

CENSUS BUREAU DATA STEWARDSHIP PROGRAM

DS014 - Data Linkage Policy

PURPOSE

To establish criteria for conducting projects that use data linkage at the U.S. Census Bureau.

BACKGROUND

Since the 1940s, the U.S. Census Bureau has routinely matched records to support its mission. Title 13, United States Code (U.S.C), Section 6 provides the Census Bureau authority to do this by directing the Census Bureau to acquire and use information in lieu of conducting direct inquires of the general public. Data linkage activities are essential for compiling, evaluating, editing, and analyzing the Census Bureau's census and survey data, as well as other data. These activities also help the Census Bureau improve the content and coverage of its data collections, permit cross-sectional and longitudinal analyses of populations and establishments, explore and improve the quality of estimates and other administrative data, and ultimately provide quality data products that would otherwise be infeasible without adding cost or burden to the public.

Over the last decades, the Census Bureau's linkage activities have increased with the availability of administrative data and computing technology. While the benefits of linkage are clear, the Census Bureau recognizes that these activities must take place in a strategic and measured manner that respects its legal authority, mission, and ethical standards with regard to individual privacy.

This policy provides guidance to Census Bureau employees and Special Sworn Status individuals who engage in data linkage activities. This policy sets the framework for the Census Bureau's corporate requirement for openness regarding linkage activities involving data on persons, households, businesses, or addresses. The Census Bureau will communicate with the public about its linkage activities, the purposes and public benefits thereof, and how the Census Bureau mitigates privacy risks by protecting data confidentiality. The policy requires six business practices:

1. **Mission Necessity** – The linkage must be necessary and consistent with the Census Bureau's legal authority and mission.

2. **Best Alternative** – The Census Bureau will examine alternatives for meeting the project objectives and determine that linkage is the best alternative given considerations of cost, respondent burden, timeliness, and data quality.
3. **Public Good Determination** – The Census Bureau will document the benefits to be gained by the information resulting from the linkage. In addition, the Census Bureau will proactively implement procedures to mitigate any risks resulting from the linkage. The confidentiality of the resulting information is protected under Title 13, U.S.C.
4. **Risk and Sensitivity Assessment** – The Census Bureau will assess the public perception of the level of risk to individual privacy of a particular linkage and create an appropriate level of review and tracking.
5. **Openness** – The Census Bureau will communicate with the public about its linkage activities, how they are conducted, and the purpose and benefits derived from them.
6. **Systematic Review and Tracking** – The Census Bureau will subject all in-scope linkage activities to a systematic review process, using the criteria set forth in this policy, and these projects will be centrally tracked.

The Policy Section outlines these six business practices in greater detail. The Census Bureau operationalizes these business practices in a *Data Linkage Checklist* (see Attachment A).

Attachments B and C provide an explanation of the legal authorities, ethical guidance, and relevant policies and procedures that interact with the *Data Linkage Policy*, as well as the definition of key terms.

SCOPE

The concept of “data linkage” describes the process of creating a one-to-one correspondence for an entity (e.g., person, household, or business/establishment) by combining two or more separate datasets (or files).

Additionally, linkages that create a connection between files for an individual, between an individual and a household, or between an individual and a business/establishment are in-scope for this policy. Furthermore, where linkage is concerned, the policy applies to the creation of a linked file from data files using personal or business identifiers, as well as linking files whose personal and/or business identifiers were previously replaced by a unique, anonymized identifier (e.g., a protected identification key, or PIK).

This policy covers data linkages that are:

- Conducted on behalf of the Census Bureau;¹
- Undertaken for statistical purposes² under the authority of Title 13, U.S.C.; and
- Protected (i.e., the linked product) under the confidentiality provisions of Title 13, U.S.C., Sections 8 and 9.

This policy applies to many types of linkages undertaken at the Census Bureau. For example, linkages that are carried out to enhance the content or coverage of data collected or acquired by the Census Bureau, to obtain cross-sectional or longitudinal information about populations and establishments, to develop new methods to build innovative infrastructures to support future censuses or surveys, and to explore and improve the quality of estimates and other administrative data.

Data linkages involve external data and/or data internal to the Census Bureau.³ Some linkages involve data external to the Census Bureau **and** data collected by the Census Bureau directly, under the authority of Title 13, U.S.C. (e.g., linking federal tax information with the decennial census). Other linkages involve external data **and** reimbursable survey data collected by the Census Bureau under Title 13 Section 8(b) (e.g., linking the SIPP with Social Security Administration data). Yet, other linkages involve only data collected directly by the Census Bureau under the authority of Title 13, U.S.C. (e.g., linking decennial census responses to the American Community Survey).

In the past, linkages within the scope of this policy tended to be carried out as part of a research agenda outside of a production environment. However, as the breadth and depth of our linkage skills and projects grow, the Census Bureau undertakes more and more experimental and/or novel methods involving linkages within a production environment, which are all in-scope for this policy.⁴

¹ Data linkages are carried out by Census Bureau employees or special sworn status (SSS) individuals because they are the only individuals allowed access to confidential data. Furthermore, linking activities are carried out at Census Bureau Headquarters and Federal Statistical Census Bureau Research Data Centers (FSRDCs), as well as other Census Bureau approved facilities.

² The Census Bureau only carries out linkages for statistical purposes, never for enforcement or other purposes.

³ In this context, the term external data is used to refer to data that the Census Bureau did not collect directly. This could include commercial data as well as administrative data produced by governmental and nongovernmental entities that the Census Bureau would acquire and use.

⁴ Novel, or new, methods involving linkage that fall under the scope of this policy would be those methods that require experimentation, testing, and vetting before they can become a permanent part of the production cycle.

Out-of-scope data linkages include:

- Linkages created in the course of production data collection and processing operations necessary to conduct censuses or surveys;⁵ and
- Linkages undertaken by the Census Bureau on behalf of another agency on a reimbursable basis that do *not* involve files that are protected by Title 13 (i.e., linking reimbursable survey data, not protected under Title 13 with an external, sponsor-provided file).

For further guidance on the types of linkages that are considered in-scope or out-of-scope, see the examples in Figure 1 on the next page.

⁵ Current linkages considered out-of-scope for this policy include those methods that have undergone experimentation, testing, vetting, and have been in place or in production prior to the signing of this policy.

Figure 1. Examples

In-scope data linkages subject to this policy:

- Linking external data to a Title 13 census or survey, for example...
 - commercial cell phone use data provided by a cell phone service provider linked to the American Community Survey (ACS)
 - data provided by the Social Security Administration (SSA) linked to the Survey of Program Participation (SIPP)
 - multiple external data sources (e.g., state-provided earnings and unemployment data, and administrative data from various sources) linked with the economic and decennial census, as well as the ACS
- Linking internal Title 13 or reimbursable data, for example...
 - decennial census data linked to the Current Population Survey (CPS)
 - demographic data from the SIPP linked with the economic census
- Linking panels of a survey, or different vintages of data, to create a longitudinal dataset, for example...
 - 5 years of historical CPS modules linked with ACS data or linking several decades of decennial census data together at the individual or housing unit level

Out-of-scope data linkages not subject to this policy:

- Linkages created in the course of production data collection and processing operations necessary to conduct censuses or surveys, for example...
 - Linkages necessary to carry out a survey or census, such as linking census data to a universe within a master database (i.e., the Master Address File, known as the MAF, or the Business Register)
 - Linkages necessary to update master databases (MAF, Business Register)
 - Linking Title 13 or reimbursable census/survey results or listings back to the data source from which they came, such as independent listing operation outcomes linked back to the decennial census for evaluation purposes
 - Using the CPS Crosswalk File when working with historical and current CPS datasets
- Linking that involves Foreign Trade Data
- Comparing aggregated data between two or more separate files that have not been linked at the person, household, or business level, for example...
 - Aggregating confidential micro-data from the ACS and the SIPP to the county level to compare estimates of poverty and educational attainment—without actually linking the ACS file to the SIPP file at the individual or household level
- Using a pre-existing, linked dataset that already has approval and will not, during the course of its reuse in a new project, be linked with other datasets. For example...
 - Using the Longitudinal Employer-Household Dynamics' Employment History File or the Employer Characteristic File

POLICY REQUIREMENTS FOR DATA LINKAGE

This policy outlines six business practices that must be followed in order to carry out data linkage activities at the Census Bureau:

1. Mission Necessity – The linkage must be necessary and consistent with the Census Bureau’s legal authority and mission.

The proposed linkage must satisfy the following criteria:

- Can be conducted under the Census Bureau’s legal authority (Title 13).
- Is conducted for statistical purposes only.
- Is conducted only by Census Bureau employees or special sworn status (SSS) individuals.
- Fulfills the Census Bureau mission to serve as the leading source of quality data about the nation's people and economy.

2. Best Alternative – The Census Bureau will examine alternatives for meeting the project objectives and determine that the linkage is the best alternative given considerations of cost, respondent burden, timeliness, and data quality.

The proposed data linkage considered the following criteria:

- Alternatives to linkage that could meet project objectives are evaluated (e.g., direct collection).
- The linkage is demonstrated to be the most cost-effective among the alternatives.
- The linkage is demonstrated as the best way to reduce public burden in accordance with the Paperwork Reduction Act.
- The linkage meets data user needs in the most timely fashion.
- The linkage will provide higher quality than, or at least comparable quality to, data from other alternatives.

3. Public Good Determination – The Census Bureau will document the benefits to be gained by the data linkage. In addition, the Census Bureau will proactively implement procedures to mitigate any risks resulting from the linkage. The confidentiality of the resulting information is protected under Title 13, U.S.C.

The proposed data linkage must satisfy the following criteria:

- The linkage will result in more accurate, more descriptive and/or higher quality data than can be produced without such a linkage.
- There is a demonstrated public need for the data produced by the linkage;
- The Census Bureau uses secondary (i.e., data from other government agencies) and third party/commercial data in a way that is consistent with the data provider's confidentiality and use provisions outlined in an interagency or other special agreement.
- The linked product is protected under the confidential provisions of Title 13, U.S.C., Sections 8 and 9.

4. Risk and Sensitivity Assessment – The Census Bureau will assess the public perception of the level of risk to individual privacy of a particular linkage and create an appropriate level of review and tracking.

This business practice is routinely satisfied during the typical project review cycle, and includes the following activities:

- The potential sensitivity that the general public would have to the linkage is identified and responded to appropriately during the project review cycle.
- A Data Linkage Checklist is completed for every linkage project. Along with concurrent reviews undertaken in other parts of the Census Bureau (i.e., scientific merit, subject matter review, and policy review) it documents a proposal's sensitivity level and compliance with the *Data Linkage Policy*.
- Sometimes the Census Bureau relies on an executive-level governing body, the Data Stewardship Executive Policy (DSEP) Committee, for guidance on the pursuit of potentially sensitive data linkage projects.
- The Census Bureau seeks the approval of the data-providing agency for any linkages carried out with records received from other entities, when required.

5. Openness – The Census Bureau will communicate with the public about its data linkage activities, how they are conducted, and the purposes and benefits derived from them.

This business practice must be satisfied in the following ways:

- The public is made aware of linkage activities at the Census Bureau through the publication of System of Records Notices (SORNs) in the Federal Register. The SORN describes the content of the system of records, the routine statistical uses of these data, and provides a period for public comment.

- The Census Bureau also publishes information about linkage activities on our [public website](https://www.census.gov/about/what/admin-data.html). (<https://www.census.gov/about/what/admin-data.html>)
- And other public education opportunities, when feasible.

6. Systematic Review and Tracking – Data linkage activities will undergo a systematic review process using the criteria set forth in this policy and be centrally tracked by the Census Bureau.

This business practice is routinely satisfied during the typical project review cycle, and includes the following activities:

- The data linkage activity meets relevant project review criteria (e.g., it demonstrates a Title 13 benefit to the Census Bureau and passes several reviews in the areas of scientific merit, and policy issues).
- Linkages involving administrative data meet the project review criteria set forth in the *Handbook for Administrative Data Projects*.
- The project information (e.g., methodology, data sets, staff, results, project number, etc.) is maintained and tracked in a centralized electronic tracking system.
- Access to data sets, including those received from other agencies, is centrally managed by the Census Bureau and Information Owners.

To assist staff in effectively analyzing proposed linkage activities, the Census Bureau created a checklist designed to help staff determine whether a project meets this policy's requirements and to gauge the sensitivity of the proposed linkage (See Attachment A for the Data Linkage Checklist).

The Data Linkage Checklist will be completed electronically in the Census Bureau's centralized electronic tracking system or completed on paper and uploaded there. In addition, for each record linkage project, a Project Description and other information must be uploaded to that electronic tracking system for review and approval by all necessary parties.

EFFECTIVE DATE

Upon signature.

LEGAL AUTHORITIES

This policy provides criteria to specify which projects are appropriate within the scope of its legal authority, as specified under Title 13, U.S.C., Sections 6, 8, and 9. Projects conducted

under Title 13 U.S.C. also are subject to the provisions of the Privacy Act and the legal authorities of the sponsoring agency. Additionally, this policy is consistent with the provisions of the Paperwork Reduction Act of 1995.

IMPLEMENTATION

Please consult the *Handbook for Administrative Data Projects* for further details on the process of approval for data linkage projects; it follows the approval process for projects that use administrative records. The process can be summarized as follows:

- Project Contacts are responsible for completing the Data Linkage Checklist as well as developing a project description for each linkage project deemed in scope. Both the project description and checklist must be uploaded to the Census Bureau's centralized electronic tracking system. Projects that lack this full documentation in the tracking system will be returned to the Project Contact for revision and resubmission.
- Division Chiefs and others specified in the Division review process are responsible for thoroughly reviewing the Data Linkage Checklist, as well as all other project materials, for each new project.
- The Information Owner must approve of the data's use, per the project's proposal, and will indicate whether additional agency approvals are required.
- The Administrative Records Coordinator, in the Policy Coordination Office, is responsible for reviewing all Data Linkage Checklists as part of the policy review for these types of projects.
- The Policy Coordination Office, in consultation with the Data Integration Policy Committee (DIPC), is responsible for ensuring an evaluation of the policy at regular intervals.

RELATED DOCUMENTS

A number of Census Bureau policies and procedures laid the foundation for the *Data Linkage Policy*, including the following *highlights*:

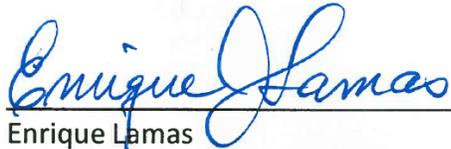
- The *Strategic Plan (FY18-22)*, Goal 1 Objectives 1.2 and 1.3, which provide direction for reducing respondent burden and collecting data once from respondents and using it for multiple data products.
- The *Privacy Principles*, which outline how the Census Bureau protects the privacy of individuals.
- The *Administrative Data Acquisition, Access, and Use Policy* establishes a mandatory business process for conducting projects that use administrative data. Its companion

piece, the *Handbook for Administrative Data Projects*, 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 proposals involving administrative data.

POLICY OWNER

Administrative Records Coordinator, Policy Coordination Office (PCO)

SIGNATURE

 11/16/2018
Enrique Lamas
Chair, Data Stewardship Executive Policy Committee

Summary Information	
Policy Title:	Data Linkage Policy
Date Signed:	November 16, 2018
Last Reviewed:	November 5, 2018
Intended Audience:	All Staff
Policy Owner:	Administrative Records Coordinator, PCO
Office Responsible for Implementation:	PCO
Office Responsible for Dissemination:	PCO
Stakeholder Vetting:	DSEP, PCO, Legal, DIPC