

Data Repositories

The Pepperdine community is encouraged to use [Pepperdine Digital Commons](#), a digital repository and publication platform designed to collect, preserve, and make accessible the academic output (such as publications, presentations, white papers, data sets, etc.) of Pepperdine faculty, students, staff, and affiliates. It is the mission of Pepperdine Digital Commons to facilitate the discovery of Pepperdine's scholarly communications, provide instant access to full text works, and preserve these materials in an open, digital environment.

To begin using this free service, offered through the Pepperdine University Libraries, please contact Katie Richardson, Archivist for Special Collections and University Archives, at katie.richardson@pepperdine.edu.

Licensing

When applying a license to your own data, you are encouraged to make it as open as appropriate to enable others to use and build on your data. It's also important to note that data is not copyrightable, but it can be licensed under [Creative Commons](#).

Funding Agency Requirements

It's important to identify what the grant funding institution requires when it comes to sharing data. Some require a data management plan like the NSF while others just encourage data sharing like the IMLS.

Pepperdine University Libraries

Katie Richardson
Archivist for Special Collections and
University Archives
Phone: 310-506-4323
Email:
katie.richardson@pepperdine.edu

<http://infoguides.pepperdine.edu/datacuration>

Managing Data Sets

The Pepperdine University Libraries are here to support your efforts in data management by providing the resources to help preserve, document, manage, and publish your data as you travel through the research life-cycle.

What is Data?

In this context, data is described as units of information observed, collected, or created during the course of research. Data typically falls into one of four categories; 1. Observational; 2. Experimental; 3. Simulation; and 4. Derived or Compiled.

Lifecycle of Data Management



Why Do I Need a Data Management Plan?

Planning for data management early on will make curation activities much easier throughout the data lifecycle.

- To comply with the federal agencies proposal requirements.
- To describe how data will be maintained and what resources will be needed to preserve it.
- To have a well described and organized data when posting a supplemental data set with your publication.
- To prepare data for review before and/or after the article is accepted for publication, as required by some publishers.
- To facilitate the re-use of data sets, open access, and data sharing.

What is a Data Management Plan?

Recommended Components:

- A description of the project. What is the purpose of the research? What organizations and personnel will be involved in the project?
- A description of the data that will be collected. What is the nature and format of the data and how will it be collected?
- The standards that will be used for saving file formats and assigning metadata.
- Plans for storage and data management. What file formats, backup procedures, and security will be in place?
- If there are any legal or ethical issues regarding the data collected. For example, if study participants require confidentiality, how will this be protected?
- Defining access policies and provisions for the data. Will any restrictions be necessary?
- Identifying a data set repository to store your data to ensure discovery and preservation.
- Assigning data management responsibilities. Who will be responsible for data management and how will compliance with this plan be monitored and ensured over time?

Citations and Recognition

It's Important to cite data in order to:

- Give the data producer appropriate credit.
- Allow easier access to the data for re-purposing or re-use.
- Enable readers to verify your results.

Typically, the following elements should be included when citing a data set:

1. Author(s)
2. Title
3. Year of Publication
4. Publisher (the data center/repository)
5. Any application identifier (including edition or version)

6. Availability and access (URL or doi/doi link)

An example of how to appropriately cite data would be: Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. Geological Institute, University of Tokyo. <http://dx.doi.org/10.1594/PANGAEA.726855>

Data Preservation

The **file format** in which you record, store, and transmit your data is a primary factor in one's ability to use your data in the future. Plan for both hardware and software obsolescence.

Formats likely to be accessible in the future are:

1. Non-proprietary (ex. use .txt rather than MS Word .doc)
2. Open, document standards
3. In common usage by the research community
4. Use standard character encodings (i.e. ASCII or UTF-8)
5. Unencrypted
6. Uncompressed

It's important to make sure your data is **backed up and secure**.

- Make 3 copies of your data (e.g. original + external/local + external/remote) and have them geographically distributed. Backup options include local storage (harddrive, CD/DVD, flash drive, departmental server) or cloud based storage (Pepperdine digital commons, Amazon S3, or Carbonite).
- Secure your data i.e. unencrypted and uncompressed
- Test your back up system periodically.

Ethical Issues

All sensitive information in your data should be redacted before depositing in a public archive or repository. Access to the data may need to be restricted for a certain number of years in order to ensure privacy.