**Efficient Data Management Practices: Tools & Services at UC-Berkeley**

**Anna Sackmann: Kresge Engineering Library**

**Jamie Wittenberg: Research Data Management**

**23 August 2016**

**Data Management and Data Sharing:**

Good Data Management is rooted in data sharing. Public funders require that research products are available to the public after a certain amount of time. They want to ensure maximum return on investment, & quality, sustainability, and openness. In addition to funders, many journals now have data sharing requirements as a condition of publishing. This is in part for transparency and ultimately it safeguards you. Finally, and perhaps most importantly, as scientists, you have the ethical responsibility to foster research transparency and integrity.

“Good data management is the compilation of a small number of routine practices that add up to good habits.” –Kristin Briney (UW-Milwaukee)

**Documentation**

The question you need to be able to answer:

What would someone unfamiliar with your data need in order to find, evaluate, understand, and reuse them?

1. Why do we document data?

* Data without notes are unusable
* You will not remember everything
* Others may need to use your data

2. How do we document data?

There are many best practices for documenting data, including keeping an excellent lab notebook. There are two additional critical tools:

* Data Dictionary: a data dictionary allows you to define and elaborate on your variables and other information that you have in spreadsheets without adding more text to rows and columns. For example:
  + What are your variables?
  + How were they measured?
  + What do they mean?

See the adjacent example.

* Readme.txt file: Your research notes and your data often live in two separate locations which makes data hard to interpret because there is no nearby documentation. A readme.txt is a simple, nonstandard way to remedy this issue. The readme.txt lives in the same zipped file as your data and describes the general content and structure of the files. A readme.txt usually contains the following information:
  + Project name
  + Project summary
  + Funding information
  + Primary contact information
  + Location of the data and other supporting materials
  + Organization and naming conventions used for the data

See the adjacent example.

3. Data Storage & Security

Follow the 3-2-1 backup rule: you should have 3 copies of your data on at least two different types of media, with at least 1 offsite copy.

Personal computer

Network servers GOOD OPTIONS

External hard drives

CDSs OK, but cumbersome

DVDs

Thumb drive/USB drive Use for data transfer ONLY

Enterprise Cloud Storage at UC-Berkeley:

UC-Berkeley has enterprise versions of Box and Drive through bConnected. Click on the respective logos to learn more about each option.

[](http://researchdata.berkeley.edu/content/berkeley-box)[](https://bconnected.berkeley.edu/)

**Preserving Data**

You need to transition your data from active data storage to preservation when you’re wrapping up with your project. Part of this will require you to review your funder requirements and the original data management plan that was written.

**Type 🡪 Format examples**

Text 🡪 .txt; .rtf; .xml; .pdf

Tabular 🡪 .csv; .tsv

Image 🡪 .tif; .jpeg

Video 🡪 .mp4; .avi;, .mov

For more information on finding the right type of repository for your data, visit the [Registry of Research Data Repositories](http://www.re3data.org/).

UC-Berkeley has its own data repository called DASH, which is designed for researchers at the 10 University of California Campuses. Click on the DASH logo below for more information.

[](https://dash.cdlib.org/)