Frequently Asked Questions about Accessing Data in the Data Linkage Infrastructure

Last revised: September 8, 2016

What needs to be included in the project proposal?

To gain approval to conduct a project, a named Project Contact must write a proposal that indicates the project's methodology and objectives, anticipated output, benefit to the Census Bureau, and datasets required (including any that need to be acquired for the current project). The proposal also includes logistical details about where the project work will take place, over what period of time, and the contact information for supporting staff who will participate. If the project plans to bring external data into the Data Linkage Infrastructure, the Project Contact should submit with the proposal a letter authorizing this use and transfer of the external data, particularly noting uses of personally identifiable information (PII).

Proposals to conduct research with administrative records are reviewed for scientific merit, feasibility, and potential risk of disclosure. Projects requesting the use of Census Bureau data, or "Title 13 data," must demonstrate a Title 13 benefit, which includes program evaluation. The approvals process is managed by the Census Bureau's Policy Coordination Office and tracked. Reviews are conducted by representatives of the Policy Coordination Office, Census Subject Matter Experts (SMEs), an Executive-level Division Chief, and a local Information Owner who is responsible for knowing restrictions and terms of use. When requesting non-Title 13 datasets in the Data Linkage Infrastructure, the Census Bureau will facilitate project approvals from other data owners and providers, as determined by their agreements with the Census Bureau.

Once my proposal is approved, how do I access the data?

Researchers on administrative records projects must either be current Census Bureau employees or obtain Special Sworn Status. The Census Bureau gives Special Sworn Status (SSS) to individuals to conduct work that specifically benefits a Census Bureau program. SSS individuals are sworn for life to protect the data as Census Bureau employees are sworn, and they are subject to the same legal obligations and penalties. Employees and SSS on administrative records projects must take annual Data Stewardship Training, Title 26/FTI training, and any additional training required by the agreement associated with the particular records being used. Users must also adhere to relevant ethics, confidentiality, and privacy protection procedures.

When an administrative records project is approved, a Census Bureau data access team confirms that all researchers have completed required trainings and then provides access within a secure computing environment. Access at one of the Federal Statistical Data Research Centers (FSRDCs) is most common, but remote access may be allowed in some cases. Use of an FSRDC or other mode may incur a charge. Users can access read-only, de-identified versions of approved files, and they conduct all work within a project-level workspace shared by the approved researchers. All analysis of microdata must take place in this computing environment.

Access is project-specific and limited to the proposed time frame of the project.

What datasets are available? For programs administered by states, which states can I use?

Access varies by data set and state, since each data set held in the Data Linkage Infrastructure has its own governance structure determined by the data owner's agreement with the Census Bureau. For some program data sets, Census has all the states; for others, just a handful. Moreover, some state data agreements already allow projects by external (non-Census Bureau) researchers, others allow approval on a case-by-case basis, while still others restrict use by external researchers. See the Administrative Data Inventory for a list of data currently held in Census' linkage infrastructure.

Can I bring in data to link for analysis?

Researchers are encouraged to bring data, particularly randomized controlled trials (RCTs), into the expanding Data Linkage Infrastructure to link with other datasets. Data external to the existing infrastructure must use secure transfer to bring in to the Census Bureau system. If the data need linkage identifiers, the Census Bureau can create them (may incur a charge). Conducting linkage services requires the data have personally (PII). Census then processes the data, removes PII, harmonizes and links the datasets, and pushes the de-identified data to the secure project workspace. In some cases, the project may require a formal agreement for the Census Bureau to acquire the data in order to process and link the data.

How are datasets linked in the infrastructure?

The Census Bureau's Center for Administrative Records Research and Applications (CARRA) has a person identification validation system that assigns a unique identifier, the Protected Identification Key (PIK), to people using personally identifiable information (PII), which does not necessarily need to include a Social Security number. CARRA uses the PIK to link people across datasets. CARRA also does address matching. The Census Bureau's Center for Economic Studies (CES) links businesses across datasets using primarily the tax ID, Employer Identification Number (EIN). The linking process itself is usually a one-time service, at least for person- and address-matching, that does not need to be repeated. Executing the data acquisition and linkage agreement can take 5-12 months.

How do I bring out my analytical results?

When researchers need to remove aggregated output, tables, or model coefficients, the output must be reviewed for disclosure to ensure that no individual or business information is revealed and that output is consistent with the original proposal. For Census Bureau data, disclosure review is conducted by a Disclosure Avoidance Officer (DAO) or by the full Disclosure Review Board (DBR).. The Census Bureau will also facilitate disclosure review by all relevant data owners. Once the results pass the DAO's review, disclosed output is provided to the research team outside of the secure computing environment, usually via email. The research team can then produce reports, presentations, and other products outside of the secure environment.

Depending on the terms of the data provider's agreement with the Census Bureau, these products may also go through a final review by the data-providing agency or Census team member to confirm that final products were consistent with the original proposed project.

I heard the Census Bureau is streamlining its access process. How is it changing?

Currently, external researcher access to Census Bureau data (Title 13) happens exclusively through Federal Statistical Research Data Centers (FSRDC) following a lengthy approval process and with limited capacity to link data sets, including data sets held in the Census Bureau's Data Linkage Infrastructure. The new model would provide an expedited approval process, much greater data linkage capacity, and potentially provide an opportunity for remote/virtual RDC access. For program evaluation projects using non-Census datasets only, the Census Bureau will facilitate project approval from data owners, data linking and provisioning, and disclosure review. The virtual access mode is uncertain due to policy constraints which vary by data owner. For instance, access to Internal Revenue Service (IRS) (Title 26) data is only allowed at the FSRDCs or Census Headquarters. Additionally, Census is currently making an effort to negotiate and include program evaluation as a Title 13 research purpose in its data acquisition agreements.