



Mastering Your NIH Data Management and Sharing Plan

Heather Owen, Data Librarian
howen@library.rochester.edu

Data Services



Service Types

- Chat & Email
- 1-on-1 Consultations
- Private Training and Workshops
- Live Public Webinars and Recorded Webinars

Webpage

<https://www.library.rochester.edu/services/data-management-and-sharing>

Specialties

- Data Management and Sharing Plans
- Data Management Best Practices
- Reproducibility
- Data Curation and Repositories
- Data Visualization
- Data Tools

Data Tool Admins

- UR Research Repository (URRR)
- LabArchives
- OSF
- protocols.io
- DMPTool
- ICPSR

Agenda



NIH Policies: Data Management and Sharing & Public Access

Creating Your DMSP

Data Services, Resources, and Tools

Q & A

NIH Policies: Data Management and Sharing & Public Access





DMS Policy

The NIH Data Management and Sharing (DMS) Policy is a set of requirements and guidelines related to the management and sharing of research data generated by work funded wholly or partially by the National Institutes of Health (NIH). This policy applies to all proposals received after January 25th, 2023, that result in the generation of scientific data.

DMSP

A Data Management and Sharing Plan (DMSP) is a document included in applicable NIH proposals that prospectively outlines how a research team intends to manage, preserve, and share the scientific data associated with the proposed work.

In Summary...

Researchers applying for NIH funding need to:

- Write a DMSP as part of their grant application
- Comply with it (therefore, share their scientific data)

What is Scientific Data?

It includes...

- Any data commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications
- Any data needed to validate and replicate research findings

It does not include...

- Laboratory notebooks
- Preliminary analyses
- Compared case report forms
- Drafts of scientific papers
- Plans for future research
- Peer reviews
- Communications with colleagues
- Physical objects such as laboratory specimens.

[NIH. \(n.d.\). Research covered under the data management & sharing policy.](#)

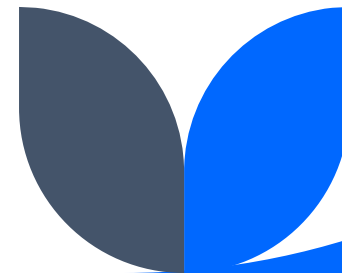


Justifiable Reasons for Limiting Sharing of Data

It is understood that some data cannot be shared openly or even at all, due to certain ethical, legal, and social issues.

Reasons include...

- Informed consent will not permit or will limit the scope or extent of sharing and future research use
- Existing consent prohibits sharing or limits the scope or extent of sharing and future research use
- Privacy or safety of research participants would be compromised and protective measures such as de-identification and Certificates of Confidentiality would be insufficient.
- Explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure
- Restrictions imposed by existing or anticipated agreements
- Datasets cannot practically be digitized with reasonable efforts



However, you MUST include a rationale for why scientific data will not be shared.

When Does it Need to be Shared?

According to the DMS Policy, scientific data needs to be shared by the earlier of two timepoints:

The time of an associated publication

OR

The end of the performance period (unless the grant enters into a no-cost extension).

In the case of a no-cost extension, data management and sharing-related requirements are extended to the end of the extended performance period.





National Institutes
of Health

Public Access Policy

The NIH Public Access Policy requires Author Accepted Manuscripts accepted for publication in a journal, **on or after December 31, 2025**, to be submitted to PubMed Central (PMC) upon acceptance for publications, for public availability without embargo upon the Official Date of Publication.

How to Satisfy this Policy?

Two Ways:

- The researcher submits an electronic version of the Author Accepted Manuscript to PubMed Central upon the Official Date of Publication.
- Journals or publishers with formal agreements with NLM submit the Final Published Article to PubMed Central upon the Official Date of Publication.

Definitions:

Official Date of Publication is “the date on which the Final Published Article is first made available in final, edited form, whether in print or electronic (i.e., online) format.”

An **Author Accepted Manuscript** is “the author’s final version that has been accepted for journal publication and includes all revisions resulting from the peer reviewed process, including all associated tables, graphics, and supplemental material.”

A **Final Published Article** is “the journal’s authoritative copy, including journal or publisher copyediting and stylistic edits, and formatting changes, even prior to the compilation of a volume or issue or the assignment of associated metadata.”



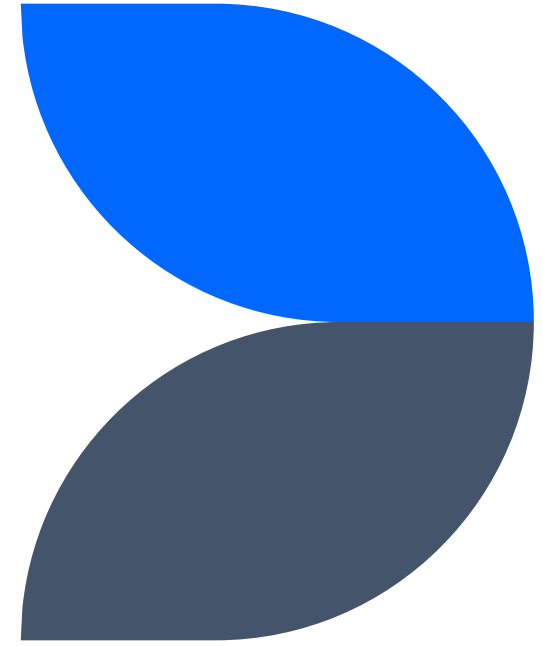
Important Things to Note

- Submitting to PubMed Central **is free**. There is no cost associated with satisfying this policy.
- This policy affects all articles where the Final Published Article is made available after December 31, 2025, regardless of when the researcher received funding.
- There is no embargo.

If you need assistance with publishing, contact [Scholarly Communications!](#)



Creating Your DMSP



General Advice from the NIH

- Two pages is just a recommendation – it can be longer.
- Don't be afraid to use tables – it is easier for the NIH to read.
- Be careful not to contradict yourself.
- Some individuals mention planning to share only through “publications” and “conferences” or “by request.” This is not enough!
- Check out the examples on the [NIH Website](#).



Reoccurring Issues Noted by the NIH

Sharing

- Vague reasons are given for not sharing
- Sharing methods are too restrictive without a justifiable reason why

Budget

- Budget does not include DMS activities

Repository

- No repository is chosen
- An ill-fitting repository is chosen
- When using multiple repositories, not specifying which data goes to which repository
- Overreliance on generalist repositories (make sure a subject-specific one doesn't exist!)

Data

- Not clear on which data will be generated and which data will be shared
- Lack of detailed description about data types (including estimated amounts).

Standards

- Researchers are being vague or not mentioning any standards at all.
- Check out NIH CDE repository, NCI-thesaurus, NNLM, and CaDSR.



DMSPs for NIH Require 6 Elements



Data Types



Related Tools, Software and/or Code



Data Standards



Data Preservation, Access, and Associated
Timelines



Data Sharing Agreements, Licenses, and
Other Use Limitations



Oversight of Data Management

[NIH. \(2020\). Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.](#)

Data Type - Summary

- Summary of the types and estimated amount of scientific data.
- Descriptions may include data modality, level of aggregation, and degree of data processing.
- Description of which scientific data from the project will be preserved and shared. Remember, NIH does not anticipate for all scientific data to be shared. You must provide a rationale for these decisions.
- A brief listing of the metadata, other relevant data, and any associated documentation that will be made accessible to ensure the scientific data is understandable and interpretable.



[NIH. \(2020\). Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.](#)

Data Type - Example

This project will produce **imaging** data generated/obtained from **ultrasonography**. Data will be collected from **10** of research participants/specimens/experiments, generating **1** datasets totaling approximately **half a GB** [amount of data] in size. The following data files will be used or produced in the course of the project: **DICOM**. Raw data will be transformed by **filtering and enhancing the images** and the subsequent processed dataset used for statistical analysis. To protect research participant identities, **individual** data will be made available for sharing.



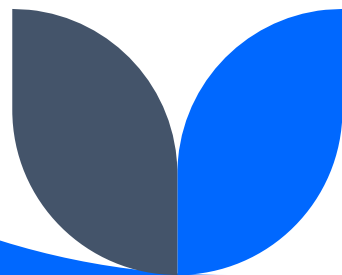
Adapted from NIH. (2022). 2023 NIH Data Management and Sharing Policy.

Data Type – Table Example

Element 1: Data Type

A. Types and amount of scientific data expected to be generated in the project:

Type	Species	Platform/ Source	Amount
Array-derived genotype data	Human	Illumina	1,000 research participants (500 cases/controls), prospective enrollment
30x whole-genome sequence data	"	"	"
RNA-seq data	"	"	"
Hi-C WGS	"	"	"
Phenotypic and clinical data	"	Institutional EHR	"
Demographic data	"	"	"



Data Type – What We've Noticed

- Researchers are forgetting to mention the file type
- Researchers are not including the amount of data (an estimate or range is ok).
- Researchers are mentioning the data they are sharing, but not being specific about the data they are not sharing (and why).
- Researchers are not including readmes and/or data dictionaries and/or methodology in the type of documentation they are sharing.



Related Tools, Software and/or Code

– Summary & Example

- List the names of **tool(s) and software** needed to access or manipulate shared scientific data to support replication or reuse.
- If necessary, specify **how these tools can be accessed**, and if the tools are likely to remain available.

Imaging data will be made available in **DICOM** format, which requires the use of **MicroDicom** to be accessed and manipulated. The **MicroDicom** tool, which can be used to **view the images** is available free of charge through <https://www.microdicom.com/>.

[NIH. \(2020\). Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.](#)

[Adapted from NIH. \(2022\). 2023 NIH Data Management and Sharing Policy.](#)



Related Tools, Software and/or Code

– What We've Noticed

- Researchers are not specifying the version of the software or tools they are using.
- Researchers are not specifying whether these tools are open or proprietary. They are not specifying how they can be accessed.
- Individuals are not giving URLs to the tools. [Note: Do not hypertext links.]



Standards - Summary

- Indication of what **standards** will be applied to the scientific data and associated metadata:
 - Data formats
 - Data dictionaries
 - Data identifiers
 - Definitions
 - Unique Identifiers
 - Other data documentation
- Plan may indicate that **no consensus** data standards exist for the scientific data and metadata



[NIH. \(2020\). Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.](#)

Standards - Examples

To facilitate their efficient use, all of our data and materials will be structured and described using the following standards: **Network Common Data Format**.

Alternatively

Formal standards for **DICOM** data have not yet been widely adopted. However, our data and other materials will be structured and described according to best practices.



[Adapted from NIH. \(2022\). 2023 NIH Data Management and Sharing Policy.](#)

Standards – Table Example

Element 3: Standards:

Data Type	Standard
Human array-derived genotype data	VCF
30x whole-genome sequence data	Sequencing data and variant calls will be shared in CRAM and VCF formats, respectively.
RNA-seq data	Data will be QCd and analyzed according to ENCODE Bulk RNA-seq Data Standards. FASTQs, BAM alignment files, and TSV transcript quantifications will be shared.
Hi-C WGS	FASTQ
Demographic, Phenotypic and Clinical Data	<ul style="list-style-type: none">• PhenX for surveys• RxNorm for meds• PCORnet CDM which is derived from OMOP for EHR data collection for secondary outcomes• Current Procedural Terminology (CPTs), Logical Observation Identifier Names and Codes (LOINCs) and diagnoses ICD10 codes
Study protocols	Customized (non-standard) & to be developed



[NIH. \(2022\). Sample DMS plan – Human genomic data project.](#)

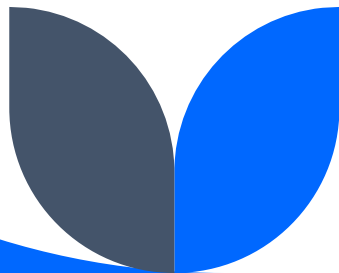
Standards – What We've Noticed

- Researchers are being ambiguous here.
- Reach out to colleagues in your department or contact your subject librarian or other people with subject expertise.
- Look into research being done in your field. What standards are they using?
- If you are using a subject-specific repository, what standards do they use or recommend?
- NIH recommends these resources to learn more about different standards that exist: NIH CDE repository, NCI-thesaurus, NNLM, and CaDSR.
- You can search for standards on [FAIRsharing.org](https://fairsharing.org).



Data Preservation, Access, and Associated Timelines - Summary

- Name of repository(ies) where data will be archived and shared
- How the scientific data will be findable and identifiable (i.e., persistent unique identifier)
- When the scientific data will be made available and for how long.
- Identify any differences in timelines for different subsets of scientific data to be shared.
- Scientific data should be shared as soon as possible, and no later than time of an associated publication or end of the performance period.



[*NIH. \(2020\). Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*](#)

Data Preservation, Access, and Associated Timelines - Example

All dataset(s) that can be shared will be deposited in **University of Rochester Research Repository (URRR)**, which is the University of Rochester's instance of Figshare.

Figshare provides metadata, **DOI** persistent identifiers, and long-term access. This repository is supported by **University of Rochester Librarians and Data Curators**, who check all data before it is shared to ensure data quality and **findability**, and dataset(s) are available under a **Creative Commons CC0 license**. The links to the datasets will be included in all **publications and in the ORCIDs** of all of the researchers.

Data will be made available as soon as possible or at the time of associated publication, whichever comes first.



[Adapted from NIH. \(2022\). 2023 NIH Data Management and Sharing Policy.](#)

Data Preservation, Access, and Associated Timelines – What We've Noticed

- Researchers do not name a specific repository
- Researchers do not specify that they will ensure their data has a persistent unique identifier (if you use Figshare, you will automatically receive a persistent unique identifier when you publish your data and can reserve it ahead of time).
- Researchers do not mention linking to their data in their publications and ORCID.



Where Should You Share Your Data?

1

Subject(Domain)-Specific Repositories

2

University of Rochester Research
Repository (URRR)

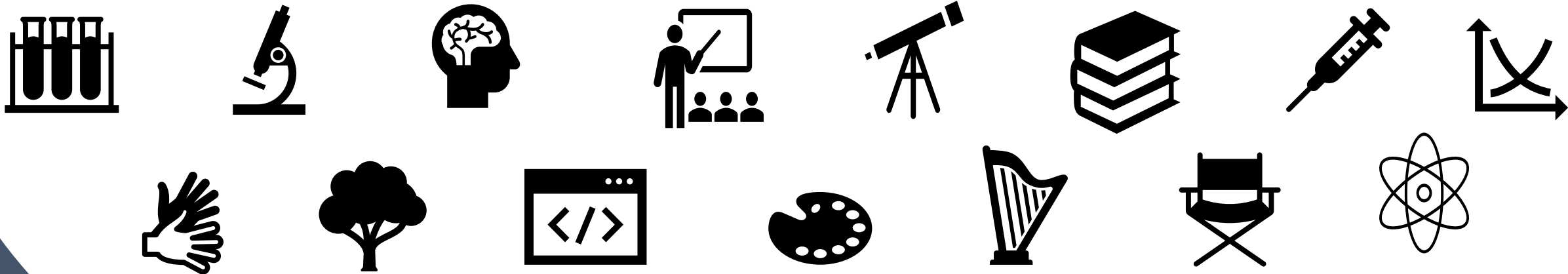
3

Other Generalist Repositories

What is a Subject-Specific Repository?

A Subject-Specific Repository (or a Domain-Specific Repository) is a repository which stores and shares data for a specific subject.

However, not all subjects have a subject-specific repository available.



How Can I Find a Subject-Specific Repository?

Where are other people in your field sharing their data?

Ask your colleagues, examine journal articles and other publications, and check funder recommendations.

Reach out to your Subject Librarian on the River Campus, at the Medical Center, or at Eastman for assistance.

NIH has a list of [subject-specific repositories](#) they recommend!

If there is no subject-specific repository...

Try the University of Rochester Research Repository
(URRR)





University of Rochester Research Repository (URRR)

Controlled access with embargo and private link

UR Instance of Figshare, a Generalist Repository

All UR members can deposit up to 10 GB of Data – contact us for more

Reserve a DOI ahead of time

Curation Service – we will check your data

Can share data, posters, articles, images, presentations, etc



Third Option: Generalist Repositories



- A generalist repositories is not subject-specific and accepts all kinds of data.
- There are 7 main generalist repositories: Harvard Dataverse, Dryad, Figshare, Mendeley Data, OSF, Vivli, and Zenodo.
- These repositories are part of GREI, the generalist repository ecosystem initiative.

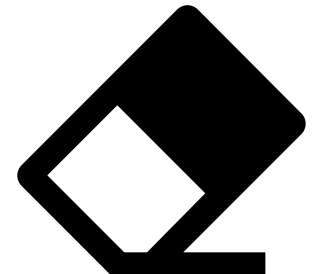
[Check out the Generalist Repository Comparison Chart to Compare the Repositories](#)

Sharing Restricted Access Data

If your data includes human participants, make sure you de-identify your data before you share it. De-identification requires you to remove any personal information from a dataset.

If de-identifying your data is not enough, you may be able to only share aggregated or summarized data. Or, you may want to share it in a repository which allows for restricted access.

[UR Data De-Identification Services](#)



ICPSR

a partner in social science research

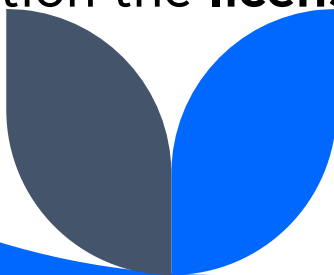
Can share data
– free and pay
options (get
quote)

Focused on
social and
behavioral
research.

Can share
restricted access
data – ICPSR
staff handle
everything

Access, Distribution, or Reuse Considerations - Summary

- Describe any **factors** affecting subsequent access, distribution or reuse of scientific data related to:
 - **Informed consent**
 - Privacy and confidentiality protections consistent with applicable federal, Tribal, state, and local **laws, regulations, and policies**
 - De-identification
 - Certificates of Confidentiality
 - Data Use Agreements
 - Other Protective Measures
 - Note whether **access** to scientific data derived from humans **will be controlled** (i.e., made available by a data repository only after approval).
 - Note any other **considerations** that may limit the extent of data sharing. You can mention the **license** you are putting your data under.



[*NIH. \(2020\). Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*](#)

Access, Distribution, or Reuse Considerations - Example

Tier	Data description	Data types
Public	Project protocols, and aggregated human subjects data	<ul style="list-style-type: none">•Interview protocols (.docx)•Analysis plans (.docx)•Aggregated survey data (.csv)•Aggregated codebook (.csv)
Restricted	Individual-level human subject data	<ul style="list-style-type: none">•De-identified survey responses (.csv)•Qualitative interview transcripts (.txt)•Interview codes (.csv)



Access, Distribution, or Reuse Considerations – No Factors Example

There are no factors that will affect subsequent access, distribution or reuse of the scientific data. All data will be de-identified and we will not place any access controls over the data.



Access, Distribution, or Reuse Considerations – What We've Noticed

- Some people do not fully explain the restrictions they place on their data. This can include:
 - De-identifying your data (Protect human participants)
 - Aggregating your data (Protect human participants; sheer size of data)
 - Adding a license restricting how users can use the data.
 - Placing restrictions on who can look at the data
 - Establishing a Data Use and Transfer Agreement (DUTA)



License

- CC0 1.0 Universal Public Domain Dedication
- CC BY 4.0
- CC BY-SA 4.0
- CC BY-NC 4.0
- CC BY-NC-SA 4.0
- CC BY-ND 4.0
- CC BY-NC-ND 4.0
- ODbL
- ODC-By
- PDDL
- MIT

Check out the [Creative Commons Website](https://creativecommons.org/) to learn more about CC Licenses.

Not all repositories will allow you to choose your own license!



Oversight of Data Management and Sharing - Summary

- Indicate how compliance in the Plan will be...
 - Monitored and managed
 - Frequency of oversight
 - By whom (e.g., titles, roles)



[NIH. \(2020\). Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.](#)

Oversight of Data Management and Sharing - Example

Key Aspects to Consider:

- Who provides oversight?
- What are the various stakeholders' roles and responsibilities?

The PI on this project will be responsible for all aspects of data management and sharing, including collecting, analyzing, and describing the data, as well as uploading the data to **[INSERT REPOSITORY NAME]**. Compliance with the Plan will be monitored and managed, [include frequency of oversight], and by whom (e.g., titles, roles).



[Adapted from Samuel J. Wood Library. \(2022\). Oversight of Data Management and Sharing.](#)

Oversight of Data Management and Sharing – What We've Noticed

- Individuals are mentioning that compliance is being checked by the university – at the moment there is no plan in place for the university to be responsible for compliance.
- Individuals are not specifying who is responsible for what individual parts of data management and sharing.
- Individuals do not mention having a backup or security plan.
- **Important Note:** The University of Rochester will maintain all study records (i.e., regulatory files, data collection forms, source documentation) for at least 3 years after the research is completed, or for a longer term if required by other contractual agreements. If the consent form includes HIPAA Authorization, it must be maintained for 6 years.



Why Manage & Share Data?

Reproducibility

**Speed Up
Medical
Research**

**Managed
Data = Less
Mistakes**

**Publications
Linked to Data
Have Increased
Citation Counts**

Open Science

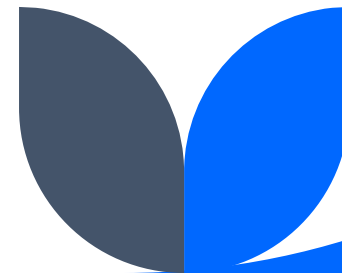
**Ensure Data is
Not Lost**

**Funded By the
Public**



Manage & Share Data Resources

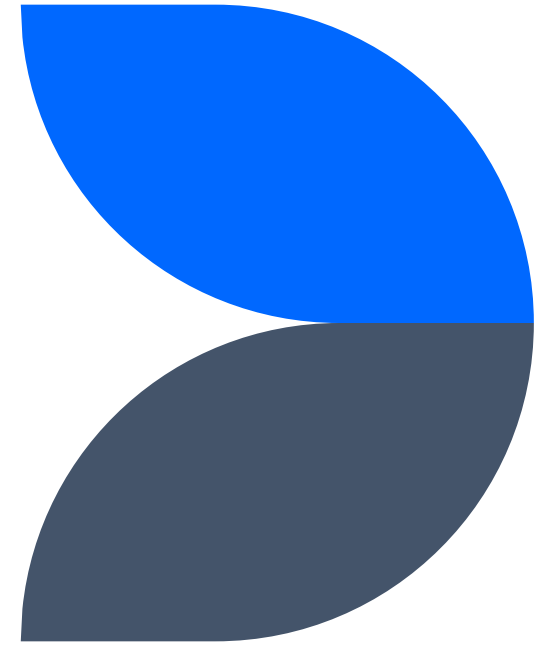
- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillibray, B. (2020). The citation advantage of linking publications to research data. *PLoS One*, 15(4). <https://doi.org/10.1371/journal.pone.0230416>
- Dattani, S. (2022, October 25). *The pandemic uncovered ways to speed up science*. WIRED. <https://www.wired.com/story/covid-19-open-science-public-health-data/>
- Kaiser, J., & Brainard. (2023, January 25). *Ready, set, share! As funders roll out new requirements for making data freely available, researchers weigh costs and benefits*. Science. <https://www.science.org/content/article/ready-set-share-researchers-brace-new-data-sharing-rules>
- Mullard, A. (2021, December 9). *Half of top cancer studies fail high-profile reproducibility effort*. Nature. <https://www.nature.com/articles/d41586-021-03691-0>



Q & A Break!



Data Services & Resources



Service Types

Answer Questions Via Chat or Email.

Offer one-on-one consultations in-person or via zoom.

Give private workshops/webinars to groups, classes, or dept. upon request

Review DMSPs for all Funders and Review Datasets

URL: <https://www.library.rochester.edu/services/data-management-and-sharing>



Areas of Specialization

DMSPs

Review plans and give advice for all funders

Reproducibility

Statistical methods, APIs, Scholarly Identity, ORCID, PIDs

Data Management Best Practices

Data and file management, documentation, readmes, data dictionaries, file naming conventions

Data Curation and Repositories

University of Rochester Research Repository (URRR), deposit and curate data, choose a repository

Data Visualization

Excel, Tableau, RStudio, Jupyter Notebooks, Python, Etc.

Data Tools

LabArchives, OSF, protocols.io, the DMPTool, ICPSR, and BioRender

URL: <https://www.library.rochester.edu/services/data-management-and-sharing>



Library Resources



NIH-Specific DMS Resource Guide – This guide offers detailed answers to common questions related to the NIH DMS Policy and has advice on how to create a DMSP for the NIH.



General DMS Resource Guide – This guide offers detailed advice on how to create DMSPs, and how to manage, organize, document, store, and share your data.



We Want You To...

Submit Your Research!

URRR can store your datasets, posters, journal articles, creative works, and **more!**

Submitting to the **University of Rochester Research Repository (URRR)** can enhance your CV and includes:

- **DOIs!**
- **Usage metrics!**
- **10 GBs of storage!**
- **Long-term preservation!**
- **Listing in Google Scholar & Datasets!**
- **And so much more!**

WOW!



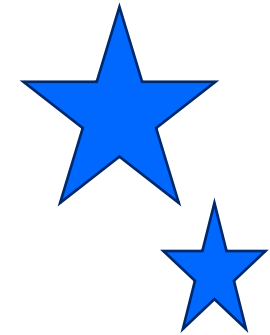
scan QR code



for more info



Reach Out to Us for Help
in Using URRR!
rochester.figshare.com





LabArchives

UR Community
Members Have
Free Access.

Use during the
research process to
keep your folders,
files, data, and
code, organized.

Optionally share
your notebook
publicly or with
guests. Optionally
publish it to get a
DOI.

An Electronic
Laboratory
Notebook (ELN).

Collaborate with
others and invite
people to your
journal. Has
version control.

Learn More on the [UR LabArchives Landing Page](#)



UR community members have
institutional access.

Can store files, data, code and other
materials. Every file gets a unique,
persistent URL, and it has version
control.

An open project management tool
that allows you to collaborate with
others on your research.

Can share all your research outputs.
Submit preregistrations and
registrations to the OSF Registry.
Share papers in OSF Preprints.

Travel to <https://osf.io/institutions/rochester>
Learn More: <https://libguides.lib.rochester.edu/osf>





- University of Rochester has **institutional access**. [Free]
- A **secure platform** where you can share methodology and protocols.
- **Version control** is built into the platform.
- Allows for **private & secure collaboration** with HIPAA compliance, audit trail, approval/signature, 2-factor authentication, encryption, VPC, and more.
- Can **share publicly, and link to** in your publications and your readme.
- Increases **reproducibility** of your research!

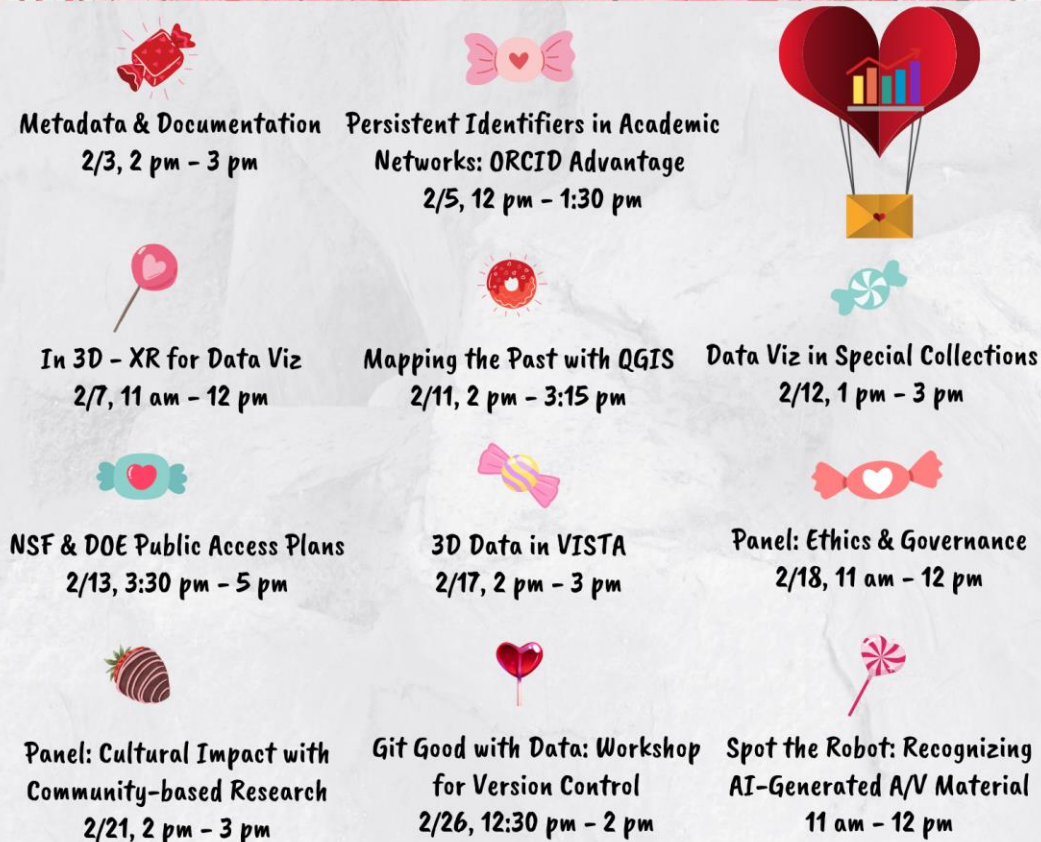














NIH Resources

- [Link to the Policy & Instructions on How to Write It](#)
- [That Constitutes as Research that Needs to Be Shared Under the NIH Policy](#)
- [Domain-Specific Data Repositories](#)
- [Draft of a Template Supplied by NIH](#)
- [Format, including citations, file names, file sizes font, hyperlinks, page limits, etc.](#)
- [6 main elements to include](#)
- [Budgeting for Data Management and Sharing](#)
- [Submitting Human Genomic Data](#)
- [NIH FAQ, including information on human data, and when data should or should not be shared](#)



Love Data Month



 Metadata & Documentation 2/3, 2 pm - 3 pm	 Persistent Identifiers in Academic Networks: ORCID Advantage 2/5, 12 pm - 1:30 pm	 Data Viz in Special Collections 2/12, 1 pm - 3 pm
 In 3D - XR for Data Viz 2/7, 11 am - 12 pm	 Mapping the Past with QGIS 2/11, 2 pm - 3:15 pm	 Data Viz in Special Collections 2/12, 1 pm - 3 pm
 NSF & DOE Public Access Plans 2/13, 3:30 pm - 5 pm	 3D Data in VISTA 2/17, 2 pm - 3 pm	 Panel: Ethics & Governance 2/18, 11 am - 12 pm
 Panel: Cultural Impact with Community-based Research 2/21, 2 pm - 3 pm	 Git Good with Data: Workshop for Version Control 2/26, 12:30 pm - 2 pm	 Spot the Robot: Recognizing AI-Generated A/V Material 11 am - 12 pm



Join us for Love Data Month 2025!

A celebration of data, aiming to raise awareness and build a community to engage on topics relating to metadata, data cleaning, analysis, and data management with some fun events to make you love your data even more!

Scan to View all Events

Check Out Love Data Month!

- Metadata & Documentation
- Persistent Identifiers in Academic Networks: ORCID Advantage
- In 3D - XR for Data Viz
- Mapping the Past with QGIS
- Data Viz in Special Collections
- NSF & DOE Public Access Plans
- 3D Data in VISTA
- Panel: Ethics & Governance
- Panel: Cultural Impact with Community-based Research
- Git Good with Data: Workshop for Version Control
- Spot the Robot: Recognizing AI-Generated A/V Material

<https://libguides.lib.rochester.edu/lovedata>

Please fill out our survey!

It will take only a few minutes.

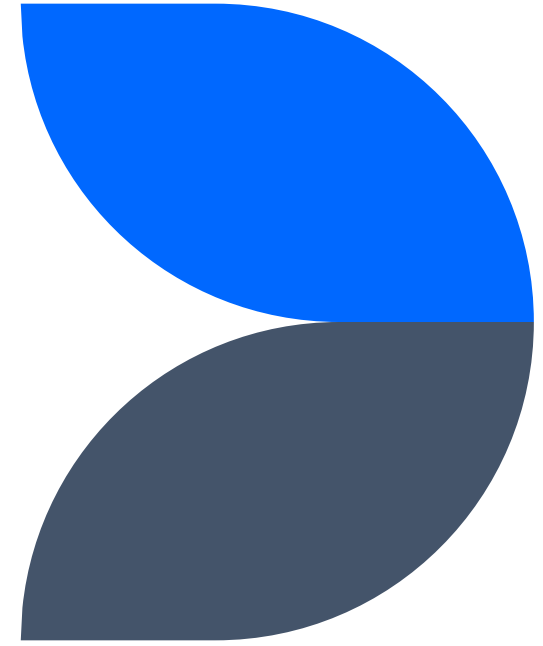


Thank you
very much!



https://rochester.libwizard.com/f/NIHDMSP_Survey

Q&A





Mastering Your NIH Data Management and Sharing Plan

Heather Owen, Data Librarian
howen@library.rochester.edu