Data Management

The collection and organization of data are an integral and critical part of the research process. This module will discuss management of data, and data management plans and tools.

Learning Objectives:

1. Identify the importance of data management
2. Learn how to properly manage data
3. Familiarize you with data management services and tools offered by Purdue

Importance of data management

Once data are collected and analyzed, it is sometimes required that the data is stored for future use or retained to satisfy a policy requirement by an institution or by a funding agency or organization. There are instances during the data analysis phase that authorized individuals other than the Principal Investigator (PI) handles the data. Each project member, regardless of role, must adhere to project guidelines and to policies for conducting quality research. If there are multiple people involved in this process, management of data becomes an even more critical issue. Similar to maintaining data quality, data management is crucial before, during, and after the data collection phase. It is an essential aspect of the Responsible Conduct of Research.

It is important that any research team addresses all issues related to data management before starting a research project. This will ensure that data are stored and managed in an easily usable format. For this purpose, a data management plan should be created at the start of a research project. Many funding agencies now require researchers to submit their data plans with requests for funding.

Data management will help researchers in various aspects. These include:

- Meeting funding agency requirements
- Protecting federal investment on research and development
- Expediting the scientific process; saving time and resources in the long run
- Ensuring that research data and records are accurate, complete, authentic, and reliable
- Ensuring research integrity and replication
- Increasing research efficiency
- Enhancing data security while minimizing the risk of data loss
- Preventing duplication of effort by allowing others to use your data
- “Complying with practices conducted in industry and commerce” (Boston University Libraries, n.d.b; The University of Edinburgh, 2013a; Research Information Network, 2007).
Data Management Plan

As noted above, several federal funding agencies require that data management plans be submitted and reviewed as components of proposals for funding. This data management plan is a document that describes the process of creating, organizing, documenting, storing, and sharing data (The University of Edinburgh, 2013b). A data management plan may start with a planning checklist, which is comprised of questions relating to one’s research and research data. Although undergraduate researchers may not be responsible for the creation of this plan, they should be aware of aspects of the data management plan so that they can implement the plan. Below are a few suggested checklist questions as this data management plan is created:

- What data will be created during the research project?
- Which file format(s) will be used for the data?
- How will the data files be named and organized?
- How will changes in the data files be tracked?
- Where will the data be stored?
- Who should have access to the data?
- How long will the data be stored? (e.g., 2 years? 10 years?)
- How will the data be documented?
- Should the data be placed in a repository or an archive for sharing and long-term storage?
- What are the funding agency requirements for data management? (Boston University Libraries, n.d.c).

Data management plans should be maintained and updated throughout the course of the research project (The University of Edinburgh, 2013b). Based on the data management planning checklist, a data management plan will address many issues, such as:

- What type of research data will be created or collected
- What funding, institution, and legal policies apply to the data
- What data management practices will be used for backups, storage, access control, and archiving
- What facilities and equipment will be required to conduct the research (e.g., hard disk space, backup server, repository)
- Who will own and have access to the data
- Who will be responsible for each aspect of the plan
- What data will be preserved at the conclusion of the project
- How reuse of data will be enabled (The University of Edinburgh, 2013b).

Data Management Planning Tools

To assist in the creation of data management plans, Purdue Libraries provides researchers’ access to an online data management tool called DMP Tool or Data Management Planning Tool. Here, you are given the option to select your institution and log on with your
Purdue Credentials. Once you gain access, you can create your own data management plan using preselected fields or templates. The DMP Tool can be accessed at https://dmptool.org/. The Purdue Libraries also provides consultation and collaboration needed for data management and planning (Purdue Libraries, 2014).

A data management planning self-assessment questionnaire is also available at: https://purr.purdue.edu/site/media/docs/dmp_self_assessment.pdf. This self-assessment is provided to assist researchers as they create a checklist and identify the key aspects in data management planning for their projects. Purdue Libraries also have several data management resources and guides that can aid researchers in their research and data planning. These and other relevant resources may be accessed online at the Purdue Libraries.

**Data Management Extras**

Other aspects of data that sometimes might not be considered while planning data management include:

1. **Documenting Data**: Make sure that the data are well organized and documented. This helps long-term and makes it easier for researchers to revisit the data and understand the dataset in preparation for future publications, research outcome reporting, and funding requests.

2. **Data storage and backup**: Often, individuals forget to backup their data once it is stored at a location (e.g., network drives, personal computers and laptops, or external disks). In cases where multiple people have access to the stored data, there is a high risk of accidental deletion or corruption of the data. Researchers need to backup the original data in various, password-protected locations in the event of network failures or accidental data losses. Depending on the sensitivity of the data, cloud resources such as Dropbox, Amazon Cloud services, or Apple iCloud may be excellent alternatives to backing up data (The University of Edinburgh, 2013c). Confirm with your Principal Investigator/research advisor and sponsor that it is acceptable to store your data in such places.

3. **Data security**: Data security is another important aspect that needs to be considered in data management. Sensitive data should be secured digitally, using security methods such as encryption protocols (The University of Edinburgh, 2013c) and only in locations for which access is known and controlled (i.e. not in cloud resources). The network on which the data is stored should also be secured and monitored for potential breaches. Access to data should be controlled, given only to select individuals, and monitored. Data collected in paper formats such as a paper-based survey should be stored and locked at a secure location.

Although different disciplines follow a variety of practices for data collection and storage, it is important to comprehend the seriousness of the data management process. If you have any
questions about this process, talk to your research advisor about ways to manage data in your group.

References:


