

Data Management

Introduction to the Command-line Interface

Instructors



Julie Goldman

Research Data Services Librarian



Meghan Kerr

Archivist and Records Manager



Radhika Khetani

Training Director, Research Scientist



Meeta Mistry

Research Scientist



Kathleen Keating

Research Computing Consultant



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

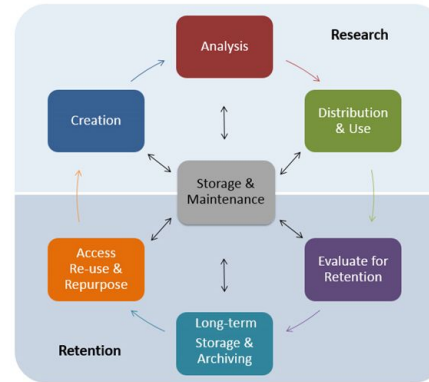
Hi+|S

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.



[Submit Questions and Feedback](#)

[Upcoming Trainings & News](#)

[Receive Data Management Updates](#)

UPCOMING EVENTS

2019
MAY
29
Responsible Conduct of
Research: Collaborative
Science

2019
JUN
03
Data Skills: Planning for
Research Success

2019
JUN
13
Data Management Working
Group Monthly Meeting

[More ▸](#)

FEATURED NEWS



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



Why Manage Data?

- Running the same workflow can be labour intensive
- Manual manipulation of data files:
 - is often not captured in documentation
 - is hard to reproduce
 - is hard to troubleshoot, review, or improve
- Hard to find poorly organized, documented data
- Hard to analyze poorly recorded workflows

Why the Command-Line?

- The shell is powerful
 - Built-in commands for easy data manipulation -- count and mine research data, even when data is spread across multiple files
 - Workflows can be automated which makes you more productive and also improves the reproducibility of your work by allowing you to save and then repeat them
- The shell can be used to copy, move, and combine files
 - Every step can be captured in the shell script and allow reproducibility and easy troubleshooting

Training Materials

<http://tinyurl.com/rdm-shell>

Workshop Outline

Lessons	Estimated Duration
<u>Introduction to the command line</u>	75 min
<u>Creating files and Searching within them</u>	15 min

Tying it Together

Why Data Management:

Not a prerequisite of working with the shell, but data should be organized in a clear and predictable manner.

Taking the time to structure your research data and filenaming conventions in a consistent and predictable manner is certainly a significant step towards getting the most out of Unix commands.

Why Command-Line:

“Because so much depends on it.”
Installing software, configuring your default editor, and controlling remote machines frequently assume a basic familiarity with the shell.

Allows users to automate repetitive tasks and capture small data manipulation steps that are normally not recorded to make research reproducible.

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.

Submit Questions and Feedback

Upcoming Trainings & News

Receive Data Management Updates

UPCOMING EVENTS

2019
MAY
29

Responsible Conduct of Research: Collaborative Science

2019
JUN
03

Data Skills: Planning for Research Success

2019
JUN
13

Data Management Working Group Monthly Meeting

[More ▶](#)

FEATURED NEWS

DMWG Featured in Nature Article: How to pick an electronic laboratory notebook

Thursday, August 9, 2018

<https://datamanagement.hms.harvard.edu>

Upcoming Workshops / Seminars

Introduction to High Performance Computing

Thursday, July 11

1:30 - 3:00 pm

Countway Library Minot Room

bit.ly/RDM-Summer19

Creating Meaningful Data: Metadata Essentials

Thursday, August 8

12:30 - 1:30 pm

Countway Library 403 Classroom

bit.ly/RDM-Summer19

bit.ly/rdm-survey