

Data Skills

Planning for Research Success

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 & Worksheet: <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

Harvard Biomedical Data Management

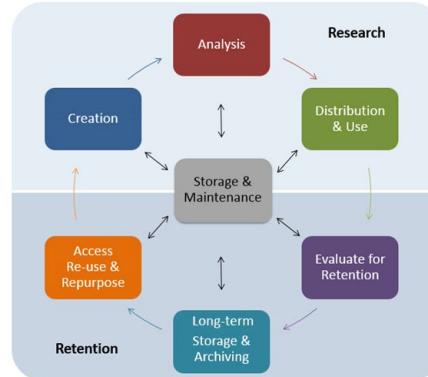
Best practices & support services for research data lifecycles

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 JULY 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

FEATURED ONLINE TRAINING:



[Best Practices for Biomedical Research Data Management](#)

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:



[Research Rigor & Reproducibility: Understanding the Data Lifecycle for Research Success](#)

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.



Licensed under the Creative Commons Attribution Non-Commercial License unless otherwise noted.

Last Updated: 2019-03-11

Harvard Biomedical Data Management Website

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

Introduce Yourself!



Name

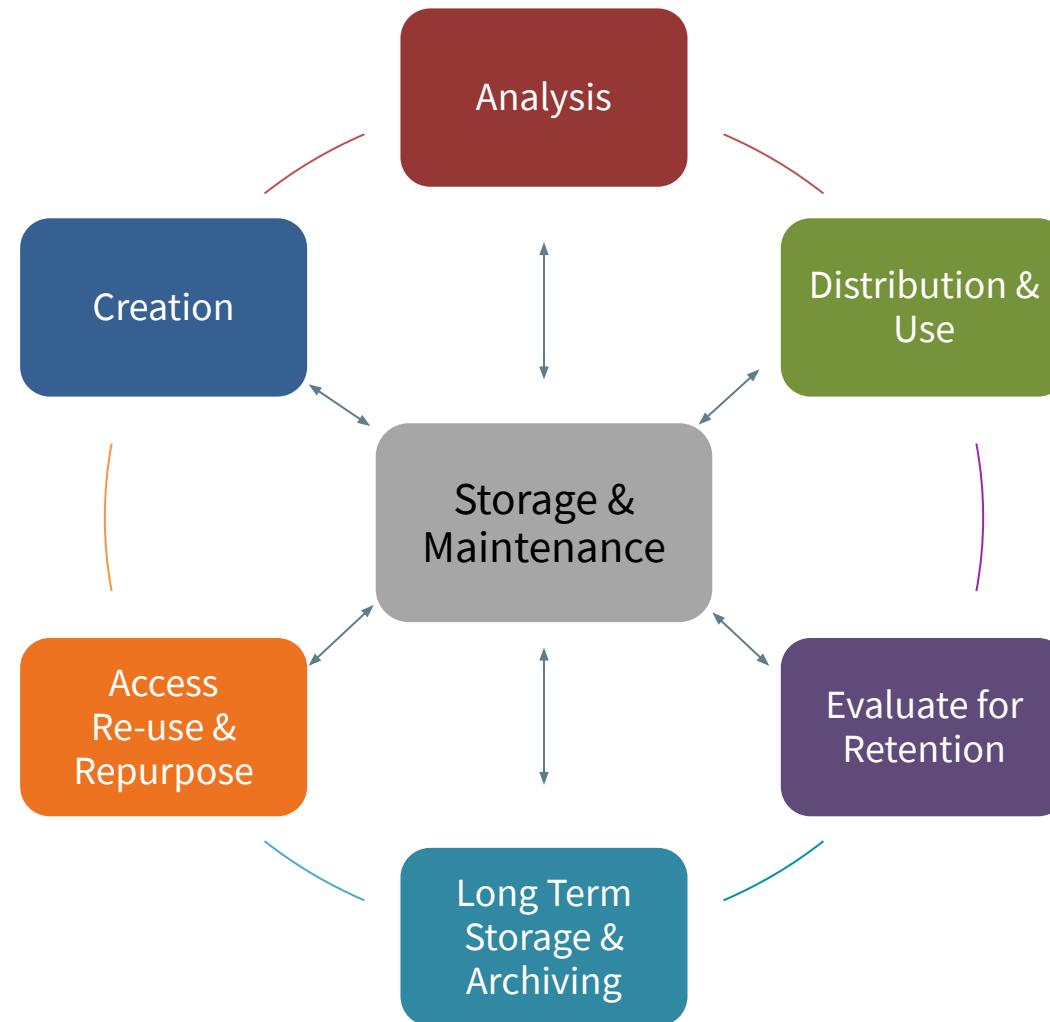
Affiliation / School / Department

Most common data format you work with?

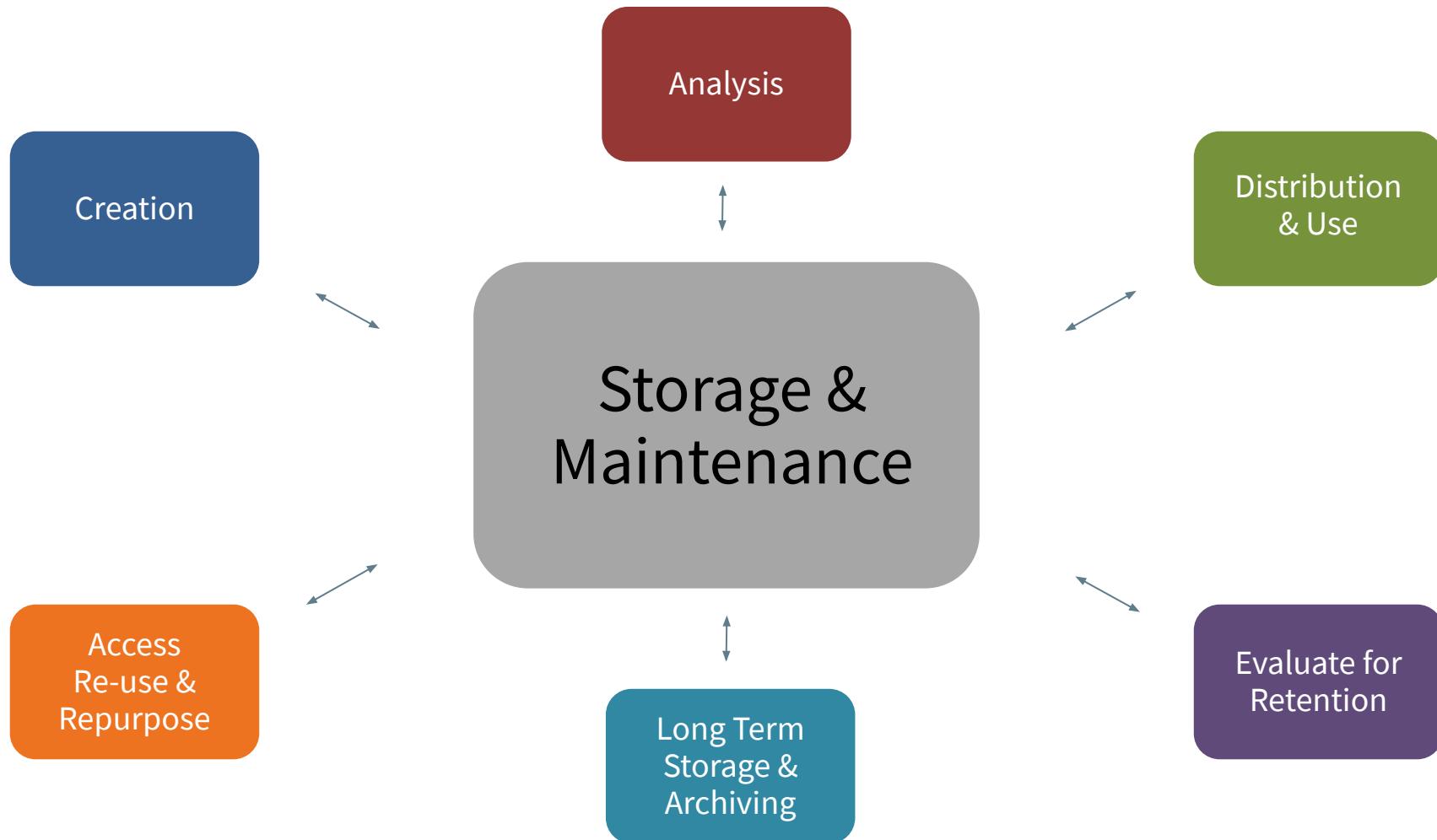
(Text, Excel, SPSS, Google Docs, etc.)

What you hope to get out of this class?

Data Lifecycle for Biomedical Data

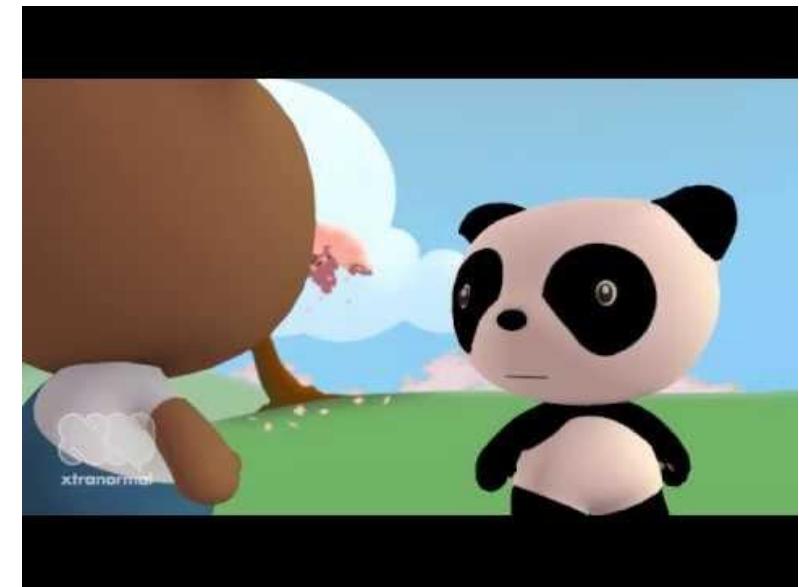


Storage affects the whole cycle



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

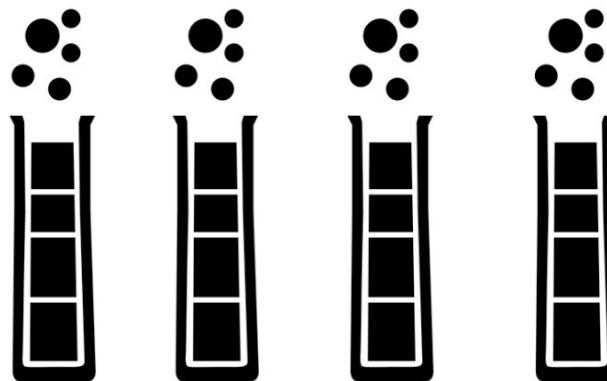


Data Sharing and Management Snafu in 3 Short Acts

Goal = Reproducibility

Considerations for reusability

- File formats
- Versioning
- Metadata
- Documentation
- Copyrights
- Ethical results

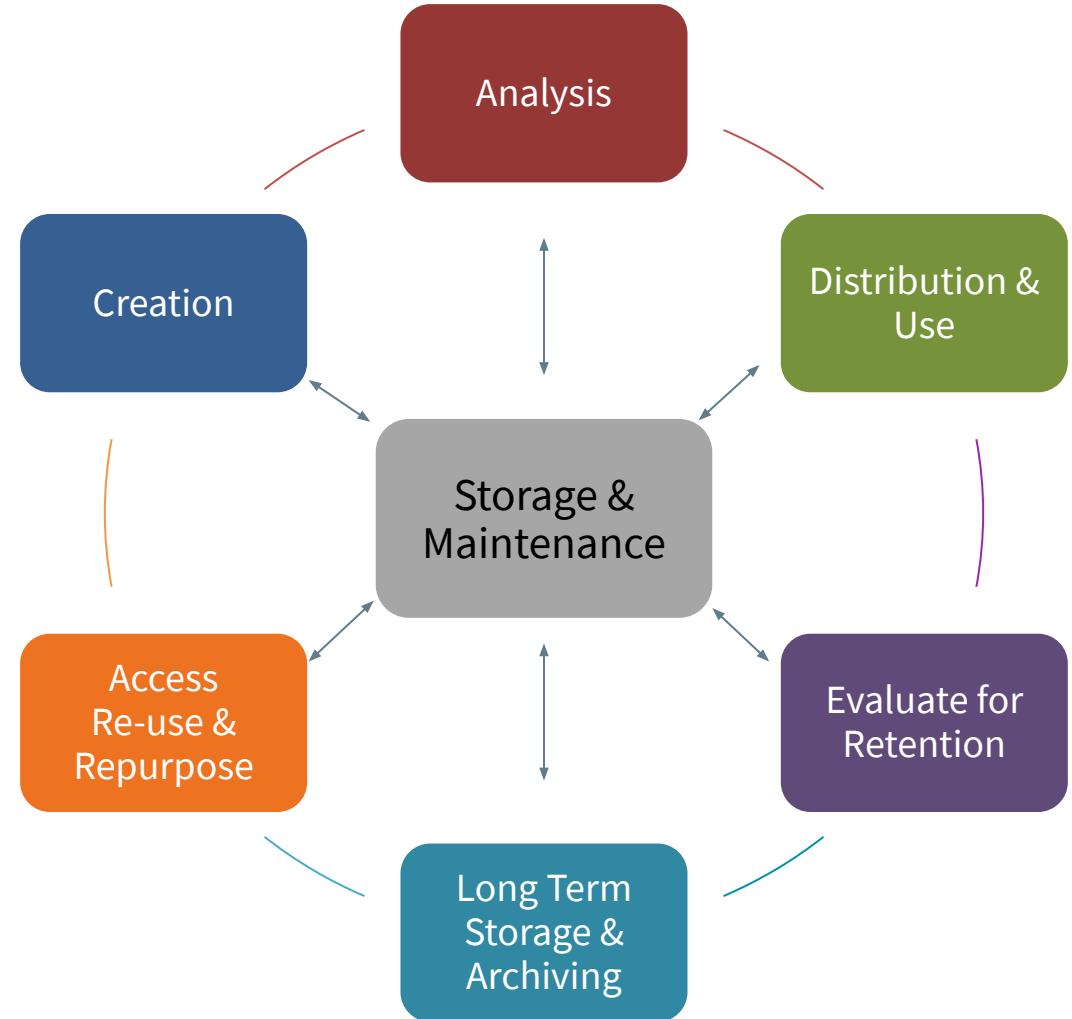


Science!



Not Science.

Let's go through the research data lifecycle



Data Stages

Raw data: What is being measured or observed?

Example: raw .txt file produced by an instrument

Processed data: How can the raw data be manipulated?

Example: data with Z-scores calculated

Analyzed data: What does the data tell us?

Example: rendered computational analysis

Finalized/Published data: How does the data support

your research question?

Example: polished figures appear in Cell

Creation

✓ Raw data

✓ Working files

Analysis

✓ Analytical
methods

✓ Analysis results

Metadata

Data documentation provides the information necessary to fully understand and interpret the data

Metadata should be standardized, consistent and interoperable, and facilitates discovery, preservation and archiving of data



Andy Warhol, *Big Torn Campbell's Soup Can (Pepper Pot)*, 1962 The Andy Warhol Museum, Pittsburgh Founding Collection, Contribution The Andy Warhol Foundation for the Visual Arts, Inc.

Your Turn!

On-Your-Own Exercise

Documentation



- What data products has your latest project produced?
- What do you anticipate generating?
- What are the different formats you use?
- What programs or code is needed to read or understand these files?
- Make a list of the most important information to record for each dataset.

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

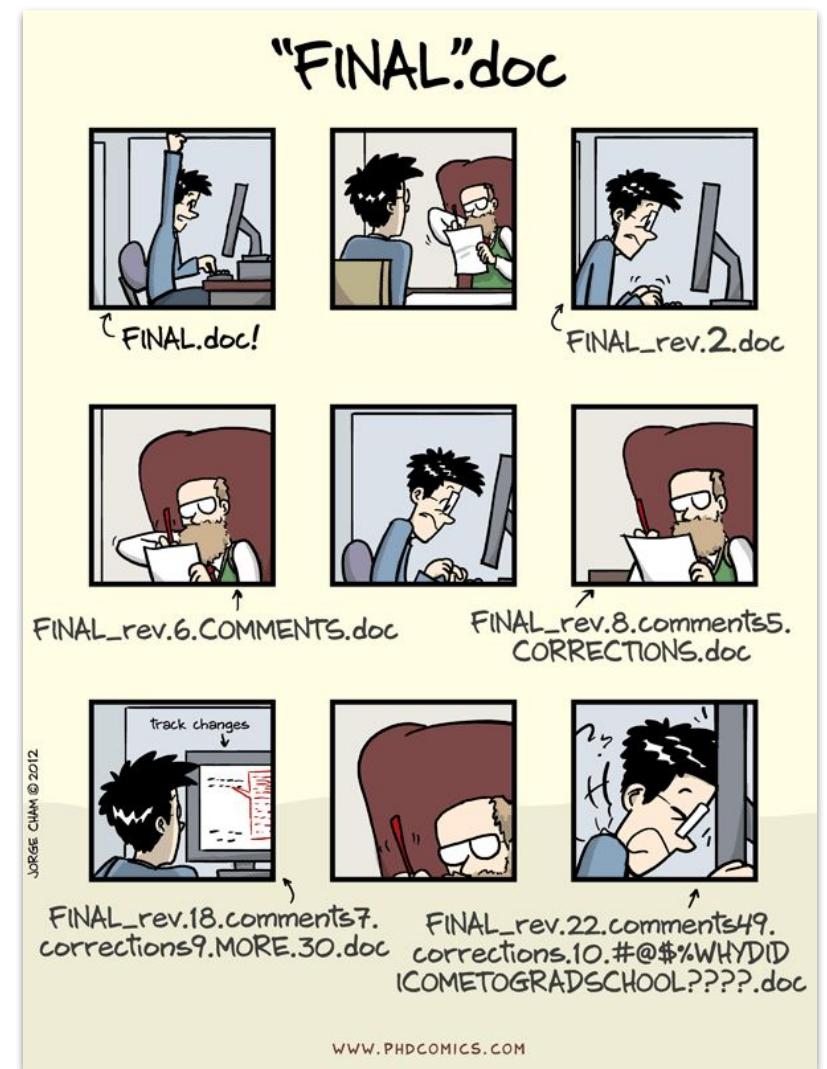
- Any system is better than none
- One project, one folder
- Separate folders for data or project stages
- Date-based folders (pairs well with lab notebook)

File Conventions

Files with naming conventions:

20161104_ProjectA_Ex1Test1_SmithE_v1.xlsx

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



Document! Document! Document!

No point to have a system without documentation!



Your Organizational System

- What are your file naming conventions?
- What is your folder hierarchy?



Your Workflow

- How did you get from raw data to the final product?

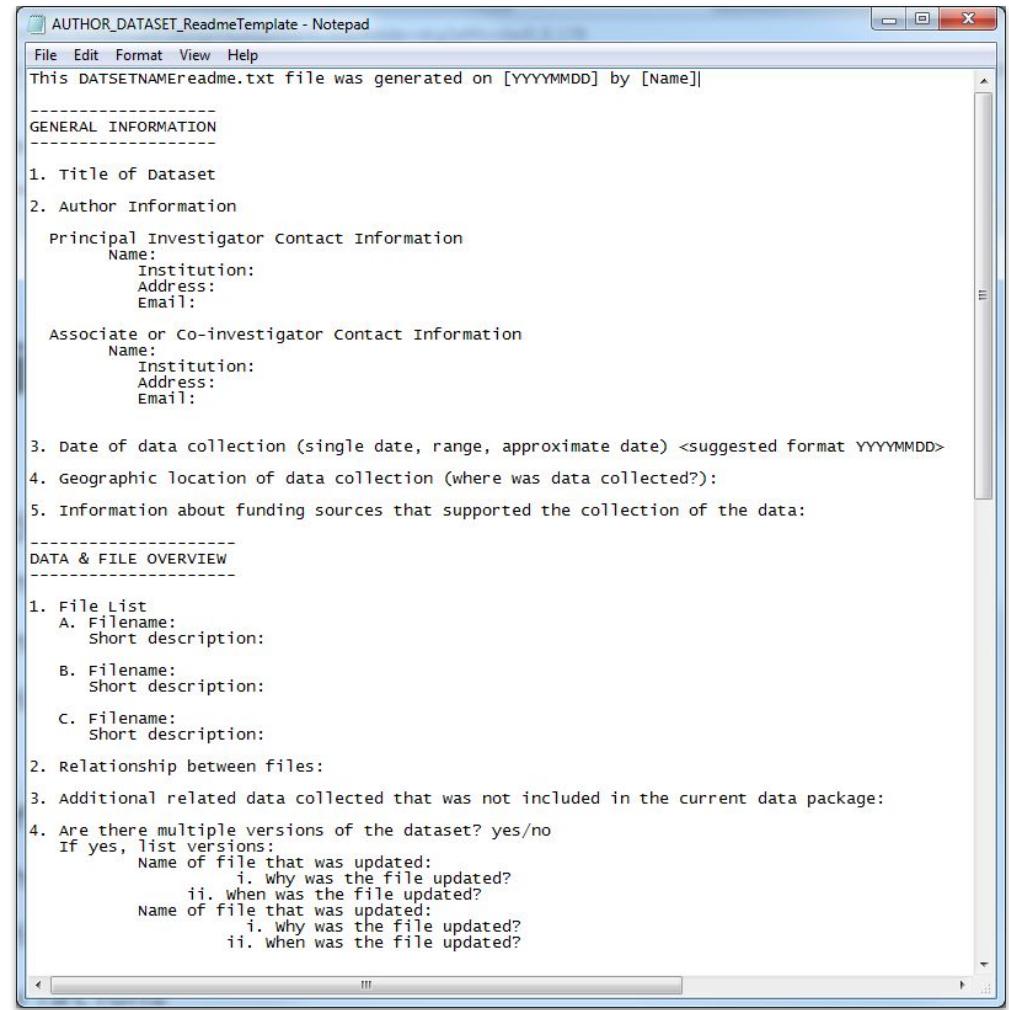


Your Code

- What is that data field? What are the units?
- What does that acronym mean?

README File

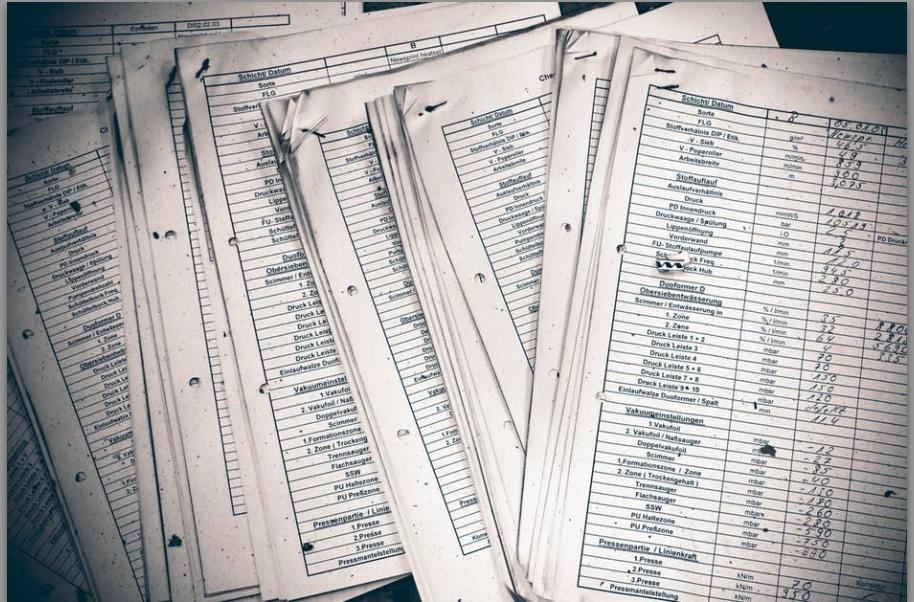
- Basic project information
- Title, Contributions, Grant Info
- Contact information
- All locations of where data live, including backups
- Useful information about the files and how they are organized
- Explain file naming conventions and abbreviations



Your Turn!

On-Your-Own Exercise

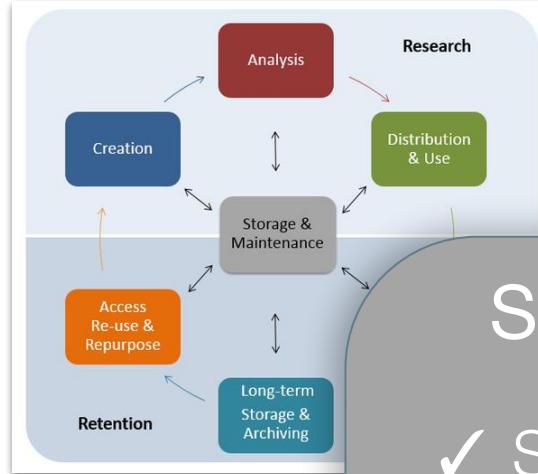
Conventions



- How do you currently organize your files and folders?
- Do you find what you want, when you need them? If not, how do you want your files to sort or be discovered?
- What metadata elements are important for understanding your data? (author, date created, type of instrument, etc.).
- Are there any standards in your field, institution, department, group, etc. for what metadata you should be collecting?

Storage

**Storage,
backup, and
security are
interrelated**



Storage & Maintenance

- ✓ Store on appropriate tier, with proper security
- ✓ Store locally on servers or in the cloud
- ✓ Plan to maintain system

Storage Options

Harvard Medical School

- HMS Tiered Storage
 - HMS offers several storage tiers that allow users to store data in different places, with varying behaviors, performance, and means of access
- HMS IT Software and Backups
- HMS Research Computing Orchestra high performance cluster

T.H. Chan School of Public Health

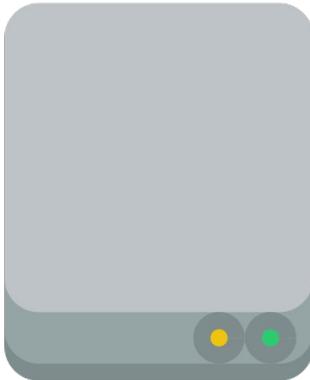
- HSPH Managed Servers
 - S: drive - department's shared storage location
 - P: drive - personal storage
- HSPH Information Security Consulting
- FAS Research Computing Center Odyssey high performance cluster

Backups = 3 Copies



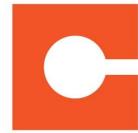
Here

lab computer



Near

portable hard drive
(stored offsite)



CODE42

Far

Harvard CrashPlan
(on lab computer)

Security

Access

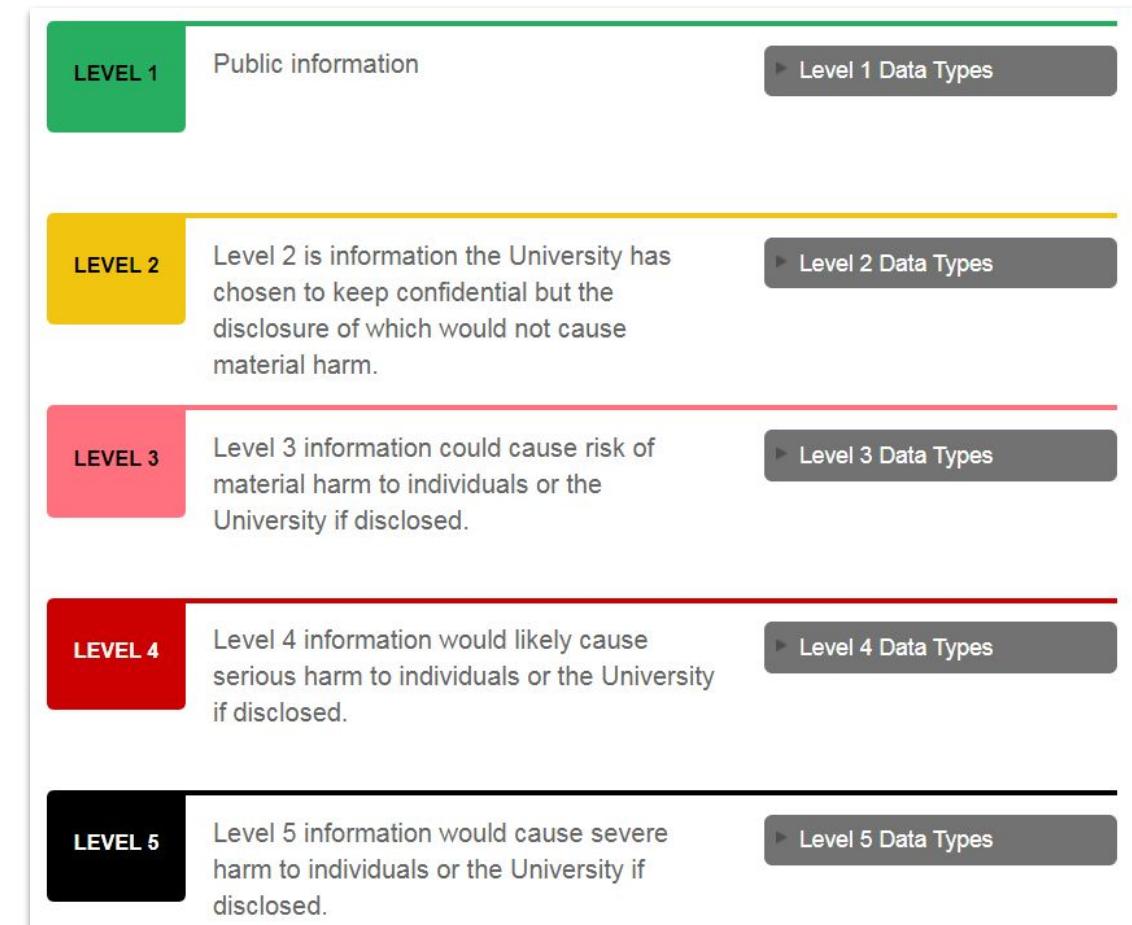
Limiting the availability of your data

Systems

Protecting your hardware and software

Data Integrity

Ensure that your data is not manipulated in an unauthorized way



Electronic Lab Notebooks

Electronic Lab Notebooks at HMS

Lab notebooks are good for writing down procedures, observations, conclusions and for drawing flow charts and diagrams by hand. However, in order to accommodate the increase of digital data collected, researchers have taped instrumentation and computer printouts onto the pages of their notebooks, or cross-referenced larger data sets by recording file names and locations in the notebook.

An ELN (electronic lab notebook) is a software tool that in its most basic form is a digital version of a page in a paper lab notebook. In this electronic notebook you can record your observations, draw graphs, store data and images, and search for data using your computer or mobile device. This offers several advantages over a paper notebook:

The number of available ELN tools is increasing and the functions they offer are becoming more advanced. However, this may be confusing to evaluate all of the advantages and limitations of each tool and how it can best serve your project.

The Electronic Lab Notebook Matrix has been created to aid HMS researchers in evaluating the various ELN tools currently available. It is designed to help researchers compare and contrast the numerous solutions available and make informed decisions about which tools best fit their needs. The matrix provides a detailed overview of the features and capabilities of each tool, along with specific details on how they can be used to support research workflows.

Questions about Electronic Lab Notebooks at HMS? Contact us at datamanagement@hms.harvard.edu.

Features	Specifications															
	Benchling	Biovia	Confluence	Doccollab	ECL	ELOG	Evernote	Exemplar	Findings	Hivebench	IDBS	LabArchives	LabCollector	LabWare	LabVantage	LabW
Interactivity																
Intuitive Interface Design	✓	No response received	*	*	No response received	*	No response received	*	*	*	*	✓	No response received	*	*	
Auto Metadata Harvest	*	No response received	✗	✓	No response received	✗	No response received	*	No response received	✗	✓	✓	No response received	✗	✓	
Search functions can search across file formats and beyond types	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
Ability to manipulate files and images	*	No response received	*	*	No response received	*	No response received	*	No response received	*	*	*	*	*	*	
Support for multiple open windows	✓	*	✓	✓	✓	No response received	✓	✓	*	✓	✓	✓	✓	✓	✓	
Ability to link out	✗	No response received	*	✓	✓	✓	✓	✓	✓	✓	✓	✓	No response received	✓	✓	
Support for Researcher Documentation																
Hyperlink support	✓	No response received	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	No response received	✓	✓	
Metadata Creation Prompts	✓	No response received	✗	✓	✓	No response received	✗	No response received	✗	✓	✓	✓	No response received	✗	✓	
Rights Management (licensing)	*	No response received	*	✓	✓	No response received	*	No response received	✓	✓	*	✓	✓	✓	✓	
Protocol Integration	✓	*	✓	✓	✓	No response received	✓	*	*	✓	✓	✓	*	*	*	
Adaptability to Lab Workflows																
Accounts/Permissions Levels	✓	No response received	*	✓	✓	✓	✓	✓	*	*	✓	✓	✓	✓	*	
Internal Data Sharing	✓	*	*	✓	✓	No response received	✓	✓	*	✓	✓	✓	✓	✓	*	
Adaptable to a Variety of Workflows	*	No response received	*	*	*	No response received	*	No response received	*	*	*	*	*	*	*	
Compatibility with authoring tools	✓	No response received	*	✓	✓	No response received	✗	No response received	✗	✓	✓	✓	No response received	✗	*	
Windows Compatible	✓	No response received	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	*	
Macintosh Compatible	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	*	
Linux Compatible	✓	✓	✗	✓	✓	✓	✓	No response received	✓	✓	✓	✓	✓	No response received	✓	
Android Compatible	✓	✓	✓	✓	✓	✓	✓	No response received	✓	✓	✓	✓	✓	✓	*	
iOS Compatible	✓	✓	✓	✓	✓	✓	✓	No response received	✓	✓	✓	✓	✓	✓	*	
Storage																
Cloud Storage	✓	No response received	✗	✓	✓	No response received	✓	No response received	✓	✓	✓	✓	✓	✓	*	
Local Storage	*	No response received	✓	✓	✗	No response received	✓	No response received	✓	✓	✓	✓	No response received	✓	✓	
Hybrid (cloud/local) Storage	*	No response received	✗	✗	No response received	✗	No response received	✗	No response received	✓	✓	✓	No response received	✗	✗	
Versioning	*	*	*	*	*	No response received	*	No response received	*	No response received	*	*	*	*	*	
File Redundancy	*	No response received	*	*	No response received	*	No response received	*	No response received	*	*	*	No response received	*	*	
Creates stable URLs or persistent identifiers for entries	✓	No response received	✓	✓	No response received	✓	No response received	✓	No response received	✓	✓	✓	No response received	✓	✓	
Can unregistered users access the data found at persistent links?	✓	No response received	✓	✗	No response received	✗	No response received	✓	No response received	*	*	*	No response received	✗	✗	
Storage Capacity - Users	*	No response received	*	*	No response received	*	No response received	*	No response received	*	*	*	No response received	*	*	
Storage Capacity - Max File Size	*	No response received	*	*	No response received	*	No response received	*	No response received	*	*	*	No response received	*	*	

Electronic Lab Notebook Matrix

<https://datamanagement.hms.harvard.edu/electronic-lab-notebooks>

Your Turn!

On-Your-Own Exercise

Storage



- Conduct a quick inventory of your data:
 - *What datasets do you have?*
 - *How big are they?*
- Inventory where your files are currently stored, including backups:
 - *How safe are your data?*
- Do you have any PHI or HRCI data?
 - *What do you need to ensure their security?*

Data Sharing

When establishing data sharing and access policies and provisions, consider *whom* you will share your data with, *how* it will be shared, and *when* in the research process you will share it.

Distribution & Use

- ✓ Share data with collaborators
- ✓ Annotate datasets & upload to public repositories
- ✓ Include in relevant publications & reports

Ownership & Data Custodians

Do you know who owns your data or the dataset you are using?



read

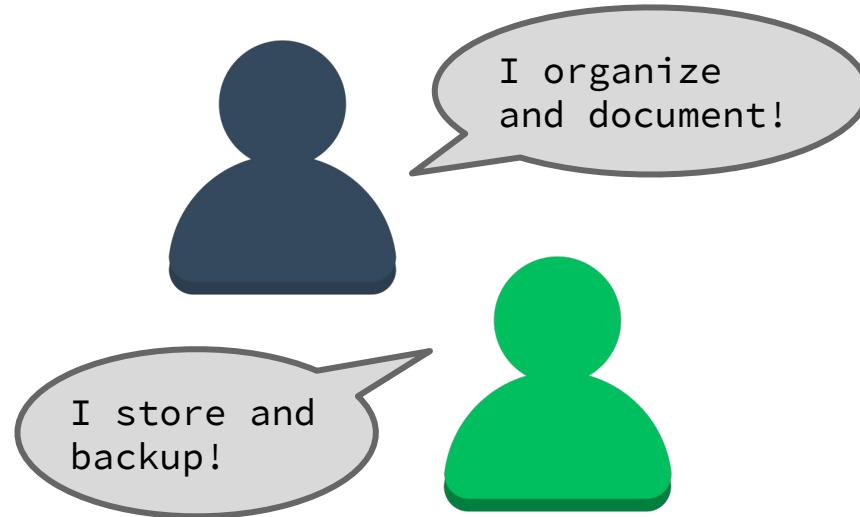


write



download

Who has access?



Who has
responsibility?



Who has IP rights?

Repositories

- Types:
 - Institutional
 - Domain or discipline-specific
 - General
- Funder specified repository
- Repository of Research Data Repositories
<https://www.re3data.org>



Look for:

- Open access
- Generates persistent IDs
- Good archival practices (Trusted Digital Repository certification)
- Flexible metadata
- Additional services (data cleanup, format migration/normalization, metadata assistance, etc.)

REPOSITORIES

Dryad
figshare
GigaScience
Harvard Dataverse
NIH and NCBI Repositories
Scientific Data
Zenodo
Additional Resources

HOME / BEST PRACTICES /

Repositories

The number of available resources for data sharing and data publication has increased substantially over the past few years, making it difficult for individual researchers to evaluate the advantages and limitations of the various options they search for the right solution to address their needs.

Here, we compare and contrast several of the general data repositories and data publication platforms available for biomedical science researchers. Click on the matrix below to see detailed descriptions of each resource.



Requirement

Yes
No

Page last updated July 2, 2018

Requirement	Dataverse	Dryad	figshare	Zenodo	GigaScience	Scientific Data
Data Size and Format						
Hosting of common file formats (e.g. csv, tsv, xls, xlsx, doc, pdf)	✓	✓	✓	✓	✓	✗
Hosting of proprietary file formats (e.g. raw image files)	✓	✓	✓	✓	✗	✗
Unlimited size per file	✗	✓	✗	✗	✓	✗
Unlimited total dataset size	✓	✓	✓	✓	✓	✗
Data Licensing						
CC0 waiver1	recommended	required	recommended	available	required	✗
Data Attribution and Citation Tools						
Assignment of dataset DOIs	✓	✓	✓	✓	✓	✗
User Access Controls						
Tiered access (e.g. administrator-level, collaborator-level, curator-level)	✓	✗	✓	✗	✗	✗
Journal-integrated, anonymous access (for peer review pre-publication)	✗	✓	✓	✗	✓	✗
Optional embargo to data release following publication	✗	✓	✓	✓	✓	✗
Data Access Tools						
Comprehensive data and metadata search tools	✓	✗	✗	✗	✗	✗
Data access via direct download	✓	✓	✓	✓	✓	✗
Data downloading via API	✓	✓	✓	✓	✓	✗
Built-in tools for reading proprietary file formats	✗	✗	✓	✗	✗	✗
Integrated data analysis tools	✓	✗	✗	✗	✓	✗
Cost						
Data deposition fees	none	tiered	none	none	none	✗
Data maintenance fees	none	none	none	none	none	✗



HARVARD MEDICAL SCHOOL

Yes
No

Page last updated July 2, 2018

Requirement	Dataverse	Dryad	figshare	Zenodo	GigaScience	Scientific Data
Data Size and Format						
Hosting of common file formats (e.g. csv, tsv, xls, xlsx, doc, pdf)	✓	✓	✓	✓	✓	✗
Hosting of proprietary file formats (e.g. raw image files)	✓	✓	✓	✓	✗	✗
Unlimited size per file	✗	✓	✗	✗	✓	✗
Unlimited total dataset size	✓	✓	✓	✓	✓	✗
Data Licensing						
CC0 waiver1	recommended	required	recommended	available	required	✗
Data Attribution and Citation Tools						
Assignment of dataset DOIs	✓	✓	✓	✓	✓	✗
User Access Controls						
Tiered access (e.g. administrator-level, collaborator-level, curator-level)	✓	✗	✓	✗	✗	✗
Journal-integrated, anonymous access (for peer review pre-publication)	✗	✓	✓	✗	✓	✗
Optional embargo to data release following publication	✗	✓	✓	✓	✓	✗
Data Access Tools						
Comprehensive data and metadata search tools	✓	✗	✗	✗	✗	✓
Data access via direct download	✓	✓	✓	✓	✓	✗
Data downloading via API	✓	✓	✓	✓	✓	✗
Built-in tools for reading proprietary file formats	✗	✗	✓	✗	✗	✗
Integrated data analysis tools	✓	✗	✗	✗	✓	✗
Cost						
Data deposition fees	none	tiered	none	none	none	✗
Data maintenance fees	none	none	none	none	none	✗

Data Repository Comparison Matrix

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

Your Turn! On-Your-Own Exercise Repositories



- What does long-term mean for your data?
- How long do you need to keep/store/share your data?
- Do your funder(s) or publisher(s) require any specific data sharing mechanism?
 - Are you required to deposit your data in a repository?
 - Is there a disciplinary repository available?
 - What repository(ies) will work for your dataset?

Data Publication: Citation & Attribution

Acknowledgement of the use of someone else's information or work is a long-accepted practice in scholarly communication.

Core elements of a data citation:

- *Author/Creator(s)*
- *Title*
- *Version*
- *Publication Date*
- *Publisher/Archive*
- *Identifier/Locator*

Access Re-use & Repurpose

- ✓ Provide access to the data
- ✓ Define sharing policies and use agreements
- ✓ Provide appropriate attribution via citation

Data Publication: Citation & Attribution

Author/Creator(s): creators of the data; can be one or more people or organizations

Title: title of the data set

Version: exact version or edition of the data set used

Publication Date: date when the data set was published or released

Publisher/Archive: data center or repository that is archiving and distributing

Identifier/Locator: URL or other linkable locator for the data; a persistent, permanent URL such as a DOI (Digital Object Identifier) or a handle is preferred

Unique Identifiers



Digital Object Identifier

Permanently assigned to an object to provide a resolvable persistent network link to current information about that object, including where the object, or information about it, can be found on the Internet

www.doi.org



Open Researcher and Contributor ID

Provides a persistent digital identifier that distinguishes you from every other researcher and supports automated linkages between you and your professional activities ensuring that your work is recognized

orcid.org

Your Turn!

On-Your-Own Exercise

ORCID

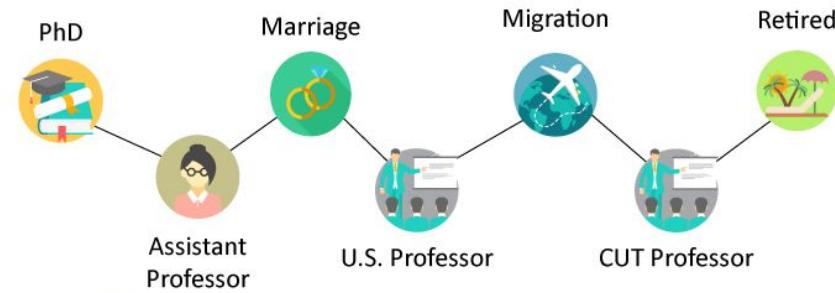


Don't have an ORCID?

Create one now!

It's free, easy and will last throughout your professional career!

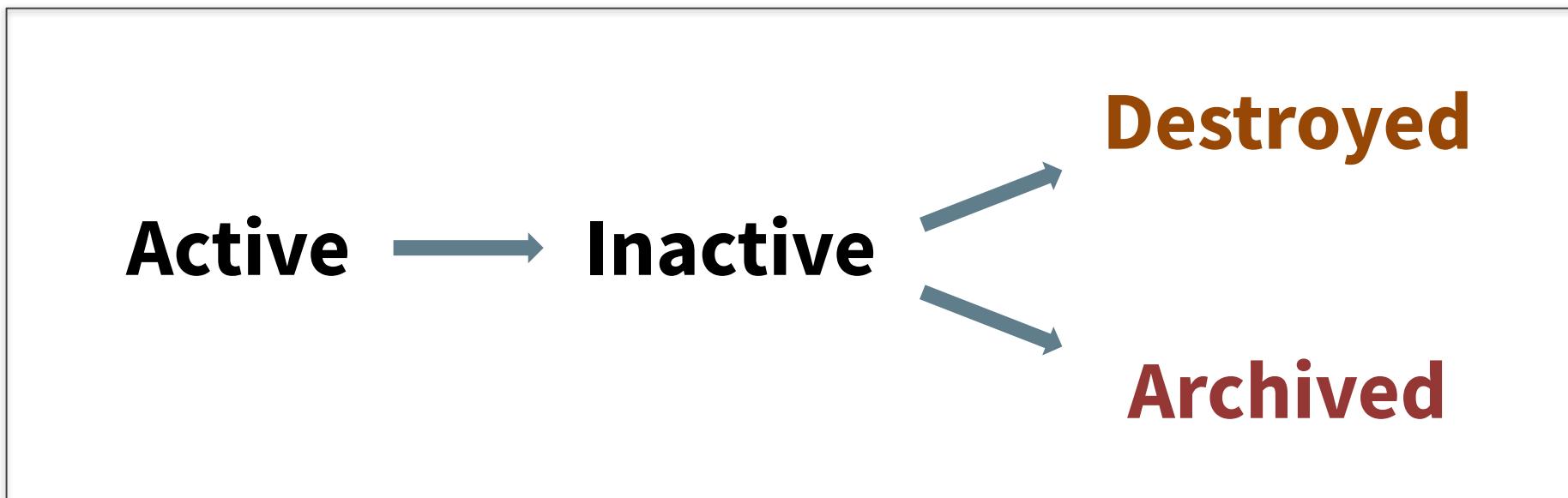
orcid.org



ID 0000 - 1234 - 5678 - 0000

Research Records

Four Types of Records



Retention

Data retention requirements are put in place by funding agencies and sponsoring institutions for a number of reasons:

- *promote the reuse of data within and across disciplines*
- *protect intellectual property rights*
- *make research findings available*
- *support open data initiatives*

Evaluate for Retention

- ✓ Identify and retain essential research records
- ✓ Organize and annotate appropriately

Appraisal & Archiving

Appraisal process for evaluating research records and data:

- ***Inventory of the records:*** volume, data types, formats, metadata, other relevant information
- ***Interview about the project:*** impact of the project, significance of the research or researcher, basic information about the grant

Long-term Storage & Archiving

- ✓ In compliance with HMS & federal policy
- ✓ As requested by investigators

Your Turn! On-Your-Own Exercise

Research Environment



Who

- Is there anyone in your lab responsible for data management?
- Is there anyone on your latest project responsible for data management? Is there a plan for when/if they leave?

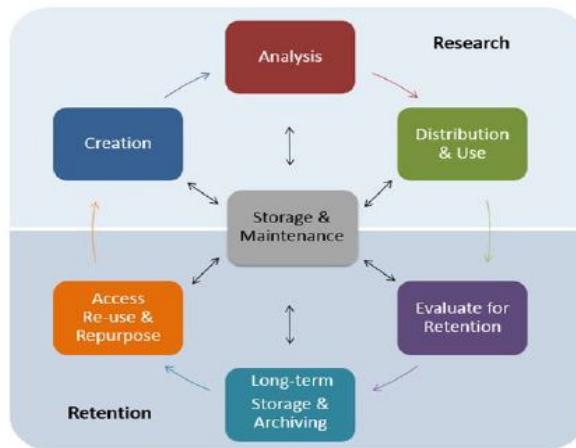
How

- Where/how is your data stored currently? Is it backed up?
- What are the plans when the grant ends?
- Does your group use collaborative tools, and if so, what works and what doesn't?

We know this is a lot...

We suggest introducing small changes starting today!

Research Data Management Checklist



This document serves as a reference checklist to keep track of the elements that make up good research data management in the RDM lifecycle.

The RDM lifecycle is not linear and you may find yourself jumping around this lifecycle throughout your project.

Begin building or locate a detailed README.txt overview of your project immediately. Examples of data documentation include lab notebooks and experimental protocols, questionnaires, codebooks, data dictionaries, software syntax and output files, information about your equipment settings and calibration, database schema, methodology reports, and provenance information.

<http://datamanagement.hms.harvard.edu/metadata-overview>

Your DMP document should describe final dataset formats, documentation, analytic tools necessary to use the data, data sharing agreements, and how and when the data will be made accessible to others.

We are open to identifying new kinds of data management practices that could benefit the biomedical sciences. If you would like to contribute to the RDM website for your field, please contact the HMS Data Management Working Group through the website link to "Submit your questions and feedback!"

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

DATA CREATION: RDM PLANNING

What does your research project look like from start to (anticipated) finish?

```

graph TD
    Create[Create] --> Analyze[Analyze]
    Analyze --> Research[Research]
    Research --> Distribute[Distribute & Use]
    Distribute --> Store[Storage & Maintenance]
    Store --> Evaluate[Evaluate for Retention]
    Evaluate --> Retain[Retain]
    Retain --> Access[Access Re-use & Repurpose]
    Access --> Create
    
```

DATA CREATION: RDM PLANNING

What does your research project look like from start to (anticipated) finish?

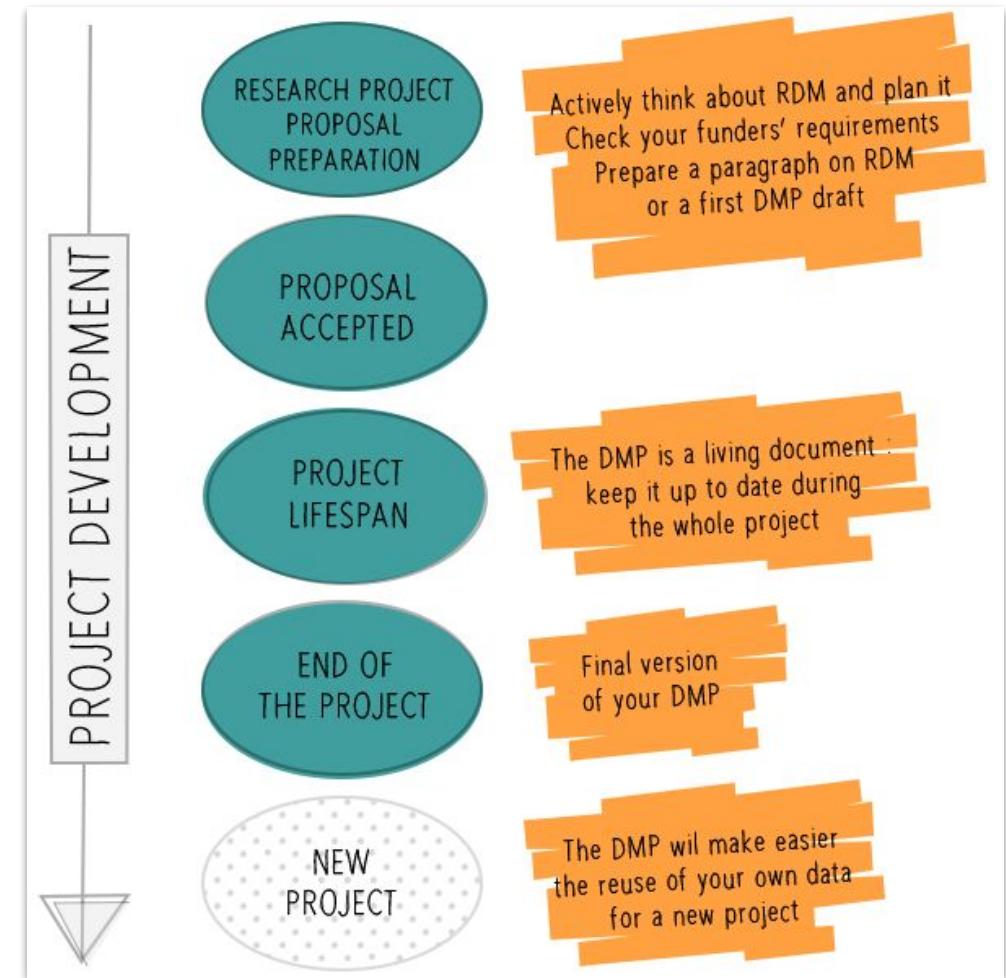
<input type="checkbox"/> ID	✓ Determined by the funder and/or institution
<input type="checkbox"/> Funder(s)	✓ Data security policy ✓ Data sharing policy ✓ Data retention policy
<input type="checkbox"/> Grant #	✓ Post award DMPs only
<input type="checkbox"/> Project name	✓ As it appears exactly as on the grant. Append to grant proposal.
<input type="checkbox"/> Project description (background/rationale)	✓ What research question(s) are you addressing? ✓ Summarize the study methods and design including data collection method(s) and purpose of collection. ✓ If creating or collecting data in the field, how will you ensure its safe transfer into your main secured systems?
<input type="checkbox"/> Data description	✓ Content description (brief) - include any value definitions, questionnaires or instruments, or analysis procedures. ✓ Type (imagine data, genomic, Qx, etc.) ✓ Format <ul style="list-style-type: none"> • Databases: XML, CSV • Geospatial: SHP, DBF, GeoTIFF, NetCDF • Moving Images: MOV, MPEG, AVI, MXF • Audio: WAVE, AIFF, MP3, MXF • Numbers/statistics: ASCII, DTA, POR, SAS, SAV • Images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP • Text: PDF/A, HTML, ASCII, XML, UTF-8 • Graphs: JSON, YAML, XML
If you need to convert or migrate your data files from one format to another, be aware of the potential risk of the loss or corruption of your data and take appropriate steps to avoid/minimize.	
✓ Briefly justify the use of format – is your chosen format open, non-proprietary and in widespread use? ✓ Estimated volume? ✓ Describe any existing data being used (citations, link and DOI).	
<input type="checkbox"/> PI	✓ Name of Principal Investigator(s) or main researcher(s) on the project.
<input type="checkbox"/> PI ORCID ID	✓ ORCID http://orcid.org/
<input type="checkbox"/> Administrative data	✓ Contacts/addresses/email details ✓ Date of first DMP ✓ Date and details for subsequent revision(s) of DMP
<input type="checkbox"/> Additional Institution(s)	

Research Data Management Checklist

datamanagement.hms.harvard.edu/hms-data-lifecycle

Data Management Plan

A data management plan (DMP) is a written document that describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data.

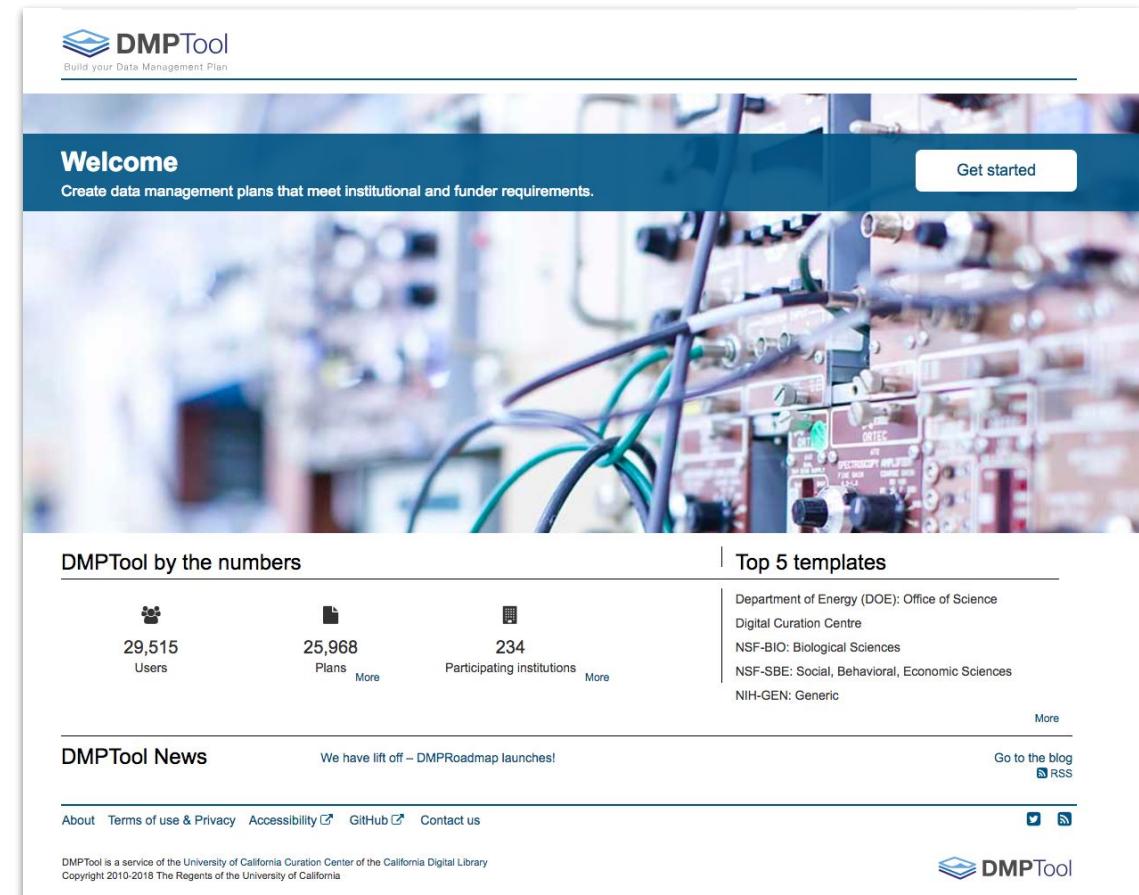


<https://researchdata.epfl.ch/plan-fund/dmp>

DMPTool

The DMPTool is an online tool that includes data management plan templates for many of the large funding agencies that require them.

Harvard is an affiliated partner institution. You can login as a user from your institution with your HarvardKey. By being affiliated Harvard, you will be presented with institution-specific guidance to help you complete your plan.



The screenshot shows the DMPTool homepage. At the top, there's a navigation bar with the DMPTool logo and a subtext "Build your Data Management Plan". Below the header is a banner featuring a blurred image of laboratory equipment, specifically a circuit board with various components and wires. The banner has a dark blue header with the word "Welcome" and a subtext "Create data management plans that meet institutional and funder requirements." On the right side of the banner is a white button labeled "Get started". Below the banner is a section titled "DMPTool by the numbers" containing three items: "29,515 Users", "25,968 Plans", and "234 Participating institutions". To the right of this is a "Top 5 templates" sidebar listing five categories: "Department of Energy (DOE): Office of Science", "Digital Curation Centre", "NSF-BIO: Biological Sciences", "NSF-SBE: Social, Behavioral, Economic Sciences", and "NIH-GEN: Generic". At the bottom of the page, there's a "DMPTool News" section with a link to "DMPRoadmap launches!", social media links for Twitter and RSS, and a footer with links to "About", "Terms of use & Privacy", "Accessibility", "GitHub", and "Contact us". The footer also includes the DMPTool logo and the URL "https://dmptool.org".

<https://datamanagement.hms.harvard.edu/data-management-plan>

Questions?

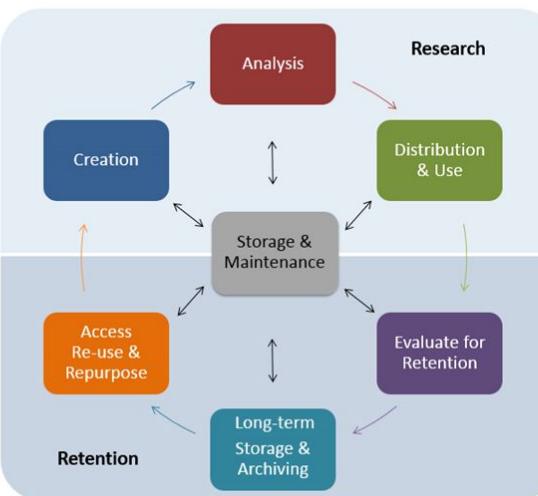
Harvard Biomedical Data Management
Best practices & support services for research data lifecycles

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 the Command-line Interface

Tuesday, June 18

1:30 – 3:00 pm

TMEC 227 Mini Amphitheater

bit.ly/RDM-Summer19

Introduction to High Performance Computing

Thursday, July 11

1:30 – 3:00 pm

Countway Library Minot Room

bit.ly/RDM-Summer19

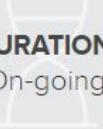
bit.ly/rdm-survey



BEST PRACTICES FOR BIOMEDICAL RESEARCH DATA MANAGEMENT

Started Jan 8, 2018

 COURSE DATE:
Starts Jan 8, 2018

 DURATION:
On-going

 COMMITMENT:
7+ hrs/week

 REQUIREMENT:
None

 COURSE TYPE:
Self-paced

 CREDENTIAL:
Badge, Certificate
(free)

Open Online Course via Canvas
bit.ly/HMS-RDM-MOOC

Key Resources

Harvard Biomedical Data Management
datamanagement.hms.harvard.edu

Center for the History of Medicine | Archives and Records Management
www.countway.harvard.edu/chom/archives-and-records-management

Research Information Technology Solutions
rits.hms.harvard.edu

Office of the Vice Provost for Research | Research Data Security & Management
vpr.harvard.edu/pages/research-data-security-and-management

Harvard Catalyst | The Harvard Clinical and Translational Science Center
catalyst.harvard.edu

Office for Scholarly Communications
osc.hul.harvard.edu/policies