Data Management and Grant Proposals

Most agencies that provide research funding request that researchers describe the process they will use to manage data throughout the proposed project. It is strongly recommended that data collected as part of a research project be deposited in an archive and be made publicly available in a reasonable amount of time.

What is Data Management?

Throughout a research project data is generated and needs to be organized, described, stored, and secured. At the end of a project the resulting data used in analysis can be deposited into an archive or shared with others; secure data needs to be stored in such a way that privacy and confidentiality are maintained. Funding agencies encourage or require researchers to develop a plan for handling data in grant proposals.

Guidelines for Data Management

NIH has developed the NIH Data Sharing Guidance and Implementation Policy (See also the section on What to Include in an NIH Application.)

NSF provides guidance and a list of FAQ's here.

ICPSR has provided links to Data Management Plans and examples.

The UK Data Archive has an excellent manual on data management. You can download the document here.

Data Management Budget

If you are applying for a grant, you should build data management tasks and personnel into your project and budget. Funding agencies allow for budget to include costs associated with data management and disposition.

If you would like help in budgeting for archiving of data, in making a data management plan, or if you have other questions, please contact the Archive.

The costs for preparing data for public use vary widely and depend on the local support and services of your department or other units at UCLA. It also depends on the complexity and format of your data and how the data were collected. For surveys and statistical files, you should at least plan to ask for funds to do the following tasks:

- Clean and process the original file (containing no identifiers of respondents); possible conversion to a statistical package or preparation of data definition statement files
- Prepare an electronic version of the questionnaire, including any details on the flow of the questions, randomization, branching, etc.
- Prepare a complete codebook containing complete coding for each variable.

Protecting Privacy and Confidentiality

At the heart of responsible use of data is respect for and protection of individuals and the confidential information they provide to researchers. Grant applicants generally need to describe data preparation procedures to ensure that data collected from respondents, including individuals, institutions or other potentially identifiable entities are followed.

As the principal investigator, it is your responsibility to ensure that respondent privacy and confidentiality is preserved. You must be able to demonstrate that no respondent can be directly or indirectly identified in the final public data you deposit in an archive or share with other investigators or use for instructional purposes.

Keep in mind that there are direct identifiers such as names, addresses, or numbers for telephone, social security, driver's license, etc. There are also indirect identifiers which, when used with other variables in the data set, may provide enough detail to identify the respondent. The handling of direct and indirect identifiers to ensure respondent privacy and confidentiality is maintained, is the responsibility of the principal investigator.

Examples of indirect identifiers:

- Geographic detail (i.e., zip codes, census tract, block number, etc.)
- Memberships (i.e., clubs, groups, organizations, etc.)
- Names of schools attended by the respondent or respondent's family
- Job titles, positions held in organizations, elected offices, etc.
- Personal information (i.e., income, events, certain medical procedures, etc.)

Data Protection Plans

Data Protection Plans are of importance in a couple of situations: Preparing grant proposals for original data collection or using an existing data file which has restricted access because it has been judged to contain identifiers of respondents.

In some research projects, the nature of the research will generate a data set containing variables which might be used to identify respondents. However, removal of the identifying characteristics would render the data set useless to further analysis. In this case, arrangements can be made to store data securely and enable reuse. Grant proposals and human subjects review applications will typically contain a data protection plan for the storage of the final research dataset in an archive or repository. Examples of what should be included in such plans are generally supplied by the funding agency. ICPSR also provides guidance on what should be a part of a plan for data protection.

Here are some specifics given by ICPSR related to Data Protection Plans:

1. List and describe all locations where copies of the data will be kept.
2. Describe the computing environment in which the data will be used:
    - Computing platform (PC, workstation, mainframe platform)
    - Number of computers on which data will be stored or analyzed

CC BY NC SA

- Whether personal computers used in the research project will be attached to a network or will operate independently (stand-alone)
- Physical environment in which computer is kept (e.g., in room with public access, in room locked when not in use by research staff)

3. List and describe how data will be stored: (e.g., on PC hard drive, on removable storage media such as CD, diskettes, or Zip(R) drive.)
4. Describe methods of data storage when data are not being used.
5. Describe methods of transmitting the data between research team members (if applicable).
6. Describe methods of storage of computer output (in electronic form as well as on paper).