Research Data Management Tools and Resources

HARVARD MEDICAL SCHOOL

Sarah Marchese, Research Data Management
Harvard Medical School, IT Research Computing

Agenda

- Who We Are
- What is RDM?
- Recommendations
- Resources
- Support
- Questions

Research Data Management (RDM) Team

Collaborate with researchers to better organize, manage, and store research data throughout the data lifecycle

- Advance and refine processes to move infrequently accessed data to long term storage
- Develop automated methods for transferring data between storage platforms
- Provide a better understanding of how to sustainably manage large scale data storage
- Provide new product and technology recommendations based on researcher needs.
- Create and maintain data management tools and resources to prepare data for sharing and reuse

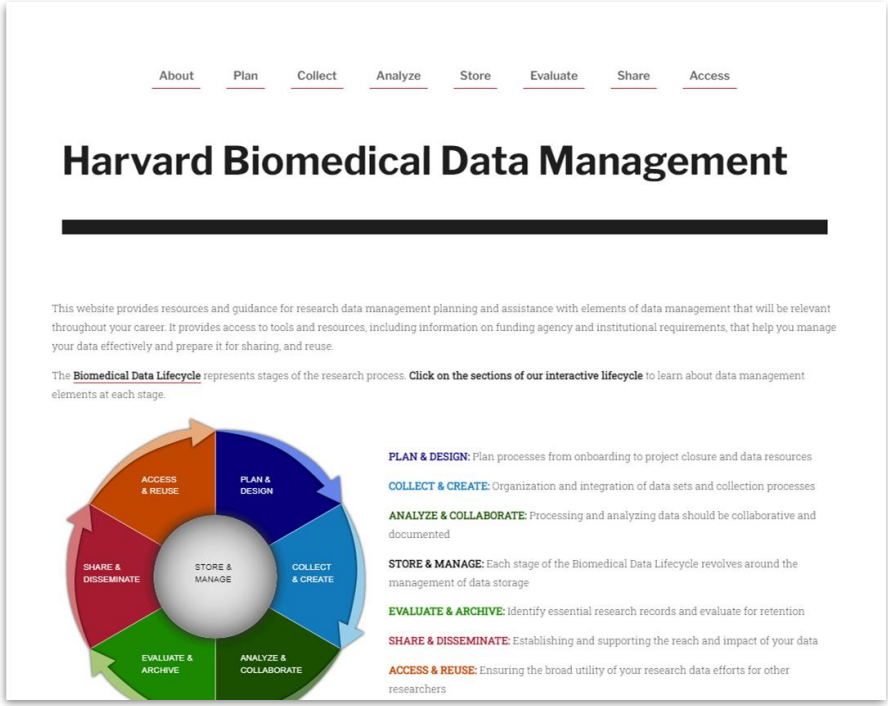
Jessica Pierce, Research Data Manager

Sarah Marchese, Senior Research Data Management Analyst

Danielle Brown, Research Data Management Support Specialist

Research Data Management Working Group

- The Harvard Longwood Medical Area Research Data Management Working Group (LMA RDMWG):
  - Provides guidance, resources and solutions
  - Develops recommendations to meet current and future data management needs
  - Offers a variety of expertise including the management of high-throughput screening and image data, research computing, educational programming, and library sciences.



https://datamanagement.hms.harvard.edu

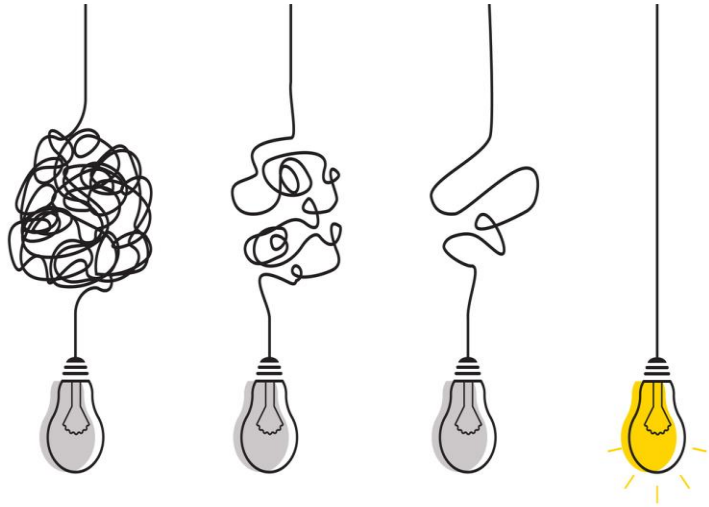
Research Data Management

"The **active and ongoing** management of data **through its lifecycle** of interest and usefulness to scholarship, science, and education."

— *The University of Illinois' Graduate School of Library and Information Science*

Why is Data Management Important?

- Essential for responsible research
- Creates lab standards for data collection, storage and sharing
- Enhances research production and consistency
- Encourages reproducibility and open science
- Prevents data from being lost or deleted
- More efficient collaborations
- Required by funding agencies and publishers



RDM Recommended Practices

1. File organization and naming techniques
2. Providing context with documentation
3. Proper storage and data security
4. Data sharing strategies
5. Data management planning

Data Management Planning

- How do you intend to gather the data?
- How will you be analyzing the data?
- Where should the data be stored?
- How will the data be used?
- Who needs access to the data? Is the data restricted?
- Do you intend to share the data? Will it be open access?

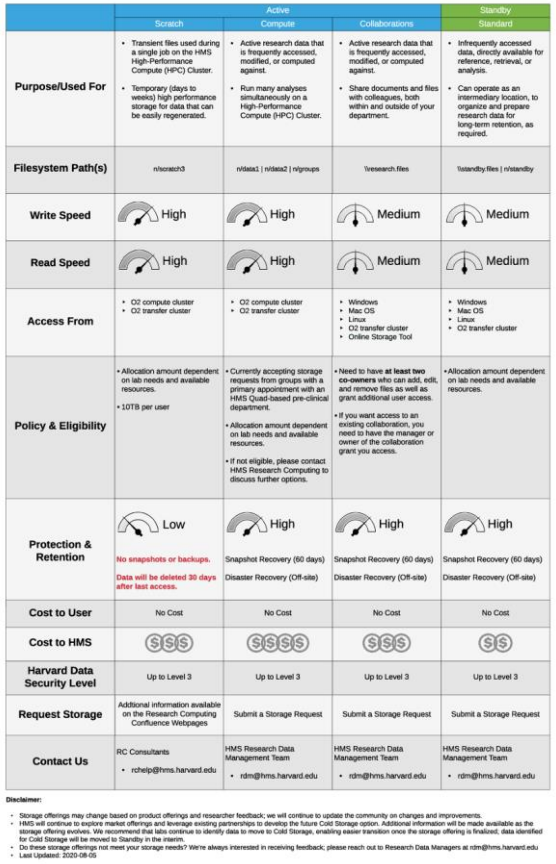
Data Management Plans

- Written document outlining plans for handling all of the data resulting from a research project
- Detailed procedures for data collection, organization and processing
- Plan for when data leaves your lab
- Frequently referred to & updated

NIH Policy for Data Management and Sharing

- The NIH Data Management and Sharing Policy is effective as of January 25th
- Replaced the 2003 NIH Data Sharing Policy
- Designed to promote positive change in data management and sharing culture
- Costs associated with data management and data sharing are allowable
- Plans can be revised throughout the project
- Plans should not include proprietary or private information
- Plan should be two pages or less
- Practices should be consistent with FAIR (findable, accessible, interoperable, reusable) data principles

Storage Offerings



	Active			Standby
	Scratch	Compute	Collaborations	Standard
Purpose/Used For	• Transient files used during a single job on the HMS High-Performance Compute (HPC) Cluster. • Temporary (days to weeks) high performance storage for data that can be easily regenerated.	• Active research data that is frequently accessed, modified, or computed against. • Run many analyses simultaneously on a High-Performance Compute (HPC) Cluster.	• Active research data that is frequently accessed, modified, or computed against. • Share documents and files with colleagues, both within and outside of your department.	• Infrequently accessed data, directly available for reference, retrieval, or analysis. • Can operate as an intermediary location, to organize and prepare research data for long-term retention, as required.
Filesystem Path(s)	n/scratch3	n/data1 | n/data2 | n/groups	%research.files	%standby.files | n/standby
Write Speed	High	High	Medium	Medium
Read Speed	High	High	Medium	Medium
Access From	• O2 compute cluster • O2 transfer cluster	• O2 compute cluster • O2 transfer cluster	• Windows • Mac OS • Linux • O2 transfer cluster • Online Storage Tool	• Windows • Mac OS • Linux • O2 transfer cluster
Policy & Eligibility	• Allocation amount dependent on lab needs and available resources. • 10TB per user	• Currently accepting storage requests from groups with a primary appointment with an HMS Quad-based pre-clinical department. • Allocation amount dependent on lab needs and available resources. • If not eligible, please contact HMS Research Computing to discuss further options.	• Need to have **at least two co-owners** who can add, edit, and remove files as well as grant additional user access. • If you want access to an existing collaboration, you need to have the manager or owner of the collaboration grant you access.	• Allocation amount dependent on lab needs and available resources.
Protection & Retention	Low No snapshots or backups. Data will be deleted 30 days after last access.	High Snapshot Recovery (60 days) Disaster Recovery (Off-site)	High Snapshot Recovery (60 days) Disaster Recovery (Off-site)	High Snapshot Recovery (60 days) Disaster Recovery (Off-site)
Cost to User	No Cost	No Cost	No Cost	No Cost
Cost to HMS	$$$	$$$$	$$$	$$
Harvard Data Security Level	Up to Level 3	Up to Level 3	Up to Level 3	Up to Level 3
Request Storage	Additional information available on the Research Computing Confluence Webpages	Submit a Storage Request	Submit a Storage Request	Submit a Storage Request
Contact Us	RC Consultants • rchelp@hms.harvard.edu	HMS Research Data Management Team • rdm@hms.harvard.edu	HMS Research Data Management Team • rdm@hms.harvard.edu	HMS Research Data Management Team • rdm@hms.harvard.edu

Disclaimer:

- Storage offerings may change based on product offerings and researcher feedback; we will continue to update the community on changes and improvements.
- HMS will continue to explore market offerings and leverage existing partnerships to develop the future Cold Storage option. Additional information will be made available as the storage offering evolves. We recommend that labs continue to identify data to move to Cold Storage, enabling easier transition once the storage offering is finalized; data identified for Cold Storage will be moved to Standby in the interim.
- Do these storage offerings not meet your storage needs? We're always interested in receiving feedback; please reach out to Research Data Managers at rdm@hms.harvard.edu
- Last Updated: 2020-08-05

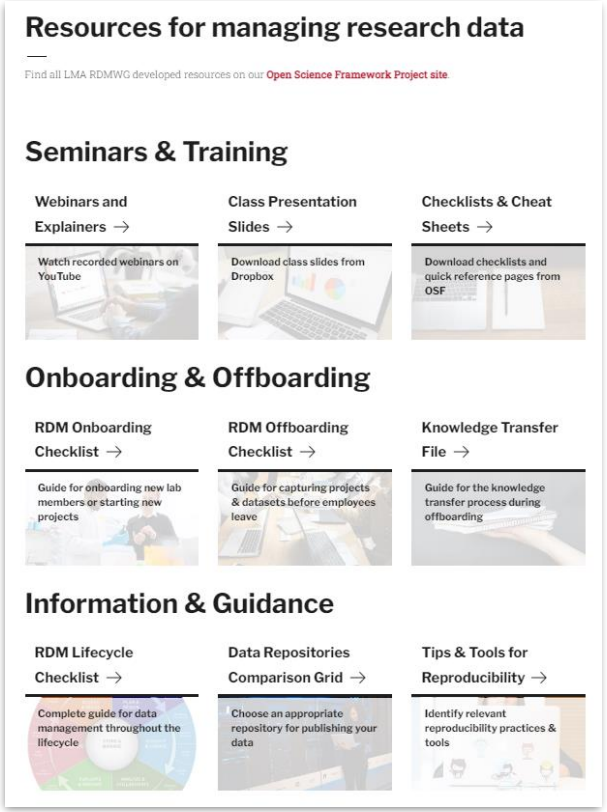
- HMS offers several storage options that allow users to store data in different places, each with distinct behaviors, performance, and means of access
- **Active**
  - Compute (O2)
  - Collaborations (research.files)
- **Standby**
- **Cold** (New!)

Data Repositories

- Repositories provide technical infrastructure to provide access and curation of research data
- Provide a persistent identifier and a citation for your data
- Provide access controls
- Are compliant with funders and journals requirements

Research Data Management Resources

- Guidance & recommended practices for the data lifecycle
- Research policies & requirements
- Data services across the LMA
- News & blog posts
- Live training sessions (virtual)
- Recorded video tutorials



about/what-research-data-management/rdm-resources

Research Data Management Resources

- HMS IT Research Computing Website
- Harvard Biomedical Research Data Management
- Harvard Biomedical Research Data Management Resources
- Countway Library Data Services
- Harvard Medical School Office of Research Administration
- Harvard Medical School Information Security

Questions/Discussion

Sarah Marchese, Research Data Management
Harvard Medical School IT Research Computing
Email: rdmhelp@hms.harvard.edu
https://datamanagement.hms.harvard.edu/