

Data Management and Sharing Plan Basics: Preparing for the new NIH Data Management and Sharing Policy

November 10, 2022

Elizabeth Roth
Research and Data Science Strategist,
Network of the National Library of Medicine
Region 2
MUSC Libraries

Content Overview

- Data Sharing Landscape
- Scope of the Policy
- Elements of the DMSP
- Tools and Resources for writing a DMSP
- Additional Resources



Data Sharing Landscape





Why does NIH Want Data to be Shared?

Advance rigorous and reproducible research

- Enable validation of research results
- Make high-value datasets accessible
- Accelerate future research directions
- Increase opportunities for citation and collaboration

Promote public trust in research

- Foster transparency and accountability
- Demonstrate stewardship over taxpayer funds
- Maximize research participants' contributions
- Support appropriate protections of research participants' data

Scope of the Policy



NIH Data Management & Sharing Policy



Requires Data Management and Sharing Plans (DMPs) for all research conducted at or funded (in whole or in part) by the NIH that **generates scientific data**.

Policy goes into effect January 25, 2023

Projects Generating Data

Includes

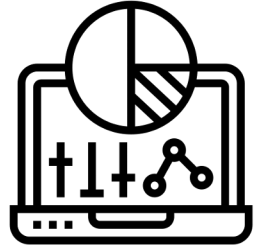
- Competing grant applications that are submitted to NIH
- Contract proposals
- NIH Intramural Research Projects
- Other funding agreements with the NIH (e.g., Other Transactions)

Excludes

- Training (T)
- Fellowships (F)
- Construction (CO6)
- Conference Grants (R13)
- Resource (Gs)
- Research-Related Infrastructure Programs (e.g., So6)



Definition of Scientific Data



“The **recorded factual material** 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.”

:



What counts as data for sharing?



- Adequate data to validate and replicate study findings
- Data resulting from the study but not necessarily supporting a publication
- Null findings that do not result in publication

Data that does NOT need to be shared

- Lab notebooks
- Preliminary analyses
- Completed case report forms
- Peer reviews
- Physical objects, such as
laboratory specimens



Data Sharing Timeline



By when do I need to share the data?

At the time of publication OR at the end of award, whichever comes first.

NIH understands that some scientific data generated with NIH funds may be proprietary and can be withheld from sharing for up to 20 years.

Allowable Costs

Reasonable, allowable costs may be included in **NIH budget requests** for:

- Curating data
- Developing supporting documentation
- Formatting data according to accepted community standards, or for transmission to and storage at a selected repository for long-term preservation and access
- De-identifying data
- Preparing metadata to foster discoverability, interpretation, and reuse
- Local data management considerations, such as unique and specialized information infrastructure necessary to provide local management and preservation (for example, before deposit into an established repository).
- Preserving and sharing data through established repositories, such as data deposit fees.

Unallowable Costs

Budget requests must NOT include:

- Infrastructure costs that are included in institutional overhead (for instance, Facilities and Administrative costs)
- Costs associated with the routine conduct of research, including costs associated with collecting or gaining access to research data.
- Costs that are double charged or inconsistently charged as both direct and indirect costs

Plan submission, review, and compliance

Submission

With application for funding in Budget Justification section

Assessment

Peer reviewers may comment on (not score) budget

NIH program staff assess Plans

Plans can be updated

Compliance

Incorporated into Terms and Conditions

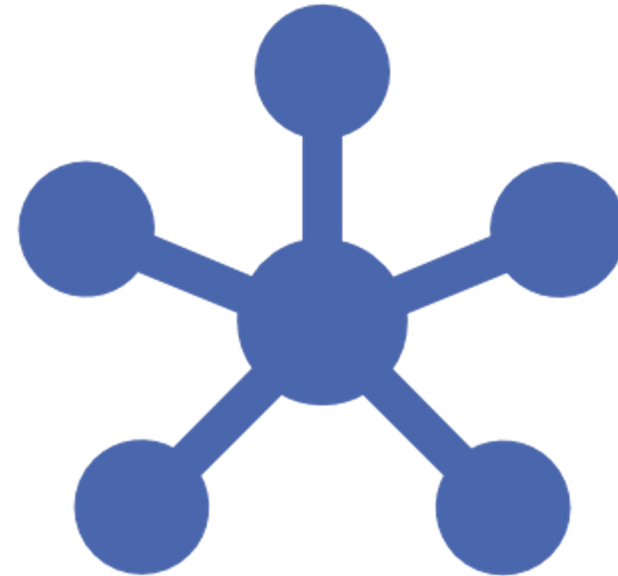
Monitored at regular reporting intervals

Compliance may factor into future funding decisions



DMSP and the broader policy landscape

- DMSP represents the minimum requirements for all NIH
- NIH Institutes, Centers, and Offices (ICOs) may provide more specific guidance
- Other policies may apply (e.g. Genomic Data Sharing policy)



Key Takeaways

- The policy requires you to write a plan. It does NOT require data sharing.
- The data management and sharing plan is NOT scored. It is reviewed for completeness.
- The plan is a summary. It is only 2 pages long.
- The plan is a living document. It will likely change and that's okay and expected.
- Maximize the amount you share, within the limits of the law.
- When in doubt, contact your Program Officer!



Elements of the DMSP



What is a Data Management and Sharing Plan (DMSP)?

- DMSPs state how scientific data will be **managed and shared**, including key procedures and responsibility over the course of the award
- DMSPs include **descriptions** of what data will be shared and what limitations there are on data sharing
- Submitted DMSPs **must comply** with other relevant policies (e.g., NIH Genomic Data Sharing Policy)
- The policy does not proscribe methods of **data sharing or management**, but does have preferences (e.g., use of NIH-supported repositories). The policy also expects PIs to follow the submitting plans.



Elements of DMSP (2023)



- 1. Data type**

Identifying data to be preserved and shared

- 2. Related tools, software, and/or code**

Tools and software needed to access and manipulate data

- 3. Data standards**

Standards to be applied to scientific data and metadata

- 4. Data preservation, access, and associated timelines**

Repository to be used, persistent unique identifier, and when/ how long data will be available

- 5. Access, distribution, or reuse considerations**

Description of factors for data access, distribution, or reuse

- 6. Oversight of data management and sharing**

Plan compliance will be monitored/ managed and by whom

1. Data Type

- A description of data, associated metadata, and documentation (e.g., data dictionary, study protocol, data collection instruments)
 - Data modality (e.g., imaging, genomic, survey)
 - Level of aggregation (e.g., individual, grouped, summarized)
 - Level of data processing (e.g., raw vs. processed data)
- What data will be shared
- What metadata and documentation will be included to facilitate interpretation

2. Related tools, software, and/or code

- Any additional tools/software are needed to access or manipulate the data (e.g., Python packages/modules)
- Names of specific software tools (e.g. Python, SPSS, etc)
- Availability of tools (e.g. open source vs purchase)
- Expected lifespan of the tools compared to length of data availability

3. Standards for the data/metadata

- Data formats
- Data dictionaries
- Common Data Elements
- Identifiers
- Definitions
- Indicate when no consensus standard exists

4. Data preservation, access, and associated timelines

- Name(s) of repository(ies) where data and metadata will be deposited
- How data will be made identifiable (e.g., Digital Object Identifier)
- When the data will be made available and for how long
 - No later than time of publication or end of performance period
 - How long data is anticipated to be available

5. Access, distribution, or reuse considerations

- Access, distribution, or reuse considerations
- Any limitations due to informed consent, privacy and confidentiality protections
- Controlled access protections (e.g., author approval, DUA)
- Restrictions imposed by federal, Tribal, or state laws, regulations, or policies (e.g. HIPAA)

6. Oversight of Data Management and Sharing

- Describe how compliance with the DMSP will be monitored and managed (e.g., frequency, person(s) responsible), including budgeting requirements
- Who will be responsible for oversight
- How often will oversight activities occur

Tools and Resources for Writing a DMSP



Selecting a Repository

NIH strongly encourages **subject-specific, open access Data Sharing Repositories** as a first choice.

https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

Datasets up to **2 gigabytes**

PubMed Central

Stores publication-related supplemental materials and datasets directly associated publications.



Datasets up to **20 gigabytes**

Generalist Repositories

Datasets associated with publications or otherwise and links to PubMed.



High priority datasets, **petabyte-scale**

Cloud Partners (STRIDES Program)

Store and manage large scale, high priority NIH datasets.



Free online tool: DMPTool



- Online application from the UC libraries that helps researchers create data management plans.
- Provides a click-through wizard for creating a DMP that complies with funder requirements.
- Used by many institutions, including Harvard, Johns Hopkins, UPenn, Stanford, and many others.
- Current community focus is on the new NIH DMSP Policy

<https://dmptool.org/>

DMPTool continued



- Supports collaborative creation of DMPs across a lab or research project, including tracking changes to the DMP and who makes them
- Template helps ensure you answer necessary questions and makes the policy requirements more comprehensible
- Includes guidance from the grant funder
- Allows for review and collaboration
- Can be downloaded in various formats
- Provides a repository of existing, public DMPs

2023 NIH DMSP Template

NIH 2023 Template

Project Details

Collaborators

Write Plan

Research Outputs

Request feedback

Download

Finalize / Publish

This plan is based on the "NIH-GEN DMSP (Forthcoming 2023) " template provided by National Institutes of Health (NIH) - (ver: 1, pub: 2021-02-01).

expand all | collapse all

0/12

+ Data Type (0 / 3)

+ Related Tools, Software and/or Code (0 / 2)

+ Standards (0 / 1)

+ Data Preservation, Access, and Associated Timelines (0 / 3)

+ Access, Distribution, or Reuse Considerations (0 / 2)

+ Oversight of Data Management and Sharing (0 / 1)



NIH Forms and Applications Library

Preview the Data Management and Sharing Plan Form (draft) that will soon be available in the NIH Forms and Applications Library

<https://grants.nih.gov/sites/default/files/DMS-Plan-blank-format-page.pdf>



Additional Resources



Staying up-to-Date: sharing.nih.gov

- Provides a central source of guidance related to multiple NIH data sharing policies
- Covers Data Management and Sharing, Genomic Data Sharing, Model Organisms, and Research Tools policies
- Content will be updated

U.S. Department of Health & Human Services | National Institutes of Health

NIH SCIENTIFIC DATA SHARING

Search

NIH Staff | FAQ | [Contacts & Help](#)

[DATA MANAGEMENT AND SHARING POLICY](#) [GENOMIC DATA SHARING POLICY](#) [OTHER SHARING POLICIES](#) [ACCESSING DATA](#) [ABOUT](#)

Expediting the Translation of Research Results to Improve Human Health.

FEATURED NEWS & EVENTS

[Gearing Up for 2023: Implementing the NIH Data Management and Sharing Policy](#)

[View More](#)

Resources about the Policy

- [NIH Data Sharing Policy](#)
 - Includes guidance on the current and new policies
- [NIH Data Sharing Policy FAQs](#)
 - Some answers to frequently asked questions about the 2023 policy
- [NIH Data Sharing Resources](#)
- [Forecasting Costs Related to Preserving Biomedical Data](#)
 - Provides guidance on costs related to preserving and sharing data

Resources about Repositories

- [List of Domain-Specific Repositories](#)
- [List of Generalist Repositories](#)
- [Registry of Research Data Repositories \(RE3Data\)](#)
- [FAIRSharing.org](#)
 - Includes information on repositories, data sharing policies, and data standards

Toolkits and Examples

- Toolkit from the Network of the National Library of Medicine (NNLM) National Center for Data Services
 - <https://www.nlm.gov/guides/nnlm-toolkit-nih-data-management-and-sharing-policy>
- Toolkit from the NIH DMSP Guidance Working Group
 - <https://osf.io/uadxr/>
 - **View an example of a completed plan!** <https://osf.io/euaty>

Credit and Thanks to:

Genevieve Milliken

Justin de la Cruz

Lisa Federer

Sara Samuels

Heather Coates

Nicole Contaxis

Nina Exner

Ariel Deardorff

Michelle Yee

Katie Pierce Farrier

Taunton Paine

Cindy Danielson

Ummea Urmi