FSRDC Research Proposal Guidelines for Projects Requesting Access to U.S. Census Bureau Data

July 24, 2024

Persons wishing to conduct research on restricted U.S. Census Bureau data at a Federal Statistical Research Data Center (FSRDC) must submit a research proposal to the Census Bureau through the Standard Application Process (SAP). The purpose of this document is to provide guidance for the various sections of the SAP. Generally, this document applies to datasets that list the Census Bureau as the source in the SAP metadata catalog. The FSRDC program has partnered with other federal statistical agencies to host projects using other agencies' data. Proposals to use data from other federal agencies are submitted to these agencies through the SAP; see Partner and Collaborating Agencies for web links to details about their proposal requirements.

More information about the FSRDC program is available here. Prospective researchers should contact their local FSRDC administrator and FSRDC executive director as early in the process as possible, well before submitting a proposal through the SAP.

All research must have a statistical purpose. The Census Bureau is required by law to protect the confidentiality of data collected under its authorizing legislation, Title 13, U.S. Code. Title 13 specifically prohibits using Census microdata for legal, regulatory, or commercial purposes.[1] Researchers should not include analysis in their proposal that has or can be perceived as having a regulatory, commercial, or other non-statistical purpose. Proposals are also expected not to make specific policy recommendations.

Census Bureau policies prohibit projects that involve ingestion of any data such as political affiliation, party registration, and political contributions that permit explicit indication or inference of individuals' party affiliation. More generally, the Census Bureau may not approve proposals that examine sensitive topics or populations. Consult your FSRDC administrator early on if your project topic may be considered sensitive.

---

[1] 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 whereby 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."

Preliminary Proposal Development

Researchers who wish 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. The researcher should discuss the proposed project with the administrator to determine whether the research fits with the Bureau's mandate, is feasible, and is likely to provide benefits to Census Bureau programs under Title 13 of the U.S. Code. To assist in this assessment, the administrator will likely request a 2- to 3-page proposal overview from the researcher. This overview should include a summary of the research questions, a list of the internal Census Bureau datasets to be requested, a description of the project's methodology, and planned statistical output. Researchers can develop this preliminary proposal document into the various components of the SAP below should they decide to move forward. Metadata is available for each dataset on the SAP website, and dataset pages often include links to related papers and technical documentation. Researchers are encouraged to read through available information about relevant data for their proposed project.

It is expected that the proposal development process will involve multiple iterations of the proposal, prepared in conjunction with the local FSRDC administrator. The FSRDC administrator is a resource for assistance in clarifying a range of potential issues, including topics related to feasibility and Census benefits. The executive director(s) for the FSRDC(s) that researchers plan to use should also have given their approval for the research team to use the FSRDC before they submit the proposal via the SAP.

SAP Proposal Submission

One member of the research team will submit the final proposal online using the SAP, and the proposal will be reviewed by the Census Bureau. The SAP will ask for information about the research team and the proposed project. Question wording and layout may not be exactly as shown below. If researchers are uncertain about how to respond to an item, the local FSRDC administrator can provide guidance.

1. Data

The first step to beginning a project proposal using Census Bureau data is to select all required datasets by clicking the "Request Access" button[2]. The Census Bureau encourages the research team to consult with the FSRDC administrator about the datasets during proposal development. Researchers can reference https://www.census.gov/topics/research/guidance/restricted-use-microdata.html or the SAP metadata catalog to see the data available through the FSRDCs and the years available for each dataset. To check whether a dataset in the SAP catalog can be accessed in the FSRDC, click on "Data Access" and verify that Access Modality lists FSRDC. After selecting all datasets needed for the project, click "Start Application."

Some datasets have multiple files/components that need to be selected and added individually. For instance, the Economic Census includes eight datasets based on industry, such as the Census of Manufactures and the Census of Services; list each individually if requested. As another example, the Longitudinal Employer-Household Dynamics (LEHD) includes several component files (e.g., Jobs Data, Person Demographics (Workers Only), etc.); the FSRDC administrator can help researchers identify the files needed.

---

[2] Unless you are already logged in, immediately after clicking "Request Access" you will be asked to log into ICPSR. The email address that you use to log in will impact your ability to find any subsequent applications. Be sure to log in using the email address that you intend to use for all future SAP applications and communications.

FSRDC datasets that are listed in the SAP catalog can be provisioned for approved projects. Data files collected by the Census Bureau that are not listed are unavailable or not prepared for FSRDC use.

2. Research Team

The proposal should include all members of the research team, including collaborators who do not plan to access an FSRDC. List the names of the principal investigator (PI) and other researchers, including research assistants (RAs), associated with the project. FSRDC researchers can be citizens of the United States, permanent residents, or foreign nationals. All researchers who will access the RDC, including U.S. citizens, must currently reside in the United States and have resided in the United States for at least 3 of the last 5 years (i.e., 36 months of the past 60 months). Doctoral students submitting a proposal are asked to include their primary advisor as a co-principal investigator. The Census Bureau recommends that the advisor apply for Special Sworn Status.

For each member of the research team, the application form requires name (full legal name), affiliation (or place of employment), title (faculty, grad student, research scientist, visiting faculty, other), email address, phone number, citizenship (U.S. citizen or not), residency status (having lived at least three of the past five years in the U.S.), whether the researcher currently has active Special Sworn Status, and if the researcher will access the confidential data or not. An FSRDC administrator can look up whether a researcher still has active Special Sworn Status from a previous project.

The researchers' email addresses should be institutional email addresses (e.g., a work or school email address) rather than personal email addresses. Please carefully note the email account of the principal investigator (and Co-PI if applicable) submitting the proposal as this is the only account that future changes to the proposal can be made. Note that the email address used in the application is not automatically updated to match the email used to log into SAP. We recommend using the same email address both to log in as well as in the application.

If new collaborators or research assistants join the project after the proposal has been submitted, the FSRDC administrator can assist with adding them after the project has received final approval or while the project is active.

3. Research Description

This portion of the project application will provide the project details. Each field or section should be written so that it can be understood by a competent social scientist who is not necessarily a specialist in the field or on the topic within the field. Please keep in mind that the audience for a Census proposal is not the same audience researchers typically address when writing research proposals or journal articles. The Census Bureau has its own set of expectations and requirements, and therefore researchers often find that they must make significant revisions to preexisting proposals written for other purposes (e.g., grant applications).

Project Title

Provide a descriptive title for the FSRDC project.

Project Duration

The Census Bureau encourages researchers to carefully assess the time period for which they request access.

Requests for extensions beyond the specified 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 project duration is 36 to 60 months. Note that researchers may discover that it takes longer than anticipated to become familiar with the data and computing environment.

Funding

Some FSRDC locations and data-providing agencies charge fees, which are described more fully on the SAP Fees page. Researchers must contact the local FSRDC executive director to determine if access fees will apply. List sources of funding for FSRDC access fees, as well as funding for the research project overall.

Timeline

Upload a table with a timeline, including a column for each year and a row for each task. A timeline template is available in the SAP application for download. The timeline should include activities such as cleaning data, merging data, and conducting specific analyses, as well as an estimated timeframe for outputs, such as publications, presentations, or any public use products. Tie the activities in the timeline back to the benefits (see Census Benefits section below). The researchers should also list FSRDC annual reports and other required documentation, such as reports about the project's deliverables and achieved benefits (i.e., Technical Reports). Refer to the timeline below as an example.

Task (Benefit Criteria)	Year				
	1	2	3	4	5
Geocode import transaction data to county level (5,7,11)	X	X			
Link foreign transactions data to Commodity Flow Survey (CFS): match CFS to export declarations dataset; identify imports to port locations (5,7,11)	X	X			
Build on the firm-level links in Longitudinal Firm Trade Transactions Database (LFTTD) to establish foreign transactions data links at the plant level (5,7,11)	X	X			
Link shipping recipients in transaction data to Longitudinal Business Database (LBD) to determine North American Industry Classification System (NAICS) code (11)	X	X			
Submit annual report to Center for Enterprise Dissemination (CED) (5,7,11)	X				
Conduct data analysis for models of industry flow (11)		X			
Draft paper of industry-level estimates (paper 1) (11)		X			
Model entrance of movement of goods flowing through US (11)		X			
Assess goodness of fit, quality of estimates for industry-level estimates (11)		X			
Submit annual report to CED (5,7,11)		X			
Conduct data analysis for regional impacts of trade flows (5,11)			X		
Draft Technical Report determining overstatement of imports at ports; (estimate parameters of misreporting) (5)			X		
Draft paper of regional impacts of trade flows (paper 2) (11)			X		
Submit annual report to CED (5,7,11)			X		

Task				
Finalize paper of regional impacts of trade flows (paper 2) and submit for journal review. (11)			X	
Revise paper of industry-level estimates (paper 1) to address referee reports from journal. (11)			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 (11)			X	
Draft Technical Report assessing US/Canada transaction data (5)			X	
Produce preliminary and final draft of geography of import flows and submit for journal review (paper 3) (11)			X	
Submit annual report to CED (5,7,11)			X	
Conduct data analysis on extent of wholesale use by industry (11)				X
Link transaction data to LBD and Economic Censuses to estimate where inputs are eventually absorbed (7)				X
Technical Report on limitations of new *State Imports* data product given results of analyses of imports to port locations (7)				X
Draft paper measuring trade flows and using transaction data (paper 4) (11)				X
Revise papers 2 and 3 to address referee reports from journal. (11)				X
Complete paper 4 and submit to journal. (11)				X
Submit final Technical Reports (5,7,11)				X
Produce final linked dataset (7)				X
Complete Post-project certification (PPC) and submit to CED (5,7,11)				X

Research Question(s)

Provide an overall description of the research questions. This section should describe the purpose of the project including any key concepts that will inform the research methods or models and a general description of the sample(s).

Demonstrated Need

Discuss why the project needs non-public data. Explain in detail why the project's research questions can only be addressed using the requested restricted-use microdata. Be as specific as possible. List key variables and methodological advantages of the restricted file(s) as compared to a public-use file (if applicable).

Study Population

Briefly describe the study population or universe and how it relates to the research question.

Project Abstract

This section should be 150-250 words in length, written in 1st person, reference major datasets and key methods, and discuss expected findings. Researchers may include one or two sentences about why the research is important to their field(s) of study. It should capture the essence of the project proposal, similar to a scientific journal abstract.

Time, Geographic, and Other Units Requested

For each requested dataset, include the years of data needed. Future years of requested data as they become available during the project's duration will automatically be added to the project. The years of requested data need to be clearly justified, particularly for proposals requesting a long time series.

Work Location(s)

Please indicate which FSRDC(s) the research team will access. The use of each FSRDC location must be approved by the respective local FSRDC executive director prior to submitting the proposed project to the SAP portal. The SAP will list other work locations used by other agencies, but only FSRDC locations should be selected for projects that access Census Bureau data.

Data linkages

Upload a document summarizing how all the data fit together, how they will be linked, and their expected overlap, including expected sample sizes (if applicable). State the datasets to be linked, the unit at which the linkages will occur (e.g., individual, establishment, geographic), the purpose of the linkage (e.g., geographic/industry context), and information as to how the linkages are performed.

It is important to describe how and at what unit of observation or aggregation datasets will be linked. For example, if one is linking the Annual Survey of Manufactures (ASM), collected at the establishment level, to the Research and Development (R&D) survey, which is firm based, this section should incorporate this discussion. Also, linking survey data (e.g., the ASM and R&D survey are both surveys) to other survey data may lead to small sample sizes (and therefore potential disclosure issues) if the survey samples do not significantly overlap. When linking confidential data to externally provided contextual variables, detailing the level of geography (e.g., county, tract, etc.) at which this link will be performed is helpful.

Economic (firm and establishment) data sets share a common set of identifiers and can be linked together. Record level linkage for economic data can be done using the County Business Patterns Business Register. Many identifiers for demographic (household and person) census and survey data sets are designed to be used within the survey year and are not longitudinal, or linkable across demographic surveys. For example, the American Community Survey has a person- (PNUM) and household-level (CMID) indicator that uniquely identifies that respondent in a single year but does not link to the same person/household in other ACS years. However, many demographic datasets can be linked together using person-level Protected Identification Keys (PIKs). The PIK variable is either found within the dataset itself or in a crosswalk file. When needed for linking, such crosswalk files need to be listed separately in the data section of the SAP application.

Individual level linkage between user-provided data and Census Bureau-provided person-level data must be performed by internal Census staff and will incur fees.[3] 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.[4] FSRDC administrators can provide more linkage details.

---

[3] 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 a few departments. The researcher must pay the cost of having PIKs assigned to individuals in user-provided data.
[4] Furthermore, for business data with FTI, a complete match results in a violation of the IRS "fact-of-filing" information protection.

User-Provided Data

If the project team plans to bring in non-FSRDC data to the project, enter information about each dataset. The list should be as comprehensive as possible. Enter the dataset name, a short description of the data, and the approximate size of the file. The application will ask for the data source and whether the data is publicly available or proprietary. If available, include the link where the data can be downloaded.

Future requests for user-provided data not listed in the proposal will require additional justification and approval. Note that public-use Census data are considered external data (sourced from public web sites).
A project cannot have a dataset that has had disclosure avoidance protections applied (i.e., the publicly available microdata) in the same project space as the data that haven't had these same methods applied (i.e., the restricted data), because it could reveal confidential information about the disclosure avoidance procedures. This applies to public and restricted data from the same year or for samples that overlap. However, projects are allowed to use aggregated estimates, including those from summary tables or public-use microdata.

The SAP form asks whether Census Protected Identification Keys (PIKs) would need to be applied. This is only necessary if the project would like to link user-provided data at the individual level to Census Bureau person-level data. Consult with the local FSRDC administrator early on if the project will need PIKs applied to user-provided data.

Software Requirements

The FSRDCs have statistical software ready to use in the computing environment, such as SAS, Stata, R, Python, and OpenGeoDa. Consult with the FSRDC administrator about other software needs. If the project team requires statistical software other than that which is currently available to FSRDC researchers, please provide that information. Individual software requirements can often be accommodated, though this section must justify the priority and inability of available packages to do the work, and the research team must pay for the license. If needed, list any add-ons or other software packages needed.

Note that the FSRDCs use a Linux-based environment and do not have internet access.

Methodology

This section describes the statistical equations that will be estimated, what key variables are needed and how they will be measured, how the data will be used, and how all datasets fit together.

It is important in this section to describe not only the methods used to answer the research questions but also the methods used in producing the other benefits the project will provide to the Census Bureau as described in the Census Benefits section. For example, if the project will evaluate the quality of certain variables in the restricted Census dataset, this section should discuss in detail how variables will be evaluated and what types of results will be reported to the Census Bureau.

All requested Census data need to be discussed in the methodology section. The purpose for each requested dataset and its role in the analysis should be clear. Data crafting is an essential element of research design. The proposal needs to be as detailed as possible in describing how measures will be constructed and how the data will fit together to construct the focal sample(s). Generally, this is accomplished by discussing how variables are

constructed and from which datasets they are retrieved.

It is important to convey to the reviewers that the researchers have read the publicly available documentation about the datasets they are requesting and have thought through some of the implications of the data in relation to their research approach. Four sources of information about datasets are often helpful: the SAP metadata catalog, survey questionnaires, Census web pages detailing survey and sample design, and working papers that Census staff and FSRDC researchers have written using the data. Working papers by series are available at [Working Papers by Series (census.gov)](#) and [Working Papers, Center for Economic Studies, U.S. Census Bureau | IDEAS/RePEc](#). Also, the FSRDC administrator can provide additional information or answer questions about the data and variables if answers are not readily available in the public documentation.

The unit(s) of observation for the analysis and the groups of units for which researchers will carry out analysis are both important. For each, be as specific as possible. Is the unit of observation an establishment, firm, household, or person? Alternatively, will the unit of observation be at a more aggregated level (e.g., Census tract or county)? Will the analysis use some combination of units? Additionally, for what groups of units (what "levels") will researchers specify their models? Will researchers estimate models at, say, national or sub-national levels? If so, what levels? In addition, will researchers run separate models by gender, race, household structure, industry, etc.? If the project uses user-provided data, how big are the samples that will be analyzed? Information that describes the units of analysis and the groups of units/levels of samples the project will be using helps Census reviewers assess both the feasibility of the project and the risk of disclosing confidential information.

Write out the equations that will be estimated. Researchers do not need to give the exact functional form they plan to use, but they should 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. Researchers do not need to list every specification, nor every variable they will be using from each dataset, but they should discuss how the left- and right-hand sides of each equation will be measured (i.e., if they are dichotomous, categorical, or continuous, etc.). Specify the dataset(s) from which these measures will come and how they will be constructed.

List of References

Provide a brief discussion of the relevant literature. A full literature review is not necessary. Instead, provide a targeted discussion of the literature that focuses on the key datasets as well as the key concepts to be examined in the analysis. This will help demonstrate scientific merit of the research question(s) as well as demonstrate a deeper knowledge of the requested data. It is also helpful, where appropriate, to connect the proposal research to activities and initiatives performed by the Census Bureau. Include a list of publications referenced in the proposal.

Project Products

Describe the deliverables that the research team anticipates producing, such as journal articles, conference presentations, Technical Reports, PhD dissertations, government reports, Center for Economic Studies (CES) working papers, and other expected products. FSRDC projects are required to produce at least one Technical Report (see Census Benefits section below for more detail) and one CES working paper. Include the names of journals targeted for publication.

Requested Output

In this section, researchers must describe the output they expect to submit for [disclosure avoidance review](#) to

release 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 FSRDCs must emphasize model-based output (regression coefficients, standard errors, and the like). Summary statistics (variable means, etc.) are allowed, but only to the extent that they support model-based output. For example, the kinds of output expected to be released are regression estimates and tables similar to those found in an article in a peer-reviewed academic journal. Researchers who desire disclosure of large volumes of tabular output should request a special tabulation from the Census Bureau rather than requesting access to microdata at an FSRDC. [5]

The Census Bureau looks for several things in evaluating proposals for disclosure risk. The project application must clearly state the proposed methods and output. Reviewers understand that, because the research team is conducting research, it is unlikely that they know all the details in advance. Nevertheless, reviewers need enough detail to assess whether the proposed project can succeed without posing undue risk. In particular:

- Reviewers need to determine whether there may be possible "thin cells" in the output; for this purpose, a cell is the group of observations (individuals or businesses) underlying any estimate researchers may release. To evaluate this, reviewers need clear and accurate information on the types of output to be requested (e.g., model tabulations, graphs), the units of analysis underlying the output (e.g., individuals and businesses, or groups of these), and the groups (levels) for which researchers will request output (e.g., industry, gender, geography, and possible crosses of these). [6] For example, detailed industry-by-geography cells become thin very rapidly.
- Reviewers 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 reviewers must carefully consider. Note that including indicator variables in the analysis creates additional "cuts" at the sample (i.e., cross tabulations), but only if the estimated coefficients associated with the indicator variables are reported. Because of this potential issue, the review team encourages including statements in the proposal, as appropriate, for example: "Specifications will include detailed industry controls. However, we will not release these fixed effect coefficients; we will only note in our results tables that we included them."
- Reviewers need to know whether researchers will request tabulations, graphs, or maps; these can have particular difficulty satisfying disclosure standards.
- Reviewers need to know whether researchers plan to produce sub-state estimates. Depending on the population of sub-state areas, noise injection may be required.
- Reviewers will also look for types of output that are relatively unfamiliar, such as from new statistical techniques. They must try to assess their disclosure risk and welcome any insights researchers may have on this.

Disclosure risk is a complex topic, and researchers will need to discuss their proposed output with their FSRDC administrator as they develop their proposal.

---

[5] Demographic tabulations: https://www.census.gov/programs-surveys/popest/data/special-tab.html
More about custom tabulations: https://www.census.gov/about/policies/privacy/statistical_safeguards/custom-tabulations.html
[6] If a project uses 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."

Census Benefits

Researchers' access to the internal microdata is covered under legal statute because of their role in assisting Census with its work.[7] Proposals must demonstrate that the predominant purpose of the research is likely to benefit Census Bureau programs by providing one or more Title 13 benefits to the Census Bureau. The Census Benefits section should address clearly and specifically how the project will provide one or more of the Title 13 benefits listed below. The text length is 10,000 characters per benefit criteria, so clear and concise presentation is critical. The benefit description for each criterion should focus on the following:
- What benefits are you producing for this criterion?
- Why are they important?
- Which datasets will you use?
- What are your proposed deliverables (papers, etc.)? If the deliverable is a research paper that you will submit for publication, please include examples of journals you are considering for submission.

FTI is protected under Title 26 of the U.S. Code; generally speaking, the establishment and firm-based surveys and censuses include FTI variables. Requests for FTI datasets 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. Requests for FTI differ from other requests in two important ways. First, criteria 1 through 4 below apply only to projects that do not include Federal Tax Information (FTI).
Second, FSRDC proposals that request FTI are required to meet at least two of the criteria listed below. Thus, a second benefit is proposed, most commonly an exercise connected to survey planning and processing activities at Census, such as an analysis about the data itself. The ideal benefit analysis leverages the same data preparation work as is intended for the estimates of population. Non-FTI proposals may propose only one benefit. The potential benefits include:

Criteria for non-FTI projects:

1. Evaluating concepts/practices: The project benefits the Census Bureau by evaluating concepts and practices underlying Census Bureau data collection and acquisition, processing, estimation and measurement, and dissemination practices, including consideration of continued relevance and appropriateness of Census Bureau procedures to changing economic and social circumstances.

   Explanation: This criterion includes projects such as those that are designed to:
   - Evaluate whether the current data contain the information required to reflect changing social and economic circumstances and whether alternative sources improve data quality or reduce burden.
   - Evaluate whether published reports and data provide information relevant to these changing circumstances, and point to new dissemination methods that would improve their relevance.

   Examples:
   - Changes in family circumstances: Grandparents rather than parents are thought

---

[7] 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."

to be primary caregivers of a growing proportion of children. Evaluating new questions about this care giving relationship can point to:
- o potential improvements in questionnaire design;
- o additional areas where new questions are needed (for example, on the grandparents' physical health, and on their responsibilities as caregivers of other family members);
- o aspects of the relationship that are most important to convey in new published statistics and reports.
- Changes in sources of health insurance coverage: the new government State Children's Health Insurance Program (SCHIP) program was intended to provide health insurance coverage to some groups of children. Because the program was new, new questions were needed to collect information on it. Evaluating the responses provides important information on the quality of the new data, and on the need for additional questions, or modifications to the questions.
- Welfare reform: The legislation was intended to change incentives for labor force participation and employment, as well as for participation in government transfer programs. Extensive analysis of post-reform data can enhance the data's relevance by providing assessments of the legislation's effectiveness. These analyses can also provide information pointing to the need to evaluate the extent of deficiencies in questionnaire content and collection practices that hinder the data's relevance for this new circumstance.

2. Increase relevancy of trends: The project benefits the Census Bureau by analyzing changing demographic, social, or economic trends that affect Census Bureau programs, especially those that evaluate or hold promise of improving the quality of Census Bureau products.

Explanation: The Census Bureau needs new data products such as new reports, on-line tools, public-use files, etc. to provide relevant information about changing demographic, social, or economic trends. New questions, surveys, data sources, and methods may be required to ensure that Census Bureau products continue to be relevant in a changing economy and society. Improving relevance enhances the quality of the Census Bureau's products.

Examples:
- Changes in the demographic composition of households -- age structure, race and ethnicity, duration of relationships -- affect assumptions underlying the specific data the Census Bureau collects. For example, assumptions about the ages at which meaningful labor force participation begins, or formal schooling begins and ends, sometimes determine which individuals are asked about such behaviors. An aging society in which people continue to participate in the labor force for many years, and so have more years to benefit from mid-life education and retraining, may invalidate those assumptions.
- New reports on, for example, numbers of workers returning to formal schooling, by age.
- Number and socioeconomic characteristics of workers who are self-

11

employed, independent contractors, etc., supplementing information currently collected or not collected at all.
- Shifts in education patterns and employment in specific economic sectors may require information about new types of education, jobs, and employment practices. Analyzing the ability of current data to address such new and emerging patterns provides information the Census Bureau needs to consider whether it needs to modify existing occupational and employment classification systems, and questions about types of employment.

3. Developing utilities: The project benefits the Census Bureau by developing a means of increasing the utility of Census Bureau data for analyzing or evaluating public programs, public policy, and/or demographic, economic, or social conditions.

   Explanation: Census Bureau data are widely used to analyze existing and proposed public programs, to inform public policy decisions, and to investigate changing demographic, economic, and societal conditions. Publications and other public use data in their original forms may not produce the required information. In order to make the existing data useful, specific variables and data structures need to be developed. Constructing these variables and data structures allows the existing data to be used to address an expanded set of concerns.

   Examples:
   - An individual's participation in public programs such as Medicaid depends on their own income and labor force participation and on family characteristics. Those family characteristics include family income, labor participation by other family members, and members' eligibility for employer-provided health insurance. Some of the raw material from which such individual and family information can be constructed is available on the original Census Bureau data products. But some information may not be available, or available at the level of detail required, on the public-use data. For example, Medicaid programs are administered at the county level. Access to geographic detail beyond what is available publicly may be crucial to improved modeling of program effects or outcomes, or to an accurate description of economic conditions.
   - The racial and socioeconomic composition of neighborhoods is thought to matter to individuals when they consider whether to move. Although potentially relevant data have been collected in the American Housing Survey to address this hypothesis, making use of them requires access to internal data. The data can be used to (a) characterize how stable neighborhoods are over time with respect to the race and economic status of residents within an area; (b) examine how the racial composition and socioeconomic status of neighbors affect whites' and blacks' development of plans to move out of the neighborhood and their actual out-migration; and (c) examine how the perceptions of the neighborhood by individuals and their neighbors, particularly with respect to crime and the quality of schools, influences the process of moving for whites and blacks.

4. Enhancing administrative functions: The project benefits the Census Bureau by facilitating data collection, processing or dissemination, including activities such as administrative support, information technology support, program oversight, or auditing under appropriate legal authorities.

   Explanation: Projects may benefit the Census Bureau by improving the mission-enabling administrative functions that support our data collection work. This may require advice, collaboration, oversight, or direct involvement of persons who are not Census Bureau employees.

   Examples:
   - While conducting the Census Bureau's programs requires many skills, some critical skills, such as administrative and information technology support, are most effectively acquired through the flexibility of contractor and other non-employee relationships.
   - Appropriate oversight of the Census Bureau's operations may require direct involvement of program sponsors or others with legal oversight responsibilities. For example, program sponsors may wish to observe data collection activities, or to review in detail proposed modifications to data processing. Such review requires access to the relevant Title 13 data, such as the specific response of the observed data collection.

The remaining nine criteria apply to all projects.

5. Improving quality of data: The project benefits the Census Bureau by helping to understand or improve the quality of data the Census Bureau collects or acquires.

   Explanation: The Census Bureau needs to understand and continually assess the quality of all the data it collects, and to seek ways to improve them. Understanding the limitations of and improving the quality of these input data can greatly expand their utility and the quality of the analyses they inform.

   Examples:
   - Examine little-used data. If the variables the project uses have not been used previously, or have been used in a very different application, then the Census Bureau benefits by having a researcher examine the data carefully. Good empirical analysis often begins with tasks such as examining where records or items are missing, where responses are extreme, or take on inconsistent values. The examination will be far more extensive than can be carried out in the routine internal consistency checks during survey processing. The examination should lead both to an assessment of this aspect of data quality, and to recommendations for directions for improvement.
   - Compare similar data from different sources. Independent sources often contain measures of similar concepts. Comparing measures from independent sources that should be similar, or that should differ in predictable ways, increases the Census Bureau's knowledge of its data collection programs.

6. Improving methods for data collection: The project benefits the Census Bureau by leading to new or improved methodology to collect or tabulate data.

   Explanation: Continual assessment of the Census Bureau's current practices and analysis of our previous practices is necessary, due to changing economic and social circumstances, and the development of new statistical and economic methodologies. Using these data to improve of our methodologies ensures the continued relevance of our data and programs.

   Examples:
   - Analysis of existing data may show that information should be collected at a different level, for example, from persons rather than households, or from plants rather than firms.
   - Analysis may show that data should be tabulated and published at a different level. For example, data tabulated at the firm level perhaps should also be tabulated at the plant level. Similarly, there may be meaningful tabulations at the household level of data currently only tabulated at the person level.

7. Enhancing data usage lifecycle: The project benefits the Census Bureau by enhancing the data it collects, acquires, or publishes. This includes, but is not limited to
   A. Improving imputations for non-response;
   B. Developing links across time or entities for data gathered in censuses and surveys;

   Explanation: The collection of Title 13 data involves a substantial investment of resources and taxpayer dollars. Applying new techniques to improve the quality of these data increases the value of these investments. Insights drawn from linked data may inform future data collections.

   Examples:
   - Analysis of important economic or social relationships can be compared with the explicit or implied relationships used in nonresponse imputation algorithms, leading to suggestions for improving the algorithms. Such improvements enhance Census Bureau data.
   - Linking external data to Census Bureau data enhances the Census Bureau data. Future researchers can make more informed inferences about economic and social relationships using these linked data. Those inferences may improve imputations for non-response or provide information about the quality of sampling frames and data collection techniques.
   - Linking existing Census Bureau data by developing longitudinal files for data from businesses or households creates new data that enhance the information collected in the survey or census. While some business data have been linked extensively, those links have not been exhaustively evaluated.
   - Links between business and person or household data, and links over time of person and household data, are not as extensive and have not been exhaustively evaluated. Such linkages enhance the information collected in each data set.

8. Improving underlying classification schemes: The project benefits the Census Bureau by identifying the limitations of, or improving the underlying Business Register, household Master Address File, and industrial and geographical classification schemes used to collect and code the data.

   Explanation: The Census Bureau must evaluate whether emerging new social and economic patterns reduce the integrity of our household or business sampling frames (such as the Business Register, or Master Address File). Sources to produce robust sampling frames and classification schemes are critical to ensuring the Census Bureau's data collections are relevant and cost-effective.

   Examples:
   - Identify errors in geographical and industrial coding, and potential systematic causes of those errors.
   - Linking data for the same households or businesses, either between two Census Bureau data sets or by linking data from other sources, can provide this kind of information.
     - an outside data set may identify businesses or individuals that should be in the Census Bureau frame but are not;
     - understanding the sources of any differences in economic or geographical coding improves the quality of Census Bureau data.

9. Identifying further data needs: The project benefits the Census Bureau by identifying the shortcomings of current data collection programs and/or documenting new data needs.

   Explanation: These projects benefit the Census Bureau by evaluating whether our programs collect sufficient information to address important social, demographic, or economic questions. These analyses are essential to ensure that the Census Bureau's data products are timely and relevant.

   Examples:
   - No information may be collected, or there may be insufficient detail to estimate important comparisons among subgroups of these populations.
   - Research may point out such deficiencies as data are used to address such questions or make such comparisons.
   - Research may identify the best directions in which the Census Bureau should begin collecting data to fill these gaps. For example, research may identify the need for additional information on the materials that businesses purchase to produce their product, and may in particular identify the most important details on which information should be gathered.

10. Improving sampling frames: The project benefits the Census Bureau by constructing, verifying, or improving the sampling frame for a Census Bureau data collection.

Explanation: Sampling frames are a fundamental building block for Census Bureau data collections. Continually evolving economic and social circumstances affect the appropriateness and completeness of existing and potential frames, and continual changes in statistical tools and methods affect the best practice techniques for using those frames.

Examples:
- Outside data sources can be used to construct alternative sampling frames. Linking the alternative and Census Bureau frame allows researchers to assess the Census Bureau frame and either verify it or suggest improvements.
- Alternative sampling strategies can be tested within existing sampling frames. Assessing the comparative characteristics of the alternative and existing samples provides information to verify the robustness and appropriateness of current practice, or to suggest improvements.

11. Describing population: The project benefits the Census Bureau by preparing population estimates and describing characteristics of the population.

    Explanation: These projects develop estimates for relevant population and subgroup characteristics that augment the Census Bureau's existing publications and projects to provide more complete and relevant data products.

    Examples:
    - Research typically yields statistics beyond those that the Census Bureau has already released. These statistics estimate specific populations and subpopulations and their characteristics.
    - Such statistics include summary statistics about specific variables (means, medians, moments), and coefficient estimates that summarize behavior of subgroups of the population. These statistics increase the information available about these populations, subpopulations, and their characteristics.

12. Improving survey estimates: The project benefits the Census Bureau by developing methods to address census or survey non-response to improve estimates.

    Explanation: Understanding patterns of nonresponse, and its sources, is of great importance to the Census Bureau because response rates are an important indicator of data quality. Nonresponse is so important to data quality that the Census Bureau initiated, participates actively in, and supports, long-standing interagency groups that jointly explore better measures of response rates, sources of nonresponse, and ways to improve response rates.

    Example:
    - Research that carefully addresses patterns of response, and the impact of nonresponse on data quality, provides important information that the Census Bureau needs to improve the quality of its data.

13. The project benefits the Census Bureau by developing statistical weights for a data collection or analysis conducted by the Census Bureau.

Explanation: Appropriate weights are essential to correctly presenting data. These weights are critical for the usability and relevance of Census Bureau data analyses.

Examples:
- The research may assess how survey weights are currently developed. Findings from such research are valuable to the Census Bureau in improving survey weighting.
- Particularly for external research projects, the research may address issues such as nonresponse that the researcher does not explicitly connect to the Census Bureau's processes for developing survey weights. However, Census Bureau staff may recognize that such factors are considered in developing survey weights, or should be considered. Such research findings will provide important information to the Census Bureau.

Most FSRDC proposals include as a benefit preparing population estimates and describing characteristics of the population (criterion 11). One can think of the Census survey life cycle, loosely speaking, as a series of four steps: planning, conducting, processing, and disseminating. Research estimates are a key form of information dissemination. FSRDC-based research leverages the 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 if and how the data can support a meaningful understanding of the underlying processes of our economy and broader society.

A high-quality Census benefit (i) establishes a connection to Census program activity, (ii) articulates a specific analysis exercise, along with its purpose and approach, that can be clearly understood by reviewers and Census staff, and (iii) is generally focused on only one of the above criteria. Discuss what might be found, why it could be informative, and, especially, specific information on how the analysis will be approached. It is strongly encouraged that researchers consult an FSRDC administrator prior to drafting the Census Benefits section.

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 Census web pages. 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.

Each benefit should produce at least one specific deliverable. These may include crosswalks, prototype data products, code, Technical Reports, or working papers. Researchers are required to submit Technical Reports documenting the production of research findings benefiting the Census Bureau to the CES Technical Notes series. Technical Reports should include documentation of programs and files used to generate programmatic as well as scientific (including Criteria 3 and 11) research findings reported in CES Working Papers and other venues. The Technical Report series serves as a repository of beneficial research findings and documentation that generally have not undergone disclosure avoidance review and are made available in a secure location upon request only to Census staff and external researchers authorized to view such findings. FSRDC projects are also required to produce at least one CES working paper.

One 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. For example, by linking Census establishment

microdata 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.

Another 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, it also provides some issues that may require adjustments 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 and the commuting transportation mode.