and, if necessary, the Title 26 Awareness training.
- As needed – note changes, in real time, (comings/goings) of staff currently working on the project and accessing files.
- Annually, or at the most, every 3 years -- assess the need to extend the agreement as the project's expiration date nears, and decide whether the agreement can be modified or if a new agreement is needed.
- As needed – any new project that requests this data must adhere to the uses originally expressed by the agreement acquiring the data, and must seek the permission of the data provider if that is specified in the original agreement.
- As needed – in coordination with PCO/DIR and Information Owners, PCs must comply with all reporting requirements to report to the provider agency on the data uses (e.g., reporting to the IRS on FTI data uses).

At the conclusion of the project, PCs are required to:

- Record a synopsis of the research results in the Census Bureau's Electronic Project Tracking System.
- Destroy or return the data, as outlined in the agreement.
- Close the project in the Census Bureau's Electronic Project Tracking System by setting it to Completed.

For more information on agreement administration, please see Chapter 4's section on *Project Implementation and Administration*.

## Chapter 7. Administrative Data and Confidentiality Laws

### Background

The U.S. Census bureau obtains administrative records from various sources including federal, state, local, and tribal government agencies and non-governmental entities. These data are primarily used in planning, conducting, processing, and developing surveys, censuses, and estimates conducted under the authority of the Census Act, Title 13, United States Code. They also may be used in the course of performing statistical services (including special censuses, surveys, classifications, geocoding, etc.) for other federal, state, local, and tribal governmental agencies, as well for the private sector.

This chapter provides guidance regarding the confidentiality protection applicable to administrative records coming into the Census Bureau's possession. A determination of applicable confidentiality protection is necessary to establish allowable uses and required safeguards that may apply to any particular data set. It should be noted at the outset that this determination can be rather complicated. It will depend upon the nature of the data being provided, the entity providing the records, the authority under which the provider collected the information, the confidentiality restrictions imposed by that authority, the disclosure/use restrictions imposed by the provider, the way the Census Bureau will use the data, and federal statutes, in addition to Title 13, that could impose confidentiality constraints.

Accordingly, these guidelines provide a process to be followed in determining whether and the extent to which a particular set of administrative records is afforded confidential treatment.

### General guidelines for determining confidentiality status

#### A. Applicability of Title 13 Confidentiality.

1. If the administrative records are in the public domain (either published by or made available to the public by the provider) then, assuming they are not commingled with identifiable Title 13 data, confidentiality protection under Title 13 generally will not apply.
2. If the administrative records are commingled with Title 13 protected data, then Title 13 confidentiality applies to the entire commingled data set.
3. Products prepared for administrative agencies in exchange for the use of the administrative data cannot contain identifiable Title 13 data. Enhancements to the source data are permitted as long as the enhancements do not compromise the confidentiality of census data. For instance, adding geographic codes to addresses from the administrative data is acceptable.
4. If the administrative records contain data that identify individuals or businesses that are to be used for an authorized Title 13 census or survey "instead of conducting direct inquiries" under Section 6 of Title 13, then Title 13 confidentiality applies.

#### B. Other Statutory or Regulatory Protection.

1. Administrative records may be subject to specific confidentiality restrictions in the authorizing legislation of the federal, state, local, or tribal government providing the data.
  - a. The agency furnishing the records should advise the Census Bureau of applicable confidentiality restrictions, if any, and provide a citation to the source law (statute/ordinance/regulation), together with a copy of the applicable language.
2. General Federal Statutes:

- a. **Freedom of Information Act (5 U.S.C. 552)** --- this law permits a requester to obtain government documents but exempts from disclosure documents that are: (i) made confidential by statute; (ii) proprietary (trade secrets, commercial or financial) information; and (iii) personnel/medical and similar files, disclosure of which would constitute an unwarranted invasion of personal privacy.
- b. **Privacy Act (5 U.S.C. 552a)** ---- this law applies only to data pertaining to individuals (as opposed to businesses) and prohibits disclosure of individual data if it is kept in a “system of records,” (retrievable by name or other symbol/number) and are required by statute to be maintained and used solely as statistical records.
- c. **Trade Secrets Act (18 U.S.C. 1905)** ---- this law prohibits the disclosure of proprietary business information of a particular establishment such as income, profits, losses, etc.

### **C. Terms of the Agreement.**

The Agreement under which the administrative records are provided to the Census Bureau may contain terms restricting use of the data. While these terms are binding upon the Census Bureau, they do not and cannot create confidentiality protection by agreement alone. To provide protection against requests for access by third parties, the data must be made confidential by law; otherwise, the data may be vulnerable to a request under the Freedom of Information Act.

1. The Agreement should reference all laws, particularly those addressing confidentiality, that authorize the provider’s collection and use of the data. Additional restrictions agreed to by the Census Bureau would be subject to applicable federal law, noted in B.2. above.

## **Specific guidelines for data sets obtained under an agreement for services**

### **A. Authority to Perform Statistical Services:**

The Census Bureau has authority to perform statistical services under Section 8 of Title 13, United States Code<sup>21</sup>. Under 13 U.S.C. Section 8(b), the Secretary may:

furnish copies of tabulations and other statistical materials that do not disclose the information reported by, or on behalf of, any particular respondent, and may make special statistical compilations and surveys, for departments, agencies, and establishments of the Federal Government, the government of the District of Columbia, the government of any possession or area (including political subdivisions thereof) referred to in Section 191(a) of this title, State or local agencies, or other public and private persons and agencies, upon payment of the actual or estimated cost of such work.

The Census Bureau can perform statistical services under Title 13 authority and, so long as no identifiable Title 13 confidential data are used in the process, provide the sponsor with identifiable data as an end product.

### **B. Confidentiality of Data Obtained/Produced in Performing Services:**

Where the Census Bureau comes into possession of data sets either provided by a sponsor or developed in the course of the project, two broad issues are presented: (1) “What confidentiality provisions apply while

---

<sup>21</sup> The International Trade Management Division (ITMD) is the only organizational component within the Census Bureau generally authorized to use 15 U.S.C. Section 1525 as authority to conduct reimbursable work and joint statistical projects. Unless inconsistent with the provision of Chapter 9 of Title 13 or Part 15 of Title 15 of the Code of Federal Regulations, the guidelines set forth in this Handbook apply to administrative data projects conducted by ITMD. The Census Bureau Legal Office should be consulted regarding any questions or concerns regarding ITMD’s authorities.

the sponsor's data are in the hands of the Census Bureau?" and (2) "What kinds of statistical services can be performed for the sponsor without subjecting the end product; i.e., the "Deliverable," to Title 13 confidentiality?"

1. Confidentiality Applicable to Sponsor's Data:

- a. As discussed earlier, the sponsor's data coming into possession of the Census Bureau, whether as a data acquisition under Section 6 of Title 13 or part of an agreement for services under either Section 8 of Title 13, generally will be subject to the confidentiality constraints imposed at the time of its creation. Accordingly, the sponsor's confidentiality statute will apply to the data, e.g., the confidentiality provisions of the Internal Revenue Code or CIPSEA. In some situations, Title 13 confidentiality may apply simultaneously with the sponsor's confidentiality.
- b. The terms of the agreement under which the data are provided and the federal statutes referenced in B.2. of the General Guidelines may impose additional limitations on use of the data.

2. Confidentiality Applicable to Deliverables:

The extent to which Title 13 confidentiality applies is dependent upon several factors, including: (1) the authority under which the Census Bureau acquires the data; (2) the purpose for which the data are provided to the Census Bureau; (3) the nature of the statistical services performed on the data base; and (4) the extent to which the data to be returned to the sponsor are identifiable.

In addressing this issue as it pertains to an agreement for services, there are a few tenets that are immutable -

- a. If the sponsor's data are commingled with Title 13 confidential data, then the entire commingled data set, in its identifiable format, is subject to Title 13 confidentiality.
  - i. Under such circumstances, only non-identifiable data could be returned to the sponsor.
- b. If Title 13 confidential data are used to draw the sample for a sponsor's survey, then the identifiable data produced by the survey are subject to Title 13 confidentiality.
- c. If Title 13 data are used to revise/enhance the sponsor's data, then no identifiable data can be returned to the sponsor.
- d. Where a database is supplied by the sponsor, and no Title 13 data are involved, then no Title 13 confidentiality applies, and identifiable data may be returned to the sponsor.
  - i. There may be some instances where the Census Bureau enhances/revises a sponsor's data set by applying statistical/geographic methodologies, which methodologies are derived from Title 13 protected data. The simplest example of this would be geocoding a sponsor's list of establishments using the sponsor's addresses. While the geocoding process was developed using Title 13 protected data, the process itself does not disclose identifiable, Title 13 protected addresses. Were the Bureau, on the other hand, asked to correct the addresses for the sponsor's list of establishments, this would constitute a disclosure of Title 13 protected data. Depending upon the nature of the statistical/geographic process involved, the sponsor's data, as enhanced, could be returned to it in identifiable form. Any agreement involving the use of Title 13 derived processes of this nature, must first be reviewed by the Policy Office and the Legal Office of the Census Bureau.

**C. Title 13 Confidentiality When Data Provided for a Dual Purpose:**

In rare instances, the Census Bureau might receive a data set for a dual purpose --- both as a data collection under 13 U.S.C. 6 and as a base structure for statistical services provided under 13 U.S.C. 8. In such

circumstances, the data would have a dual nature, and the confidentiality attending the data would be determined by the authorities under which the data were obtained and the uses to which put.

1. As used for a Title 13 data collection purpose, the data would be subject to Title 13 confidentiality.
2. The data set, while serving as the base structure for statistical services being worked on under either 13 U.S.C. 8, assuming no Title 13 data are implicated in the services performed, would remain the property of the sponsor, and Title 13 confidentiality would not apply.
  - i. Under the usage described in 2. above, the data set would be protected only under the confidentiality provisions applying at the time of its creation, if any, and the protection provided by the Trade Secrets Act, the Freedom of Information Act, and the Privacy Act.
  - ii. Where data are provided for a dual purpose, Census Bureau personnel should maintain the distinction and assure that the data are treated separately.

### **Making a determination**

The following questions represent decision points for determining what protections apply to administrative records obtained by the Census Bureau:

1. What are the legal (statutory/regulatory) requirements applicable to the agency providing the data? (The Census Bureau generally agrees to comply with the provider's legal restrictions, as well as its own.)
2. What are the terms of the Agreement with the provider as they define the Census Bureau's access to and use of the administrative records? Does the Agreement permit re-release to a third party, or are they to be returned to the provider? (Generally, data cannot be re-released or returned if enhanced with identifiable Title 13 data.)
3. What is the base unit of the file data ---- person/household, establishment, locator (information that may locate a specific person or business)? See Appendix A (Glossary) A for a complete definition of Locator data). (Data on individuals may be covered by the Privacy Act, but establishment (business) data will not. Locator data will generally be public record.)
4. Will the data be used for a Title 13 purpose and in what manner? (The answer to this question may depend upon the kind of data being provided. An address list purchased from a commercial source or obtained from the Postal Service may not be covered by Title 13 until actually commingled with Census Bureau confidential data. IRS data to be used, instead of collecting the same information under Title 13, will automatically be covered by Title 13, as well as the Internal Revenue Code.)

### **Resolution**

Given the many factors that can enter into a particular determination, it may not be possible for a program area to reach a conclusion. In such cases, any questions about access and use of particular administrative records should be referred to the Policy Office, which will consult with the Legal Office in making a determination.

## Chapter 8. Security Protocol

### Security Controls for Information Systems That Process Sensitive Data and Administratively Restricted Data

Census Bureau information systems and processes must satisfy requirements in the Federal Information Security Management Act, Title III of Public Law 107-347 and the Office of Management and budget (OMB) in Circular A-130, Appendix III, *Security of Federal Automated Information Resources*, in which the head of each federal agency is to establish and maintain an active security program for managing computers and automated information, as well as a minimum set of controls to be included in federal automated information security programs. These controls meet standards in the National Institute of Standards and Technology (NIST) Special Publication (SP) 800-53, “Security and Privacy Controls for Federal Information Systems and Organizations.” The Census Bureau uses the NIST Special Publication 800-64, *Security Considerations in the System Development Life Cycle*, and the NIST Special Publication 800-37, “Guide for Applying the Risk Management Framework (RMF) to Federal Information Systems” to implement and manage security risks to information systems via manual and automated continuous monitoring of information security systems that enables near real-time assessment and remediation and ongoing authorization of information systems.

System security packages for each information system containing sensitive and administratively restricted data are prepared in accordance with NIST SP 800-18, “*Guide for Developing Security Plans for Federal Information Systems*,” under the direction of the Census Bureau Office of Information Security in compliance with the most recent version of the “*Census Bureau IT Security Program Policy*,” “*Required Security Controls for Census Bureau Information Systems*,” the “*Department of Commerce (DOC) IT Security Program Policy*,” and “*Commerce Interim Technical Regulations (CITRs)*,” System security plans address minimum security requirements by selecting security controls based on the security impact of the information system (low, medium, or high). The Census Bureau tailors security control baselines for information systems based on scoping guidance and common controls. Each sensitive information system security plan contains technical control implementation for identification and authentication, access controls, audit and accountability, and system and communication protection, as well as operational security control implementation for physical environment protection, personnel security, secure configuration management, system and information integrity, system maintenance, information security awareness and training, incident response and contingency planning. Please refer to the Office of Information Security website “*Access Control Procedures*” for general information, and each program area’s information system documentation for specific security control implementation and compliance.

### System Security Officers, System Owners and Authorizing Officials

Authorizing Officials (Associate Directors for Census program areas) formally designate Information System Owners (SOs) for information systems, and System Owners formally designate Information System Security Officers (ISSOs). The personnel in these significant information security roles are responsible for coordinating the security efforts of their information systems, for implementing security in accordance with information security policies. Among other things, they are responsible for implementing and monitoring security control compliance to ensure proper safeguards are in place for the protection of sensitive data, ensure personnel are trained on information security regulations and procedures, and ensure personnel review and acknowledge information security policies (such as acceptable use, rules of behavior, e-mail, and Internet policies). Authorizing Officials are responsible for new and ongoing information system Authorization to Operate and acceptance of risk. Plans of action and milestones (POA&Ms) are developed and remediated for vulnerabilities resulting from assessments, audits, and automated scans.

## Telecommunications

All Census Bureau communication with computer processing systems that store administrative data about businesses or people and household use either point-to-point dedicated lines and/or one-time pad or FIPS encryption.

## Auditing

Information system logs record user access to online storage of sensitive and administratively restricted data at the file and database level. These logs meet guidelines established by the “*Census Bureau IT Security Program Policy*,” and IRS Publication 1075, “*Internal Revenue Service Tax Information Security Guidelines for Federal, State, and Local Agencies*.”

## Labeling of Data Displays

Printed material, screen displays, and other media containing sensitive and administratively restricted data are marked with the appropriate label (i.e. “Disclosure Prohibited Title 13, U.S.C.” or “Disclosure Prohibited Title 26, U.S.C.”), warning banner, or directory/file name in accordance with “*Census Bureau IT Security Program Policy*,” and IRS Publication 1075, “*Internal Revenue Service Tax Information Security Guidelines for Federal, State, and Local Agencies*.” Electronic and printed materials are kept in secured Census Bureau facilities under restricted, authorized access/lock and are destroyed according to Media Protection security controls described in each program area’s sensitive System Security Plans, and as outlined in the “*Census IT Security Program Policy*” and “*Media Protection Guidelines*” on the Office of Security Information website.

## Electronic Storage

All electronic media containing sensitive and administratively restricted data are stored in secured Census Bureau facilities in accordance with Census and DOC IT Security Program Policies, IRS Publication 1075, and the Acceptable Use Policy for U.S. Census Bureau Information Technology Resources. Electronic media is destroyed according to Media Protection security controls described in each program area’s sensitive System Security Plans, and as outlined in the “*Census IT Security Program Policy*” and “*Media Sanitization Procedures and Guidelines*” on the Office of Security Information website.

# Glossary

---

**Active project status** refers to projects in which activities are occurring at the present time. The only projects that are classified as “active” are those approved through the U.S. Census Bureau’s project review process. Active projects are not subject to additional internal Census Bureau review, but must be included and updated as needed in the Data Management System (DMS).

**An activity** is an element of work that is required by the project. It requires resources and takes time to complete. Activities have expected durations, costs, and resource requirements and may be subdivided into tasks. For example, a project may require the acquisition of data from two sources. The project phase is "execution", the work package is “acquisition”; and the activities might include the acquisition of data from source 1 and the acquisition of data from source 2. Activities can further be divided into tasks and subtasks.

**Address data** are a subset of locator data. It describes one or more data fields that denote the physical or mailing location of an individual or business entity. Address information includes, but is not limited to, rural and STAR route designators, P.O. Box or Postal Drawer designators, city-style addresses with street name and number; apartment, unit, or lot number; and post office and postal state information and ZIP Codes.

**Administrative data** refer to micro data records contained in files collected and maintained by administrative (i.e., program) agencies and commercial entities. Government and commercial entities maintain these files for the purpose of administering programs and providing services. Administrative data are distinct from systems of information collected exclusively for statistical purposes, such as those the Census Bureau produces under authority of Title 13 of the United States Code (U.S.C). For the most part, the Census Bureau uses, and seeks to use, administrative data developed by federal agencies, but information from state, local, and tribal governments, as well as commercial entities, are also used.

**The Administrative Records Coordinator (ARC)** in the Policy Coordination Office (PCO/DIR) provides overall direction and coordination to program areas regarding internal administrative data policy and external, data-supplier partnerships.

**Administrative Data Projects** are those that make use of administrative data. An example of such a project is “Current Population Survey – SSN Validation.”

**Agreements** for administrative data are signed documents between the Census Bureau and other agencies to acquire restricted state or federal data or data from vendors.

**Approved projects** are projects that have been reviewed and approved using the Census Bureau’s project review process specified in Section 3 of this handbook. All approved projects are documented and tracked in the DMS.

**“Business” or federal economic data** refer to data collected on economic entities that produce goods or services. The Census Bureau’s economic programs use the business concept broadly to include commercial enterprises and noncommercial organizations or institutions, employers and nonemployers, and private-sector and government entities. The business concept *excludes* private households with paid employees whenever those employees primarily support the operation of a residence.

A business organized as a sole proprietorship and the individual who owns it are one and the same. Therefore, administrative information about such a business also identifies the owner. The economic programs' statistical activities are concerned exclusively with the entity's business activities and characteristics. The disposition of information about sole proprietorships, including name, address, and taxpayer and other identification numbers, is economic rather than personal.

Economic statistics are the collection of numerical aggregates (macrodata) that describe the operating, financial, ownership, and other characteristics of businesses. When compiled for the purpose of describing populations of economic units, this concept includes data about the demographic characteristics of legal entities (natural or legal persons) that own businesses.

**Business Identifier** – A data element that identifies a business establishment, i.e., the name of the establishment or the Employer Identification Number (EIN) on a specific record. In the case of non-employer establishments, an individual's name and Social Security Number (SSN) are considered business identifiers (as opposed to personal identifiers, defined below).

**Business Register** – The business frame for most economic censuses and surveys, and it is comprised of administrative data from the Internal Revenue Service (IRS), the Social Security Administration (SSA), and the Bureau of Labor Statistics (BLS). In the course of building the Business Register, some of the administrative data the Census Bureau receives also contain information about individuals.

**Center for Administrative Records Research and Applications (CARRA)** staff provides central coordination for primarily demographic administrative data file acquisition, custodianship, and preprocessing.

**Commercial Data** – Also referred to as third-party data. These data are demographic, economic, or geographic data that the Census Bureau acquired directly from a commercial source, such as a frame of telephone numbers and email addresses.

**Commingled data sets** are files that contain administrative data, such as tax data, along with Title 13-protected data. Such files remain commingled even if the administrative data use was limited to the sample selection phase. They are typically subject to both Title 13 and any additional, data-supplier imposed restrictions.

**Completed projects** are no longer active, and no additional activity is indicated.

**Data Integration Policy Committee (DIPC)** is the primary group that monitors and evaluates administrative data procedures and activities. The Committee is responsible for keeping this handbook up-to-date and for considering any changes to related procedures. Members represent each directorate with a role in administrative data research or use.

**Data Linkage (also known as record linkage)** – For the purpose of the *Data Linkage Policy*, data linkage includes joining individual, household, or business records in two or more files.

**The Data Management System** is the system of record for Administrative Data Projects for the Census Bureau. All projects that use Administrative Data must be recorded in the system. Datasets, work locations, project methodology, project outputs, Census employees and SSS and servers where projects are conducted are included in each project record.

**Data Stewardship Executive Policy (DSEP) Committee** – The DSEP Committee ensures the Census Bureau maintains its commitment, by fulfilling the legal, ethical, and reporting obligations levied by the Title 13 of the U.S. Code, the Privacy Act, and other applicable statutes, including those of

governmental and other suppliers of data to the Census Bureau. This committee serves as the focal point for decision-making and communication on policy issues related to privacy, security, confidentiality and administrative records. It oversees the activities of several staff committees that focus on these important issues.

**The Disclosure Review Board (DRB)** – sets disclosure avoidance methodology and policy for statistical projects released by the Census Bureau. The DRB is consulted on the feasibility of administrative data project proposals and reviews summarized statistical products for the correct application of disclosure proofing methodologies.

**Disapproved projects** receive disapproval during the project review process. Once a project has been disapproved, the researcher is invited to submit a different proposal.

**Economic Statistical Methods Division (ESMD)** – ESMD’s Register Operations Branch provides central coordination of Economic Directorate administrative data file acquisition and pre-processing to support the Business Register. In addition, the **International Trade Management Division’s (ITMD) Data Management and Stewardship Staff (DMSS)** coordinate the secure transfer of administrative data files and is responsible for proposal review and data custodianship.

**External projects** predominately meet a Title 13 purpose, but are managed by individuals granted special sworn status by the Census Bureau. They are typically funded by external sources. These projects are usually conducted at the Center for Economic Studies or a Federal Statistical Research Data Center.

**Federal Statistical Research Data Center (FSRDC):** Secure facilities which provide the data user community with limited access to non publicly available data collected through a variety of Census Bureau surveys and censuses. Each FSRDC is a partnership between the Census Bureau's Center for Economic Studies (CES) and a local-area non-profit organization. The FSRDCs were established to improve the quality and utility of Census Bureau data by encouraging and supporting partnerships with researchers throughout government, academia, and business to conduct analytic research of mutual benefit and meeting statutory requirements.

**Federal Tax Returns and Return Information (FTI)** is defined in IRS Publication 1075 and 26 U.S.C., Section 6103. It refers to all data about businesses or people received from the IRS. Census Bureau projects using FTI are subject to the disclosure provisions of the Internal Revenue Code — Title 26, U.S.C., as well as the Census Bureau’s authorizing statute — Title 13, U.S.C. Files that contain any FTI, along with Title 13-protected data, are referred to as “commingled” data sets.

**Information Owner** – The Information Owner is an agency official with operational authority for specified information. The Information Owner is responsible for establishing the rules for appropriate use and protection of the subject information (e.g., rules of behavior) and retains that responsibility even when the information is shared. Ownership is at the survey and program level. The Information Owner approves or disapproves use of the data for both production and research uses. Information that is co-mingled may have more than one owner.

**Internal projects** are carried out in Census Bureau facilities by Census Bureau employees. In rare instances, an internal project may also involve individuals working on the project who are not Census Bureau employees, but who have Special Sworn Status. The majority of internal projects are routine, ongoing projects essential to the planning, execution and analysis of the Census Bureau’s core programs. Internal projects also develop new data products to measure the U.S. population and economy through data linkage and modeling.

**Joint projects** are those projects the Census Bureau conducts in collaboration with other federal agencies to predominately accomplish work under Title 13 while also meeting the needs of another federal statistical agency. A team consisting of a representative from each contributing agency typically manages joint projects. Both agencies contribute data and other resources. Joint projects directly enhance the Census Bureau's ability to conduct its censuses and surveys. Other federal agencies provide access to their own databases as part of these projects. Joint projects differ from reimbursable projects in that both employees of the Census Bureau and the other agency may have direct access to Title 13 data. Employees of the other agency must have Special Sworn Status.

**Locator data** or address data, alone or in combination with other data, may be used to locate a specific person or business. These identifiers include latitude and longitude points, telephone numbers, addresses, and address information.

**Master Address File (MAF)** – This is the decennial census and demographic survey frame, comprised of locator administrative data.

**Ongoing projects** – are typically sets of discrete production activities that are periodically repeated within a program. The project cycles are defined at their inception as having specific starting and ending points, with the associated activities repeated periodically. Predefined production cycles could be quarterly, annually, biennially, quinquennially, decennially, or some other period. As long as the methodologies, types of data, and data linkage strategies used to create a product are unchanged over time, the project is not “new” despite the fact that it occurs more than one time. However, all ongoing projects have a not to exceed date of 5 years. Every 5 years, the project must be renewed. If the project uses FTI, the IRS must approve the renewal of the project and a new 5 year period is set. Annual reporting to IRS on the status of these projects is required as well. Examples of ongoing projects include production of the annual Intercensal Population Estimates and maintenance of the Business Register.

**One-time projects** are not ongoing. From their inception, it is anticipated that these projects will be executed only once. Research projects are often one-time activities. If such a project is successful, and the program area chooses to execute it again, it would be considered a new project and would therefore require review and approval.

**Pending projects** are those awaiting project approval.

**“Person” or demographic data** are data about people or households, including counts, and demographic and housing characteristics. Personal identifiers include individual and householder names and social security numbers (SSN). These two types of identifiers allow identification of specific individuals. Demographic (person/household) statistics are the collection of numerical aggregates (macrodata) that describe the social, economic, and other characteristics of human populations.

**Personal identifiers** are defined as (1) individual and householder names and (2) SSN. These two types of identifiers easily allow identification of specific individuals. However, other personal identifiers can work in concert to achieve the same result (e.g., address, phone number, birth date, etc.)

**A Project** is a set of activities that has a distinct mission and clear starting and ending points. A project has a life cycle that proceeds from conception to planning, execution, and termination. There is generally a single point of contact for each project: the Project Contact. A project is often part of a broader program and is a building block in the design and execution of program goals. Dependent relationships may exist between projects. For example, one project may require output from another. An example of a project is “Conduct 2001 Current Population Survey.”

**The Project Contact (PC)** is the individual responsible for documenting a project proposal and shepherding it from the inception of the project to the review stage through implementation. The PC is also required to conduct annual reviews and update the project as needed to keep it current.

**Project Review Coordinator (PRC)** –The PRC is a role in the Data Management System (DMS). Project Review Coordinators coordinate the review and approval process for new administrative data projects and for amendments to administrative data projects. By contrast, Research Project Coordinator (RPC) is a formal job title in the Research and Methodology Directorate. This job title applies to several individuals in ADRM's Planning, Coordination, and Outreach Office. Research Project Coordinators coordinate the review process for all external Census Bureau projects, and serve as PRCs for administrative data projects in the DMS.

**Record Linkage (also known as data linkage)** – For the purpose of the *Data Linkage Policy*, record linkage includes joining individual, household, or business records in two or more files.

**Reimbursable projects** – Reimbursable projects are requested and funded by, in whole or in part, another agency or organization. Reimbursable projects are usually conducted to provide a more complete picture of the topics covered by the Census Bureau's censuses. Reimbursable projects must support the mission of the Census Bureau and be authorized by Title 13. As with all projects using FTI provided by the IRS, reimbursable projects using FTI may only be conducted under the authorization of Title 13.

**Research Project Coordinator (RPC)** – The RPC is a formal job title in the Research and Methodology Directorate. This job title applies to several individuals in ADRM's Planning, Coordination, and Outreach Office. Research Project Coordinators coordinate the review process for all external Census Bureau projects, and serve as Project Review Coordinators (PRCs) for administrative data projects in the DMS. The PRC is a role in the Data Management System (DMS). Project Review Coordinators coordinate the review and approval process for new administrative data projects and for amendments to administrative data projects.

**Special sworn status (SSS)** is conferred upon individuals to whom the Census Bureau approves access to confidential Census Bureau data in furtherance of Title 13 work. Project Contacts must ensure that specific background check and training requirements are met prior to SSS individuals accessing Administrative data.

**Terminated before completed** refers to projects that have been halted due to a variety of reasons such as lack of resources or the project was abandoned and is not expected to resume.

**Work Packages** are deliverables. Work packages may be divided into activities and used to identify and control work flows in the organization. Each project phase can be broken into specific actions required for completion. Work packages are SMART: "Specific, Measurable, Agreed upon, Realistic, and Time Bound Actions." As an example, the project phase "execution" can be divided into work packages exhibited below:

- Procure materials
- Build and test operations
- Develop a support structure
- Produce a system
- Verify performance
- Modify the system as required

- Work packages often represent a group of activities.

# Example of Annual Review Instructions

---

FROM THE DESK OF HAMPTON WILSON III, ADMINISTRATIVE RECORDS COORDINATOR

## To Administrative Records Project Contacts:

As we rapidly approach the end of another reporting year (December 2016), it is time to complete the annual review of all of your administrative records projects. Please share this message with your Alternate PCs and work together to ensure that the project information is up-to-date.

**All project reviews and updates must be completed by November 1, 2016.**

## General Instructions for Key Updates:

To complete the annual review, you must access the DMS and review all projects for which you are the designated PC or Alternate PC. We ask that you verify all information in your project description, but most importantly review and, if necessary, update the following items:

**BOC Employees Supporting the Project, Special Sworn Status (SSS) Supporting the Project.** . Please ensure that these listings include the names of the Project Contact and Alternate PC. For each employee supporting the project, you must specify start and end dates. For each SSS supporting the project, you must specify start and end dates, as well as function performed, whether a government employee and/or compensated, work location, and affiliation.

**Ensure that the Affiliation for ALL SSS is correct.** Every SSS must have the Affiliation field completed. We must report to the IRS based on the Affiliation field so it is very important that the correct affiliation is completed in the DMS for accurate reporting.

**Start and End Dates.** Note that if the project has an end date prior to 12/31/2016, and the status is Approved, the project must be completed/terminated or the end date needs to be extended. To close out or terminate a project, scroll to the bottom of the project and use the drop down box to select "Project Complete" or "Terminate Project". Time extensions require, at a minimum, formal Census Bureau approval. For projects that use IRS data, the extension must be approved by IRS. If you need an extension, contact Brian Holly (x31883) immediately for assistance.

**Computers Housing the Project,** and associated security plan number. If you need to add a computer location and it is not on the selection list, please use the "Report a Problem" tool on the DMS teamsite and provide the name and security plan number to add the computer to the list.

Once you have verified that all project information is up-to-date, **record an annual review of your project.** Do this by selecting "Record an annual review of this project." A pop up screen will remind you to verify all pertinent information. If completed, select "Yes" then **SAVE** your project.

## Detailed Instructions for Further Project Updates:

Project Contacts have the ability to update some project information on an ongoing basis. However, some changes require approval by the Administrative Records Coordinator (ARC) or the Project Review Coordinator (PRC) for this update.

Items that you can update independently include: Project Contact, Alternate PC, PC's Supervisor(s), Computers Housing the Project, and Employees or Special Sworn Status individuals working on the project. We do, however, request that you **review all elements** of your project for accuracy at this time. If you need to make any other changes

or additions to an already approved project, please use the “Request Amendment” feature at the bottom of the project form. Click the “Submit Request” button and complete the form indicating what you would like to change in your project that will require the amendment. If the changes or additions are deemed appropriate, we will unlock your project so you can make the necessary updates. **(Projects will remain unlocked for no more than 14 days.)** Once you have finished updating your project, submit your project for review. If your project uses IRS data, any changes in scope to the project may need IRS review and approval.

Please note that any changes to the scope of the project (i.e., approved project requires a new dataset, previously approved methodology requires modification, etc.) may constitute a need for creating a new project (see the Administrative Records Handbook, Section 1.3.2) and will require re-authorization before the work can proceed.

Thank you for your assistance with this vital activity!

# Checklist of Responsibilities for Project Contacts

---

## 1. Initial Project Development

- Write up project proposal.
- Identify datasets.
- Identify dataset availability at Census.
- Request new dataset if not currently at Census.
- Record Linkage Checklist completed, if required.
- Predominant Purpose Statement completed, if required.
- Submit project for review.

## 2. Agreement Negotiation & New Data Acquisition

- Coordinated through the Administrative Records Coordinator in PCO/DIR.
- All agreements must go through the proper channels for legal approval, if required by PCO/DIR.
- Administrative data set must be tied to an agreement or cannot be used in the project.
- Work with Information Owner through data acquisition process.

## 3. Project Review & Approval

- PC may select a Division Reviewer or Special Reviewer (Reviewer outside of Division if deemed necessary).
- Answer questions from the Division Reviewer, Division Chief, Project Review Coordinator, Project Review Team, or other agency as needed.
- Update the PPS and/or project as needed and participate in meetings or conference calls by other agency as needed.

## 4. Project Implementation

- Monitor Data Stewardship training of project supporters.
- Monitor Title 26 training status of new and project supporters.
- If training is NOT current, data access will be revoked (automatically or at PC's request).
- Responsible for ensuring that research on the project conforms to scope and datasets approved for the project.
- Inform PRCs of any plans for a Public Use File (PUF) or synthetic data early in project development because these may require special approvals.
- Ensure SSS are renewed as needed with help of COR.
- Ensure server list(s) on the project is current.
- Ensure project supporter list is current.
- Ensure project information is current.
- Conduct an Annual Review on the project.
- Ensure Disclosure Avoidance.
- Review and clear all interim papers/presentations through the DRB that may be written during the course of the project.

Submit annual progress reviews.

**5. Post-Project Analysis & Closeout**

- Upload to the DMS any results documents, papers, request for disclosure review by the DRB, disclosure-proofed output or any other pertinent documents.
- Upload to the DMS the completed Post Project Certification (PPC).
- Close out project in DMS.
- Ensure that data disposition is carried out as required by the agreement and Census Bureau Policy (i.e., data is destroyed or returned to owner).

# FSRDC Project Research Proposal Guidelines

---

February 2, 2015

Persons wishing to conduct research at a Federal Statistical Research Data Center (FSRDC) must submit a research proposal to the Center for Economic Studies (CES). The following guidelines describe the research proposal submission process.

Each FSRDC is a joint venture between the Census Bureau and a research institution. All FSRDCs charge project fees for cost recovery. Some institutions maintain affiliations with their local FSRDC that cover fees for their faculty, staff and/or students. Researchers will want to discuss the fees with the FSRDC Administrator and/or Executive Director for the FSRDC(s) they intend to use.

After you and your administrator have developed your proposal, the proposal will be submitted to the CES for Census review. Researchers on approved projects must also complete a background investigation, including fingerprinting. It can take a number of months to complete these steps before you might expect to begin work.

Census has also partnered with the Agency for Health Care Research (AHRQ) and National Center for Healthcare Statistics (NCHS) to host health data projects in the FSRDCs. Proposals to use AHRQ or NCHS data are submitted to these agencies (see Appendix A for web links).

## Table of Contents

I.	The Proposal Process .....	1
A.	Preliminary Proposal Development.....	1
B.	Final Proposal Submission.....	1
C.	Proposal Content Guidelines.....	1
1.	Project Description.....	2
a)	Introduction.....	2
b)	Methodology .....	3
c)	Data Sources .....	4
d)	Project Output and Disclosure Risk Section .....	5
e)	Project Duration and Funding Source .....	6
2.	Benefits Statement .....	6
3.	Abstract.....	10
II.	Census Review.....	10
A.	Proposal Review Process .....	10
B.	Post Census Approval.....	12
	Attachment 1: Preliminary Proposal Template.....	14

Instructions for Preliminary Proposal ..... 15  
Example Preliminary Proposal..... 17  
Attachment 2: Predominant Purpose Statement ..... 19  
Appendix A: FSRDC Data..... 23  
Appendix B: Useful Links..... 24

## I. The Proposal Process

### *A. Preliminary Proposal Development*

Researchers who wish to develop a proposal to conduct research at a Federal Statistical Research Data Center (FSRDC) should first contact the FSRDC administrator at the primary center where the research will be conducted (see list of [FSRDC contacts](#)). The researcher should discuss the proposed project with the administrator to determine whether the research fits with the Bureau's mandate, is feasible, is likely to provide benefits to Census Bureau programs under Title 13 of the U.S. Code, is designed to produce model based statistical estimates, and that the research output will not pose undue risk of disclosing information about survey respondents.

Once the FSRDC administrator and the researcher have agreed it is sensible to proceed, the researcher submits to the administrator a completed Attachment 1: Preliminary Proposal. The administrator will review the summary, advise the researcher of any suggestions for refinement, and submit it to CES in order to generate a project tracking number.

CES encourages researchers to assess carefully the time period over which they request access. Requests for time extensions beyond the agreement end date undergo careful scrutiny, must be justified, and generally are granted for circumstances beyond the control of the researchers (e.g. unexpected illness). A typical proposal duration is 36 months. Requests for longer duration need to be well justified.

The proposal should include the PI's names as well as the names of all researchers, including RAs, associated with the project. We can host foreign citizens in the FSRDCs if they live in the U.S. Researchers that wish to work in the FSRDC that do not reside in the U.S. should discuss their circumstances with the FSRDC administrator early on. Doctoral students submitting a proposal for dissertation work are asked to include their primary advisor as a co-principal investigator. CES recommends that the advisor also apply for Special Sworn Status (see below).

### *B. Final Proposal Submission*

The final proposal consists of three separate documents in rtf (rich text) or doc/docx (MS-Word) format: project description, benefits statement, and abstract. These documents are submitted to CES through your FSRDC administrator (e.g. email). It is expected that the proposal development process will include exchange of multiple iterations of the proposal and, in particular, the benefit statement (called the Predominant Purpose Statement, PPS) with your FSRDC administrator before the final proposal is ready to be submitted. Your administrator is a resource for assistance in clarifying a range of potential issues, including topics related to feasibility and Census benefits.

### *C. Proposal Content Guidelines*

The three document files should conform to the following specifications.

## 1. Project Description

This document should describe the nature of the research question(s), the methodology (including model equations to be estimated, how model variables will be measured, and hypotheses to be tested), Census and non-Census datasets to be used, expected outcomes, and a list of references cited. The proposal should be aimed at a competent social scientist who is not necessarily a specialist in the field or on the topic within the field. Though the benefits to Census are presented in a separate proposal document (discussed in the next section), you should integrate into the proposal as appropriate how your research benefits the Census Bureau. Please keep in mind that this is a proposal for secure access for statistical analysis using internal Census data; proposals written for other purposes (e.g., funding grants) may require significant adaptation. The proposal should:

- Contain a title and the names of all researchers at the top of the first page.
- Include appropriate headings and subheadings throughout the document to assist reviewers in following the proposal narrative.
- Not include any appendices unless approved in advance by the FSRDC administrator.
- Be limited to no more than fifteen (15) single-spaced pages or thirty (30) double-spaced pages (inclusive of references).
- Use a font size of at least 11 point and should have at least one-half inch margins all around. Twelve point font and one-inch margins are preferred.
- Include page numbers.
- Should be written in revisable form text: rtf, doc, or docx format.<sup>1</sup>

The full proposal should contain the following sections; however, researchers are free to add additional sections as they see appropriate.

### a) Introduction

This section should motivate interest in your project and provide an overall description of your research. A full literature review is not necessary; some references are fine. Make a convincing statement about why you need non-public data and briefly describe what you will do with the data. The introduction should also state that the primary purpose of this research is to benefit the

U.S. Census Bureau. It is also helpful, where appropriate, to connect the proposal research to activities and initiatives performed by the Census Bureau.

---

<sup>1</sup>Some data partners request revisable form documents in order to perform electronic comparison in case they receive multiple versions.

All research must have a statistical purpose. In particular, Title 13 statutes specifically prohibit the use of census micro-data for legal or regulatory purpose.<sup>2</sup> Please do not include analysis in your proposal that has or can be perceived to have a regulatory purpose. Proposals are also not expected to make specific policy recommendations.<sup>3</sup>

**a) Methodology**

This section describes your empirical specification, the statistical equations you will estimate, why the Census data is needed, what key variables are needed, how the data will be used, and how all datasets fit together.

The requested Census data needs to clearly map into the methodology section. Even though “Census Provided Data” is a separate section of the proposal (below), there should be a very clear connection between the methodology and the data. The purpose for each requested dataset and its role in the analysis should be clear. Many researchers overlook the “data crafting” aspect of the methodology. You need to be as detailed as possible in describing how measures will be constructed, and how the data will fit together. This doesn’t need to be a separate section or separate discussion per se, but it needs to be discussed where appropriate.

Describing the linking of datasets is important, and at what unit of observation or aggregation the linking will occur. For example, if one is linking the Annual Survey of Manufacturers (ASM), collected at the establishment level, to the R&D survey which is firm based, the research methodology should incorporate this. Also, linking survey data (e.g., the ASM and R&D are both surveys) to other survey data may lead to small sample size (and therefore potential disclosure) issues if the survey samples don’t significantly overlap. For demographic data, for example linking the SIPP to externally provided contextual variables, detailing the level of geography at which this link will be performed is helpful.

---

<sup>3</sup> In addition, for proposals using establishment and firm data, we recommend avoidance of the word “policy” altogether. For example, substitute “decision maker” for “policy maker.”

<sup>2</sup> Title 13, Chapter 1, Subchapter 1, section 9, Information as confidential. “Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof... may... use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or make any publication where by the data furnished by any particular establishment or individual under this title can be identified... Copies of census reports which have been so retained shall be immune from legal process, and shall not, without the consent of the individual or establishment concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding.”

Generally speaking, economic (firm and establishment) data sets share a common set of identifiers and can be linked together. The identifiers for demographic (household and person) census and survey data sets are most often designed to be used within the survey year and are not longitudinal, or linkable, across demographic surveys. Record level linkage for economic data can be performed by approved projects in the FSRDC using the Business Register. Record level linkage for demographic data must be performed by internal Census staff.<sup>4</sup> Note that linking cannot generate complete matches of an internal dataset to an external one; this would reveal existence of specific respondents in the Census data, which is a violation of Title 13 protections.<sup>5</sup>

It is important to convey to the reviewers that you have read the documentation that is publicly available about the datasets you are requesting and have thought through some of the implications of the data in relation to your research approach. A list of restricted-use data commonly used by approved projects in the FSRDCs is included in the links given in Appendix A.

Often three sources of information about datasets are helpful: survey questionnaires, Census web pages detailing survey and sample design, and working papers that prior CES and FSRDC researchers have written using the data. See Appendix B. Also, ask your FSRDC administrator to provide additional information or answer questions.

The unit(s) of observation for the analysis and the groups of units for which you will carry out analysis are both important. For each, be as specific as possible. Do you foresee the unit of observation being the establishment, firm, household, or person? Or will your unit of observation be at a more aggregated level (e.g., Census tract or county)? Will you use some combination of units? Additionally, for what groups of units (what "levels") will you specify your models? Will you estimate models at (say) national or sub-national levels; if so, what levels? In addition, will you run separate models by gender, race, household structure, industry, etc.? Information that describes the units of analysis and the groups of units/levels of samples you will be using for your analysis helps Census reviewers assess both the feasibility of your project and the risk of disclosing confidential information. For more on the latter, see section (d) following, and II.A below.

Write out the equations you plan to estimate. You do not need to give the exact functional form you plan to use, but write out in as much detail as possible the general equation and relevant variables, and talk about how the data will fit into the equation. You do not need to list every specification, nor every variable you will be using from each dataset, but discuss how the left and right hand sides of each equation will be measured, and if they are dichotomous, categorical, continuous, etc. Specify the dataset(s) from which these measures will come and how they will be constructed.

---

<sup>4</sup> Circulation within Census of personally identifiable information (PII) such as name, address and SSN data is intentionally very limited, and excluded from the FSRDCs. Record level linkage is performed at Census by internal specialists within only a few departments. Approval for record level linkage to demographic surveys, especially mandatory surveys, can be very difficult to justify, as well as expensive to perform.

<sup>5</sup> Furthermore, for business data with FTI, a complete match results in a violation of the IRS "fact-of-filing" information protection.

Finally, it is important in this section to describe not only the methodology of your research questions but also the methodology of other benefits your project will provide to the Census Bureau. For example, if your project will evaluate the quality of certain variables in the restricted Census dataset, you should discuss in detail how you will evaluate these variables and what types of results you will report to Census.

#### **b) Data Sources**

Data sources should be divided into two sections: “Census Provided Data” and “Researcher Provided Data.”

In the “Census Provided Data” section, provide a bulleted or numbered list (or summary table) outlining the exact data you are requesting. Please consult the online tables in the links in Appendix A to see the data available through the FSRDCs and the years available for each dataset. If a dataset is listed as a separate entity in the table, it needs to be listed separately in your proposal. As examples, “LEHD Data” is insufficient information, instead list the requested datasets individually (e.g., EHF, ICF, etc.); the “Economic Census” includes eight datasets, list individually those requested; for the Longitudinal Research Database (LRD) please instead list the ASM, CMF and LBD.

The years of data needed for each requested dataset must also be stated. If you are requesting years of data not yet released, detail these with an “if available” caveat (e.g. CMF 2002, 2007, and if available 2012). The years of requested data need to be clearly justified, particularly for proposals requesting a long time series. You can either do this in the methodology section or in a paragraph at the end of the data section. A brief summary of this justification is also inserted in section B of the benefits statement (discussed below) for datasets that include Federal Tax Information (FTI).

It is straightforward for the Census Bureau to provide the standard FSRDC datasets that are listed in the links in Appendix A for an approved project. Data collected by the Census Bureau that is not listed is generally unavailable, not prepared for FSRDC use, or procedurally problematic in some way. If a need for non-standard data exists, it is crucially important to discuss this with your FSRDC administrator, and important to incorporate this into your timing plans as requests for non-standard datasets considerably lengthen proposal evaluation.

If you plan to bring in non-FSRDC data to the project, list these datasets in a separate section called “Researcher Provided Data.” The source for your external data should be indicated. Note: public use Census data are considered external data (sourced from public web sites).

After the list of data, it is often useful to include a paragraph summarizing how all the data fit together, and how they will be linked and their expected overlap (if applicable), if not already covered in the methodology section.

### **c) Project Output and Disclosure Risk Section**

In this section, you must describe the output you expect to ask to remove from the FSRDC. This description is important in assessing both the substance of the proposal and the risk of disclosing confidential information. Output from the Census FSRDCs must emphasize modeling output (regression coefficients, standard errors, and the like). Summary statistics (variable means, etc.) are allowed, but only to the extent that they support modeling output. For example, the kinds of output expected to be released are regressions and tables as found in an article in a peer-reviewed academic journal. You should make the point that your project will emphasize modeling output.

Researchers who desire disclosure of large volumes of tabular output should request a special tabulation from the Census Bureau<sup>6</sup> rather than requesting access to micro data at the FSRDC.

Be clear in describing the unit(s) of analysis you will use and the groups of units ("levels") which you will create research results. Are you only using certain industries or demographic groups? Your work will be cleared for release only if it passes certain disclosure criteria,

including minimum sample size restrictions. For example, for a regression, the underlying sample must be large enough, as well as those subsamples as defined by indicator variables.<sup>7</sup>

Note that including indicator variables in your analysis creates additional "cuts" at the sample (i.e., cross tabs), but only if you report the estimated coefficients associated with the indicator variables. Because of this potential issue, we encourage including statements in your proposal, as appropriate, for example: "Specifications will include detailed industry controls; however we will not release these fixed effect coefficients, as all we need is to note in our results tables that we included them."

Disclosure risk is a complex topic, and you will need to discuss your proposed output with your FSRDC administrator as you develop your proposal. See below Section II.A "Confidentiality protection and disclosure avoidance review requirements" for more specifics on what we look for in assessing proposals for disclosure risk.

### **d) Project Duration and Funding Source**

This section should state the proposed duration of the project and expected intensity of use (e.g., 15 hours per week). List sources of funding for the FSRDC lab fees, as well as funding for the research project overall.

---

<sup>6</sup>Demographic tabulations: <http://www.census.gov/population/www/cen2000/sptabs/main.html>  
Economic tabulations: [http://www.census.gov/econ/susb/special\\_tabulations.html](http://www.census.gov/econ/susb/special_tabulations.html)

Researchers often discover that it takes longer than anticipated to become familiar with the data and computing environment. For most projects, duration of at least 24 or 36 months is recommended. Sixty months is the maximum length of a project, however requests for long duration projects will need to be justified. All projects are required to submit a yearly progress report/interim benefits statement.

## 2. Benefits Statement

The benefits statement is contained in the Predominant Purpose Statement (PPS). Proposals must demonstrate that the research is likely to provide one or more Title 13 benefits to the Census Bureau. A research project must demonstrate that its predominant purpose is to benefit Census Bureau programs. If a project has as its predominant purpose one, or any combination, of the following criteria, then it will be considered to have as its predominant purpose increasing the utility of Title 13, Chapter 5 data. Researchers should consult an FSRDC administrator for more information. Any proposals seeking to use datasets that contain Federal Tax Information (FTI), generally speaking the establishment and firm based surveys and census, must also be reviewed by the Internal Revenue Service to ensure that the predominant purpose of the research is to contribute to Census Bureau programs under Title 13, Chapter 5 of the U.S. Code. The following are a list of approved Title 13 benefits.

The first four criteria apply only to projects that do not include Federal Tax Information (FTI), generally speaking the establishment and firm based surveys and census. Projects requesting only demographic data (data for persons and/or households) can include these first four criteria.

1. Evaluating concepts and practices underlying Census Bureau statistical data collection and dissemination practices, including consideration of continued relevance and appropriateness of past Census Bureau procedures to changing economic and social circumstances;
2. Analyzing demographic and social or economic processes that affect Census Bureau programs, especially those that evaluate or hold promise of improving the quality of products issued by the Census Bureau;
3. Developing means of increasing the utility of Census Bureau data for analyzing public programs, public policy, and/or demographic, economic, or social conditions; and
4. Conducting or facilitating census and survey data collection, processing or dissemination, including through activities such as administrative support, information technology support, program oversight, or auditing under appropriate legal authority.

---

<sup>7</sup> If you are using LEHD data, output must have been estimated using a sample that includes at least three states. The following language for LEHD proposals should be used: "We understand disclosed results using LEHD data must have been generated using multiple states."

The remaining nine criteria apply to all projects.

5. Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census, or estimate;
6. Leading to new or improved methodology to collect, measure, or tabulate a Title 13, Chapter 5 survey, census, or estimate;
7. Enhancing the data collected in a Title 13, Chapter 5 survey or census. For example: improving imputations for non-response; developing links across time or entities for data gathered in censuses and surveys authorized by Title 13, Chapter 5;
8. Identifying the limitations of, or improving, the underlying Business Register, Master Address File, and industrial and geographical classification schemes used to collect the data;
9. Identifying shortcomings of current data, collection programs and/or documenting new data collection needs;
10. Constructing, verifying, or improving the sampling frame for a census or survey authorized under Title 13, Chapter 5;
11. Preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5;
12. Developing a methodology for estimating non-response to a census or survey authorized under Title 13, Chapter 5;
13. Developing statistical weights for a survey authorized under Title 13, Chapter 5.

The PPS should address clearly and specifically how the project will provide one or more of the Title 13 benefits listed above. It has no length limitation, however clear and concise presentation is critical. Most commonly, researchers have written the benefits to address two criteria, Criterion 11 and one other.

Most FSRDC proposals include as a benefit preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5 (criterion 11). This is because, if one thinks of the Census survey life cycle, loosely speaking, as a series of four steps: planning, conduct, processing, and dissemination, then your research estimates are a key form of information dissemination. FSRDC based research leverages unique skills of experts in a joint venture to generate new estimates of the population of individuals and firms to complement the variety of Census-produced publications, and further informs Census analysts of if and how the data can support meaningful understanding of the underlying processes of our economy. In writing benefits for criterion 11, include all of the analyses you plan to perform in summary form.<sup>8</sup>

---

<sup>8</sup> For some federal agencies, including the IRS, the PPS is the document and project scope they review and approve.

Your access to the internal microdata is covered under legal statute because of your role as a consultant to assist Census with its work.<sup>9</sup> From a practical perspective, the implementation of this has included a norm for FSRDC proposals to meet at least two of the criteria listed above. Thus, a second benefit is proposed, most commonly an exercise connected to survey planning and processing activities at Census, for example an analysis about the data itself. The ideal benefit analysis leverages the same data preparation work as is intended for the estimates of population.

A good Census benefit accomplishes three elements: (i) a connection to Census program activity, (ii) articulates a specific analysis exercise, its purpose and approach, that can be clearly understood by reviewers and Census staff, and (iii) is focused, generally on only one of the above criteria. Discuss what might be found, why it could be informative, and, in particular, some specifics on how the analysis will be approached.

Related to (i) above, a benefit analysis that can be shown to have a relationship to activities Census is working on or potentially thinking about strengthens its relevance.

Census program staff describe much about their activities on the Census web pages, easily accessible through searches.<sup>10</sup> Note that contributions that benefit other researchers, including Census researchers, are welcome (e.g., documentation contributions, or methods contributions that are of potential interest to academic researchers); however, these might not necessarily have application to Census activities.

### *Census Benefit Example 1*

An example of a Census benefit is to benchmark Census collected data to that of an outside source, to understand the quality of data produced through a Title 13, Chapter 5 survey. By linking Census establishment micro data to external sources of merger and acquisition activity as reported by firms to other federal agencies, projects have investigated the extent to which known acquisition events as identified in the external data are reflected in the Census business register. The terms of an acquisition can vary considerably, and these differences may be related to differences in reporting behavior. Cases of mismatch were explored and documented in a technical memorandum to Census Bureau staff, focusing specifically on the ability of the register to capture ownership dynamics.

---

<sup>9</sup> Title 13, Chapter 1, Subchapter II, Section 23 states “The Secretary may utilize temporary staff, including employees of Federal, State, or local agencies or instrumentalities, and employees of private organizations to assist the Bureau in performing the work authorized by this title, but only if such temporary staff is sworn to observe the limitations imposed by section 9 [Information as confidential] of this title.”

<sup>10</sup> Proposals requesting LEHD data are expected to include benefit to the LEHD program.

The PPS document is formatted in a specific way. See Attachment 2 for the standard template. Unlike Attachment 1, only researchers who will be applying for SSS and accessing the data in the FSRDC should be listed under the heading “Researchers” in the PPS. In the section entitled “Criteria”, specify which of the 13 benefit criteria listed above are relevant to your project. Then enter text under each relevant criterion describing your project’s benefits. Note that it is important not to change the format of the document, nor edit the paragraphs that are pre-printed there. Your FSRDC administrator will assist you with the formatting and content of the PPS.

In the section “Requested Datasets” you should provide information on the datasets you will use in your project. If your project includes FTI datasets, then that section is filled out as follows: (i) enter each FTI dataset and the years, (ii) enter the FTI variables included in the FTI datasets (these are a standardized set for each data file – your administrator will help you with this), and (iii) if your project asks for a long time series, include a paragraph at the beginning of the section justifying the request.

An example for the Census of Manufactures is as follows:

**Census of Manufactures (CMF), 2002, 2007 and if available 2012**

FTI:	Description:
ST	State
IND	Industrial Classification
(SIC/NAICS) LFO	Legal Form of Organization
CFN	Census File Number (includes EIN)
EIN	Employer Identification Number

*Census Benefit Example 2*

An example of a Census benefit centers on the transition of the Decennial Census long form to the American Community Survey (ACS). This change provides advantages, such as ongoing updated information more frequently than once every ten years, however also provides some issues that may require adjustment to population estimates.

Projects have investigated the effect of changes in question design, comparing data from the same question in cases of different wording, and seasonality issues due to the change from data collected in April of the decennial year to ACS continuously collected data for variables such as journey time to work

COU	County
SW	Salaries and Wages (IRS Payroll)
TE	Total Employment
TVS	Shipments (IRS Receipts)

Under the section “Census Bureau Datasets”, list any internal, restricted Census Bureau datasets that your project will use, along with the years of data. Under the section “User Provided Datasets,” list both external proprietary datasets that you will use as well as public use datasets (including public use Census datasets). If you will be linking any of the “User Provided Datasets” to the FTI datasets, describe at what level that linkage will occur (e.g., establishment level, Census tract level, etc.)

If your project includes FTI datasets, then in the “Need and Use of FTI” section, you must list every FTI variable in these datasets and describe why your project needs that variable and how it will be used. An example is provided below.

<b>FTI</b>	<b>Datasets requested that include the FTI</b>	<b>Description</b>
CFN	ASM, CCN, CFI, CMF, CMI, CRT, CSR, CUT, CWH, LBD	Census File Number (Contains EIN)

This field is used to link establishments across Census datasets within a year. For 2001 and earlier years, the CFN is also used to recover the Census firm identifier, embedded within the CFN. These links will be performed to construct a panel of establishments incorporating variables from the ASM, LBD and Economic Census files.

The benefits document/PPS should not be repeated in the proposal. In addition, the PPS and the proposal should not reference each other; they are stand-alone documents. The two documents together with the abstract constitute the proposal submission. In addition, the PPS should be in executive summary form: no equations, no footnotes, and no references/bibliography. Finally, the PPS must be written in the third-person.

Appendix B includes some helpful links to source documents that can provide input to drafting benefit statements. It is important to discuss proposed benefits to Census with your administrator.

### 3. Abstract

This document should be no longer than one single-spaced page (or two double-spaced pages) and it should capture the essence of the project proposal: one or two sentences should succinctly state what the project will do and the data it will use, including all restricted Census datasets and the years of data coverage. The abstract must also address the proposed benefits to the Census Bureau. At the top of the first page, the project title and the names of all researchers must be stated.

## II. Census Review

### *A. Proposal Review Process*

The Center for Economic Studies accepts proposals for review at any time during the year. Both Census Bureau and external experts on subject matter, datasets and disclosure risk review all proposals. Relevant data sponsors and data custodians also review proposals that request certain datasets, for example other federal and state agencies. See the sponsor column in the data tables in the links in Appendix A. For example, proposals seeking to use datasets that contain FTI must also be reviewed by the Internal Revenue Service. LEHD data are sponsored by the 50 states; many states require state review of proposals that request use of their data.

Research proposals submitted to CES are reviewed based on five major criteria.

#### ***Benefit to Census Bureau programs under Title 13***

While each of the criteria is important in determining whether to approve a proposal, the project's potential to benefit Census Bureau programs carries the greatest weight.

#### ***Scientific merit***

This criterion relates to the project's likelihood of contributing to existing knowledge. When available, evidence that a Federal-funding agency such as NSF or NIH has approved the proposed research constitutes one indication of scientific merit.

#### ***Clear need for non-public data***

The proposal should demonstrate the need for and importance of non-public data. The proposal should explain why publicly available data sources are not sufficient to meet the proposal's objectives.

### ***Feasibility***

Feasibility criteria include data availability and statute consistency. Internal Census reviewers will be focused on the ability of the data to support the proposed research, and will be interested in the proposal's discussions of measures and how the researchers plan to combine the data. The proposal must show that the research can be conducted successfully given the methodology and requested data. Here, researchers need to convince reviewers that they sufficiently understand the data they are requesting (unit of measurement, available measures, etc.), that their research questions can be answered in a disclosure-compliant way, that they have thought through and anticipated potential sample size, level of aggregation or other feasibility issues, and that the number of datasets requested is reasonable and well justified.

### ***Confidentiality protection and disclosure avoidance review requirements***

Output from all research projects must undergo and pass disclosure review.

- The Census Bureau is required by law to protect the confidentiality of data collected under its authorizing legislation, Title 13, U.S. Code.
- Tabular and graphical output presents a higher risk of disclosure of confidential information than do coefficients from statistical models.
- Some data files are collected under the sponsorship of other agencies. In providing restricted access to these data, CES must adhere to all applicable laws and regulations.
- Researchers may be required to sign non-disclosure documents of survey sponsors or other agencies that provide data for their research projects.

We look for several things in evaluating proposals for disclosure risk. To do this, the proposed methodology and output must be clearly stated in the proposal, especially the information requested above in Sections I.C.1.b and I.C.1.d. We realize that, because you are conducting research, you probably don't know all the details in advance. Nevertheless, we need enough detail to assess whether the proposed project can succeed without posing undue risk. In particular:

- We need to determine whether there may be possible thin "cells" in your output, where for this purpose, a "cell" is the group of observations (individuals or businesses) underlying any result you may release. To evaluate this, we need clear and accurate information on the types of output to be requested (e.g., models tabulations, graphs) and the units of analysis underlying the output (e.g., individuals and businesses, or groups of these); and the groups (levels) for which you will request output (e.g. industry, gender, geography, and possible crosses of these). For example, detailed industry-by-geography cells become thin very rapidly.
- We need to know whether variables included in models are discrete or continuous; both are allowable, but discrete variables (especially dummy (0, 1) variables) define "cells" that we must take into account.
- We need to know whether you will request graphs or maps, which can be especially problematic.
- We will also look for types of output that are relatively unfamiliar to us, such as from

new statistical techniques. We must try to assess their disclosure risk, and we welcome any insights you may have on this.

Reviewed proposals receive one of two ratings:

- **Approved.** The proposal successfully addresses all of the review criteria mentioned above.
- **Not Approved.** The proposal fails to meet most or all review criteria. Subject to approval of the FSRDC administrator, not-approved proposals may be resubmitted as a new proposal, after addressing reviewers' concerns in a substantial revision.

The Project Review Coordinator (PRC) communicates the outcome of the review process to the researcher, which includes a review synopsis, an explanation for the decision, and copies of the expert reviews.

### ***B. Post Census Approval***

Following Census approval, the PRC will coordinate reviews from data sponsors and other federal agencies (including the IRS) as needed. Approval of research proposals by CES (and the IRS if FTI is requested) is but one step in the process before research can commence. In many instances, CES must obtain permissions to access certain data. This process can range from a few weeks to many months depending upon the data requested.

Once a project has all necessary approvals, all researchers who expect to access FSRDC data must undergo a background investigation including fingerprinting. After completion of the background check, the Census Bureau grants Special Sworn Status (SSS) to each researcher, which subjects them to incarceration of up to five years and/or fines of up to \$250,000 if they knowingly or inadvertently disclose confidential information on individuals, households or businesses. All SSS individuals must take annual trainings in proper Data Stewardship and the use and protection of Title 26 data.

All approved research projects are governed by a written agreement between the researcher(s) and the Census Bureau. The agreement stipulates the start and end dates for the project, and responsibilities of both parties with respect to procedures and practices. Each researcher on the project must sign an agreement with the Census Bureau. If researchers are added to the project after the project starts, these researchers will also each sign an agreement. Separately, an agreement with the FSRDC partner institution may be required, as well as arrangements for FSRDC fee payments.

### **Timing**

The review process is both lengthy and rigorous, requiring researchers to exhibit patience throughout. We continually strive to simplify and compress proposal review times, and are committed to process improvement for proposal review. However, experience has demonstrated that due to the number of steps, participants and data permissions involved, elapse time between the final proposal submission and the actual commencement of research generally includes multiple months. This duration also varies substantially by individual proposal depending on the

scope and complexity of the proposal, the number and variety of datasets, and other agencies involved.

Researchers can help speed up the process by:

- Adhering closely to all practices and procedures for proposal submission.
- Rationalizing and simplifying the scope and data requested in the proposal where appropriate.
- Working closely with their FSRDC administrator on proposal development and on any requested revisions or clarifications to proposals or predominant purpose statements.
- Providing CES with the terms of use for any proprietary external datasets.
- Processing their SSS paperwork quickly and completely.

We are proud of the many research projects conducted over the last three decades throughout the FSRDC network, and are grateful for your interest.

## Attachment 1: Preliminary Proposal Template

Title:

Principal Investigator:

Date:

### *Prerequisites Checklist*

1. I have read these research proposal guidelines.
  2. I have corresponded with an FSRDC administrator about my prospective project.
  3. I understand that the predominant purpose of this project must benefit the Census Bureau data programs.
  4. I understand that there are usually lab fees for using an FSRDC.
  5. I understand that work is performed in a restricted access environment and only cleared results are allowed to leave the FSRDC.
  6. I understand that the approval process may take a considerable amount of time, for example, if review by other agencies or data owners is required or if data availability issues exist.
- I have performed each checklist item above: Please initial \_\_\_\_\_

Duration:

Funding Source:

Description:

Census Datasets:

Years of Data:

Primary FSRDC Location:

Secondary FSRDC Locations:

Benefits:

### *Research Team*

Name	Affiliation	Email	Type	SSS	Citizenship	FSRDC

## **Instructions for Preliminary Proposal**

These instructions provide some context for entering the information about your project in the Preliminary Proposal form. Note that it is not necessary to have all information exact to fill out this form, as the information can be updated any time up until the proposal is submitted for Census review. It is expected that you will update the information as you go along as part of the proposal development activity. The final version of this form will be submitted along with your proposal.

**Title:** The title of your proposal.

**Principal Investigator:** Researcher that will submit the proposal for Census review, and will be the central focal point of communication with Census on proposal and project affairs.

**Date:** Today's date

Please perform each checklist item and type your initials to indicate agreement.

**Duration (months):** Number of months of access to the data in the FSRDC starting from the first day of logon (after project approval) to FSRDC project completion.

**Funding Source:** List sources of funding for the FSRDC lab fees, as well as funding for the research project overall.

**Description:** The description is intended to give Census staff a short idea of what your research project is about, as well as the potential Census benefits.

**Census Datasets:** A list of the internal census datasets requested in the proposal. Please list the acronyms for the datasets from the links in Appendix A. The acronyms are in parenthesis after the dataset names in the tables.

**Years of Data:** The time span of the data you will use for your research project.

**Primary FSRDC Location:** The main FSRDC location from which the project will be administered.

**Secondary FSRDC Locations:** Any additional FSRDC locations researchers from the project will use regularly.

**Benefits:** The relevant benefit criteria the project will address. Please list the benefit numbers as taken from the benefit criteria listed in Attachment 2.

*Research Team*

List the full name, institutional affiliation, email address, and other requested information for all researchers affiliated with the project.

Name	Affiliation	Email	Type	SSS	Citizenship	FSRDC

Type: PI = Principal Investigator, CI = Co-investigator, RA = Research Assistant.

SSS: Y/N. Indicate if the researcher is expecting to undergo the Special Sworn Status background check to enter and use the FSRDC.

Citizenship: Indicate country of citizenship.

FSRDC: Indicate the FSRDC the researcher is planning to use.

## Example Preliminary Proposal

Title: R&D Investment and Plant Productivity

Principal Investigator: John Smith

Date: 03/13/2012

### *Prerequisites Checklist*

1. I have read these research proposal guidelines.
2. I have corresponded with an FSRDC administrator about my prospective project.
3. I understand that the predominant purpose of this project must benefit the Census Bureau data programs.
4. I understand that there are usually lab fees for using an FSRDC.
5. I understand that work is performed in a restricted access environment and only cleared results are allowed to leave the FSRDC.
6. I understand that the approval process may take a considerable amount of time, for example, if review by other agencies or data owners is required or if data availability issues exist.

I have performed each checklist item above: Please initial   JBS                  

Duration (months): 36 months

Funding Source: National Science Foundation

Description: The project will investigate the effect of investment in research and innovation on the evolution of plant productivity over time. Benefits to Census include linking external patent data to explore the extent to which R&D activity is captured in the BRDIS and SIRD surveys.

Census Datasets: BRDIS, SIRD, ASM, CMF

Years of Data: 1992-2009, and if available 2010-2012

Primary FSRDC Location: Boston

Secondary FSRDC Locations: Chicago

Benefits: 5, 11

*Research Team*

Name	Affiliation	Email	Type	SSS	Citizenship	FSRDC
John B. Smith	Belmont University	<a href="mailto:smithj@belmont.edu">smithj@belmont.edu</a>	PI	Y	USA	Boston
Jane F. Parker	Northwestern University	<a href="mailto:parkerj@northwestern.edu">parkerj@northwestern.edu</a>	CI	Y	Canada	Chicago

## Attachment 2: Predominant Purpose Statement

### Predominant Purpose Statement (PPS)

Draft, [insert date here]

9999 - "[Title]"

#### Persons Requiring Special Sworn Status<sup>11</sup>

Researcher 1            Name – Affiliation

Researcher 2            Name– Affiliation

.

.

Researcher n            Name– Affiliation

Additional research assistants may be added to this project after it commences as needs dictate. In accordance with the Census Bureau – IRS Criteria Agreement of June 2012, the Census Bureau will notify IRS Statistics of Income of any additions and/or subtractions of Special Sworn Status individuals within five business days of the action.

**Duration** – [months]

#### Locations

[Name] Census Research Data Center

[Name] Census Research Data Center

**Funding** – [Source(s)]

---

<sup>11</sup> If there are any Census Bureau employees on the project, replace “Persons Requiring Special Sworn Status” with “Persons with FTI Access” and list the employee’s Census Bureau division after the employee’s name.

**Background** [If necessary]

[Description]

**Criteria**

The predominant purpose of this study is to increase the utility of Title 13, Chapter 5 data of the Census Bureau by meeting the criteria listed below and as described.

*Criterion x*                    *[Title]*

[Description]

*Criterion y*                    *[Title]*

[Description]

**Requested Datasets**

**FTI Datasets**

Only the following Federal Tax return Information (FTI) items from the following files will be needed by the researcher(s) to conduct this project.

Dataset 1

Dataset 2

.

.

Dataset n

## **Census Bureau Datasets**

Dataset 1

Dataset 2

.

.

Dataset n

## **User Provided Datasets**

Dataset 1

Dataset 2

.

.

Dataset n

## **Need and Use of FTI**

[Description]

## **Project Completion**

The researcher(s) will provide the Census Bureau with annual progress reports on this multi-year project. After this project is completed, the researcher(s) will produce a CES Working Paper and related research papers for conference presentation and/or journal publication. They will also

produce a CES technical memorandum summarizing the key findings pertinent to Census Bureau Title 13, Chapter 5 programs. In addition, the researcher(s) will produce a Post Project Certification (PPC) summarizing whether and how the criteria described in the project proposal's Predominant Purpose Statement (PPS), as well as any others, were or were not met.

### Project Timeline

The proposed duration of the project is X years. [Additional text here only if necessary]

[An example is given in the table below. **You must include a separate task for completing annual reports at the end of each project year and for completing the post-project**

**certification at the end of the project.]**

Task	Year				
	1	2	3	4	5
Geocode import transaction data to county level	X	X			
Link foreign transactions data to CFS: match CFS to export declarations dataset; identify imports to port locations	X	X			
Build on the firm-level links in LFTTD to establish foreign transactions data links at the plant level	X	X			
Link shipping recipients in transaction data to LBD (to determine NAICS code)	X	X			
Submit annual report to CES	X				
Conduct data analysis for models of industry flow		X			
Draft paper of industry-level estimates (paper 1)		X			
Model entrance of movement of goods flowing through US		X			
Assess goodness of fit, quality of estimates for industry-level estimates		X			
Submit annual report to CES		X			
Conduct data analysis for regional impacts of trade flows			X		
Determine overstatement of imports at ports; (estimate parameters of misreporting)			X		
Draft paper of regional impacts of trade flows (paper 2)			X		
Submit annual report to CES			X		
Finalize paper of regional impacts of trade flows (paper 2) and submit for journal review.				X	
Revise paper of industry-level estimates (paper 1) to address referee reports from journal.				X	

Conduct data analysis for geography of import flows (add endogenous variables to industry-level estimates) – requires feedback from models developed and peer-reviewed in paper 1				X	
Assess US/Canada transaction data				X	
Produce preliminary and final draft of geography of import flows and submit for journal review (paper 3)				X	
Submit annual report to CES				X	
Conduct data analysis on extent of wholesale use by industry					X
Link transaction data to LBD and Economic Censuses to estimate where inputs are eventually absorbed					X
Report limitations of new <i>State Imports</i> data product given results of analyses of imports to port locations					X
Draft paper measuring trade flows and using transaction data(paper 4)					X
Revise papers 2 and 3 to address referee reports from journal.					X
Complete paper 4 and submit to journal.					X
Produce final linked dataset					X
Complete Post-project certification (PPC) and submit to CES					X

## Appendix A: FSRDC Data

CES Data Products: <http://www.census.gov/ces/dataproducts/index.html>

Economic Data: <http://www.census.gov/ces/dataproducts/economicdata.html>

Demographic Data: <http://www.census.gov/ces/dataproducts/demographicdata.html>

Longitudinal Employer-Household Dynamics (LEHD) Data:  
<http://www.census.gov/ces/dataproducts/lehddata.html>

Agency for Health Care Research (AHRQ) Data:  
[http://meps.ahrq.gov/mepsweb/data\\_stats/onsite\\_datacenter.jsp](http://meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp)

National Center for Healthcare Statistics (NCHS) Data:  
<http://www.cdc.gov/FSRDC/>  
<http://www.cdc.gov/FSRDC/B2AccessMod/ACs220.htm>

## **Appendix B: Useful Links**

### **General**

CES Website: <http://www.census.gov/ces/index.html>

FSRDC Locations and Contact Information: <http://www.census.gov/ces/main/contact.html>

IRS Criteria Document: [http://www.census.gov/ces/pdf/IRS\\_Criteria\\_Document.pdf](http://www.census.gov/ces/pdf/IRS_Criteria_Document.pdf)

CES Working Papers: <http://ideas.repec.org/s/cen/wpaper.html>

and older CES Working Papers: <http://www2.census.gov/ces/wp/>

### **Census Benefits**

Writing Benefits Statements: [http://www.census.gov/ces/pdf/Writing\\_Benefit\\_Statements.pdf](http://www.census.gov/ces/pdf/Writing_Benefit_Statements.pdf)

Potential FSRDC Methodological Topics:

[http://www.isr.umich.edu/src/mcFSRDC/Potential  
Research%20Data%20Center%20Methodological%20Topics.htm](http://www.isr.umich.edu/src/mcFSRDC/Potential%20Research%20Data%20Center%20Methodological%20Topics.htm)

Data Collection, Analysis and Dissemination Process Topics (NSF-Census Solicitation):  
[http://www.nsf.gov/pubs/2010/nsf10621/nsf10621.htm#pgm\\_desc\\_txt](http://www.nsf.gov/pubs/2010/nsf10621/nsf10621.htm#pgm_desc_txt)

Census Research Topics: <http://www.census.gov/srd/researchtopics.pdf>

Abowd & Stinson: <http://www.census.gov/sipp/workpapr/abowd-stinson-ME-ss-20110707-sehdsworking.pdf>

Census Research Reports and Studies: <http://www.census.gov/srd/www/byyear.html>

Research Opportunities at the Census Bureau:

<http://www.census.gov/srd/research.pdf>

<http://www.census.gov/prod/2001pubs/dir-01rfp.pdf>

Census Strategic Plan: <http://www.census.gov/main/www/strategicplan/strategicplan.html>

Census Performance and Accountability Report:

[http://www.census.gov/aboutus/pdf/USCensusB](http://www.census.gov/aboutus/pdf/USCensusBureauFY2010_PAR.pdf)

[ureauFY2010\\_PAR.pdf](http://www.census.gov/aboutus/pdf/USCensusBureauFY2010_PAR.pdf)

[http://www.census.gov/aboutus/pdf/USCensusBureauFY2009\\_PAR.pdf](http://www.census.gov/aboutus/pdf/USCensusBureauFY2009_PAR.pdf)

## **Some Questionnaires**

*Firm and Establishment Census and Surveys*

<http://bhs.econ.census.gov/bhs/pages/formarchive.html>

<http://bhs.econ.census.gov/bhs/index.html>

[http://bhs.econ.census.gov/BHS/CEN\\_2007FORMS.html](http://bhs.econ.census.gov/BHS/CEN_2007FORMS.html)

<http://www.census.gov/epcd/www/ec97form.html>

<http://www.census.gov/econ/overview/FORMPROG.HTM>

*Person and Household Census and Surveys:*

American Community Survey (ACS)

[http://www.census.gov/acs/www/methodology/questionnaire\\_archive/](http://www.census.gov/acs/www/methodology/questionnaire_archive/)

2000 Decennial Census

<http://www.census.gov/dmd/www/2000quest.html>

1990 Decennial Census

<http://www.census.gov/prod/1/90dec/cph4/appdxe.pdf>

Survey of Income and Program Participation (SIPP)

<http://www.census.gov/programs-surveys/sipp/tech-documentation/questionnaires.html>

Current Population Survey (CPS)

<http://www.census.gov/cps/>

American Housing Survey (AHS)

<http://www.huduser.org/portal/datasets/ahs.html>

# Predominant Purpose Statement Template

---

## Predominant Purpose Statement (PPS) Draft, [insert date here]

9999 - "[Title]"

### Persons Requiring Special Sworn Status<sup>22</sup>

Researcher 1            Name – Affiliation  
Researcher 2            Name– Affiliation  
.  
.  
Researcher n            Name– Affiliation

Additional research assistants may be added to this project after it commences as needs dictate. In accordance with the Census Bureau – IRS Criteria Agreement of June 2012, the Census Bureau will notify IRS Statistics of Income of any additions and/or subtractions of Special Sworn Status individuals within five business days of the action.

**Duration** – [months]

### Locations

[Name] Census Research Data Center  
[Name] Census Research Data Center

**Funding** – [Source(s)]

**Background** [If necessary]

[Description]

### Criteria

The predominant purpose of this study is to increase the utility of Title 13, Chapter 5 data of the Census Bureau by meeting the criteria listed below and as described.

---

<sup>22</sup> If there are any Census Bureau employees on the project, replace “Persons Requiring Special Sworn Status” with “Persons with FTI Access” and list the employee’s Census Bureau division after the employee’s name.

*Criterion x*                    *[Title]*  
[Description]

*Criterion y*                    *[Title]*  
[Description]

**Requested Datasets**

**FTI Datasets**

Only the following Federal Tax return Information (FTI) items from the following files will be needed by the researcher(s) to conduct this project.

Dataset 1

Dataset 2

.

.

Dataset n

**Census Bureau Datasets**

Dataset 1

Dataset 2

.

.

Dataset n

**User Provided Datasets**

Dataset 1

Dataset 2

.

.

Dataset n

**Need and Use of FTI**

[Description]

**Project Completion**

The researcher(s) will provide the Census Bureau with annual progress reports on this multi-year

project. After this project is completed, the researcher(s) will produce a CES Working Paper and related research papers for conference presentation and/or journal publication. They will also produce a CES technical memorandum summarizing the key findings pertinent to Census Bureau Title 13, Chapter 5 programs. In addition, the researcher(s) will produce a Post Project Certification (PPC) summarizing whether and how the criteria described in the project proposal's Predominant Purpose Statement (PPS), as well as any others, were or were not met.

**Project Timeline**

The proposed duration of the project is X years. [Additional text here only if necessary]

[An example is given in the table below. **You must include a separate task for completing annual reports at the end of each project year and for completing the post-project certification at the end of the project.**]

Task	Year				
	1	2	3	4	5
Geocode import transaction data to county level	X	X			
Link foreign transactions data to CFS: match CFS to export declarations dataset; identify imports to port locations	X	X			
Build on the firm-level links in LFTTD to establish foreign transactions data links at the plant level	X	X			
Link shipping recipients in transaction data to LBD (to determine NAICS code)	X	X			
Submit annual report to CES	X				
Conduct data analysis for models of industry flow		X			
Draft paper of industry-level estimates (paper 1)		X			
Model entrance of movement of goods flowing through US		X			
Assess goodness of fit, quality of estimates for industry-level estimates		X			
Submit annual report to CES		X			
Conduct data analysis for regional impacts of trade flows			X		

Task	Year			
Determine overstatement of imports at ports; (estimate parameters of misreporting)			X	
Draft paper of regional impacts of trade flows (paper 2)			X	
Submit annual report to CES			X	
Finalize paper of regional impacts of trade flows (paper 2) and submit for journal review.				X
Revise paper of industry-level estimates (paper 1) to address referee reports from journal.				X
Conduct data analysis for geography of import flows (add endogenous variables to industry-level estimates) – requires feedback from models developed and peer-reviewed in paper 1				X
Assess US/Canada transaction data				X
Produce preliminary and final draft of geography of import flows and submit for journal review (paper 3)				X
Submit annual report to CES				
Conduct data analysis on extent of wholesale use by industry				X
Link transaction data to LBD and Economic Censuses to estimate where inputs are eventually absorbed				X
Report limitations of new <i>State Imports</i> data product given results of analyses of imports to port locations				X
Draft paper measuring trade flows and using transaction data(paper 4)				X
Revise papers 2 and 3 to address referee reports from journal.				X

# Examples: Predominant Purpose Statements

---

## **Predominant Purpose Statement (PPS) #1**

Draft, August 3, 2015

**EWD-5994100 - Conduct the Annual Survey of Entrepreneurs for Reference Years 2014-2016 and Provide Aggregate Estimates at the National Level for Reference Years 2013-2016**

### **Persons with FTI Access**

#### Economy Wide Statistics Division

Kevin Deardorff – Chief, Economy Wide Statistics Division  
Cynthia Davis Hollingsworth – Economy Wide Statistics Division  
Patrice Norman - Economy Wide Statistics Division

#### Economic Applications Division

Sam Jones – Chief, Economic Applications Division  
Christopher Rill – Economic Applications Division  
Joseph Talbot - Economic Applications Division

#### Economic Statistical Methods Division

Carol Caldwell – Chief, Economic Statistical Methods Division  
Xijian Liu - Economic Statistical Methods Division  
Richard A Moore Jr. - Economic Statistical Methods Division  
James Hunt – Economic Statistical Methods Division

#### Center for Economic Studies

Lucia Foster – Chief Economist, Center for Economic Studies  
Cheryl Grim – Center for Economic Studies  
Scott Ohlmacher – Center for Economic Studies  
Javier Miranda – Center for Economic Studies

#### Special Sworn Status Individuals

Alicia Robb – Ewing Marion Kauffman Foundation  
Sari Kerr – Ewing Marion Kauffman Foundation Researcher  
Robert Fairlie - Ewing Marion Kauffman Foundation Researcher  
Audrey Kindlon- National Science Foundation

Additional research staff with Special Sworn Status may be added to this project after it commences as needs dictate. In accordance with the Census Bureau – IRS Criteria Agreement of June 2012, the Census Bureau will notify IRS Statistics of Income of any additions and/or subtractions of Special Sworn Status individuals within five business days of the action.

**Duration** – 48 months

### **Locations**

Suitland Federal Complex

**Funding** – Funded as a joint project between the U.S. Census Bureau, the Ewing Marion Kauffman Foundation, and the Minority Business Development Agency

### **Background**

The Annual Survey of Entrepreneurs (ASE) will be modeled after the Survey of Business Owners (SBO). Currently, the Census Bureau conducts the Survey of Business Owners every five years as part of the Economic Census. The SBO provides the only comprehensive source of information on selected economic and demographic characteristics for businesses and business owners. Estimates include number of firms (employer and nonemployer), sales/receipts, annual payroll, and employment by gender, ethnicity, race, and veteran status. This project will provide more frequent and timely data on businesses and business owners and thus help contribute to the measure of the U.S. economy. Timely information at higher frequencies is critical for decision and policy makers and the business community. This project addresses the suggestions concerning the SBO that were made in *Understanding Business Dynamics: An Integrated Data System for America's Future* (National Academies Press, 2007).

This project has three complementary components: an annual survey of the employer universe (2014-2016), modeling of the employer universe for 2013, and modeling of the nonemployer universe (2013-2016).

For the first component, the project will result in the annual survey of employer firms for reference years 2014-2016 using the core questions from the 5-year SBO. These core questions will be supplemented with questions concerning the financial challenges faced by the businesses. In addition, a new module consisting of roughly 7-10 questions on special topics will be introduced each year. The module for the first year will collect information on innovation and research and development activities. The modules for the second and third years will collect information on topics such as management practices, globalization or some other topic of interest. This project will be a funded joint effort between the Census Bureau (42 percent) and the Ewing Marion Kauffman Foundation (58 percent). Kauffman Foundation staff support will be directed towards review and quality checks on response data and estimates.

Although the ASE will be a Census Bureau initiative, over which the Census Bureau shall have full control and responsibility, the Census Bureau intends to collaborate with the Kauffman Foundation and other agencies to ensure the ASE is as robust and effective as possible.

The second part of the project will cover reference year 2013. In order to complete the annual time series from 2012 through 2017 for each demographic group, the Minority Business Development Agency (MBDA) will fund summary statistics at the national level for each group for the calendar year 2013.

Finally, for the third part of the project, the Census Bureau will produce summary statistics for nonemployer businesses for reference years 2013 through 2016.

The quinquennial SBO is a large-scale survey of businesses with and without paid employees across the United States. It collects data about the racial, gender, and Hispanic origin of the ownership make-up of a representative sample of businesses with and without paid employees. The survey was mandated by an executive order of President Richard M. Nixon on March 9, 1969 and is authorized by both Titles 13 and 26 U.S.C. Since 1977, the survey has been conducted every five years as part of the Economic Census. In addition to the demographics of the ownership composition, SBO collects characteristics of the business owners (e.g., age and education level) and of the business (e.g., years in operation and sources of funding). The information collected is analyzed and tabulated by demographic category. It is used by legislators, decision makers, and researchers to identify problem areas, propose legislation, and formulate policy.

The largest challenge of the quinquennial SBO is identifying the small proportion of minority businesses without using an extremely large and costly mailout process. Due to the scarce proportion of minority-owned businesses, the Census Bureau attempts to pre-identify as many potential minority-owned firms as possible before the selection of the sample. This is primarily done by using a sample design that consists of stratifying by geographic, industry, and race(s) of the owner(s). The employer and nonemployer registers contain the geographic and industry codes. Accurate demographic data on the race of each owner is essential. Social Security Number information, in the form of Protected Identification Keys (PIKs), is obtained from the Census Bureau Register. These PIKs are then matched to the Social Security Numident and the Title 13 Race/Ethnicity file to obtain demographic information about each known owner. This process allows the survey to assign the most likely race of the owner(s) to each business enterprise and thus to identify, before mailout, a large portion of the businesses owned by each demographic group and to select a sample that is sufficient to make accurate estimates for each. The data have also proven to be invaluable in the procedure for the adjustment of the estimates due to non-response. The five-year survey has successfully used it for over 20 years.

The quinquennial SBO produces the most comprehensive set of ownership demographic characteristics and it has become clear that this information is needed on a more timely basis. Policymakers, businesses, and researchers have indicated a need for estimates for the intermediate years. In addition, these data user groups have noted that they need this information more quickly. Currently, there is a long lag until information is published from these collections. For example, the 2012 survey was not completely mailed until mid-2014 and the characteristic tables will not be tabulated until late 2015. For these reasons, the Ewing Marion Kauffman Foundation and the U.S. Bureau of the Census have agreed to conduct a complementary survey on an annual basis, beginning with the 2014 reference year. For the 2014 ASE, mailout is

scheduled to start September 2015. The mailout process will follow similar procedures as the SBO. This requires approval to use the FTI and Title 13 administrative data on an annual basis, similar to the one previously approved for the five-year cycle.

The new annual survey would consist of approximately 290,000 businesses with paid employees. It would contain questions on owner demographics, owner characteristics, business characteristics, and financing. Because it is restricted to businesses with paid employees, the survey would not be able to produce any statistics for businesses without paid employees (these businesses are referred to as nonemployers).

The second part of the project continues the focus on the employer universe, but uses modeling to create annual data before the survey commences. Thus, this part of the project focuses solely on 2013 reference year.

The third part of the project focuses on the nonemployer universe and thus complements the survey. The employer statistics alone would be insufficient to many data users (in particular the MBDA), because the nonemployer segment comprises almost 80 percent of the universe. For many demographic groups, the percentage ownership of businesses without paid employees is substantially higher than the percentage of businesses with payroll. In 2007, for example, African Americans owned less than 2 percent of all firms with paid employees and over 8 percent of the firms without paid employees. For this reason, Census would like to produce national level summary nonemployer statistics for each demographic group. The effective use of administrative data allows us to better model race, gender, and Hispanic origin for each known owner of each business. Although this process is not exact, at the U.S. level estimates are substantially better than intra-economic-census projections made by researchers or policymakers. In addition, a Census Bureau produced measure of variability would accompany each estimate, thus allowing other data users to evaluate the credibility of any argument, which is based on these statistics. This is usually not the case with estimates produced by external groups. The use of these data would also allow the Census Bureau to produce summary estimates for employer businesses in 2013, thus completing the time series from 2012 estimates from the quinquennial SBO through 2014-2016 from the annual mailout.

## **Criteria**

The predominant purpose of this study is to increase the utility of Title 13, Chapter 5 data of the Census Bureau by meeting the criteria listed below and as described.

- *Criterion 5: Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census, or estimate*

As stated in the “Background”, SBO estimates are currently produced once every five years. While the published data do convey a considerable amount of information on the net change in business operations over the previous five-year period, a yearly time series would show effect of economic events, such as the 2008 recession. In the absence of yearly data, users often produce their own time series. They use estimation methods that are inconsistent with the assumptions used to create the SBO sampling frame, select the

sample, and produce the estimates, including the adjustments for non-response. If allowed to produce these estimates, SBO staff could use procedures that are more consistent with those employed by the quinquennial SBO.

The construction of the SBO sampling frame involves using FTI and demographic data to segment the businesses into mutually exclusive groups. For example, a business with several owners, all of whom reported themselves as Asian in the 2010 Decennial Census of Population and Housing Units, will be classified as “Very Likely Asian-owned”, while a single owner, for which we have no race or Hispanic origin information, but whose last name (Lopez, Rodriguez) suggests that he may be Hispanic will be in the segment as “Possibly Hispanic-owned”. It is more accurate to estimate and measure the variability of each segment, than attempt to make an estimate for the universe as a whole. This is made possible with these datasets.

The availability of these data also increases the quality of the published estimates. The 2007 SBO mailed to about 2.3 million businesses out of a universe of 27.1 million businesses. Although the survey is mandated by law, only about 60 percent of the mailed cases returned the questionnaire with a response. The response rate for firms with paid employees was 75 percent and firms without paid employees were about 50 percent. Response rates are also not uniform across businesses within the demographic ownership classifications. About 67 percent of the businesses classified as “Likely White non-Hispanic-owned” responded, compared to only about 40 percent of those classified as “Likely Hispanic-owned”. These two segments are quite heterogeneous, but the businesses within each segment are relatively homogeneous. When SBO adjusts for non-response, the same procedure is used on each segment, but the two results are substantially different because the segments are heterogeneous. Because this procedure yields a much more accurate adjusted estimate at the segment level, the overall estimate, which is an aggregation of the estimates of the individual segments, is more accurate.

Currently, when policymakers and researchers need intercensal year estimates, they create their own time series. Their assumptions often lead to erroneous estimates. These estimates are rarely accompanied by the corresponding measure of variability. For the modeled estimates, the Annual Survey of Entrepreneurs would

- create and segment the universe and sampling frame as described above;
- assume that the mailout consisted of every business that either
  - was mailed a questionnaire in the previous SBO and is still in operation, or
  - birth businesses, i.e. those which started operations since the previous SBO and were not subjected to selection for the sample;
- assume the only respondents to the mailout are those who responded to the previous SBO, and
- assume that most of these respondents did not have ownership changes since the last SBO.

This means the non-respondents would be

- those who were selected, but did not respond in the previous SBO, and
- all of the birth businesses.

Since the sampling frame has been segmented, and the segments are homogeneous, we are able to use our in place non-response adjustment techniques to make

- segment level estimates, after non-responses adjustments, for each segment,
- segment level total errors, sampling error and the bias in the non-response adjustment procedure, for each of the segment estimates,
- national level estimates by aggregating the segment level estimates, and
- total errors on the national level estimates in a similar error.

The amount of rigor should provide a better national level time series than produced by data users, who do not have access to the requested information. The number of cases and the percentage of the estimate represented by the birth segment may be very large for each year. For example, for 2016, the SBO staff realizes that nonemployer births may constitute as much as 70 percent, the percentage of nonemployer births between 2007 and 2012. The total error will reflect that such a large portion of each estimate is attributed to “non-response”. Not only should the resulting time series have more accurate estimates, but also the measure of variability validates (or invalidates) the time series,

Whether estimating from a probabilistic survey --- the 2014, 2015, and 2016 business with paid employees estimates --- or through modeling for the 2013 employer and intercensal nonemployer summary statistics, use of the requested FTI and Census Bureau data sets is critical to providing reliable time series estimates. The Kauffman Foundation will also use Special Sworn Status researchers for review and quality checks of the data based upon their expertise with similar data and topics.

- *Criterion 7: Enhancing the data collected in a Title 13, Chapter 5 survey or census. For example, improving imputation for non-response, or developing links across time or entities for data gathered in censuses and surveys authorized by Title 13, Chapter 5.*

Under Criterion 5 we discuss improving the quality of the data product. The second paragraph of the bullet discussed the use of FTI and Census Bureau data sets to segment the universe. The examples given are two of the more obvious cases. There are cases more difficult to classify. For these we use logistic regressions with inputs such as (1) the country of birth, (2) the mother’s maiden name, (3) the racial and Hispanic composition of the ZIP Code in the last Decennial Census. This allows us to divide the universe into homogeneous segments. The operations in each segment, that have reported in previous quinquennial SBO’s, are much more similar to the others, than had they been randomly or haphazardly assigned to that segment. Once this is completed, there are several options on how to proceed.

- Select a probabilistic sample and use the standard expansion methods to make estimates and variances on that estimate.
- Use hot-deck imputation for each operation in the segment; using the cases with responses from a previous SBO as donors. Measure the relative bias of the imputation and apply that to the estimate to measure the estimation error.

- Assume that the distribution of the cases with responses is representative of the distribution of the entire segment. With a simple weight adjustment to the cases with a response, a variance for the segment estimate can be obtained.
- Assume that the distribution of the cases with responses is representative of the distribution of the entire segment. Assign responses to the other cases in such a manner that the distribution of the entire segment is equivalent to the distribution of the cases in the segment with responses.

Although any of the methods can be employed regardless of the quality of the segmentation, non-homogenous segments lead to less precise estimates and less accurate measures of variability. The assumption of homogeneity of the segments cannot be made without use of the requested data.

- *Criterion 10: Constructing, verifying/, or improving the sampling frame for a census or survey authorized under Title 13, Chapter 5.*

If allowed to use the requested data sets for the annual survey, we can construct a frame from which we can make reasonable estimates for each demographic group at the U.S. level. As noted in the background, there is considerable overlap in the universe of the SBO and the ASE. To the extent that the ASE sample overlaps with that of the SBO, yearly updates to the ASE will refresh the SBO sample. Thus the ASE will provide a benefit to the SBO sample.

- *Criterion 11: Preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5.*

The ASE will provide high quality, timely information on demographics of business owners and businesses in the employer universe. A Special Sworn Status researcher from NSF with experience with NSF’s measures of innovation will examine the correlation between business and owner characteristics and innovation activities as measured in the first year’s module in order to validate the results.

## **Requested Datasets**

### **FTI Datasets**

Only the following Federal Tax return Information (FTI) items from the following files will be needed by the researcher(s) to conduct this project.

### **Business Register Dataset (BR) 2012-2016**

#### **2012 through 2015 EMPUNIT\_FINAL:**

<b>FTI:</b>	<b>Description</b>
HIST_CFN	Historic Census File Number
EINUNIT_ID_CHAR	9-digit Employer Identification Number
LGL_FORM_CODE	Legal Form Code

NAICS_CODE	6-digit North American Industry Code
Classification	
NAME1	Name of Tax Filer and
Spouse	
NAME2	Continuation of Name
STREET_MAIL	Street Address for Mailing
STATE_ABBR_MAIL	2-character State Abbreviation for Mailing
ZIP_MAIL	5-digit ZIP Code for Mailing
STREET_PHY	Street Address for physical location
STATE_ABBR_PHY	2-character State Abbreviation for physical
location	
ZIP_PHY	5-digit ZIP Code for physical location
EMP_MAR12_20XX	March 12 employment in year 20XX (XX=
12, 13, 14, 15)	
PAY_ANN_20XX	Annual Payroll in year 20XX (XX= 12, 13,
14, 15)	
RCPT_20XX	Sales and receipts in year 20XX
(XX= 12, 13, 14, 15)	

**2012 through 2015 EINUNIT\_FINAL**

(used to get establishment data for companies that operate multiple establishments)

**FTI:**

**Description:**

HIST_CFN	Historic Census File Number
EINUNIT_ID_CHAR	9-digit Employer Identification Number
BMF_PIK	Owner's Protected Identification Key (based on SSN)
LGL_FORM_CODE	Legal Form Code
NAICS_ADMIN_CODE	6-digit North American Industry Code
Classification	
BMF_NAME_PRIM	Primary Name under which tax return is filed
BMF_NAME_PRIMCONT	Other Name or Overflow
STREET_MAIL	Street Address for Mailing
STATE_ABBR_MAIL	2-character State Abbreviation for Mailing
ZIP_MAIL	5-digit ZIP Code for Mailing
STREET_PHY	Street Address for physical location
STATE_ABBR_PHY	2-character State Abbreviation for physical
location	
ZIP_PHY	5-digit ZIP Code for physical location
EMP_MAR12_20XX	March 12 employment in year 20XX (XX=
12, 13, 14, 15)	
PAY_IRS_CURR_20XXQ1-Q4	Quarterly Payroll in 2012, 2013, 2014, 2015
RCPT_IRS_CURR_20XXQ1-Q4	Annual Sales and Receipts in 2012, 2013, 2014, 2015)

**2012 through 2015 FINAL NONEMPLOYER DATA SET**

<b>FTI:</b>	<b>Description:</b>
NONEMPID_CHAR	Identification Number based on EIN/PIK
PIK_FILER	1040 Filer Protected Identification Key (based on SSN)
PIK_SPOUSE	1040 PIK (based on SSN) of spouse for Joint Returns
PIK_OWNER	1040 PIK the joint return listed as owner on Schedule C
LGL_FORM_CODE	Legal Form Code
NAICS_CODE	6-digit North American Industry Code
Classification	
NAME1	Name of Tax Filer and Spouse
NAME2	Continuation of Name
STREET_MAIL	Street Address for Mailing
STATE_ABBR_MAIL	2-character State Abbreviation for Mailing
ZIP_MAIL	5-digit ZIP Code for Mailing
RCPT	Annual Sales and Receipts in 2012, 2013, 2014, 2015

### **2012 Survey of Business Owners Tabulation Register**

<b>FTI:</b>	<b>Description:</b>
HIST_CFN	Historic Census File Number
EINUNIT_ID_CHAR	9-digit Employer Identification Number
NONEMPID_CHAR	Identification Number based on EIN/PIK
BMF_PIK	Owner's Protected Identification Key (based on SSN)
PIK_FILER	1040 Filer Protected Identification Key (based on SSN)
PIK_SPOUSE	1040 PIK (based on SSN) of spouse for Joint Returns
PIK_OWNER	1040 PIK the joint return listed as owner on Schedule C
NAME	Primary Name under which tax return is filed
STREET_MAIL	Street Address for Mailing
STATE_ABBR_MAIL	2-character State Abbreviation for Mailing
ZIP_MAIL	5-digit ZIP Code for Mailing
N12_LFO	2012 Legal Form Code
N12_NAICS4	2012 6-digit North American Industry Code Classification
N12_EMPLOYMENT	March 12 employment in year 2012
N12_PAYROLL	Annual Payroll in year 2012
N12_RECEIPTS	Annual Sales and Receipts in year 2012

### **Census Bureau Datasets**

2012 through 2015 Census Numident File, Family Version

2012 through 2015 Title 13 Race/Ethnicity File, Family Version

## User Provided Datasets

None

## Need and Use of FTI

<u>FTI</u>	<u>Datasets</u>
EIN SBO TAB REG	EINUNIT_FINAL, EMPUNIT_FINAL K1,

The 9-character Employer Identification is used to link data between the files listed above.

BMF_PIK	EINUNIT_FINAL
PIK_SPOUSE	NONEMPLOYER DATA SET
PIK_FILER	NONEMPLOYER DATA SET
PIK_OWNER	NONEMPLOYER DATA SET

All Protected Identification Keys are the Census Bureau's encryption of Social Security Numbers. They are used to link FTI data to ownership demographic information from the Social Security Administration NUMIDENT extract and the Title 13 Race and Ethnicity file.

BMF_NAME_PRIM	EINUNIT_FINAL
BMF_NAME_PRIMCONT NAME1 DATA SET	EMPUNIT_FINAL, NONEMPLOYER
NAME2 DATA SET	EMPUNIT_FINAL, NONEMPLOYER
NAME	2012 SBO TABULATION REGISTER

The business or individual names are used if a case is selected for mailing. In addition, the name strings are parsed. Strings which are possibly surnames are used to identify cases which indicate the potential of having owners with either Hispanic or Asian descent.

STREET_MAIL NONEMPLOYER, SBO	EINUNIT_FINAL, EMPUNIT_FINAL,
CITY_MAIL	EINUNIT_FINAL, EMPUNIT_FINAL, NONEMPLOYER, SBO
STATE_ABBR_MAIL	EINUNIT_FINAL, EMPUNIT_FINAL, NONEMPLOYER, SBO
ZIP_MAIL	EINUNIT_FINAL, EMPUNIT_FINAL, NONEMPLOYER, SBO
STREET_PHY NONEMPLOYER, SBO	EINUNIT_FINAL, EMPUNIT_FINAL,
CITY_PHY	EINUNIT_FINAL, EMPUNIT_FINAL, NONEMPLOYER, SBO
STATE_ABBR_PHY	EINUNIT_FINAL, EMPUNIT_FINAL, NONEMPLOYER, SBO
ZIP_PHY NONEMPLOYER, SBO	EINUNIT_FINAL, EMPUNIT_FINAL,

The business and individual mailing address information is used for mailing in case the case is selected in sample. In addition, business and individual mailing and physical address information are used in parsing the name strings. Strings that exist in both one of the NAME and at least one of the ADDRESS fields are eliminated from the strings considered as surnames. Physical state codes are also used in the stratification for the sample design.

<b>FTI</b>	<b>Datasets</b>
LGL_FORM_CODE	EINUNIT_FINAL, EMPUNIT_FINAL, NONEMPLOYER_FIN
N12_LFO	2012 SBO TABULATION REGISTER

Legal Form Codes are used as input in the modeling of the race and the Hispanic origin likelihoods. They are also used as keys for the hot deck imputation procedure.

NAICS_ADMIN_CODE	EINUNITS_FINAL
NAICS_CODE	EMPUNITS_FINAL, NONEMPLOYER_FINAL
N12_NAICS4	2012 SBO TABULATION REGISTER

Industry codes are used as stratification variables in the sample design. They are also used as input in the modeling of the race and the Hispanic origin likelihoods and as keys for the hot deck imputation procedure.

EMP_MAR12_CURR	EINUNITS_FINAL
EMP_MAR12_	EMPUNITS_FINAL
N12_EMPLOYMENT	2012 SBO TABULATION REGISTER
PAY_IRS_CURR_Q1-Q4	EINUNITS_FINAL
PAY_ANN_	EMPUNITS_FINAL
PAY_NET_IRS_CURRQ1Q4	EINUNITS_FINAL
N12_PAYROLL	2012 SBO TABULATION REGISTER
RCPT_NET_IRS_CURR	EINUNITS_FINAL
RCPT_	EMPUNITS_FINAL
RCPT	NONEMPLOYER_FINAL
N12_RECEIPTS	2012 SBO TABULATION REGISTER

Employment is used in conjunction with Payroll to determine which operations have and do not have paid employees. This variable is then used for stratification and as a key for imputation. Receipts size classes (based on the magnitude of the receipts value) are also used as a key for imputation.

Employment, payroll, and receipts are used to produce tabulations.

### **Project Completion**

Output from the ASE will be provided in the following products:

Company Summary  
Characteristics of Businesses  
Characteristics of Business Owners  
Special tabulations

The following products may also be provided:

Women-Owned Firms  
Black-Owned Firms  
Hispanic-Owned Firms  
Asian-Owned Firms  
Native Hawaiian- and Other Pacific Islander-Owned Firms  
American Indian- and Alaska Native-Owned Firms  
Veteran-Owned Firms

Special Topics – Should the Census Bureau conduct any special studies under this project the Internal Revenue Service will be notified by means of a revised Predominant Purpose Statement (PPS). Such special activity might include new supplemental surveys, statistics, or Public Use Micro Data Samples (PUMS).

Disclosure Avoidance Practices – Public releases of all statistics by the Census Bureau conform to policies and rules described in the Federal Register Notice, Vol. 68, N. 86, Monday, May 5, 2003 and in the May 27, 2003 letter from Census Bureau Director Louis Kincannon to IRS RAS Director Mark Mazur.

## Project Timeline

The proposed duration of the project is 4 years.

Task	Year			
	1	2	3	4
Extract the universe from the employer and nonemployer data sets	X	X		
Identify the owners of each operation (where possible)	X	X		
Obtain demographic owner information from Numident Title 13 Race/Ethnicity file	X	X		
Link owner information by PIK match to each operation	X	X		
Link other administrative information needed for modeling likelihoods of each gender-race-Hispanic origin combination	X	X		
Create frame by segmenting universe	X	X		
Model responses for businesses with paid employees	X			
Model responses for businesses without paid employees	X	X	X	X
Determine sampling parameters for employer segment		X	X	X
Select sample for businesses with paid employees		X	X	X
Mail sample questionnaire		X	X	X
Edit responses		X	X	X
Produce estimates and mean square errors at the US level for businesses with paid employees	X	X	X	X
Produce estimates and mean square errors at the US level for businesses without paid employees	X	X	X	X
Submit Post-Project Certification Document				X

**Predominant Purpose Statement (PPS) #2**  
Final, September 15, 2014

**SEHSD-892 – “Imputation of Missing Data in SIPP Using Administrative Records”**

**Researchers**

Gary Benedetto—U.S. Census Bureau  
Daniel Doyle—U.S. Census Bureau  
Jason Fields—U.S. Census Bureau  
Graton Gathright—U.S. Census Bureau  
Elaine Hock—U.S. Census Bureau  
Joanna Kling—U.S. Census Bureau  
Matthew Marlay—U.S. Census Bureau  
Huong Nguyen—U.S. Census Bureau  
Martha Stinson—U.S. Census Bureau  
Thomas Welton—U.S. Census Bureau  
Nathaniel McKee – U.S. Census Bureau  
Imelda Severdia – U.S. Census Bureau

No Special Sworn Employees support this project.

**Duration** – 4 SIPP Processing Cycles (calendar years 2014-2015, 2015-2016, 2016-2017, 2017-2018) for waves 1-4 of the SIPP 2014 panel. Census will request permission to continue using FTI for imputation purposes at the beginning of the next SIPP panel.

**Locations**

Suitland Federal Complex

**Funding** – None (internal project)

**Background**

This project requests permission to use administrative records from the Social Security Administration (SSA) and the Internal Revenue Service (IRS) to model responses to questions from the Survey of Income and Program Participation (SIPP) in order to impute missing data. Specifically, the Census Bureau will use data from W-2 tax forms, as authorized for conducting demographic, economic, and agricultural statistics programs and censuses and related program evaluation. (See 26 CFR, Part 301, 6013(j), 2691-2695). This project helps fulfill the Census Bureau's mandate to provide high-quality data products that describe the people and the economy of the United States through improvements to the imputation of missing data. The Title 13 benefit of this project falls under Criteria 7 “Enhancing the data collected in a Title 13, Chapter 5 survey or census. For example: Improving imputations for non-response.”

Academic and government researchers have long used the SIPP to study social and economic well-being among American families, in particular to look at transitions over time by exploiting the panel nature of the SIPP. The Census Bureau creates public-use micro-data through an extensive processing system for editing the raw data and imputing missing values. Since 2005, we have been engaged in an effort to redesign the SIPP. As part of the effort, the processing system is being completely re-built with the twin goals of producing more accurate data in a more timely manner.

One of the key improvements in the new processing system is new imputation methods. In current SIPP processing, all imputations are done using hot deck methods. Census now plans to switch to a model-based approach, initially for a limited number of variables. This will allow the imputation to take account of relationships between more variables and to bring in outside sources of data. We seek permission to include SSA/IRS W-2 information in our production models as explanatory variables.

The research behind this request has been on-going for many years. Working under ARTS Project 782 HHES Survey Improvement, Census staff have linked SIPP and W-2 data from SSA's Detailed Earnings Record extract from the Master Earnings File. Using this link, Abowd and Stinson (2013, Review of Economics and Statistics) found that highly educated white males had lower reported earnings in the SIPP than in W-2 tax records, largely due to men with missing data receiving imputed values that were too low. Benedetto and Stinson (2009, APPAM presentation) tested the feasibility of using model-based imputation for monthly earnings data in the 2004 SIPP panel and found that for sub-groups with high rates of missing data, the model-based approach could reduce the variance of key statistics and help with concerns about data not missing at random. More recently under ARTS Project 645, Stinson, Gathright, and Skog have linked field test data collected in 2010 and 2011 using the new SIPP instrument to IRS-delivered W-2 records to compare employment reports in the survey to tax reports (2012, FCSM presentation). These links have proven very successful in that they have shown how we might use administrative data to account for respondent differences that are unobserved in the survey.

This project requests explicit permission to use our accumulated knowledge about the link between the SIPP and W-2 records for production purposes. ARTS Project 645 stipulated "[a]s programmatic uses of W-2 information are identified and tested, the Census Bureau will seek IRS concurrence through modification to existing approved projects or the creation of new projects." Hence, this new project is the outcome of the research conducted under both ARTS 782 and ARTS 645.

### *Methods*

Data collected by the Census Bureau using the SIPP survey instrument will be transmitted to Census headquarters from Regional Offices and some basic initial processing will be performed. The data will then be transferred to a special partitioned space on the SIPP production server,

created to explicitly separate processing that requires IRS data and processing that does not. In this secure partitioned space, we will merge W-2 variables to the SIPP data and will estimate imputation models for a subset of SIPP variables. After the imputations for these variables are complete, the W-2 variables will be stripped from the data and only the original survey data and imputations will be transferred back to the regular processing space. This process will keep the administrative data separate from the main SIPP processing system and will enable the SIPP to remain a Title 13-only protected data set.

Our first plans are to use W-2 data to impute “presence of topic” indicators for respondents who missed entire sections of the survey. For example, for those respondents who did not answer the labor force section of the survey, we will impute whether they held any jobs at any point during the reference year. As an explanatory variable in this model, we would like to include indicators for having a W-2 in the reference year and the year prior to the reference year, as well as W-2 indicators for the spouse. After the imputation is completed, these topic flags (minus the W-2 indicators) will then be fed into the main SIPP processing system and will indicate the need for further imputation of variables specific to a given topic. For example if a respondent is imputed to have held one job in the reference year, the topic flag will indicate that downstream processing should impute earnings at this job. Initially 30 SIPP topics will be included in this imputation process: education enrollment, employment, SSI, SNAP, TANF, general assistance, WIC, private health insurance, Medicare, Medicaid, military health insurance, dependent care, VA benefits, OASDI, worker’s compensation, unemployment compensation, energy assistance, school lunch/breakfast, lump sum payments, disability, disability payments, retirement, retirement payments, survivor payments, foster child care payments, child support payments received, alimony payments, child support payments made, payments to non-household members, functional limitations. As we develop successful models for these initial topics, we would like to expand our use of W-2 data to imputation models for other SIPP topics and for other variables within a given topic. For example, eventually we would like to use W-2 data to impute missing earnings data. Hence, we are requesting permission to use the linked W-2 data in the imputation of any SIPP variable that we deem to be a good candidate for a model-based approach. However, we will not under any circumstance spread the use of IRS data throughout the entire SIPP processing sequence. All processing using tax data will be confined to the secure partitioned space and all tax data will be stripped from the SIPP files once the imputations are complete.

We also propose to include SSA variables from the Numident, the Master Beneficiary Record, the Payment History Update System, and the Supplemental Security Record as explanatory variables in our models. The SSA variables will be matched to the SIPP at the same time as the W-2 data and in the same secure partitioned space on the SIPP processing server.

### *Output*

The main output of this project will be public-use SIPP files that will be disseminated through the Census website. This project proposes only to use IRS data as explanatory variables in modeling and does not propose to add W-2 variables directly to the public-use SIPP files being created for this web-based dissemination<sup>23</sup>. This project will also not replace missing SIPP values directly with IRS data. W-2 variables will only be used to predict missing responses.

In addition to the public-use files, we will produce some research reports on the results from these new imputation models. In particular, for reference year 2013, we will be able to compare SIPP data collected and processed under the old survey regime to the data from the new regime.<sup>24</sup> These topic-level comparisons will enable us to determine the impact of our models and the use of tax data.

While this project is not explicitly a research project and seeks to use the IRS W-2 data for a very specific production purpose, it will nonetheless serve as a model for other surveys by demonstrating what is possible and reporting on the successes and failures of this approach. If these imputation methods and the use of administrative data prove viable in the long run for the SIPP, other larger surveys will be able to consider implementing similar methods. However, no imputation for any other survey will be performed under this Census ARTS project. Requests for permission to use tax data for the imputation of variables in other surveys will be sent to the IRS by the Census Bureau as separate projects.

## **Criteria**

The predominant purpose of this study is to increase the utility of Title 13, Chapter 5 data of the Census Bureau by meeting the criteria listed below and as described.

*Criterion 7* Enhancing the data collected in a Title 13, Chapter 5 survey or census. For example, improving imputations for non-response or developing links across time or entities for data gathered in censuses and surveys authorized by Title 13, Chapter 5.

This project seeks to employ administrative records from the Social Security Administration (SSA) and from the Internal Revenue Service (IRS) to model responses to questions from the Survey of Income and Program Participation (SIPP) in order to impute missing data.

Missing data continues to be a difficult problem in survey data, including the SIPP. The largest concern is that those survey respondents who fail to answer questions are different from other survey respondents in unobservable ways. This problem becomes particularly acute when a

---

<sup>23</sup> Census ARTS Project 782 does add W-2 variables to the SIPP and creates the public use file called the SIPP Synthetic Beta (SSB) by synthesizing both SIPP and IRS variables. However, ARTS Project 782 is completely separate from the regular SIPP processing that takes place immediately after data collection and hence will be separate from the project described in this PPS, ARTS 892.

<sup>24</sup> The 2008 SIPP panel has continued to collect data through most of calendar year 2013. The 2014 SIPP panel will be fielded in January of 2014 and will collect data for all of calendar year 2013.

respondent fails to answer large sections of the survey, leaving little information available to impute non-responses. Administrative data have the potential to play a large and helpful role in addressing these concerns. As an outside source of data, they provide another way to identify similar individuals, independent of any collected data. Administrative data can also be obtained for most individuals who provide a name, gender, and date of birth and hence are available even for individuals with large amounts of missing data. They also can provide information about the time period before the survey reference period, information which is useful for predicting labor force and program participation since these decisions are highly correlated over time. Finally administrative data can serve as a useful evaluation tool for imputations since they provide benchmark comparisons for statistics calculated from survey data after imputations have been performed.

Research conducted under ARTS 782 and ARTS 645 shows that administrative records from the IRS and SSA are very useful in models to impute missing data in the SIPP. This project seeks to apply knowledge and techniques developed under ARTS 782 and ARTS 645 to SIPP production work.

**Requested Datasets**

**FTI Datasets**

Only the following Federal Tax return Information (FTI) items from the following files will be needed by the researcher(s) to conduct this project.

**Detailed Earnings Record Extract from the Social Security Administration (SSA) Master Earnings File (MEF) (for 2014 SIPP Panel respondents only), TY 1978-2012 and future years as available**

<b>FTI:</b>	<b>Description:</b>
EIN Identification Number	Employer
NON_FICA	Non-covered earnings
PAYMENT_457 Wage Payment	Plan 457 distribution or Special
PENSION sum pension distribution	Interest, dividends, and lump
PLAN_401 401(k)	Current W2 Non FICA for
PLAN_403 403(b)	Current W2 Non FICA for
PLAN_408 408(k)	Current W2 Non FICA for
PLAN_457 457(b)	Current W2 Non FICA for
PLAN_501 501(c)	Current W2 Non FICA for

PLAN_HSA	Current W2 Non FICA for
Health Saving Accounts	
SSN	Social Security
Number	
SWP_457	Non Plan 457 Special Wage
Payment	
TIPS	FICA covered TIPS
TYPE_RECORD	Type of Record
Year	Year of Earnings
W2_BOX1_WTOTCMP	Box 1: Wages, Tips, and Other
Compensation	
W2_BOX3_WGE_FICA	Box 3: Social Security Wages
W2_DEF_COMP	Box 13: Deferred Wages
SEI_FICA	Self-Employment Income for
Social Security Purposes	
SEI_MED	Self-Employment Income for
Medicare Purposes	

**Summary Earnings Record Extract from the SSA MEF (for 2014 SIPP Panel respondents only), TY 1951-2012 and future years as available**

<b>FTI:</b>	<b>Description:</b>
SSN	Social Security
Number	
EARNINGS DATA (ADJ)	Annual Total Social Security Taxable
Earnings	
WQC51- WQC12	Annual Number of Quarters of
Coverage	

**IRS Information Returns (Form W2) TY 2005-2012, and future years as available**

<b>FTI:</b>	<b>Description:</b>
DC—Total codes D,E,F,G,H	Deferred Compensation, Total
EIN	Employer
Identification Number	
SSN	Social Security
Number	
SSW	Social Security
Wages	
TY	Tax Year
W, T, and C	Wages, Tips, and Other
Compensation	

**Census Bureau and Other Datasets**

Census Numident (SSA Numident File enhanced with Census Bureau information)  
 SSA Master Beneficiary Record (for 2014 SIPP Panel respondents only), 1962-2014 (or most

current year available)  
SSA Payment History Update System (PHUS) (for 2014 SIPP Panel respondents only), 1962-2014 (or most current year available)  
SSA Supplemental Security Records (SSR) (for 2014 SIPP Panel respondents only), 1962-2014 (or most current year available)  
SIPP Crosswalks

### **User Provided Datasets**

None

### **Need and Use of FTI**

We are requesting W-2 information from three sources: the Detailed Earnings Record (DER) and Summary Earnings Record (SER) Extracts from the Master Earnings File maintained by SSA and the Information Returns (Form W2) sent directly to Census by IRS. The direct IRS data has the major advantage of arriving at Census six months earlier than the SSA data. This early arrival will enable us to use administrative data from the survey reference year in our processing. For example, the 2014 SIPP panel will first ask respondents about 2013, making this the first reference year. The majority of Information Returns/W-2s for tax year 2013 can be sent from IRS to Census in the fall of 2014 whereas these data will only be available from SSA in April 2015.

While not as timely, the DER and SER have other advantages because of the long history and the presence of self-employment earnings. Thus, while in the fall of 2014 Census will only be able to receive DER and SER extracts through tax year 2012, these data will nonetheless allow us to create summary measures of a person's work history that will be very valuable in the modeling process.

Below we list each variable and describe how we anticipate using it to model SIPP variables and create imputations. We list DER and IRS W-2 variables together when they have the same definition and request the use of both due to the timing issues explained above. The DER and the SER are longitudinal files that contain records across many years for SIPP respondents (1951-present for SER; 1978-present for DER). We request permission to use the full history of a respondent's earnings in order to build a profile of how many years the individual has worked, how much they earned on average over prior years, what year they first had a W-2 filed by an employer, and what year they last had a W-2 filed. This profile will be very valuable in predicting an individual's work status in the SIPP reference year. For example if an individual has worked every year for the last 10 years, the probability of him or her working in the SIPP reference year is much higher than someone who has worked more sporadically. Thus for the variables below we are requesting an annual time series.

- **EIN** – This variable identifies separate jobs/employers. We will use EIN to count the number of unique jobs in a year, which will then serve as an explanatory variable in our imputation modeling. We will also use EIN to calculate the tenure at each job held during the year by counting how many years an EIN/employer filed a W-2 record for a person. Tenure will also be an explanatory variable in our modeling.

- **TY and Year** – This variable will enable us to identify which W-2 records correspond to the year of our survey and to the years immediately preceding the survey. Thus year will allow us to match the timing between the survey and the administrative data and to identify the values that will be most useful in predicting SIPP variables.
- **W, T, and C and W2\_BOX1\_WTOTCMP** – This variable contains non-deferred wages (not top-coded) and will be used to create a total earnings variable that will be an explanatory variable in all of our modeling. We expect total earnings to have explanatory power in predicting such things as who participates in means-tested programs, who has health insurance, who is employed, and who is enrolled in school.
- **SSW and W2\_BOX3\_WGE\_FICA** – We will use the presence of positive SS earnings as an indicator to classify the job as FICA-covered and will use this job characteristic as an explanatory variable in the modeling process.
- **DC (Total codes D,E,F,G,H) and W2\_DEF\_COMP** – We will sum deferred compensation together with Box 1 compensation in order to create total annual earnings. As explained under “W2\_BOX1” this total earnings variable will be an explanatory variable in our modeling.
- **SEI\_FICA** – The presence of positive FICA self-employment earnings in this field indicates that the self-employment income was taxable under FICA. This variable will be used to describe the characteristics of the self-employment spell and will be used as an explanatory variable.
- **SEI\_MED** – This variable contains self-employment earnings (not top-coded) and will be summed together with any positive amounts in W2\_Box1 and W2\_DER\_COMP from W-2 records in the same year to create total annual earnings.
- **EARNINGS DATA (ADJ)** – The Summary Earnings Record (SER) provides earnings from 1951-2012 and allows us to calculate years of total labor force experience by counting years of positive earnings during this time frame. Many SIPP respondents are old enough that their careers began before 1978 (the beginning of the DER) and hence the SER is essential to counting the total number of years worked. We will use the labor force experience variable as a predictor for many of our topics, including employment, retirement, disability, and program participation.
- **WQC51- WQC12** – This variable counts total quarters of retirement/disability insurance coverage that a person earned from working in a given year. Using these quarters of coverage variables we can calculate whether an individual is eligible to receive Social Security benefits in any given year. We will include eligibility flags as explanatory variables in our modeling of SIPP reported receipt of Social Security benefits.
- **NON\_FICA** – We will use the presence of positive non-fica earnings as an indicator to classify the job as not FICA-covered and will use this job characteristic as an explanatory variable in the modeling process.
- **PAYMENT\_457** – These special wage payments will be summed together with other earnings types to create total earnings for the calendar year.
- **PENSION** – The distribution of pension money will be included as an explanatory variable in our models to predict whether a person was working or retired.
- **PLAN\_401, PLAN\_403, PLAN\_408, PLAN\_457, PLAN\_501, PLAN\_HSA** – These breakouts of deferred wages by type of plan will be used to predict the presence of

household assets, in particular retirement assets.

- **SSN** – The Social Security Number is the primary means of linking survey respondents to the DER.
- **SWP\_457** - Non Plan 457 Special Wage Payment will be summed together with other earnings types to create total annual earnings for each respondent.
- **TIPS** – While tips are included in Box 1 total compensation, we will use this broken-out piece of earnings to create an indicator for the presence of tips in total compensation and will use this as an explanatory variable for modeling characteristics of employment and health insurance.
- **TYPE\_RECORD** – This is a utility variable created by SSA to indicate different types of records: regular employment, self-employment, and tips. We will use this variable in processing to identify the earnings source for each record and correctly handle the reported amounts in our accounting of total earnings.

### Project Completion

The researcher(s) will provide the Census Bureau with annual progress reports on this multi-year project. In addition, the researcher(s) will produce a Post Project Certification (PPC) summarizing whether and how the criteria described in the project proposal's Predominant Purpose Statement (PPS), as well as any others, were or were not met.

### Project Timeline

The project timeline below shows the activities for a single production cycle, anticipated to last approximately 18 months. The 2014 SIPP panel will have 4 production cycles as respondents are re-interviewed four times, providing data about 4 calendar years. The Census Bureau began the first production cycle in January 2014.

Task	Month									
	1-6	7	8	9	10	11	12	13	14-17	18
SIPP interviews	X									
All data transmitted to Census HQ	X									
PVS used to obtain SSNs for SIPP respondents		X								
SSN list sent to SSA		X								
SSA creates SER, DER, MBR, PHUS, SSR extracts			X	X						

Task	Month									
SSA returns extracts to Census				X						
Extracts matched to SIPP data				X	X					
Initial processing of SIPP variables that will receive model-based imputations		X	X	X	X					
Create FTI explanatory variables for imputation models					X	X	X	X		
Estimate imputation models using FTI explanatory variables					X	X	X	X		
Evaluate quality of imputations					X	X	X	X		
SIPP variables with modeled imputations move to downstream processing of remaining SIPP variables								X		
Processing of remaining SIPP variables; no FTI used								X	X	
Release of public use SIPP; no FTI variables included										X

Census will prepare an annual report for IRS to be delivered at the end of each processing cycle with the first report delivered by June 2015. This report will describe routine, on-going project activities for the year, as well as highlight and explain significant proposed or actual methodological changes, the addition or deletion of variables, and whether the number of variables being imputed has changed. The purpose of this report will be to keep IRS informed of how FTI data are being used as part of ARTS 892 in order to facilitate the renewal of the project at the beginning of the next SIPP panel. No new FTI variables not listed in this PPS will be added to SIPP processing without explicit permission from IRS to modify this PPS.