

Office of Statistical Consulting
Division of Nutritional Sciences, Colleges of Human Ecology and Agriculture and Life Sciences
Cornell University

Guidelines for Data Acquisition

Occasionally researchers need to access a data set for secondary analysis that is not available at their institution. In cases when data are to be borrowed from another institution, it is necessary to protect the interests of both the lending and recipient institutions with a written agreement. Having such an agreement in place before the data are transferred can reduce the risk of misunderstandings and conflict later on.

The following information should be included in a letter of agreement between a data lending institution and a recipient institution. The information is intentionally very general, and researchers should modify it to their own situation. Researchers at Cornell are encouraged to consult with Karen Grace-Martin in the Office of Statistical Consulting in drafting their lending agreements. The agreement should include:

1. The title of the database requested.
2. The format of the database requested, including the physical medium on which the data are stored and the electronic database type. The borrower should contact the lending institution to get as much information as possible about their data lending policies and their technology limitations before a written document is sent. It is often a good idea to request a format that they can accommodate well, even if it is not the one you want. It can ultimately save you time and frustration because you are assured that the data are usable.

Physical media include (but are not limited to):

- Tapes
- Disks
- CD-ROM
- Electronic internet transfer

Electronic database types include:

- ASCII files
- DBase, Access or another database format
- SAS, SPSS, Minitab, or another statistical program format
- Excel, Lotus 1-2-3, or another spreadsheet format

Electronic platforms include:

- Unix
- Macintosh
- PC

3. A description of the statistical research or teaching project that necessitates using the requested database. This should include:
 - a description of the final research or teaching objective, or use of the data;
 - a description of the sector(s) of the community that will be served by the product.
4. The name and title of the Senior Official at the recipient institution having the authority to bind the recipient institution to the provisions of the agreement. The Senior Official will generally be an administrator at the institution. Both the Senior Official and the Principal Project Officer should sign the letter of agreement.

It is especially important to find out about the security procedures required by the lending institution. Some data lending institutions are very concerned about the security of their data. In certain situations, the letter of agreement should first be sent to the recipient institution's attorney's office for review. A signature by the Senior Official can hold that person criminally liable for the data. In these situations, a dean or vice-president should serve as the Senior Official.

5. The name and title of the Principal Project Officer who will oversee the daily operations. The PPO is the most senior officer in charge of the day-to-day operations involving the use of subject data and is responsible for liaison with the lending institution. At a university, this will most often be a faculty member.
6. The name and title of a contact person at the lending institution. A fax number and email address are very useful if the lending institution is located in a different time zone.
7. Restrictions to the data's usage, including:
 - the names and titles of the professional, technical, and support staff who will be accessing the data;
 - the purposes for which the data may be used beyond the original project; and
 - whether and to whom the data can be redistributed.
8. Security Procedures for protecting the integrity of the data, including
 - the physical location in which the data will be stored;
 - the personnel who have access to the storage location; and
 - processes designed to protect the security of the data while in use.
9. The terms of the loan, including:
 - the length of time of the loan;
 - whether data will be returned to the lending institution at the end of the loan term; and
 - any fees to be paid and the method of payment.
10. Any and all documentation that will be sent with the database. A database without documentation is usually useless. Documentation includes, but is not limited to a:
 - Codebook,
 - Data dictionary, and
 - Copy of surveys or experimental materials.
11. The way in which credit should be given to the lending institution in publications and public presentations stemming from use of the data. Possibilities include:
 - Authorship,
 - Citation, and
 - Acknowledgement.

Acknowledgement: Thanks to Ann Grey of the Cornell Institute for Social and Economic Research for her assistance in the preparation of this document.