

Where's My Data?!

File Organization for Research



Instructors



Julie Goldman

Research Data Services Librarian
Countway Library of Medicine
Julie_Goldman@hms.harvard.edu



Meghan Kerr

Archivist and Records Manager
Center for the History of Medicine
Meghan_Kerr@hms.harvard.edu



Slides: <https://datamanagement.hms.harvard.edu/class-materials>



HARVARD
MEDICAL SCHOOL

Data Management
Working Group



Countway Library of Medicine

An Alliance of the Harvard Medical School and Boston Medical Library



Center *for the* History of Medicine

**Harvard Chan Bioinformatics
Core**



hms | hsdm

office for postdoctoral fellows



HARVARD
MEDICAL SCHOOL

OFFICE FOR
Academic and
Research Integrity



Department of
Systems Biology



HARVARD
MEDICAL SCHOOL

Research Information Technology Solutions - RITS

HMS Information Technology

ICCB-Longwood Screening Facility

DRSC/TRiP Functional Genomics

The Neurobiology Imaging Facility

in the Neurobiology Department of Harvard Medical School

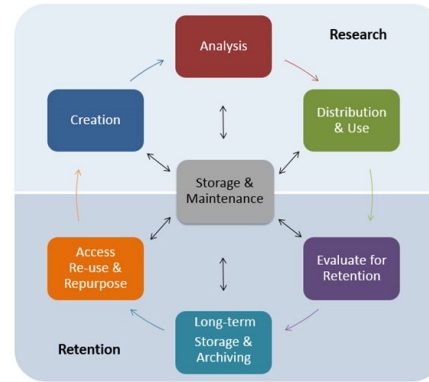
Hi+|S

Harvard Program in Therapeutic Science

Data Management

Data Management is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project. We recognize that best practices for each of these aspects of data management can and often do change over time, and are different for different stages in the data lifecycle.

Early and attentive management at each step of the data lifecycle will ensure the discoverability and longevity of your research.



FEATURED ONLINE TRAINING:



An open online course aimed at a broad audience on recommended practices for managing research data. Take at your own pace, earn badges and interact with students from around the world!

FEATURED ONLINE TRAINING:



An online supplement to an in-person workshop, specifically tailored for Post-Docs. If you are affiliated with Harvard, you may receive a course certificate to promote your time taken on this topic.

← January 2019 →

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

FEATURED NEWS



DMWG Featured in Nature Article: How to pick an electronic laboratory notebook
Thursday, August 9, 2018

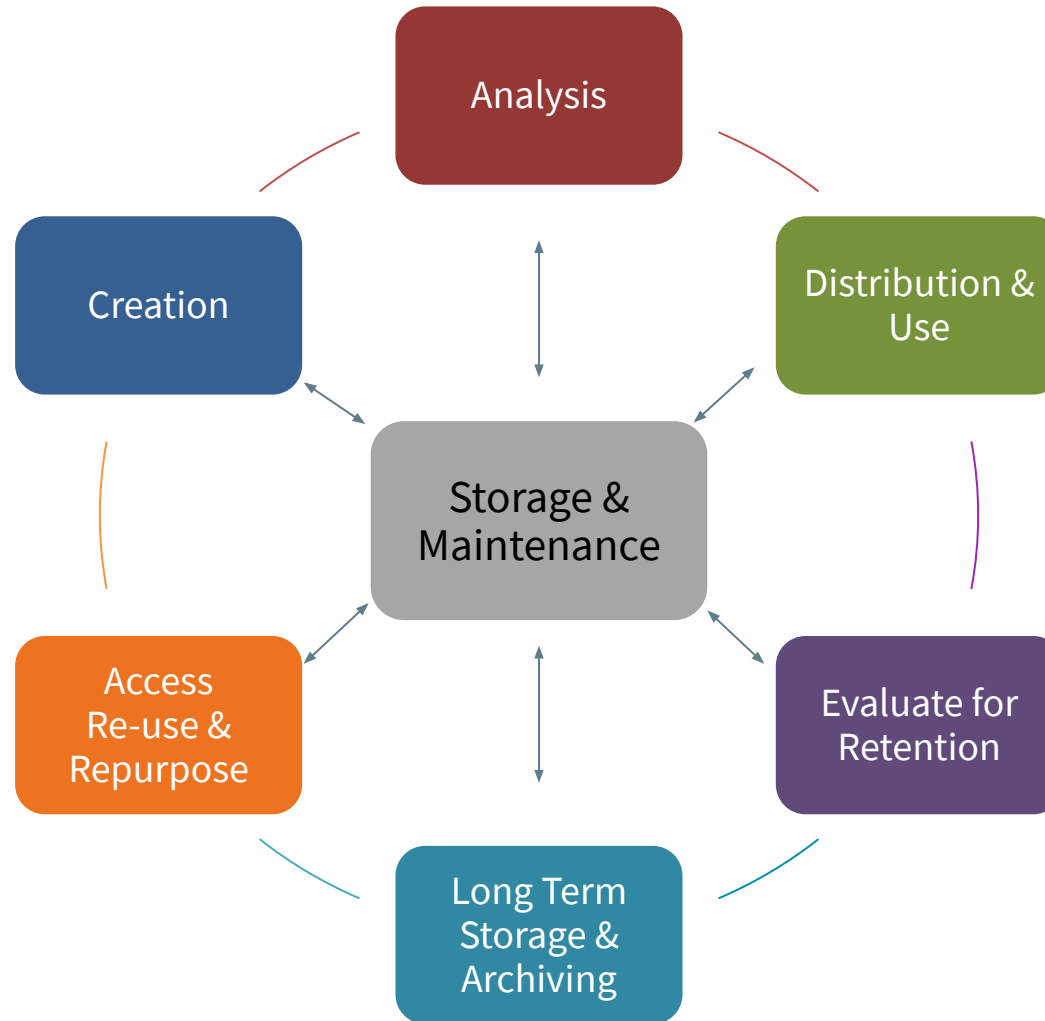
[Submit Questions and Feedback](#)

[Upcoming Trainings & News](#)

[Receive Data Management Updates](#)



Data Lifecycle for Biomedical Data



Why Manage Data?

- Easier to analyze organized, documented data
- Find data more easily
- Don't drown in irrelevant data
- Don't lose data
- Get credit for your data
- Avoid accusations of misconduct



Data Sharing and Management Snafu in 3 Short Acts

Speakers



Stephanie Mohr

Director of DRSC/TRiP-FGR
Department of Genetics



Michelle Ocana

Senior Imaging Specialist
Neurobiology Department

File Conventions



Versioning

- For analyzed data use version numbers
- Save files often to a new version
- Label the final version FINAL
- Consider GIT or SVN



Organization

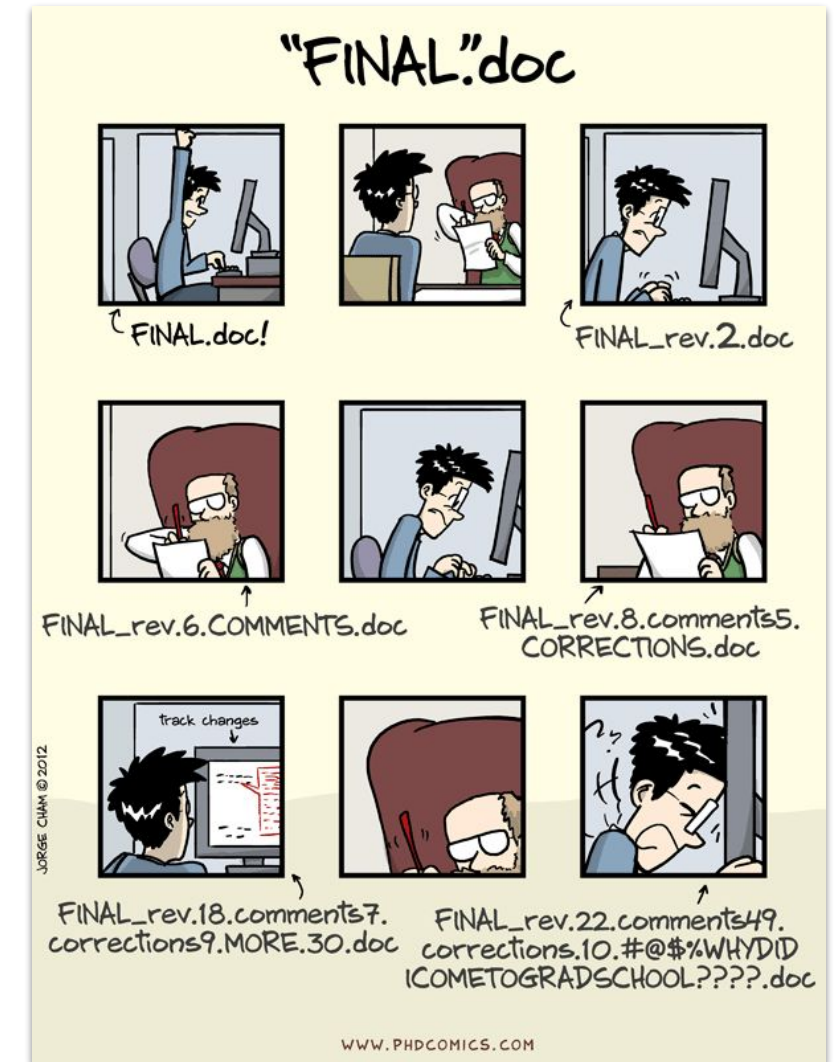
- Any system is better than none
- One project, one folder
- Separate folders for data or project stages
- Date-based folders (pairs well with lab notebook)

File Conventions

Files with naming conventions:

20161104_ProjectA_Ex1Test1_SmithE_v1.xlsx

20180204-ProjectA-Report-SmithE-v5-FINAL.docx



<http://phdcomics.com/comics/archive.php?comicid=1531>

File Organization: Scenario 1

You are a new postdoc in a large, well-established lab. A previous postdoc left a project 80% complete. Your new PI asks that you pick up the project and take it to completion. There's a lot riding on it for you—publishing with the new lab would help your postdoc grant applications. To get started, you need to find a set of image files for a particular cell line stained with a particular antibody. Your PI points you to the computer that the previous postdoc left behind and his paper notebook, then goes on a vacation in a remote location without internet access. She expects you to have found the files and begun the project by the time she gets back.

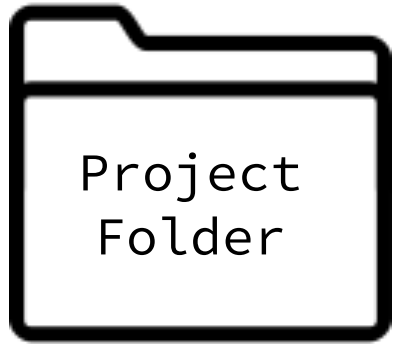
Starting point: You have the email address of the previous postdoc.

File Organization: Scenario 2

You are a mid-career PI. You published a paper that includes findings that contradict the work of another lab. You are confident in your results—your graduate student did many replicates and another trainee repeated the tests independently. The other lab contacted the journal that published your paper and your institution to challenge your results. You need to provide your university's office of research integrity and the journal the raw data, as well as demonstrate that the results can be reproduced using your analysis methods. You are desperate to find the relevant research files and analysis software code. You know that if you can find them quickly, you can avoid a retraction that would jeopardize your funding, your career, and your grad student's PhD thesis.

Starting point: You have a shared folder on a university-supported file system.

Ideal Directory Structure Example



201902_Fly005_CRISPR_SMohr



RawData



DraftFigs



AnalyzedData



FinalFigs



DraftManuscript



Final

Questions?

Harvard Biomedical Data Management

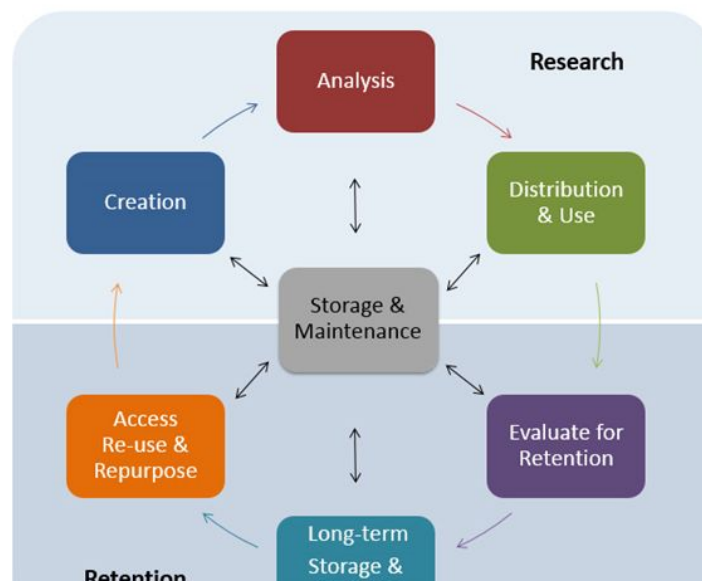
Best practices & support services for research data lifecycles

[About](#) ▾ [Best Practices](#) ▾ [Plan](#) ▾ [Store](#) ▾ [Share](#) ▾ [Resources](#) [Support](#)

Data Management

Data Management is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project. We recognize that best practices for each of these aspects of data management can and often do change over time, and are different for different stages in the data lifecycle.

Early and attentive management at each step of the data lifecycle will ensure the discoverability and longevity of your research.



← January 2019 →

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

FEATURED NEWS



DMWG Featured in Nature Article: How to pick an electronic laboratory notebook
Thursday, August 9, 2018

[Submit Questions and Feedback](#)

[Upcoming Trainings & News](#)

[Receive Data Management Updates](#)

Upcoming Workshops / Seminars

Research Management Using the Open Science Framework

Thursday, February 14

2:00 - 4:00 pm

HMS TMEC 104 Classroom

Register: bit.ly/RDM-Winter19

Getting Started with Data Management Plans

Wednesday, March 20

1:00 - 2:00 pm

Countway Library Ballard Room 503

Register: bit.ly/RDM-Winter19

bit.ly/rdm-survey