

Research Data Management

What is Data Management?

The University of Pittsburgh Library (<https://pitt.libguides.com/managedata>) defines research data management as “the organization, storage, preservation, and sharing of data collected and used in a research project. It involves the everyday management of research data during the lifetime of a research project (for example, using consistent file naming conventions). It also involves decisions about how data will be preserved and shared after the project is completed (for example, depositing the data in a repository for long-term archiving and access).”

In short, data management ensures the quality, accuracy, and findability of data throughout the whole research cycle.

Data Management for Funded Research Projects

Many institutions that provide research grants require that funded projects include a data management plan. A data management plan is a written document that details the data researchers plan to collect in their research and how those researchers will manage, archive, and share the data.

The Inter-university Consortium for Political and Social Research (ICPSR) offers detailed resources on how to create a data management plan, including a video overview, a list of elements in a data management plan, a framework for creating a plan, and a resources and examples list. These are excellent resources for researchers with funded projects. Click on the following link to access resources from ICPSR:

<https://bit.ly/ICPSRB2B>

For more support in creating a data management plan, contact data@library.columbia.edu to set up a consultation.

Data Management for Class Assignments and other Non-Funded Projects

For student projects that are not funded, such as class assignments, it is still worthwhile to think about how you will manage your data, although it will probably not be necessary to write a formal data management plan. Deciding early how you will keep your data organized and well-labeled will save you time in the long run and better allow you to come back to and continue your work later.

Data Management: Questions to Consider

Answering these questions will help you to plan your data management strategy. Keep in mind that the best time to plan for data management is before you start the project, not during or after.

- If you are using data that someone else collected, have you read all documentation and metadata? Do you know what each variable means and how the data was collected?
- In what format will the data be generated and stored?
- How will you describe your data so that you and others understand each variable later?
- Are you naming your data clearly?
- Do you have clear naming conventions for documents?
- How will you track versions?
- How will you store and backup your data?
- With whom will the data you are using be shared, and will you publish the data?
- Do you have an informed consent process in place for any human subject data?
- Has any human subject data been anonymized?

For support in creating a data management system for your project, contact the librarians at Research Data Services as early as possible. The earlier you create a data management system, the smoother your data management will go. Email data@library.columbia.edu to get started.
