SIF Secondary/Administrative Data Use: Accessing Restricted-Use Data

Restricted-Use Data

Secondary/administrative data, such as test scores from schools, or job and medical records, can be a useful resource and cost effective for use in an evaluation study under the right circumstances. However, secondary or administrative data (e.g. school academic records, unemployment insurance data) may be difficult to obtain due to an agency's commitment to protecting and preserving individuals' confidentiality.
Secondary/administrative data may have already been modified so that sensitive data cannot be connected to a particular person or establishment, and steps are necessary to make it possible to link this data back to other data collected on participants during the research study.

You can prepare your organization/team for the challenge of using sensitive restricted-use data already available from other sources by: understanding the dataset your organization/team is interested in (e.g. what variables it contains, who the respondents are); developing a cooperative relationship with the organization or people who generated the data source; and implementing security measures to assure that the data will be kept secure and minimize any security breach where individuals personally identifiable information is released.

Restricted-use data contain personally identifiable information (PII). This could include:

- ❑ Age
- ❑ Birthdate
- ❑ Marital Status
- ❑ Spouse Name
- ❑ Home Telephone Numbers
- ❑ Education History
- ❑ Biometric Identifiers
- ❑ Medical Information
- ❑ Financial Information
- ❑ Employment Information
- ❑ Criminal History
- ❑ Social Security Number
- ❑ Credit Card or Bank Account Numbers

Be aware that some information not considered PII can become PII when data are combined.

For example, a ZIP code field in itself does not identify an individual. However, when combined with race and annual income, it could point to unique individuals (e.g., extremely wealthy or poor) within that ZIP code.

Knowledge of the Dataset

Prior to deciding if a particular existing dataset is a useful resource for your study, a good understanding of the information available in that dataset as well as the data composition and structure will help to determine if it is compatible with the data your study has collected, and thus useful for an impact study (e.g. randomized, quasi-experimental designs). For example:

- Can the necessary data be accessed and stored during the time frame of the study, and given the available means?
- Do team members have the appropriate experience and expertise to effectively use the data?
- Is the secondary/admin data from an appropriate and useful sample?
- Were the data collected at a useful and appropriate time? Are there a sufficient number of data collection time points for each respondent or studied unit (e.g. family, school, site)?
- Were the data collected in a manner comparable to any primary data collection the organization/team conducts?
- Are all the data elements that your study needs available (e.g. demographic information)?
- What is the unit? Is data maintained by cohort, family, site, or individual, and is this appropriate for your study?

Partnership with the Organization/People Collecting the Secondary/Administrative Data

Accessing restricted-use data is a negotiated process. It is important to find out who has control of the data and convince the agency that your organization and your team will secure, respect and protect the privacy of the individuals in the data. It is also important to find common ground with the agency that your study will address questions that are relevant to their mission and goals. Some of the key points your organization and your team should address during the negotiation:

- Your organization and team is a legitimate user of the data (i.e. a member of or affiliated with an organization that has previously collaborate with the agency for research purposes);
- The data will be used for legitimate research for a program or population their state/agency is interested in (or at least that is not going to cause problems for their state/agency);

- Your organization and team are experienced at handling secure data and are prepared and committed to keep data safe and protect the privacy of the individuals;
- Sharing data is in the interest of (or at least not counter to the interests of) their state/agency;
- Your organization is willing to partner with the agency (e.g. answer some of their questions in addition to your own) and assure the agency that your team will not use the data inappropriately;
- Influential stakeholders whose opinions matter to the agency support your organization/team in getting this data.

Of course, the agency still may decline, but at least these things set your team up for the best likelihood of success. If the agency is willing, it is a good idea to develop a letter of agreement, especially if your study will request multiple years of data. Your organization/team may be asked to complete a data protection plan.

Develop contacts with individuals at different levels within the agency so that staff turnover will not hinder your ability to access the data you need.

Infrastructure that Supports Security Requirements

Your organization/team is more likely to get access to restricted-use data if there is a basic infrastructure that addresses security concerns. One of the basic requirements is to use a standalone computer workstation; the data do not leave the secure workstation and are accessible only to authorized staff.

SIF Secondary/Administrative Data Use: Accessing Restricted-Use Data

Standalone Computer Workstation: Minimum Security Requirements

- ❑ Data are used at a standalone workstation that is not connected to any network or other computers.
- ❑ Laptop computers cannot be used.
- ❑ The standalone workstation is in a locked room.
- ❑ Use a locked cabinet to store the original data received. The cabinet must be in a locked room accessible only by personnel authorized to use the data.
- ❑ Use strong passwords (6-8 characters with one non-alphanumeric).
- ❑ Passwords are changed at least every 3 months. Change staff passwords accordingly when staff changes.
- ❑ Read-only access to the original data files.
- ❑ Lock the computer and/or workstation room when away from computer, and enable automatic "shutdown" after 3-5 minutes of inactivity.
- ❑ No routine backups of the data.
- ❑ Remove data by overwriting at the end of the project or prior to the computer needing repair.

Resources

Before you request data, it is helpful to understand the **privacy laws** that restrict data sharing. For example, health information is protected through the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (http://privacyruleandresearch.nih.gov/). Education data is protected by the Family Educational Rights and Privacy Act (FERPA) (http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html).

A **data management plan** outlines what you will do with your data during and after a research project. DMPTool https://dmptool.org/) is a website maintained by the University of California for developing data management plans. The site offers templates and guidance for developing a plan (https://dmptool.org/dm_guidance).

Workstations should have **data encryption software** installed for sensitive data. Examples of common software include Microsoft Bitlocker http://windows.microsoft.com/en-us/windows7/products/features/bitlocker and Symantec Encryption http://www.symantec.com/encryption).

At some point you will need to **prepare datasets for archiving** and possibly sharing. The Inter-university Consortium for Political and Social Research (ICPSR) has a resource, Guide to Social Science Data Preparation and Archiving. See Phase 5: Preparing Data for Sharing. http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/index.html.