

RESEARCH DATA MANAGEMENT ONBOARDING CHECKLIST

This document serves as a general research data management-focused guide to employee/trainee onboarding as they join a new lab or begin a new project. The document provides two checklists: Checklist for joining a new lab and Checklist for Joining a New Project. Follow one or both of these checklists as they apply to your situation. For more specific information, please see other resources provided by Research Data Services at UW-Madison. Internal and external links are provided throughout the document as supplementary resources and to point users to other expert groups on campus, including a glossary of terms.

This checklist is adapted from the Research Data Management Onboarding Checklist by Harvard Medical School Data Management Working Group is licensed under a Creative Commons Attribution 4.0 International License.

CHECKLIST FOR JOINING A NEW LAB

PLANNING

1. Review Laboratory, Department, and University Data Management Policies

Policies and Procedures

Additional Resources

Contact the PI and department Research Administrator for laboratory and department-specific data management policies

UW-Madison Research Data Management Best Practices

UW-Madison Data Security, Management and Retention Guidelines (these apply when collecting and managing sensitive data)

UW System Information Security: Data Classification Policy

[UW-Madison Research Data Services: Learn About Data Management](#)

[Introduction to Research Data Management Micro-Course](#)

[Responsible Data Planning, Use, and Sharing Micro-Course](#)

[Data Security, Management, and Retention](#)

[UW System Information Security: Data Classification](#)

2. Create a Preliminary Data Workflow

Review existing lab workflows, directory structures, and documentation standards

Develop a preliminary organizational workflow for your research project, 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

[Best Practices for Data Organization, including folder structure and file naming](#)

STORAGE

1. Review Storage Options

Determine the best storage and backup options for your project or lab based on your data needs

[Best Practices for Data Storage and Backup](#)

		<i>UW-Madison storage infrastructure will be changing in the near future, and RDS will be updating resources related to storage and backup as the changes are rolled out.</i>	
SHARING			
1. Review Available Collaborative Tools			
		UW G Suite	Best Practices for Data Storage and Backup
		<i>Google Drive provides a cloud-based solution to collaboration and unlimited storage to UW-Madison faculty, staff, researchers, and students</i>	
		UW-Madison Box	UW-Madison Box
		<i>Box is a cloud solution for storing, managing, and sharing files, but should only be used for regularly accessed, working files</i>	
		Center for High Throughput Computing	Center for High Throughput Computing
		<i>Provides scalable computing resources and services</i>	
2. Review Potential Data Repositories			
		Compare several general or discipline-specific data repositories and data resources	UW-Madison Research Data Services Data Sharing Essentials
3. Review Electronic Lab Notebook Resources			
		LabArchives	
		<i>Software with an interface that replicates the pages of a lab notebook, and allows users to enter protocols, observations, notes, and other data using a computer or mobile device</i>	UW-Madison Instance of LabArchives

CHECKLIST FOR STARTING OR JOINING A NEW PROJECT

PLANNING

1. Transfer Prior Data and Related Records

Contact the Office of the Vice Chancellor for Research and Graduate Education

[Office of the Vice Chancellor for Research and Graduate Education](#)

2. Write a Data Management Plan or Review Existing Data Management Plan

Construct a Data Management Plan (DMP)

[Research Data Services: How to Create a DMP](#)

A DMP is a brief but comprehensive document that describes how you will manage the data collected and generated during a research project. It should provide information about the types and formats of the data being collected, how the data will be documented, how the data will be shared and accessed during the project by those involved, and if it will be made publicly accessible at the end of the project, how it will be stored and backed up throughout the project, how it will be archived and preserved at the end of the project, and the roles and responsibilities of the researchers and assistants involved in the project.

Creating a DMP is increasingly required when applying for funding

Following a DMP can assist with creating standards and workflows within labs and research groups

Many publishers and funders have requirements about data management and sharing that you will need to take into account when creating your DMP.

3. Create a Data Workflow or Review Existing Data Workflow

Review existing project workflows and directory structures, OR

Develop a new organizational workflow, including an established file/directory structure

Following a data management workflow can reduce file redundancy, saving storage space for your lab

Consider using an Electronic Lab Notebook, which allows users to enter protocols, observations, notes, and other data

[UW-Madison Instance of LabArchives](#)

4. Establish a Metadata Standard or Review Existing Project Metadata Standards

Metadata is a way of describing your data, and is often created in a format that is machine-readable, and also understood by humans.

Determine where metadata files are stored/should be stored for the project so that they are appropriately linked to the data that they describe

Determine whether a metadata standard is already in place for the project or whether a new standard will need to be established

	If establishing the use of a new standard in your research group, review existing standards that would be relevant to the project or that are commonly used in your field	RDS Guide to Metadata
	Establish a controlled vocabulary or adapt existing controlled vocabulary for the project	Guide to Controlled Vocabulary from Jisc
	<i>Well documented data facilitates better understanding, use, and sharing of experimental data now and in the future, helping researchers discover, access, use, reuse, and cite data. It can also serve as a guide for your own research group.</i>	
STORAGE		
1. Review Storage Options		
	Review storage resources that are already in place for the existing project, or choose relevant storage options for a new project	
	Storage and Backup Options for UW-Madison researchers, staff, faculty, and students	Storage and Backup Recommendations
	<i>It is also recommended to meet with the your department's IT staff about your storage and backup needs</i>	
SHARING		
1. Review Collaborative Tools Available		
	Review collaborative tools already in use for an existing project, or choose relevant tools for a new project	Best Practices for Data Storage and Backup
	UW G Suite	UW G Suite
	<i>Google Drive provides a cloud-based solution to collaboration and unlimited storage to UW-Madison faculty, staff, researchers, and students</i>	
	UW-Madison Box	UW Box
	<i>Box is a cloud solution for storing, managing, and sharing files, but should only be used for regularly accessed, working files</i>	
	Center for High Throughput Computing	Center for High Throughput Computing
	<i>Provides scalable computing resources and services</i>	
2. Review Potential Data Repositories		
	Compare several general or discipline-specific data repositories and data resources	UW-Madison Research Data Services Data Sharing Essentials

		<i>Your funder may have expectations or requirements about where you deposit your data</i>	
3. Review Publication Requirements			
		The journal in which you publish your research may have data management and data sharing requirements.	
4. Consult or Initiate Data Use Agreements			
		Data Use Agreements (DUAs) govern access to and treatment of data.	DUA Evaluation Form
		<i>I. Data is provided by an outside organization to UW-Madison for use in research at UW-Madison</i>	DUA when UW-Madison receives a dataset
		<i>II. Data is provided by UW-Madison to an outside organization for use in its research</i>	DUA when UW-Madison discloses a dataset

GLOSSARY OF TERMS

Archive	The transfer of materials or data to a facility/site authorized to appraise, preserve, and provide access to the information
Data Lifecycle	The stages of data throughout its life, or the course of a project, from its creation to analysis, storage and backup, distribution, preservation, and reuse
Data Management Plan (DMP)	A two-page document that articulates how the data will be treated during its collection, processing, analysis, preservation, and use/reuse over time. It should include comprehensive information about the types and formats of data, metadata standards or other methods of data documentation, policies for access and sharing, and plans for archiving and preserving for long-term access. A DMP ensures data will be properly documented and available for use by researchers in the future. DMPs are becoming increasingly required by funding agencies when applying for funding.
Data Repository	A place to hold data, make data available for reuse, and organize data in a logical manner. Data repositories are often subject-specific and allow for researchers to self-submit data. Data repositories may have requirements regarding subject or research domain, data re-use and access, file format and data structure, and the types of metadata that can be used.
Data Security	The 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.
Data Use Agreement	A Data Use Agreement (DUA) governs access to and treatment of data. DUAs are required when data is provided by an outside organization to UW-Madison for use in UW-Madison research, or when data from UW-Madison research is provided to an outside organization for their use.
Metadata	Structured information about a resource that describes, explains, locates, or otherwise make it easier to understand, retrieve, use, or manage that resource. It ensures that the context for how data was created, analyzed, or stored is clear, detailed, and reproducible.
Research Data Management	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 to ensure that their data will be organized, easily shareable with other researchers, and archived for long-term preservation and access.