NIH Data Management & Sharing Policy (Effective 2023): A Guide for Pitt Health Sciences Researchers

Helenmary Sheridan, Data Services Librarian
Health Sciences Library System, University of Pittsburgh
Email: helenmary@pitt.edu (individual), HSLSDATA@pitt.edu (team)
Website: Data Management at the Health Sciences Library System
Last updated: 2022-10-26

About this Handout

The National Institutes of Health's new Policy for Data Management and Sharing (DMS Policy), which goes into effect January 25, 2023, will require NIH-funded researchers to submit a plan outlining how scientific data from their research will be managed and shared within their funding application. The policy includes an expectation that researchers will maximize their data sharing within ethical, legal, or technical constraints, and explicitly encourages researchers to incorporate data sharing via deposit into a public repository into their standard research process.

To help University of Pittsburgh researchers comply with the new policy, HSLS Data Services has put together this guidance for the new policy along with best practices for data management and data sharing. This document mirrors the text content available at the NIH Data Management & Sharing Policy (Effective 2023) Libguide. If you have a question, desire one-on-one consultation, or would like us to review your data management and sharing (DMS) plan, please don't hesitate to contact HSLS Data Services.

Related resources for this topic as a whole

- Policy for Data Management and Sharing (DMS Policy): Official text of the policy as issued by the Office of the Director, National Institutes of Health, October 29, 2020 (effective January 25, 2023)
- NIH Scientific Data Sharing: The NIH's clearinghouse for all announcements, policies, and other resources related to scientific data sharing, covering the new DMS Policy as well as the Genomic Data Sharing Policy, Model Organism Sharing Policy, Research Tools Policy, Clinical Trial Dissemination Policy, Public Access Policy, and NIH Institute and Center Data Sharing Policies
- NIH Contacts & Help: The Contacts and Help page on the Scientific Data Sharing website provides contact information for program staff by institute or center; an email address for data sharing questions; and email addresses and request forms related to genomic data sharing

NIH Data Management & Sharing Policy Overview

In 2020, the National Institutes of Health released a [Policy for Data Management and Sharing (DMS Policy)](#) (also called the DMSP) that updated the institutes' data sharing policy from 2003. After years of public comment and revision, the final policy will go into effect on **January 25, 2023**.

What does the policy require?

The policy requires three things: that researchers think about how they will manage, document, and share their scientific research data before beginning data collection; that they show the NIH their thought process in a formal Data Management and Sharing Plan in their funding application (with an accompanying budget detailing data management costs), and that they make their research data as publicly available as possible within a reasonable time frame.

Which researchers must comply with the policy?

All researchers who are funded in whole or in part by the NIH must comply with the policy, regardless of funding level. However, the policy only applies to activities that generate scientific research data according to a specific definition given below. If you are funded through a Training (T), Fellowship (F), Construction (C06), Conference (R13), Resource (G), or Research-Related Infrastructure Program (e.g., S06) grant, the policy will not apply to your work. See the complete list of [NIH activity codes](#) for more information.

What types of data are covered by the policy?

The DMS Policy applies to scientific research data, defined as "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." This covers both quantitative and qualitative data, and applies to all data types including image, audio, and video data of all file formats and sizes. It explicitly does not apply to "laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects such as laboratory specimens" (source: [Research Covered Under the Data Management & Sharing Policy](#)), as those may contain or produce data but do not constitute scientific data themselves.

Determining which data fall under the policy will depend on your own judgment and the norms of your field. If you generate a lot of calibration, null, duplicate, or "junk" data and would not normally consider that as necessary to replicate findings, you do not need to address that data in your DMS plan. Conversely, you may consider some research material to be "data" that is not explicitly addressed in the policy. The policy does not address scripts, code, algorithms, or computational models as data products in and of themselves, but if you consider those to be necessary for replication of your work, you should discuss them in your DMS plan in order to make your management and sharing easier. (The policy does explicitly ask researchers to describe any software tools necessary to access or manipulate the data.)

The ultimate goal of the DMS policy is to encourage researchers to think about data management and data sharing in concrete terms before executing their research plan, because experience has shown that it's much harder to do after the conclusion of a project. We recommend erring on the side of addressing too much rather than too little.

How is this different from the previous policy?

The [2003 NIH Data Sharing Policy](#) only applied to investigators seeking single-year direct costs of $500,000 or more, and it required researchers to either outline their plans for sharing their data or to explain why they could not. In contrast, the new policy requires *all* investigators seeking any level of funding to submit a plan, and it asks researchers to outline their plans for organizing, storing, and documenting their data (AKA data management) as well as for sharing it. While the new policy acknowledges that legitimate technical, legal, and privacy-related barriers to complete data sharing still exist, it more explicitly urges researchers to share their data promptly in order to increase the overall reproducibility of science.

How will this policy be enforced?

NIH program staff will evaluate the DMS plans and recommend them for approval or denial. The plans will not be shared with peer reviewers and will not be factored into a submission's overall score; however, the budget including costs for data management will be shared with peer reviewers. Once approved, the plan will become a Term and Condition of the award and compliance will be evaluated through annual progress reports. If an investigator does not follow the approved plan, it may affect renewal of the grant and/or future funding decisions. However, the plan can be modified if necessary by working with NIH program officers.

On a local level, ensuring compliance with the plan is left to the investigator. The University of Pittsburgh has not established a person or team who is responsible for making sure researchers comply with their approved plans. Contact [HSLS Data Services](#) to discuss best practices and workflows for setting research teams up for success.

Related Resources

- [Research Covered Under the Data Management & Sharing Policy](#): Overview of included research types from the NIH Scientific Data Sharing website
- [Implementation Details for the NIH Data Management and Sharing Policy](#): Implementation details for the NIH DMS Policy going into effect January 25, 2023, which applies to all NIH extramural grant and cooperative agreement activities
- [Data Management & Sharing Policy Overview](#): Introduction to the new policy from the NIH Scientific Data Sharing website

Complying with the Policy: General Workflow and Timeline

January 25, 2023 is the date on or after which all NIH applicants must plan and budget for managing and sharing data, submit a data management and sharing (DMS) plan when applying for funding, and comply with any approved DMS plan. The steps below outline a general workflow for preparing for and complying with the new policy. Since articulating a DMS plan can be a lengthy process, we recommend starting as early as possible.

Before submission:

1. Become familiar with policy requirements, the required elements of a data management and sharing plan, and the allowable costs that can be included in a budget request by reading this page and associated materials

2. Contact other Pitt resources if necessary: the [Pitt Office of Sponsored Programs (OSP)](#) for information regarding Data Use Agreements, and the [Pitt Human Research Protection Office (HRPO)](#) if you have questions about IRB approval or how to include data sharing in your informed consent forms
3. Choose one or more repositories, or other data-sharing platforms, to share your data; these should be specified in the DMS plan, so select them first and become familiar with their requirements and potential costs
4. Write the DMS plan and the budget. The DMS plan can be submitted through [DMPTool](#) for informal feedback from HSLS before final submission to the NIH, but please allow librarians time to read and offer comments on the application

After submission, during review:

5. If changes to the plan are requested by NIH program officers, make modifications through Just-In-Time (JIT) procedures

After submission and approval:

6. As you conduct research and after, manage your data and create documentation (metadata, data dictionaries, protocols etc.) as outlined in the project's approved DMS plan
7. Provide updates on data management and sharing in annual [progress reports](#) to the NIH, and submit modifications for review/approval as necessary
8. Share data (ideally through a repository) in accordance with the approved DMS plan as soon as possible, **no later than the time of an associated publication or end of performance period**, whichever comes first.

Related resources

- [Pitt Office of Sponsored Programs (OSP)](#): University of Pittsburgh's central office that provides administrative support and oversight related to funding proposals and awards, including Data Use Agreements
- [Pitt Human Research Protection Office (HRPO)](#): University of Pittsburgh's central office that maintains the Institutional Review Board (IRB) and oversight of human subject research
- [DMPTool](#): DMPTool helps researchers create, review, and share data management plans that meet institutional and funder requirements. You can sign on with your University of Pittsburgh e-mail address.

Writing the NIH Data Management & Sharing Plan

The Data Management and Sharing (DMS) plan should be a one- to two-page document submitted with the general funding application as a PDF attachment. If a researcher's application is also subject to the [Genomic Data Sharing Policy](#), they should address GDS-specific topics within this general DMS plan and will no longer submit a separate GDS plan; see the call-out boxes in the NIH's Writing a Data Management & Sharing Plan page for more information on incorporating GDS.

What should be included in a DMS Plan?

The topics to be included in a DMS Plan are laid out in the NIH's Supplemental Policy Information: Elements of an NIH Data Management and Sharing Plan. Briefly, the elements are:

1. **Data types**: the general kinds, approximate amounts or file sizes, and anticipated file formats of data produced in the course of research; which of these data types will be shared and why or why not; and the metadata and explanatory documentation that will help others find, understand, and reuse the data
   - For more information on metadata and documentation, see the HSLS Data Services "Principles of Data Management" guide
2. **Related tools, software, and/or code**: the names of any specialized tools needed to access or manipulate the data, with access instructions if they are researcher-produced tools
3. **Standards**: the names of any data collection instruments, metadata standards, data formats, Common Data Elements, or other data standards that will apply to the collected data, if applicable
   - For more on Common Data Elements, read a brief overview from the National Library of Medicine, "Health Data Standards: A Common Language to Support Research and Health Care," and browse the NIH Common Data Elements (CDE) Repository
4. **Data preservation, access, and associated timelines**: general plans for how, where, and by whom the data will be safely maintained (the preservation aspect), and for how, where, and with whom the data will be shared (the access aspect). Specificity is important in this element, as the NIH guidelines explicitly recommend listing the name of one or more repositories where the data will be preserved and/or made publicly available
5. **Access, distribution, or reuse considerations**: general discussion of any legal, technical, consent, or privacy-related reason preventing the full and timely sharing of all data resulting from the research
6. **Oversight of data management and sharing**: which member of the research team will monitor compliance with the plan (e.g., by title or role), with what frequency, and using which methods

How specific should a DMS Plan be?

In general, more specificity is better because a comprehensive plan requires less spur-of-the-moment decision-making later. This is meant to be a helpful tool for researchers to organize and share their data, not merely a hoop to jump through! That being said, DMS Plans should be under two pages. This rule is true for all NIH applications, but some institutes or centers may have more specific requirements outlined in this list of NIH Institute and Center Data Sharing Policies.

If you are not sure about an element of the Plan yet because it will depend on emergent needs during the course of the study, you can address this uncertainty within the relevant section. For example, if you are developing computational tools to work with your data, you can reference these without specifically naming software that doesn't yet exist. However, you should demonstrate that you are proactively considering these questions, especially when it comes to sharing your data through a repository or other data-sharing platform. If you specify something in your Plan that you want to change later, you may do so with approval of NIH staff.

Are there templates or examples of a DMS Plan?

The NIH is developing a format page that aligns with the elements given above. Currently (as of 2022-10-09), a [preview version of the format page](#) is available, with a final fillable format expected later this fall.

The best way to create a DMS Plan is by using the templates available through [DMPTool](#), a service that allows you to submit draft DMS plans to HSLS librarians feedback before official submission. The DMPTool team has updated their NIH template to reflect the requirements of the new policy, and will keep it up to date as new guidance is announced. For one-on-one help, contact [HSLS Data Services](#).

Related resources

- [DMPTool](#): DMPTool helps researchers create, review, and share data management plans that meet institutional and funder requirements. You can sign on with your University of Pittsburgh e-mail address.
- [Implementation Changes for Genomic Data Sharing Plans Included with Applications Due on or after January 25, 2023](#): Announcement of the NIH's intention to "establish a single Plan submission requirement for research subject to both the [GDS Policy and the NIH DMS Policy] ... [that] will take effect for Plan submission due dates on or after January 25, 2023"
- [NIH Common Data Elements (CDE) Repository](#): The NIH CDE Repository hosts over 29,000 standardized, structured definitions of data elements used in scientific research
- [Implementation Details for the NIH Data Management and Sharing Policy](#): Implementation details for the NIH DMS Policy going into effect January 25, 2023, which applies to all NIH extramural grant and cooperative agreement activities
- [Supplemental Information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan](#): Brief description of the Elements to be addressed within the short Data Management and Sharing Plan submitted with NIH funding applications after January 25, 2023
- [Writing a Data Management & Sharing Plan](#): Guidance on writing a DMSP from the NIH Scientific Data Sharing website

Costs for Data Management and Sharing

Managing your data and making it publicly available in an easy-to-use format with clear documentation costs money, both real and as time spent. The NIH allows investigators to request money for data management and sharing in the budget and budget justification sections of their applications. Specific information is included in the NIH's [Supplemental Policy Information: Allowable Costs for Data Management and Sharing](#).

In general, reasonable costs for the following are allowable:

- Curating data, which could include evaluating it for completeness, harmonizing datasets, and performing various storage- and maintenance-related tasks
- Creating documentation such as READMEs, data dictionaries, and codebooks, and metadata to help make the datasets discoverable on the web by both humans and machines
- Cleaning, de-identifying, and formatting data
- Specific local data management considerations

- Fees for repositories or other data-sharing platforms, such as data-processing charges, membership charges, or cloud storage costs

The following costs are **not** allowable:

- Infrastructure costs included in institutional overhead
- Costs associated with the routine conduct of research, including costs of collecting or accessing data
- Costs that are double charged as both direct and indirect costs

All costs that are included in the budget must be incurred during the performance period regardless of how long the data will remain available. Budgets, but not the associated Data Management & Sharing Plans, will be shared with peer reviewers for assessment of their reasonableness. The NIH has not announced a separate source of money specifically for data management and sharing costs, and these costs should be treated as a part of the overall budget proposal.

The publications within "[Forecasting Costs for Preserving, Archiving, and Promoting Access to Biomedical Data](#)," a project of the National Academies, may be useful when determining costs. However, we highly recommend contacting HSLS Data Services for help anticipating the costs of storing or sharing data in a repository. Many data repositories are free, but some have storage caps (either per-file or overall) that require data processing charges not otherwise advertised. We can help you determine the most appropriate home for your data before you name it in your DMS Plan and request funds in your budget.

Related resources
- [Supplemental Information to the NIH Policy for Data Management and Sharing: Allowable Costs for Data Management and Sharing](#): Summary of the types of costs related to data management and sharing that researchers may include in their budget proposal after January 25, 2023
- [Budgeting for Data Management & Sharing](#): Overview of allowable costs from the NIH Scientific Data Sharing website

Selecting a Repository or Data-Sharing Platform

The NIH strongly recommends that researchers deposit their data into a public data repository for long-term storage and access. If access to the data needs to be restricted, controlled-access repositories or other data-sharing platforms are available, but sharing data via email by request or hosting it on a lab server will not meet the policy's requirements in most cases.

What is a data repository and why should I use one?
A data repository is a platform for hosting research datasets that enables them to be findable, accessible, interoperable, and reusable by researchers and the public all around the world. (See the [FAIR Data Principles](#), which were developed to optimize the reuse of scientific research data). There are many reasons to use a data repository instead of a lab website, FTP server, or cloud storage like Google Drive:

- Data repositories are less work for researchers. They have dedicated systems and staff to keep all the uploaded data organized, and provide instructions on finding and downloading data. It's someone else's job to run a repository, which means it doesn't have to be yours!
- Data repositories contribute to the scientific record. All submissions to a (good) data repository are time-stamped and issued version numbers, so subsequent changes and re-uploads can be tracked. Most repositories also issue DOIs, or Digital Object Identifiers, which can establish a particular file as the "version of record" and enable authors to cite datasets in journal articles. This makes it easier to demonstrate priority, provenance, and compliance with data governance policies.
- Data repositories are reliable. They monitor files for corruption and back up all data in case of disaster. A good data repository will make their preservation plan publicly available and tell you how long they promise to keep your data. They don't depend on one key staff member to keep the whole thing running, like a lab website might, and they won't suddenly break all their links because a local IT department changed strategy.
- Data repositories are becoming part of standard scientific practice. When researchers and reviewers look for data, they increasingly turn first to public data repositories. A growing number of journals are integrating with repositories to link to supporting data for publications, and may automatically deposit your supplementary data in a repository if you haven't already uploaded it there.

Which data repository should I use for my data?

There are many, many repositories, some of which take all kinds of research material (publications, data, posters, etc.) and some of which specialize in a content type or discipline. There isn't one "best" repository, but rather one "best suited for your particular data."

Here are some questions to lead you to a possible repository match:

- Does your funding announcement or program specify a particular data repository to use? (For example, the NIH Common Fund's Stimulating Peripheral Activity to Relieve Conditions (SPARC) program has a dedicated SPARC Portal.) If so, use that repository.
- Does your research area have a dedicated repository where everyone in your field puts their data? Here are a few ways to find domain-specific repositories:
    - Browse the HSLS Data Services Locating and Sharing Data guide for an overview of repository types and examples
    - Look through the NIH's list of discipline-specific repositories for sharing scientific data, which includes repositories that are open for anyone to submit/download, and data repositories that restrict access in some way
    - Search Re3Data, the Registry of Research Data Repositories, for terms related to your field. This will probably turn up smaller repositories, or repositories focused on a particular geographic region. Make sure they seem robust enough to host your data for the long term.
    - Ask colleagues where they are sharing their data. Sometimes big generalist repositories (repositories which take all disciplines) have sub-collections focusing on particular fields, and word of mouth is the best way to discover them.
- Some repositories specialize in a particular type of data rather than data from a specific field. Examples include OpenNeuro, which accepts BIDS-compliant MRI, PET, MEG, EEG, and iEEG data

regardless of specialization, and the Cancer Imaging Archive, which hosts de-identified medical images related to cancer from a huge variety of disciplines. Is there a repository for your modality of data?

- Use a well-resourced generalist repository that takes data from all disciplines. See the NIH's short list of generalist repositories, some of which also take non-data materials like publications. In some cases, that may be a bonus: in Zenodo, for example, you can share your analysis code alongside your data files by linking your data submission with your Github repository.
- The NIH policy also allows for using institutional repositories, which are run by a researcher's affiliated institution. D-Scholarship, Pitt's institutional repository, is not optimized for data but is available as an option.

It can be difficult to choose among the large generalist repositories, although all of the options on the NIH list meet the agency's Desirable Characteristics for All Data Repositories checklist. This Generalist Repository Comparison Chart provides a quick reference for the repositories' costs, storage caps, and limitations, although it may not include all repositories currently available.

Licenses and use restrictions

Most repositories will ask you to apply a license to your uploaded data so that users know what they may and may not do with your work. Since the purpose of sharing data is to facilitate data reuse and increase reproducibility, some repositories specifically require you to apply a Creative Commons 0 "No Rights Reserved" license. This means that other people may download, re-analyze, re-share, and otherwise re-use your data, but it does not exempt them from the standard expectations of citation and giving academic credit.

Note that choosing not to apply a license is the same thing, legally speaking, as stating "all rights reserved." In the strictest interpretation, this means that a user wouldn't even be allowed to email a copy of your dataset to themselves because it might be unauthorized copying! However, people may interpret a lack of license as meaning that they can do *anything* with your data. To prevent confusion, it is best to choose an appropriate license and clearly attach it to your files as a text file bundled with a data download or as text in a metadata record.

Software and code, including research analysis scripts, also benefit from clear licenses. Github has a simple, interactive tool that suggests and explains appropriate code licenses at ChooseALicense.com.

Related resources

- Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data: Best practices for managing and sharing human research subject data, addressing both identifiable and de-identified data
- Supplemental Information to the NIH Policy for Data Management and Sharing: Responsible Management and Sharing of American Indian/Alaska Native Participant Data: Recommendations for protecting human subject data from American Indian and Alaska Native participants and for working with tribal governments and organizations
- Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research: Guidance on choosing a data repository for sharing scientific research data, including desirable characteristics for all data repositories and specific considerations for protecting human data

- [Selecting a Data Repository](#): Guidance on evaluating repositories from the NIH Scientific Data Sharing website
- [Repositories for Sharing Scientific Data](#): List of NIH-supported domain-specific repositories from the NIH Scientific Data Sharing website, with links to generalist repositories
- [re3data](#): Registry of Research Data Repositories. Searchable interface but can also browse by subject, content type and country.

Frequently Asked Questions (FAQs)

Q: If I submit my data to a repository, can I remove it later?
A: In general, no. If you make a mistake or want to re-upload a new version, you can resubmit your files, but the old files will remain visible as a previous version in order to preserve the scientific record. If you encounter legal or confidentiality issues, you can request that your files be withdrawn from the repository, but usually a metadata-only record will remain that describes the files (in terms of author, title, etc.) that used to be there.

Q: When I submit my data to a repository, am I giving up any rights?
A: In general, no. Anyone who uses your data that they found in a repository should acknowledge/cite/credit you appropriately. (Many repositories actually make that easy by issuing DOIs to all data submissions, which are frequently required to cite datasets.) In some cases, repositories do require you to license your submitted data with a Creative Commons Zero (CC0) license, essentially putting it in the public domain. This helps enable replication and reuse, and data is generally not protected by copyright anyway. A CC0 license does not exempt anyone from the normal expectations of scholarly credit as mentioned above.

Q: My data files are really big. Will a repository accept my data for deposit?
A: Each repository has different file size limits, both per-file and per-user. Dryad is one of the biggest, accepting up to 300GB, but they charge additional storage fees over 50GB. If you have truly massive datasets (especially common with images or video), contact [HSLS Data Services](#) and we will help you find a solution.

Q: My data includes protected health information. Can I share/do I still have to share it?
A: Yes, you can share it (and the NIH expects you to) if you can reasonably de-identify everything in your data and you have informed consent to do so. Most repositories state that they do not accept personally identifiable health information, and that by uploading your data you are certifying that it is appropriately scrubbed. In some cases, you may still want to restrict access to your data only to qualified or vetted researchers. Contact [HSLS Data Services](#) and we can help you navigate next steps.

Q: My data is managed by a Data Use Agreement. Who do I contact to find out more about how this works with the NIH policy?
A: Data Use Agreements (DUAs) are handled by the [Pitt Office of Sponsored Programs (OSP)](#). See their website for more information or contact them for assistance.

Q: I am concerned about whether data sharing is allowed by the informed consent forms my research participants signed. How do I find out more about what I am allowed to share?
A: Contact the IRB (Institutional Review Board) help line through the [Pitt Human Research Protection](#)

Office (HRPO). They can help you understand what is and is not allowed in your current informed consent forms, and write language that aligns with the new NIH requirements for your future forms.

A: I have made a sincere effort to share my data in accordance with the policy, but there simply isn't a solution that fits all of my requirements (due to size, privacy requirements, etc.) Is there any way I can share my data upon request with vetted researchers without going through a repository, and will that satisfy the NIH requirements?

A: The NIH understands that there will be cases in which data can't be completely shared. In that situation, it's important to demonstrate in your DMS Plan that you have considered all of the available options and still have specific constraints or concerns. If you would like to publicly describe your data and invite qualified researchers to apply for full or increased data access, we can help you create a metadata-only record in a data catalog like the Pitt Data Catalog. This strategy can also help you increase the findability of large datasets hosted on a cloud storage service. Contact HSLS Data Services for more information.