

Data Management for Labs

How to Hit the Ground Running

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

Sarah Hauserman

Research Data Analyst
Research Information Technology Solutions
Sarah.Hauserman@hms.harvard.edu



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

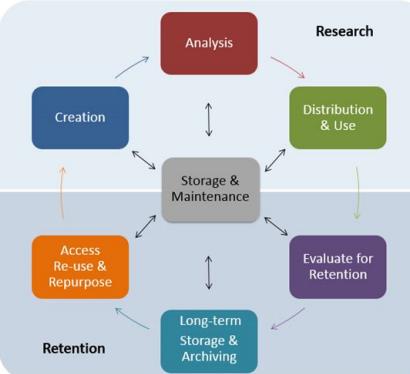
Harvard Biomedical Data Management
Best practices & support services for research data lifecycles

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 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.



UPCOMING EVENTS

- 2019 APR 11** Data Management for Labs: How to Hit the Ground Running
- 2019 MAY 02** Data Management Working Group Monthly Meeting
- 2019 MAY 07** Getting Started with Data Management Plans

[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:


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!

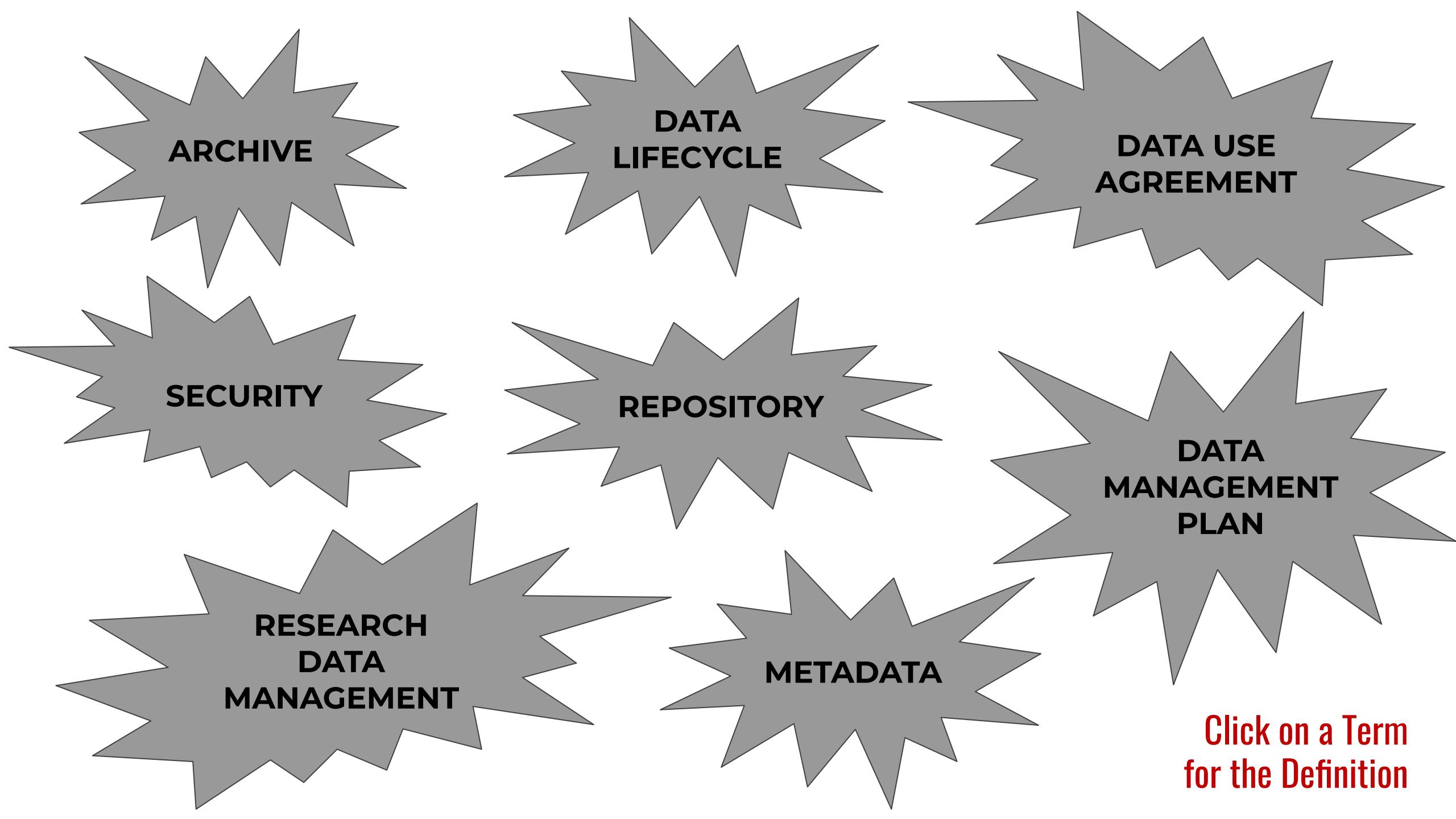
Name

School / Department / Lab

**What type of on-boarding training have you
gone through for a lab or project?**

JARGON BUSTING





ARCHIVE

**DATA
LIFECYCLE**

**DATA USE
AGREEMENT**

SECURITY

REPOSITORY

**DATA
MANAGEMENT
PLAN**

**RESEARCH
DATA
MANAGEMENT**

METADATA

**Click on a Term
for the Definition**



ARCHIVE

The transfer of materials or data to a facility/site authorized to appraise, preserve, and provide access to the information.



DATA LIFECYCLE

The data lifecycle represents all of the stages of data throughout its life from creation to distribution and reuse.



DATA USE AGREEMENT

A Data Use Agreement (DUA) governs access to and treatment of data: (i) provided by an outside organization to Harvard for use in Harvard research, or (ii) provided by Harvard to an outside organization for use in its research.



SECURITY

Data security refers to ways in which data is kept safe from harm, alteration, or unauthorized access during gathering, analysis, storage, and transmission. Computer systems used to store data should have security measures such as firewalls, virus protection and strong password protection.



REPOSITORY

A place to hold data, make data available for use, and organize data in a logical manner. Also, an appropriate, subject-specific location where researchers can submit their data. Data repositories may have specific requirements concerning subject or research domain, data re-use and access, file format and data structure, and the types of metadata that can be used.



DATA MANAGEMENT PLAN

A data management plan determines how data should be collected, normalized, processed, analyzed, preserved, used and re-used over its lifetime. A DMP associated with a research study can include comprehensive information such as the types of data, metadata standards used, policies for access and sharing, and plans for archiving and preserving data to make accessible over time. A DMP ensures data will be properly documented and available for use by researchers in the future and are often required by grant funding agencies such as the National Science Foundation.



RESEARCH DATA MANAGEMENT

Research data management is a concept used to describe the managing, sharing, and archiving of research data to make it more accessible to the broader research community. Research data management provides an opportunity for researchers to create a plan ensuring their data will be organized, easily shareable with other researchers, and archived for long term preservation.

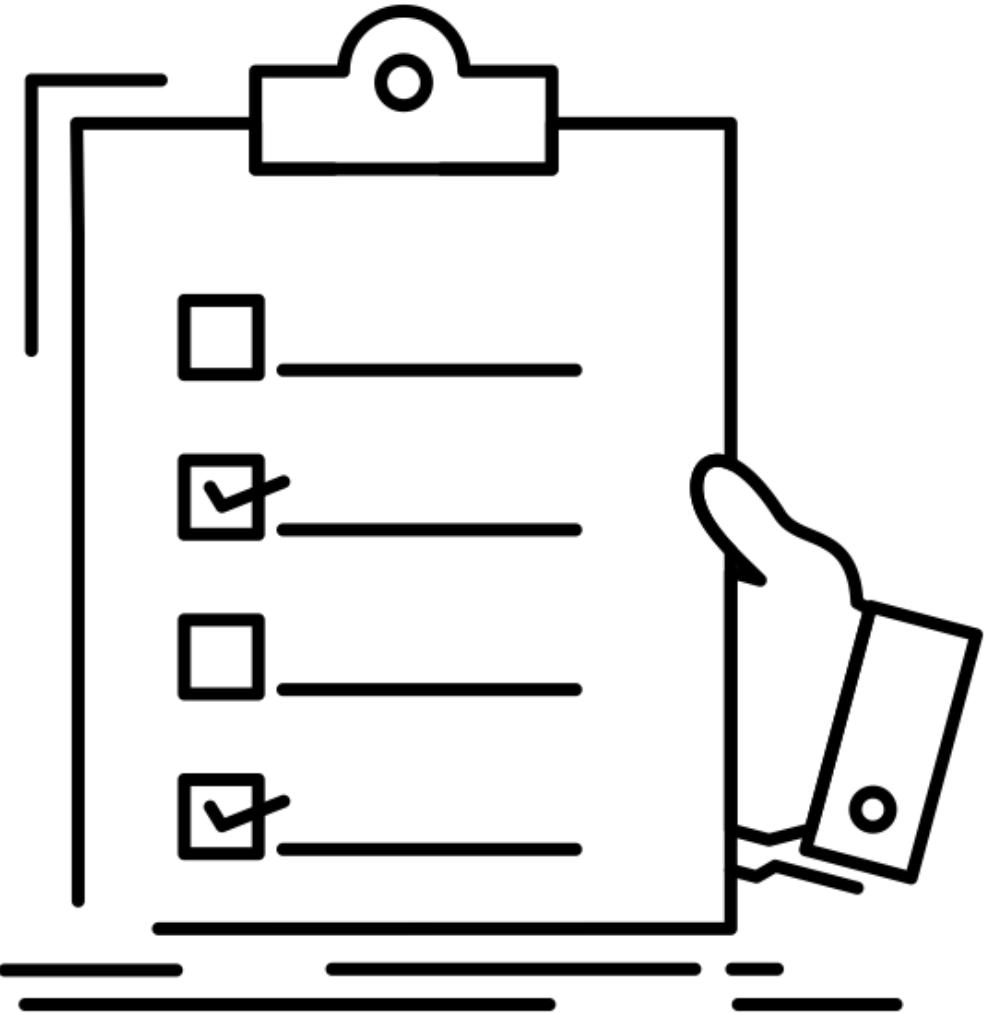


METADATA

Structured information about a resource that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage that resource. It ensures that the context for how your data was created, analyzed or stored is clear, detailed, and therefore, reproducible.

ACTIVITY

creating a checklist



Created by Flatart
from Noun Project

PLAN

Review Storage Options

Review Electronic Lab Notebook (ELN) Resources

STORE

Review Available Collaborative Tools

Review Potential Data Repositories

SHARE

Create a Preliminary Data Workflow

Review Laboratory, Department, and University Data Management Policies

PLAN



Review Laboratory, Department, and University Data Management Policies



Create a Preliminary Data Workflow

STORE



Review Storage Options

SHARE



Review Available Collaborative Tools



Review Potential Data Repositories



Review Electronic Lab Notebook (ELN) Resources

Data Management vs Project Management

Data Management

- Data Sources
- Data Acquisition
- Standards
- Data Processing
- Data Analysis Steps
- Metadata / Documentation
- Long-term Storage and Backups
- Preservation & Archiving Data
- Data Sharing, Access, Release
- Persistent Identifier Acquisition

Project Management

- Project Purpose
- General Data Management
- Explanation of significance
- Methodology
- Project Budget
- Project Staffing/Roles
- Acquisition of equipment, tools, and software
- Project Timeline and Milestones
- Project Deliverables



Albert and Neil discuss Data Management

RESEARCH DATA MANAGEMENT ONBOARDING CHECKLIST

Employee/Trainee onboarding to new labs and projects

This document serves as a general, research data management-focused guide to employee/trainee onboarding as they join a new lab or begin new projects. The document provides two checklists: *Checklist for Joining a New Lab* and *Checklist for Starting or Joining a New Project*. Follow one or both of these checklists as they apply to your situation. For more specific information, please see the [Research Data Management Checklist](#) provided by Countway Library. Internal and external links have been provided throughout the document as supplementary resources, including a glossary of terms. For additional assistance with terminology, visit [Data Management Terminology](#).

CHECKLIST FOR JOINING A NEW LAB

PLANNING		
1) Review Laboratory, Department, and University Data Management Policies:		
	Policies and Procedures <ul style="list-style-type: none">Contact the PI and department Research Administrator for laboratory and department-specific data management policies	
	<ul style="list-style-type: none">HMS Research Data Management Guidelines	Harvard Biomedical Data Management
	<ul style="list-style-type: none">Harvard Retention and Maintenance of Research Records and Data Principles (these principles apply to the Longwood Medical Campus Schools – HMS, HSPH, HSDM)	Retention and Maintenance of Research Records and Data: Principles and FAQs
	<ul style="list-style-type: none">For questions about data retention, contact the Longwood Campus Archives and Records Management Program	Longwood Campus Archives and Records Management
	<ul style="list-style-type: none">Harvard Research Data Security PolicyHarvard Information Security Policy	Harvard Research Data Security Policy Harvard Information Security Policy
2) Create a Preliminary Data Workflow:		

CHECKLIST FOR STARTING OR JOINING A NEW PROJECT

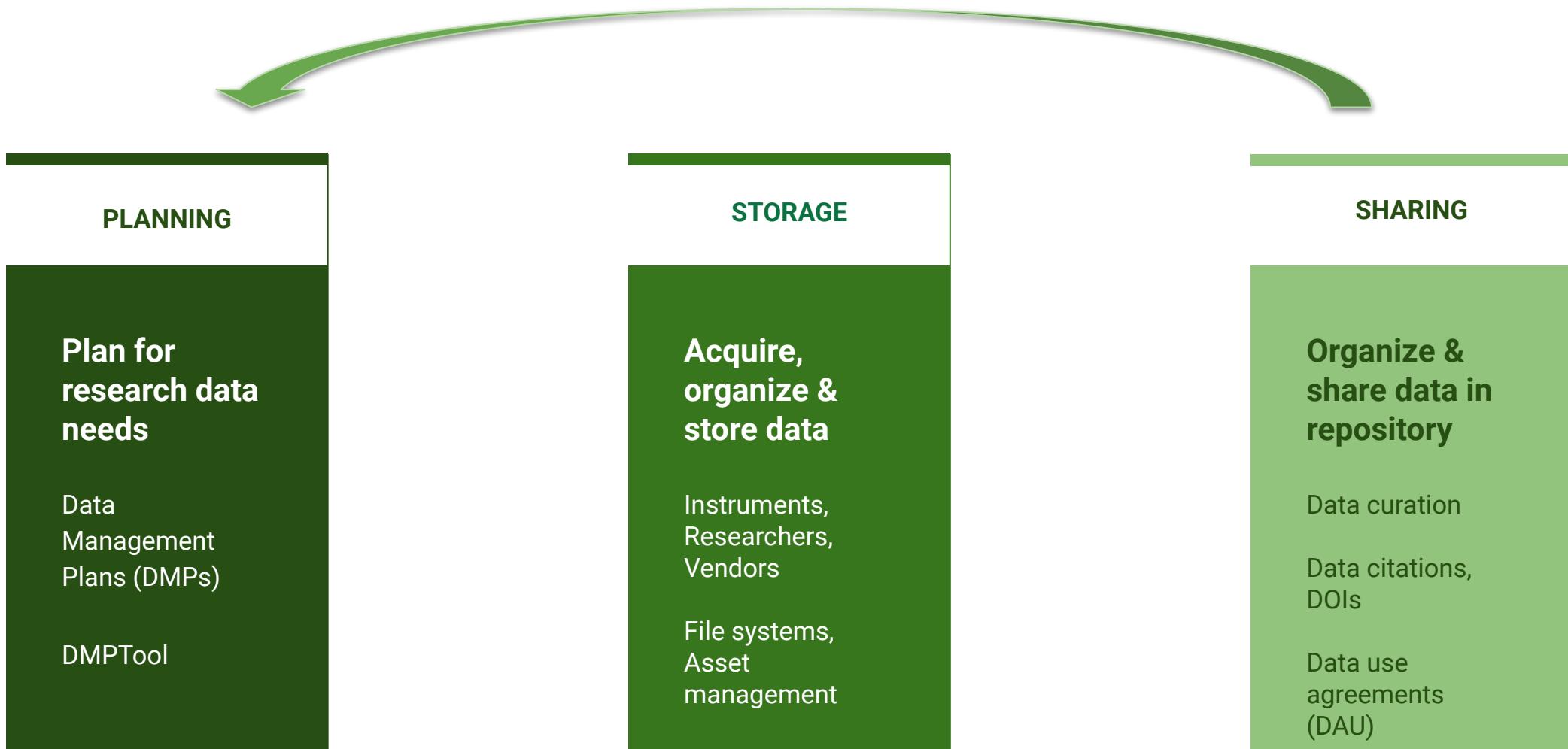
PLANNING		
1) Transfer Prior Data and Related Records to HMS (if relevant):	Contact the Office of Research Administration <ul style="list-style-type: none"> • The department Research Administrator will need to obtain Chair or Institute Director approval • Data Use Agreements (DUAs) govern access to and treatment of data: (i) provided by an outside organization to Harvard for use in Harvard research, or (ii) provided by Harvard to an outside organization for use in its research • When required, contact the Office of Research Administration to obtain a Data Use Agreement (DUA). For detailed instructions, visit the HMS Data Use Agreements webpage 	HMS Office of Research Administration HMS Data Use Agreements
2) Write a Data Management Plan or Review Existing Data Management Plan:	Construct a Data Management Plan (DMP) for the project <ul style="list-style-type: none"> • The document should describe data organization, storage, data security, final dataset formats, documentation, analytic tools necessary to use the data, data sharing agreements, retention plans, and how and when the data will be made accessible to others • Creating and following a DMP can substantially reduce the amount of storage needed by the lab by removing 	Biomedical Data Management Planning MIT Libraries: Data Management (external resource)

RDM Workflow



PLANNING	STORAGE	ANALYZE	SHARING	ARCHIVE	REUSE
Plan for research data needs Data Management Plans (DMPs) DMPTool	Acquire, organize & store data Instruments, Researchers, Vendors File systems, Asset management	Process data for current use R, Python, OpenRefine Statistical software Tableau, d3.js	Organize & share data in repository Data curation Data citations, DOIs Data use agreements (DAU)	Appraise & steward data Appraise for enduring value Migrate data to preservation repository	Discover & reuse data Locate data for new project Data repository

Checklist Workflow



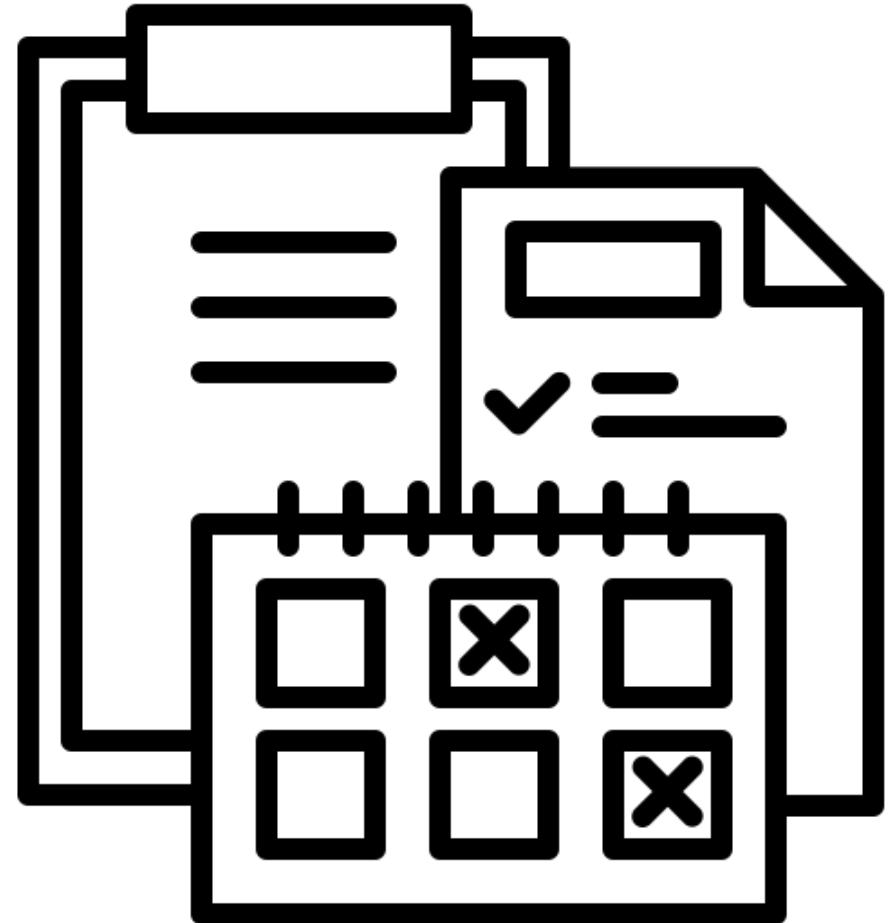
Joining a New Lab



Created by PJ Souders
from Noun Project

PLANNING

joining a new lab



Created by Maria Kislitsina
from Noun Project

Review Laboratory, Department, and University Data Management Policies



Created by PJ Souders
from Noun Project

- Contact the PI and department Research Administrator for laboratory and department-specific data management policies
- Harvard Retention and Maintenance of Research Records and Data Principles
- For questions about data retention, contact the Longwood Campus Archives and Records Management Program
- Harvard Research Data Security Policy
- Harvard Information Security Policy

Data 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*



Created by PJ Souders
from Noun Project

Create a Preliminary Data Workflow

- Review existing lab workflows, directory structures, and metadata standards
- Develop a preliminary organizational workflow for your research, including an established file (or directory) structure
 - Creating and following a data management workflow can substantially reduce the amount of storage needed by the lab by removing unnecessary files and avoiding file redundancy

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

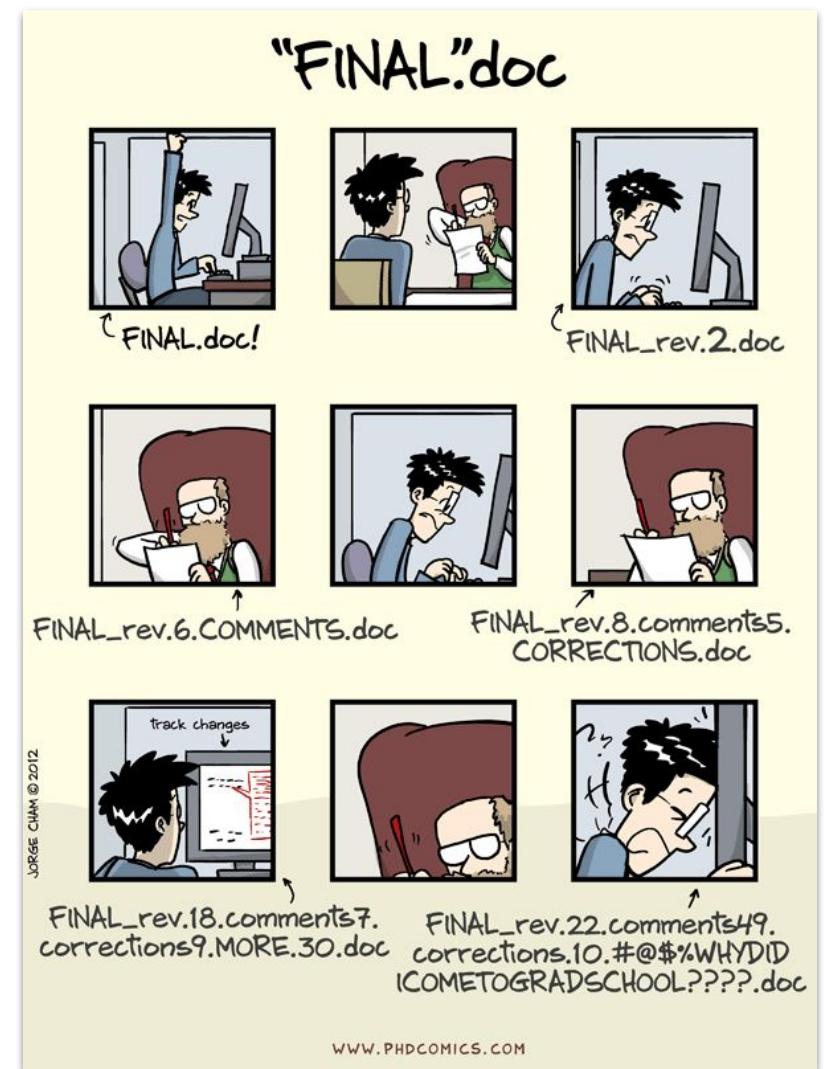
- Any system is better than none
- One project, one folder
- Separate folders for data or project stages
- Date-based folders (pairs well with lab notebook)

File Conventions

Files with naming conventions:

20161104_ProjectA_Ex1Test1_SmithE_v1.xlsx

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



Metadata

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

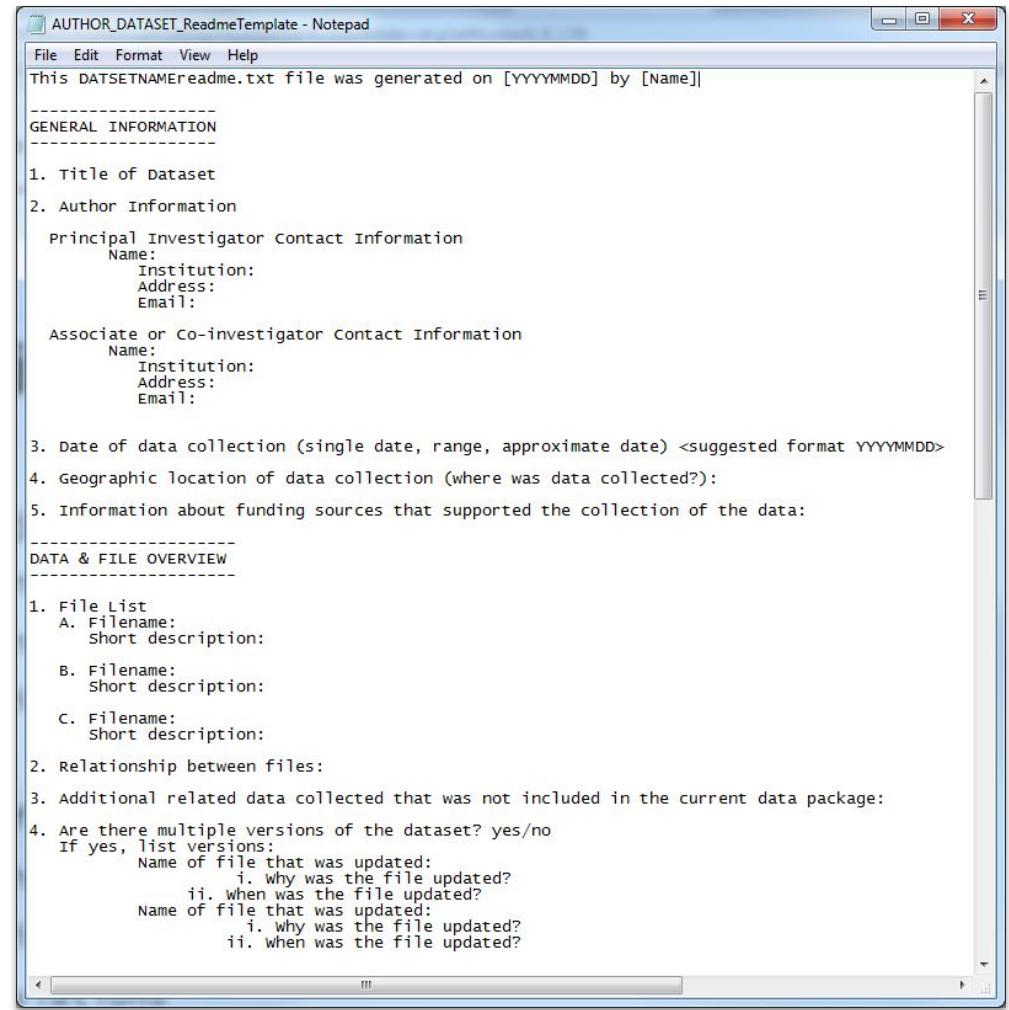
Metadata should be standardized, consistent and interoperable, and facilitates discovery, preservation and archiving of data



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.

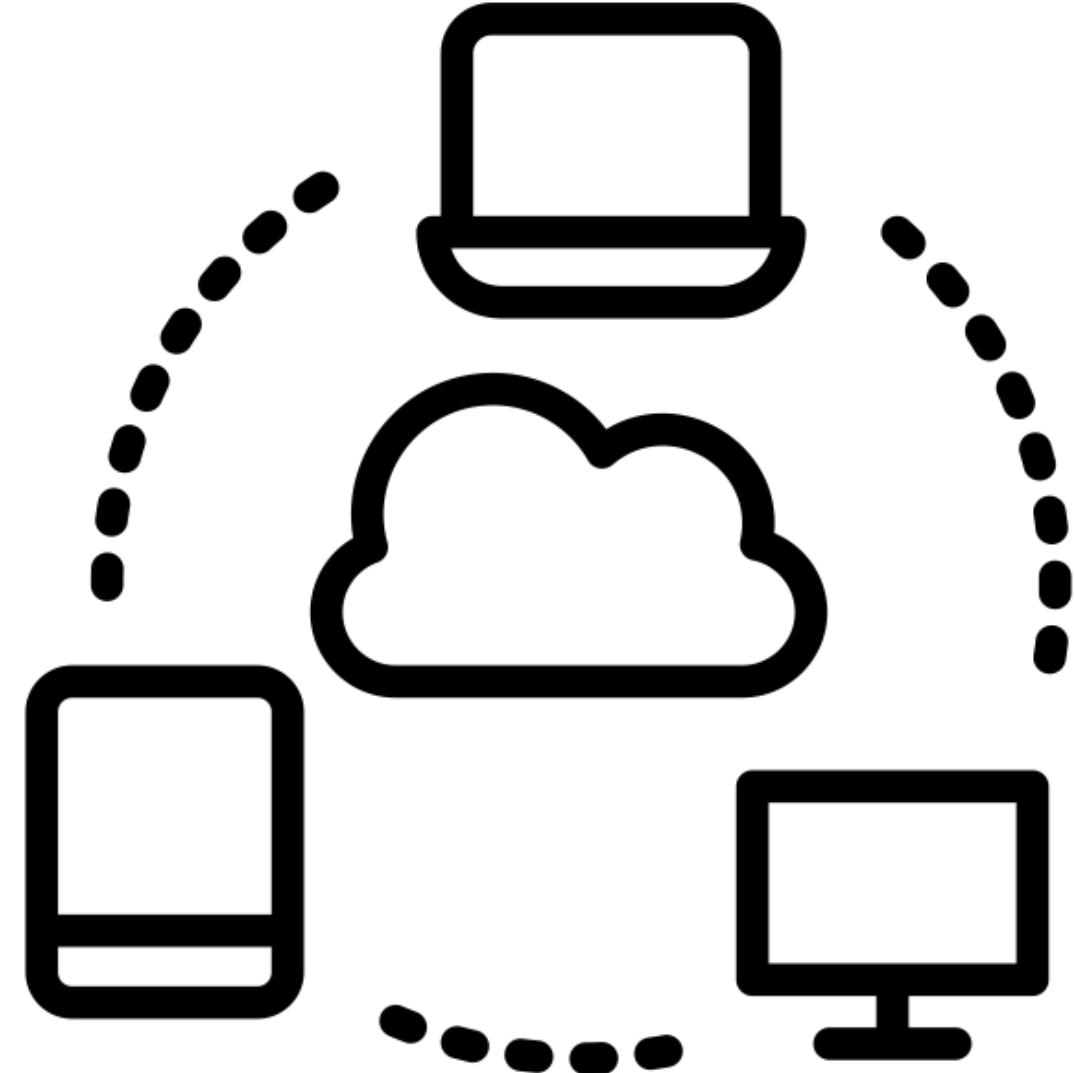
README File

- 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



STORAGE

joining a new lab



Created by ProSymbols
from Noun Project



Created by PJ Souders
from Noun Project

Review 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 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

Security

Access

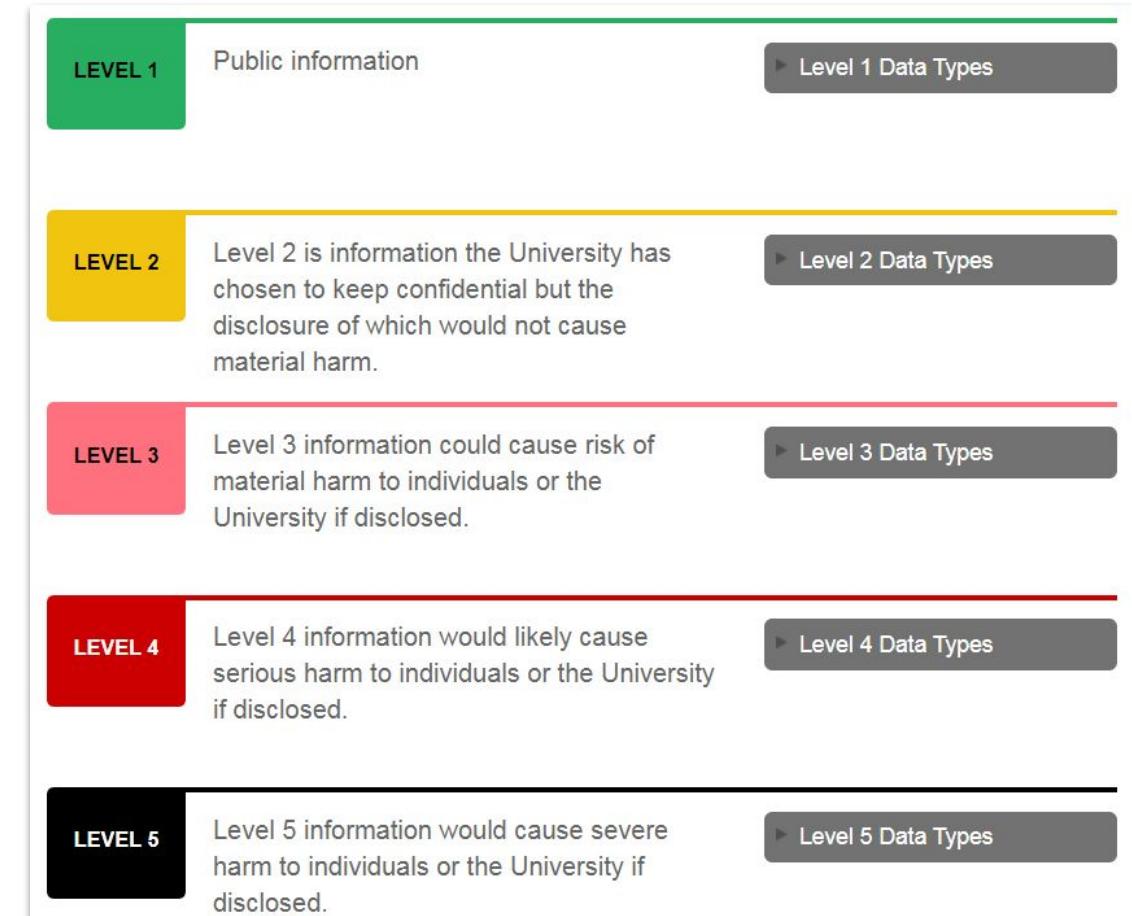
Limiting the availability of your data

Systems

Protecting your hardware and software

Data Integrity

Ensure that your data is not manipulated in an unauthorized way



HSPH Collaboration Tools: Data Security, Privacy, and Ownership



Collaboration	Tool	Level 1 Data	Level 2 Data	Level 3 Data	Level 4 Data	Level 5 Data
HSPH, HU, external users	Consumer Products (Google Drive, Gmail, DropBox, Evernote, etc.)	✓				
HSPH, HU	Harvard (IT provided) email (jharvard@hspph.harvard.edu)	✓	✓	✓		
HSPH, HU	Harvard Qualtrics or Harvard Canvas	✓	✓	✓		
HSPH, HU, external users	Harvard Dropbox	✓	✓	✓		
HSPH, HU	Harvard Office 365 OneDrive	✓	✓	✓		
HSPH, HU	Harvard Office 365 Share Point (sites)	✓	✓	✓	✓ **	
HSPH	Chan School Network File Share (P: and S: drives)	✓	✓	✓	✓ **	
HSPH, HU, external users	Harvard Amazon Web Services (AWS)	✓	✓ **	✓ **	✓ **	
HSPH, HU, external users (temporary storage)	HSPH Secure File Transfer (Accelion.sph.harvard.edu)	✓	✓	✓	✓	
HSPH	FAS Odyssey Cluster (shared high-performance computing)	✓	✓	✓ **	✓ **	

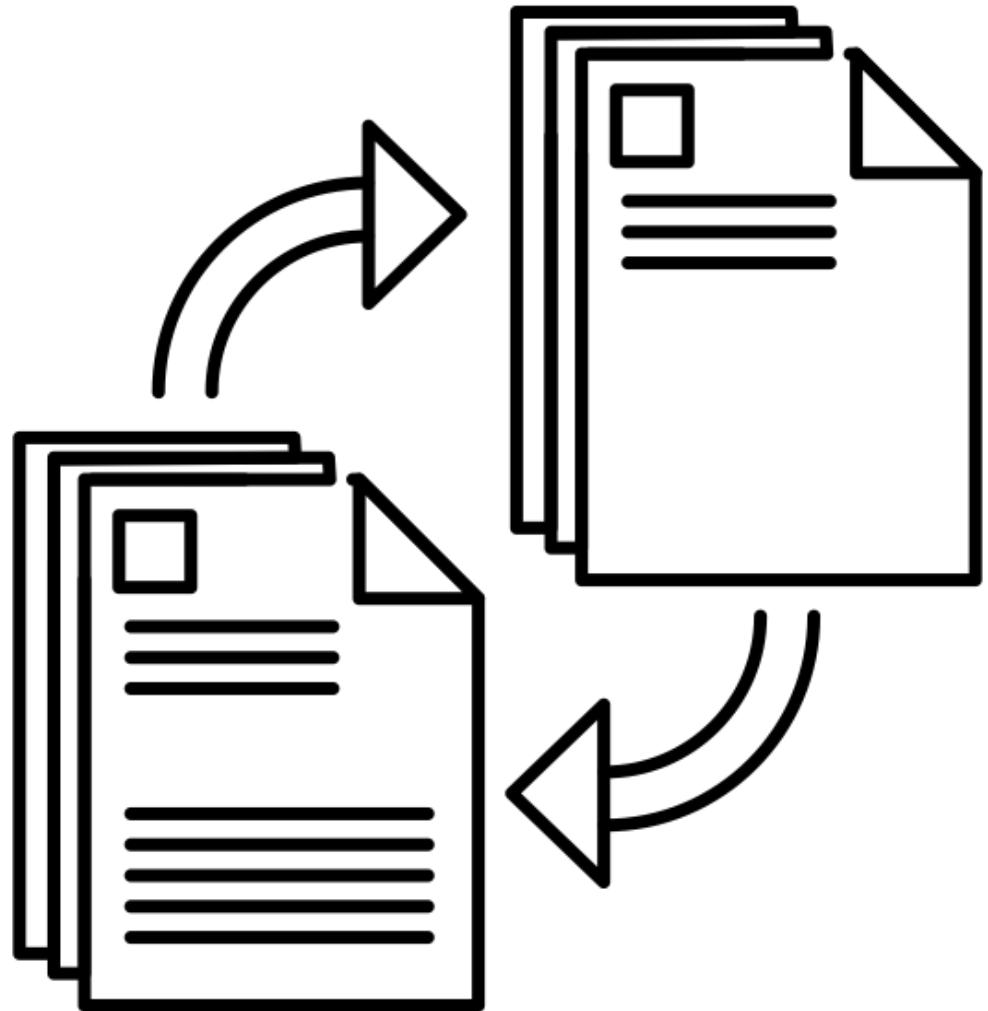
Consumer grade tools and services are **not approved** for Harvard business

** With special controls – contact SPH IT for assistance in setting up appropriate controls

For examples of Level 1-5 data, visit <http://security.harvard.edu/dct>

SHARING

joining a new lab



Created by Flatart
from Noun Project

Data Sharing

When establishing data sharing and access policies and provisions, consider:

- *whom* you will share it with
- *how* it will be shared
- *when* in the research process you will share it

Compliance:

- Funding organizations that require data management plans and data accessibility
- Journals that require submission of supporting data files to accompany manuscripts

But also:

- Find your own data years later
- Enable others to replicate and reuse your work for new analyses

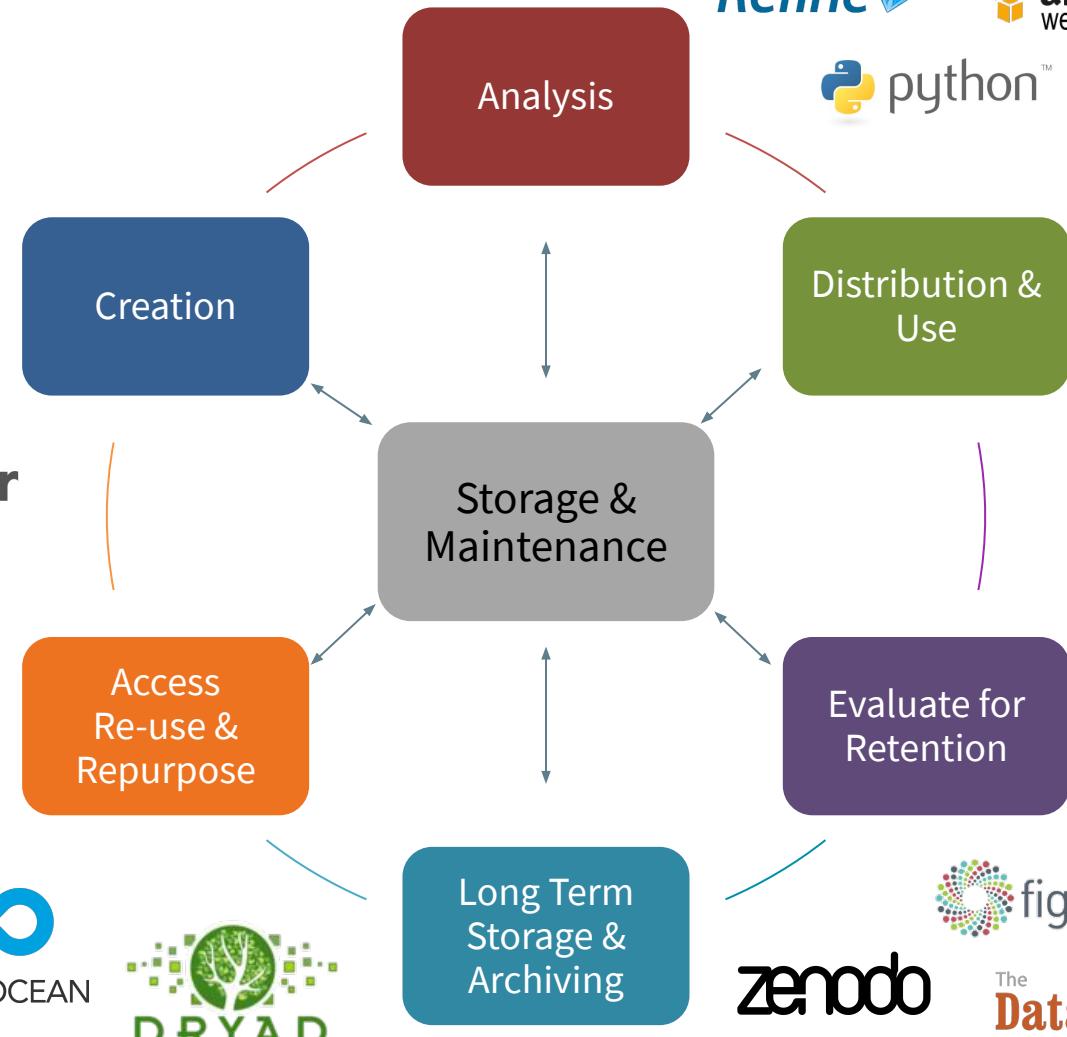


Created by PJ Souders
from Noun Project

Review Available Collaborative Tools

- Dropbox (HMS & HSPH)
 - Offers unrestricted data storage capacity and unlimited version history
 - Before leaving the lab, all files in the Dropbox account should be organized, with applicable metadata and transferred to lab accounts
- HMS Shared (Collaboration) Folders
 - HMS IT centralized file servers are secure and backed up nightly
 - Storing personal and departmental documents on HMS servers protects against data loss
- High Performance Computing
 - HMS: O2 platform for Linux-based HPC
 - HSPH: Odyssey3 based on CentOS 7
- HMS RITS Sharehost
 - Provides a way to host files that users would like to publish to the public HMS RITS Sharehost
- HMS Collaboration Spaces (Atlassian)
 - An interactive wiki-type platform that enables users to easily create a collaboration space

Open & Collaborative Workspaces





Created by PJ Souders
from Noun Project

Review Potential Data Repositories

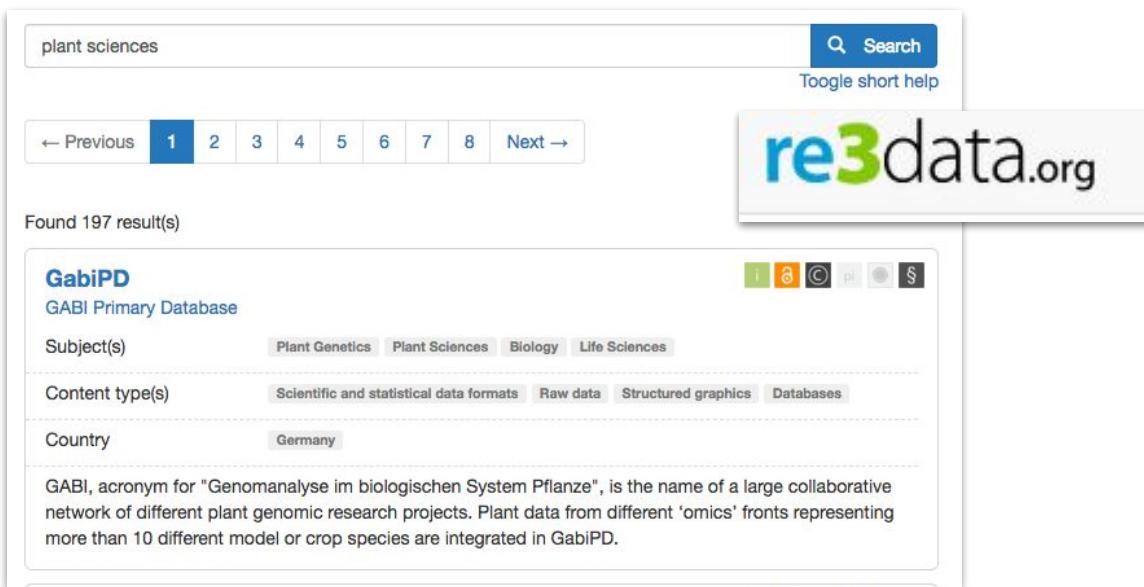
Compare and contrast several of the general data repositories and data publication resources currently available for biomedical science researchers

- Review public data repositories already established for an existing project, or choose a relevant data repository for a new project
- For some projects or scientific areas, established data repositories are not available for the data types being produced – you may need to develop a new data repository for your project

Repositories

- Funder specified repository
- Institutionally specified data repository
- Domain or discipline-specific data repository
- Repository of Research Data Repositories

<https://www.re3data.org>



The screenshot shows a search results page for 'plant sciences' on re3data.org. At the top, there's a search bar with 'plant sciences', a 'Search' button, and a 'Toggle short help' link. Below the search bar is a navigation menu with links for '← Previous', page numbers 1 through 8, and 'Next →'. A large 're3data.org' logo is centered below the search bar. The main content area displays 'Found 197 result(s)'. Below this, there's a detailed search interface with sections for 'GabiPD' (GABI Primary Database), 'Subject(s)' (Plant Genetics, Plant Sciences, Biology, Life Sciences), 'Content type(s)' (Scientific and statistical data formats, Raw data, Structured graphics, Databases), and 'Country' (Germany). A note at the bottom states: 'GABI, acronym for "Genomanalyse im biologischen System Pflanze", is the name of a large collaborative network of different plant genomic research projects. Plant data from different 'omics' fronts representing more than 10 different model or crop species are integrated in GabiPD.'

In addition to a specified data repository, you can make a deposit to a general purpose repository:

- DataDryad <http://datadryad.org>
- Figshare <https://figshare.com>
- Zenodo <https://zenodo.org>



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>

Reagent Management & Sharing

Wasted time, money, resources
when reagents are recreated.

Labs need support to:

- Keep track of reagents created
- Consistently validate all reagents in the lab
- Properly label and store reagents
- (Legally) distribute all reagents to interested researchers

Reagent repositories help solve all of these logistical problems:

- Addgene <https://www.addgene.org>
- CiteAb <https://www.citeab.com>
- Quartzy <https://www.quartzy.com>
- Resource Identification Portal <https://scicrunch.org/resources>





Created by PJ Souders
from Noun Project

Review Electronic Lab Notebook (ELN) Resources

- An ELN (electronic lab notebook) is a software tool that replicates an interface much like a page in a paper lab notebook
- Allow users to enter protocols, observations, notes, and other data using a computer or mobile device
- Considerations:
 - Cost
 - Storage
 - Interactivity
 - Software
 - Security
 - Migration (in & out)

Electronic Lab Notebooks

Benefits:

- Searchable
- Export data
- Easily shareable
- Easily accessible
- Use the mobile App
- Embed images, protocols, etc



Marta Teperek
@martateperek

Follow

Trying to find information in colleagues' hard copy lab books is a task for a detective ...
@HyvonenGroup #elncam

Notebook no.	Date 13 January 2017	101
Title Work for Sherlock Holmes		Continued from page

Marko Hyvonen
mh256@cam.ac.uk

5:33 AM - 13 Jan 2017

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>

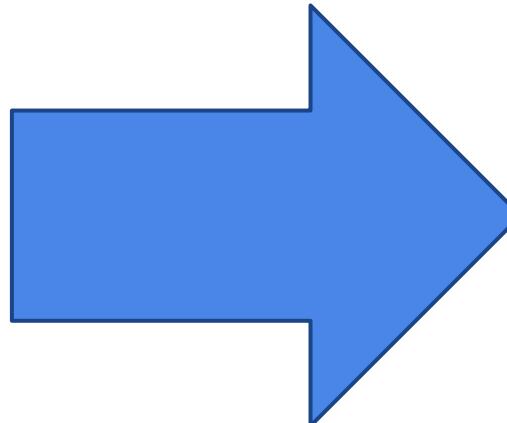
CASE STUDY

joining a new lab



Created by Creative Stall
from Noun Project

Case Study: Louisa Joins the Smith Lab



Janet Smith

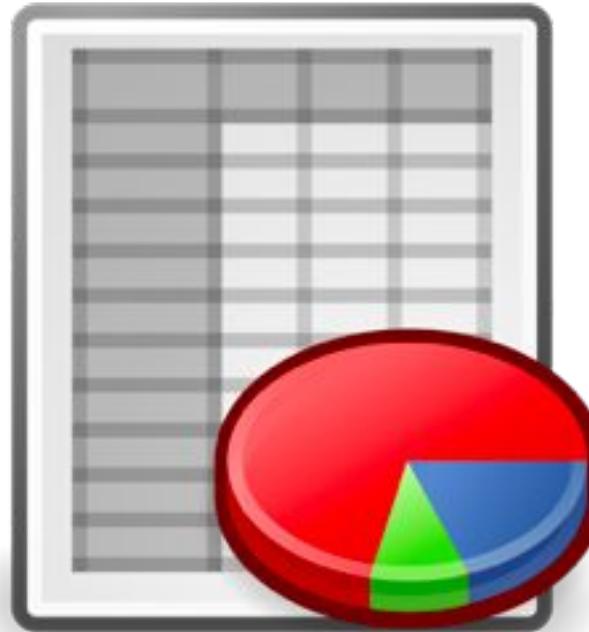
- Recently joined Harvard Medical School
- Establishment of lab protocols and procedures





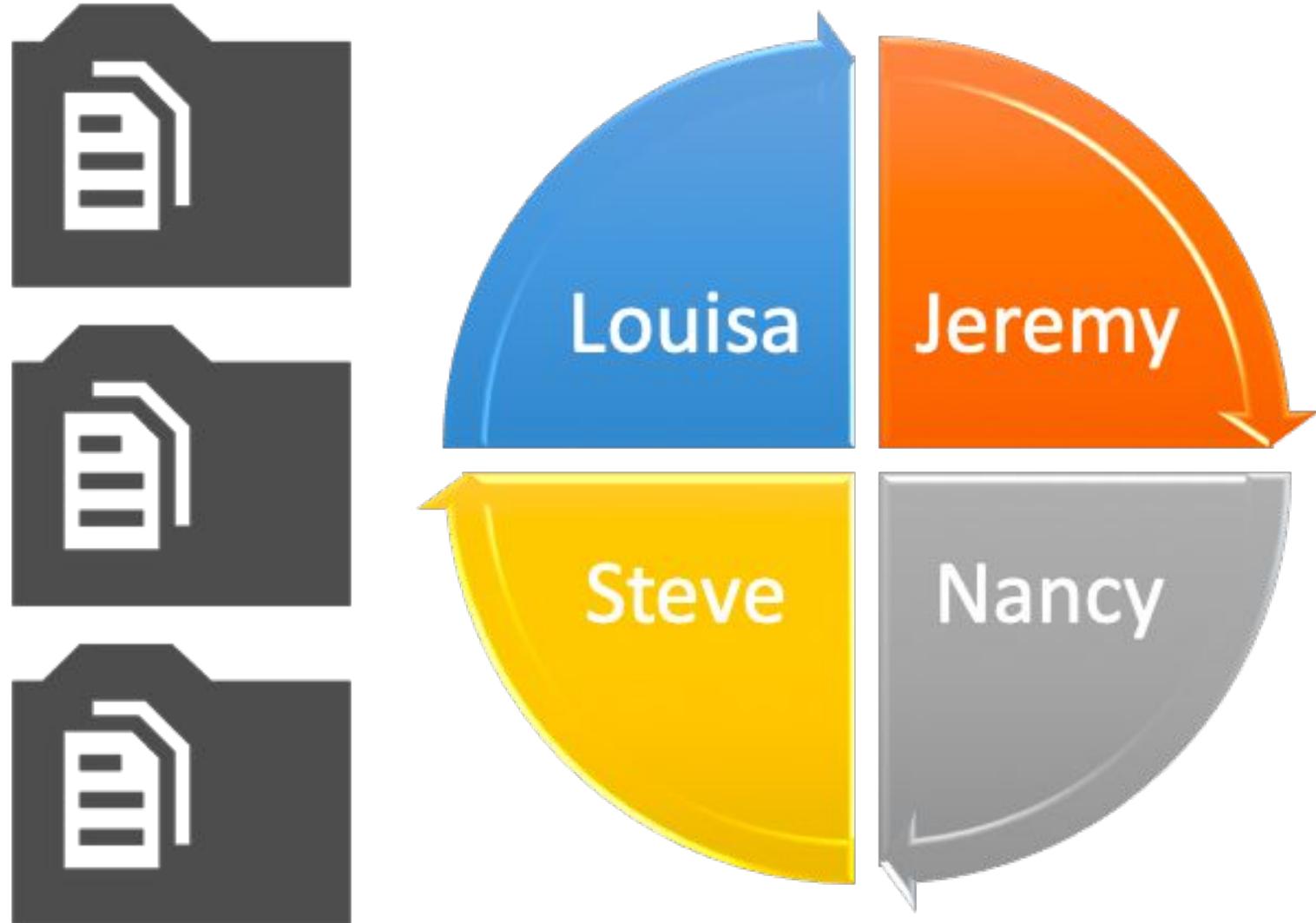
Storage

- Raw data is stored on a local hard drive
- Backed up every two weeks
- Hard Drive reaching capacity
- Louisa needs the data to run her analyses
- Data duplication



Collaboration

- Four lab members
- Separate organizational workflows
- Missing file information
- Electronic Lab Notebooks



Case Study: Questions

1. What are some of the issues that you noticed happening within this case study?
2. What are some research data management problems that you can identify?
3. How would some of these problems be solved?
4. What could Louisa have done when entering the lab to ensure a smoother transition?



Case Study: Issues

- Lack of established protocols and procedures within the lab and little oversight from PI
- No onboarding documentation provided and no place for centralized procedures or protocols
- Unsure of the existing policies at HMS
- Unaware of the grant funding requirements
- Lack of a data management plan
- Storing data on a local hard drive causing issues with file size and collaboration
- Only backed up to an external hard drive every two weeks
- Running out of available storage due to a lack of organization, including duplication
 - Thinking of deleting data to make more space available, but the data is needed for the project
 - No established file naming conventions or organized file structure, no systematic way to locate documentation
- No ReadMe files including metadata about file naming conventions, instruments, software & analyses
- Members using their own lab notebook solution leading to issues with sharing and collaboration



Data Sharing and Management Snafu in 3 Short Acts

Why should you be concerned with managing your data?

1. Enable continuity of research projects

- Easier for data producer, PI, and collaborators to find data
- If you revisit the data in six months, will you be able to remember all of the details? the parameters? Test runs vs actual experiment

2. Facilitate data sharing and re-use

- Increases visibility of research

3. Reduce research and data storage costs

- Minimize storage of duplicate files, increase ability to re-use data

4. Promote data reproducibility

- Well-documented reagents, protocols, & datasets are available
- Granting agencies and journals are increasingly requiring this documentation, including formalized data management plans

How can we help?

PLANNING	STORAGE	ANALYZE	SHARING	ARCHIVE	REUSE
Data Management Plans (DMPs) consultation, review, training	Electronic lab notebook support Metadata and data documentation services File organization and asset management training	Referrals to IT, Research computing, and research support services Promotion of research support services Comparisons of data cleaning, analysis & visualization tools	Data repository administration Referrals to specialized subject data repositories Data sharing and publishing training (e.g. DOIs, data citations) Data use agreements (DUAs)	Archival processing & appraisal Digital preservation and stewardship services Digital preservation and stewardship	Locate data for new project/data discovery Best practices for data reuse training RDM lifecycle training
Referrals to IT & Research Computing					
Navigation of institutional RDM services and resources					

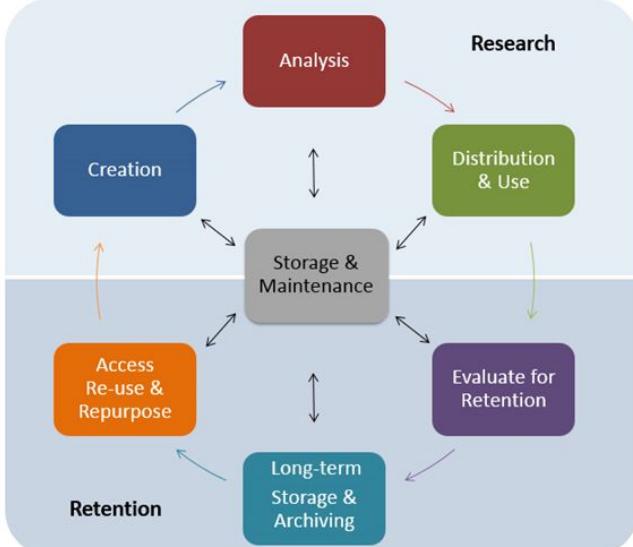
Questions?

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 APR 11 Data Management for Labs: How to Hit the Ground Running

2019 MAY 02 Data Management Working Group Monthly Meeting

2019 MAY 07 Getting Started with Data Management Plans

[More ▶](#)

FEATURED NEWS



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

Upcoming Seminars

Upcoming Summer Seminars:

Version Control with Git
Introduction to Bash
High Performance Computing

datamanagement.hms.harvard.edu

Get Upcoming Class Alerts:

Subscribe to the DMWG quarterly newsletter and monthly class announcements!

datamanagement.hms.harvard.edu/dmwg-newsletter

bit.ly/rdm-survey

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