Smart and Simple Data Management

Sarah Williams (scwillms@illinois.edu), Interim Head, Funk ACES Library; Life Sciences Data Services Librarian
Heidi Imker (imker@illinois.edu), Director, Research Data Service
Elizabeth Wickes (wickes1@illinois.edu), Data Curation Specialist, Research Data Service

About this workshop

This session aims to provide you with data management best practices and tools to increase your research efficiency and impact. We'll present a basic introduction to data management using a data management plan framework, hands on activities, and discuss how to find and vet resources for making data publicly accessible.

Upcoming workshops

Creating Data Documentation (Data Management Workshop 2)
March 30, 1-2 pm, Grainger Library 335
Registration link: https://illinois.edu/calendar/detail/4068?eventId=33254151

Writing project, code, and data documentation doesn't need to be the worst part of your day. This hands on workshop will give you experience using various types of documentation, discuss strategies for writing documentation, and get you started writing a template for your projects. Bring a dataset you'd like to work with but examples will be provided.

Data Sharing (Data Management Workshop 3)
February 21, 10-11am, Main Library 314
Registration link: https://illinois.edu/calendar/detail/4068?eventId=33254160
April 6, 1-2pm, Grainger Library 335
Registration link: https://illinois.edu/calendar/detail/4068?eventId=33254161

Making research data public is becoming a reality for many disciplines, but for many researchers and disciplines there is a complicated set of issues to consider before publication or sharing data. This workshop will cover the basics steps of research data publication, from considerations to depositing. Participants will work through guidance to help them make decisions about when and how to publish or share data.

Data Workflows
March 1, 1-2pm, Grainger Library 335
Registration link: http://illinois.edu/calendar/detail/4068?eventId=33254157

Workflow mapping is a useful tool for teams of all sizes to understand how data, code, and other resources are being shared and passed around. Like retracing your steps after losing something, tracing a project through workflows identifies all the essential products and dependencies of the research process, and can be one of the most useful places to get started with data management. This workshop focuses specifically on computational research processes, but can be adapted to most types of projects.

Activity 1: Workflow Mapping

Like retracting your steps, mapping out your project's workflow can be the best place to start with getting your data organized.

1. Think of your normal analysis workflow or use a specific project you work on.
2. Set out five post-it notes in a row.
3. Try to encapsulate your project into five high level steps. Use one post-it note per step. Add or subtract post-it notes as needed, but try to keep it minimal.
4. Once you've written out each step, think of any data, scripts, or other files that you touch, both items you use and items you make. Write those down on the bottom of the post it.
5. Write down the location of those files and if/where they are backed up.

Draw a diagram of your workflow below when you are done or place the post-it notes onto the page in order.

Take home activity: Data Inventory – Part 1

Step 1	Step 2		Step 3		
Project or data name	Type of project?	Type of data?	Where is it?	How do you access it?	Is it backed up, and how?

Take home activity: Data Inventory – Part 2

Sketch out a new structure for organizing your files and folders. As you complete it, keep a few questions in mind:

1. Is this possible to implement for completed or current projects? Just focus on the future if you need to.
2. What kind of data pipeline or workflow would this structure require? How would your current workflow need to change?
3. Is this structure and level of detail something that can be maintained over time or during busy periods?

Other important considerations:

- Strike a **balance between consistency and uniqueness** in your file/folder names. Include important metadata, such as author, date, location, version number, etc. in the file name when possible.
- Use **YYYY-MM-DD format** for your dates in all places, and enforce this in your labs, project groups, etc.
- **Discuss and document standardized or controlled vocabularies** for acronyms, project names, and especially any other qualitative coding project.
- **Create project documentation**, even if you only have time to make it brief. Develop templates for yourself or members of your lab to fill out for every project. Follow domain-specific metadata schemas when appropriate.
- **Keep your raw data protected and unchanged**. At a minimum, keep an extra copy of the original raw data somewhere in case of accidental changes or deletions. Keep separate copies of every data transformation made during your processing stages.
- **Backup your data and projects!** Follow the 3-2-1 rule for backups, meaning that you should have three copies: using two media types and one remote copy. Utilize an external hard drive and your Box account (or other applicable cloud data storage) to make regular backups of your projects. Develop backup access plans for any proprietary or legacy software required for your research.
- **Use version control for all code.** GitHub Desktop is a lightweight application that can track changes to files within a folder. Projects can also be committed to GitHub repositories (private options are available) for additional backup. Several analytics platforms, including R Studio, have git and GitHub integration built in.

Need more information about any of the above?

The Research Data Service (http://researchdataservice.illinois.edu/) team is happy to provide more information about all data management topics and offers personal data consultations. They can be reached at researchdata@library.illinois.edu or on Twitter @ILresearchdata. All Research Data Service consultations are free and confidential to all members of the Urbana-Champaign campus community.

Would you like a custom workshop for your research team or class?

This workshop and any others can be run by request. Our full menu of workshops and links to materials can be found at http://researchdataservice.illinois.edu/workshops/ and can be remixed into a variety of domain flavors.