

Coordinating Data Services in the Lab: Connecting The Research Lifecycle

*Medical Library Association
2022 Annual Meeting*



Julie Goldman (she/her)
julie_goldman@harvard.edu

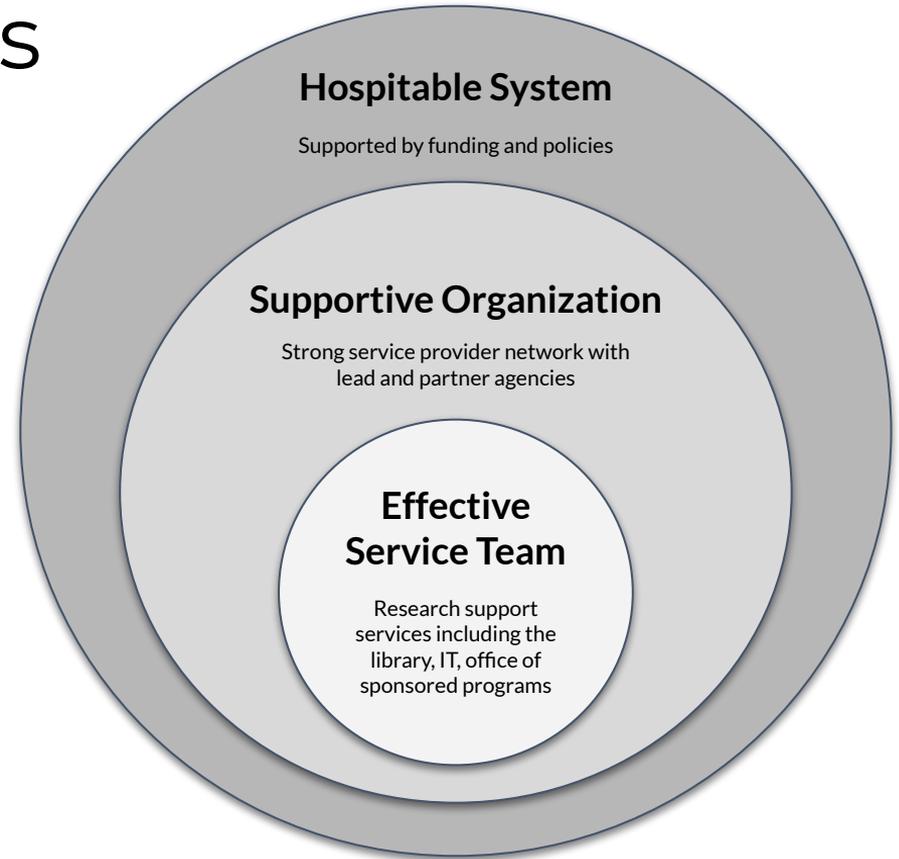


Slides in Zenodo
[10.5281/zenodo.6462246](https://zenodo.org/doi/10.5281/zenodo.6462246)

Coordinated Services

Three Levels of Supports:

- Effective Service Team
 - Library, IT, Office of Research, Research Labs & Cores, Archives
- Supportive Organization
 - Strong service network with open communication avenues
- Hospitable System
 - Funding, Policies



Service Partners Across the Lifecycle



[Research Data Lifecycle by LMA RDMWG](#)

Service Partners Across the Lifecycle



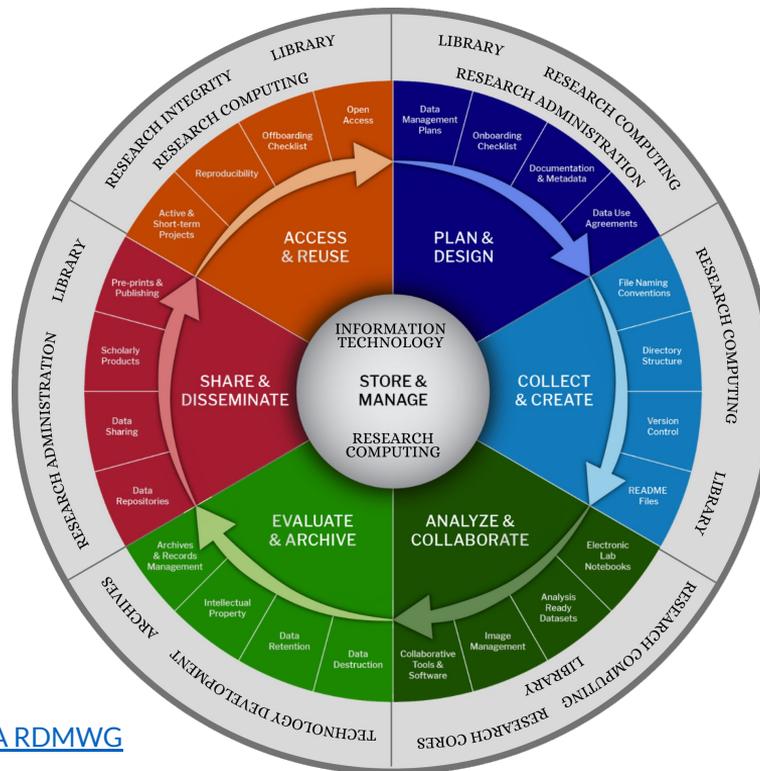
[Research Data Lifecycle by LMA RDMWG](#)

Service Partners Across the Lifecycle

DATA SHARING PRACTICES

Library assists with curating, sharing, and depositing data in a repository

Research Administration ensures compliance with regulations & policies



DATA COLLECTION PRACTICES

Library helps with data organization and documentation practices

Research Computing assists with storage of data files and offers classes on HPC



[Research Data Lifecycle by LMA RDMWG](#)

Service Offering: Tailored Resources

Why Tailored Resources?

- Captures employee institutional (or project) knowledge
- Documents essential information related to projects and datasets to enhance research production and consistency
- Reduces the amount of wasted time spent by others in the lab and duplication or unnecessary retention of immaterial data

Why Onboarding & Offboarding?

- Traditional employee orientations do not discuss data
- Data accessibility and findability are hindered with improper offboarding
- An employee's departure can negatively impact segments of the data lifecycle
- Creates lab standards for data collection and storage
- Motivating researchers and PIs is easier with prevailing policies and procedures

RESEARCH DATA MANAGEMENT ONBOARDING CHECKLIST: ABRIDGED VERSION

RESEARCH DATA MANAGEMENT ONBOARDING CHECKLIST: ABRIDGED VERSION
Employee/trainee onboarding to new labs and projects

This document serves as a general research data management-focused guide for employees/trainees 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. Internal and external links have been provided throughout the document as supplementary resources.

CHECKLIST FOR JOINING A NEW LAB

PLANNING

<input type="checkbox"/> Review Data Management Policies:	<input type="checkbox"/> HMS Research Data Management Guidelines	<input type="checkbox"/> Harvard Biomedical Data Management
	<input type="checkbox"/> Harvard Retention and Maintenance of Research Records and Data Principles	<input type="checkbox"/> Retention and Management of Research Records and Data Principles and FAADs
	<input type="checkbox"/> Longwood Campus Archives and Records Management Program	<input type="checkbox"/> Longwood Campus Archives and Records Management
	<input type="checkbox"/> Harvard Research Data Security Policy	<input type="checkbox"/> Harvard Research Data Security Policy
	<input type="checkbox"/> Harvard Information Security Policy	<input type="checkbox"/> Harvard Information Security Policy
<input type="checkbox"/> Create a Preliminary Data Workflow:	<input type="checkbox"/> Review existing lab workflows, directory structures, and metadata standards	
	<input type="checkbox"/> Develop a preliminary organizational workflow for your research, including a file (or directory) structure.	<input type="checkbox"/> Best Practices: Directory Structure
		<input type="checkbox"/> Best Practices: File Naming Conventions
<input type="checkbox"/> Create Preliminary README Files for Each Dataset:	<input type="checkbox"/> Document your data workflow(s) in a README file. The metadata you record will help ensure your data are	<input type="checkbox"/> Best Practices: README Files

Name: _____
Position/Title: _____
Department: _____
Lab/Group: _____
Dates Affiliates: _____
Date Submitted: _____

Research Data Management Onboarding Checklist: Abridged Version by [Harvard Biomedical Data Management Working Group](#) is licensed under a [Creative Commons Attribution 4.0 International License](#). [View our open access digital copy on the Harvard Biomedical Data Management website](#) or [data-management@hsb.harvard.edu](#) (PDF) (11/20/2016)

Overall guidance: In this document and accompanying tables, please document who has access to the data and lab materials that will remain in the lab. Annotate the data appropriately, referencing additional sources, if needed. If specific data files were included as part of a scientific publication, please cite the publication, with a DOI or PubMed ID, if possible. If proprietary data remains inaccessible to collaborators, ensure the data is converted to an open or common file format (e.g., .csv, .txt) and exported to a shared location, prior to your departure.

Please complete all relevant fields below.

Category:	Question:	Response:
1. Lab Notebooks, Collaborators, and Data Storage Locations:	Properly protecting research data and materials is a fundamental obligation, grounded in the values of stewardship, integrity, and commitments to the providers and sources of the data and reagents. Additional information about data management can be found on the Harvard Biomedical Research Data Management website.	
	1.1 Was your work grant funded? If so, provide grant information such as funding agency, grant title, dates, or grant number, if available.	
	1.2 List collaborators who directly contributed to datasets and materials remaining in the lab. Include name and email address(es) for each individual.	
	1.3 Will the project remain ongoing following your departure? If so, who will continue the project and have access to the data? Indicate names and email address(es) for each individual and what data they will have access to.	
	1.4 Did you utilize an electronic lab notebook	

Service Offering: Tailored Resources

Why Tailored Resources?

- Captures employee needs
- Documenting datasets
- Reducing and organizing

The transition to the new checklist has been seamless, and feedback from lab members on the new forms has been overwhelmingly positive.

Jessica Weber

Lab Data Manager, HMS Church Lab

Why Onboard?

- Traditional
- Data and offboarding
- An employee the data lifecycle
- Creates lab standards re: storage
- Motivating researchers and PIs is easier with prevailing policies and procedures

RESEARCH DATA MANAGEMENT ONBOARDING CHECKLIST: ABBRIDGED VERSION

RESEARCH DATA MANAGEMENT ONBOARDING CHECKLIST: ABBRIDGED VERSION
Employee/trainee onboarding to new labs and projects

This document serves as a general, research data management-focused guide for 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. Internal and external links have been provided throughout the document as supplementary resources.

CHECKLIST FOR JOINING A NEW LAB

PLANNING

<input type="checkbox"/> Review Data Management Policies:	<ul style="list-style-type: none"> <input type="checkbox"/> HMS Research Data Management Guidelines <input type="checkbox"/> Harvard Retention and Maintenance of Research Records and Data Principles <input type="checkbox"/> Longwood Campus Archives and Records Management Program <input type="checkbox"/> Harvard Research Data Security Policy <input type="checkbox"/> Harvard Information Security Policy 	<ul style="list-style-type: none"> <input type="checkbox"/> Harvard Biomedical Data Management <input type="checkbox"/> Retention and Maintenance of Research Records and Data Principles and FAAD <input type="checkbox"/> Longwood Campus Archives and Records Management <input type="checkbox"/> Harvard Research Data Security Policy <input type="checkbox"/> Harvard Information Security Policy
<input type="checkbox"/> Create a Preliminary Data Workflows:	<ul style="list-style-type: none"> <input type="checkbox"/> Review existing lab workflows, directory structures, and metadata standards <input type="checkbox"/> Develop a preliminary organizational workflow for your research, including a file (or directory) structure: 	<ul style="list-style-type: none"> <input type="checkbox"/> Best Practices: Directory Structure <input type="checkbox"/> Best Practices: File Naming Conventions
<input type="checkbox"/> Create Preliminary README Files for Each Dataset:	<ul style="list-style-type: none"> <input type="checkbox"/> Document your data workflow(s) in a README file. The metadata you record will help ensure your data are 	<ul style="list-style-type: none"> <input type="checkbox"/> Best Practices: README Files

Name: _____
Position/Title: _____
Department: _____
Lab/Group: _____
Dates Affiliated: _____
Dates Submitted: _____

Research Data Management Onboarding Checklist: Abridged Version by Jessica L. Campbell, Medical Area Research Data Management Working Group in partnership with Harvard University's Institutional Repository and the Harvard Biomedical Research Data Management website. © 2014. <https://www.harvard.edu/bioinformatics>

Overall guidance: In this document and accompanying tables, please document what you have access to the data and lab materials that will remain in the lab. Annotate the data appropriately, referencing additional sources, if needed. If specific data files were included as part of a scientific publication, please cite the publication, with a DOI or PubMed ID, if possible. If proprietary data remains inaccessible to collaborators, ensure the data is converted to an open or common file format (e.g., .csv, .txt) and exported to a shared location, prior to your departure.

Please complete all relevant fields below.

Category:	Question:	Response:
1. Lab Notebooks, Collaborators, and Data Storage Locations:	Properly protecting research data and materials is a fundamental obligation, grounded in the values of stewardship, integrity, and commitments to the providers and sources of the data and reagents. Additional information about data management can be found on the Harvard Biomedical Research Data Management website.	
	1.1 Was your work grant funded? If so, provide grant information such as funding agency, grant title, dates, or grant number, if available.	
	1.2 List collaborators who directly contributed to datasets and materials remaining in the lab. Include name and email address(es) for each individual.	
	1.3 Will the project remain ongoing following your departure? If so, who will continue the project and have access to the data? Indicate names and email address(es) for each individual and what data they will have access to.	
	1.4 Did you utilize an electronic lab notebook	

Planning Ahead: NIH DMSP

Cohesive and coordinated messaging:

- Created institutionally supported templates in **DMPTool** with organization-specific guidance, resources, and examples
- Coordinating Harvard University and also Harvard Medical School responses to the 2023 forthcoming **NIH Policy for Data Management and Sharing**
- While we communicate the seriousness of this policy, we are also **leveraging existing resources and expertise** to lessen the burden to this culture changing policy

This plan is based on the "Harvard Medical School Biomedical DMP Template" template provided by Harvard University - (ver: 5, pub: 2021-10-25).

expand all | collapse all 0/14

+ Data Collection (0 / 2)

Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.

What data will you collect or create?

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

B / I / [] / [] / [] / []

Save

Harvard example answer

Below are sample texts that can be modified (as relevant) to answer the question

Data Formats for Harvard Dataverse	Immediately after collection, quantitative data will be converted to Stata, SPSS, R, Excel, CSV formats. These formats are fully supported by the Harvard Dataverse, which will perform archival format migration; metadata extraction; and validity checks. Deposit in these formats will also enable online analysis; variable-level search; data extraction and re-formatting; and other enhanced access capabilities. Documentation will be deposited in PDFs, or plain-text formats, to ensure long-term accessibility, with any accompanying sound (in WAV), video, or images separate from the documentation deposited as JPEG 2000 files (with lossless compression) or uncompressed TIFF files.
General	This project will produce primarily scientific data, on an ongoing basis, that will consist of reported publications and intellectual property rights filings. Data types include (1) experimental procedures utilized to obtain data; (2) materials utilized during experimental procedures; (3) methods used for data measurements; (4) data measurements and analysis and DNA and protein sequences; (5) publications in the form of scientific manuscripts; and, (6) patents.
General	We will be working with [sequencing/imaging/mass spectrometry/structure/etc.] data. This data is to be generated [in-house/by a core facility/etc.], and we expect [N] datasets totaling at approximately [size] [units] total.

Guidance Comments

Harvard DMPTool

- Harvard Biomedical Data Management: Data Management Plans
- Webinar Video: Let's Talk Data

expand all | collapse all

Data format +

Find support for working with your data: Research Data and Scholarship

- When saving data, it is important to not only consider where you are saving it but also how you are saving.
- As you shift from saving your data as you work on it to saving it to preserve it long term, something to consider is your file formats. When saving data over the long term, we recommend saving in formats that are as open, lossless, and unencrypted as possible.
- Open (non-proprietary) formats are those that can be used and implemented by anyone. In practice, this means that files stored in open formats can be opened and used by a variety of proprietary, free, and open-source software tools rather than just a single piece of software. Open, non-proprietary, formats are far

Coordination & Collaboration is Key

Universities need to collaborate between research, libraries, and IT:

- Catalog all the services offered to support research data
- Centralize costly and complex services
- Build workflows to integrate services, research tools, computing, and repositories
- Offer resources and expertise to meet researcher's challenges and needs



All resources created by the LMA Research Data Management Working Group are available in [Open Science Framework](#)

Coordinating Data Services in the Lab: Connecting The Research Lifecycle

*Medical Library Association
2022 Annual Meeting*



Julie Goldman (she/her)
julie_goldman@harvard.edu



Slides in Zenodo
[10.5281/zenodo.6462246](https://doi.org/10.5281/zenodo.6462246)