FAQ for the 2023 NIH Policy

The National Institutes of Health (NIH) has issued a new Policy for Data Management and Sharing (DMS) when research funded or conducted by NIH results in the generation of scientific data. This policy is scheduled to go into effect on January 25, 2023.

The NIH is viewing this as a pilot program and we fully expect this to be an iterative process. This Frequently Asked Questions (FAQs) document was developed in consultation with our Affiliate Institutions and includes Broad-focused answers based on current NIH guidance.

These are intended to help clarify the implementation of the NIH Policy for Data Management and Sharing at The Broad Institute, **and will be updated on an ongoing basis**.

More information can be found on The Broad Institute Data Management Sharing Plan (DMSP) Briefing Sheet. Genomic researchers have a long tradition of data sharing through mechanisms like dbGaP. It is often required for the journals Broad Investigators want to publish in. While this is a new requirement and we encourage everyone to read this entire document and investigators must follow the new template, this is not a major departure from what most investigators have done in the past.

For more FAQs, please refer to the FAQ list compiled by NIH.

1) **What is considered "scientific data" for the purposes of this plan?**

   The final NIH Policy defines Scientific Data as: "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. Scientific data do not include 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." Even the scientific data that are not used to support a publication are considered scientific data and within the final DMS Policy's scope.

   What does that mean? In consultation with our peer institutions we're interpreting this to mean that, at a minimum, the data behind any tables or figures in a publication need to be part of the DMSP, as well as any data collected alongside it.

2) **What are the FAIR Data Principles?**

   NIH encourages data management and sharing practices to be consistent with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles and reflective of practices within specific research communities. In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in *Scientific Data*. The principles emphasize machine-actionability because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

3) **Does this new policy apply to grants already in progress?**

   No. The policy only applies to new competing grant applications, it does not retroactively apply. The effective date of the DMS Policy is January 25, 2023, including for:

   - Competing grant applications that are submitted to NIH for the January 25, 2023 and subsequent receipt dates
   - Proposals for contracts that are submitted to NIH on or after January 25, 2023
   - NIH Intramural Research Projects conducted on or after January 25, 2023
   - Other funding agreements (e.g., Other Transactions) that are executed on or after January 25, 2023, unless otherwise stipulated by NIH

   For guidance on applications for receipt dates **BEFORE** January 25, 2023, refer to the 2003 NIH

Data Sharing Policy.

4) **How will the plans be assessed?**

NIH program staff will assess the DMSPs, but peer reviewers may comment on the proposed budget for data management and sharing. "The final DMS Policy maintains NIH Program Staff assessments of Plans' merits. However, peer reviewers may comment on the proposed budget for data management and sharing, although these comments will not impact the overall score…Over time, and through these reviews, we hope to learn more about what constitutes reasonable costs for various data management and sharing activities across the NIH portfolio of research." See more under Section VI of the Final NIH Policy.

5) **Is the Data Safety Application the same as a Data Management and Sharing Plan?**

No. Submission in the Data Safety Application is not equivalent to the DMSP required by the NIH (but is still a requirement).

A DMSP is required for any proposal involving Scientific Data and includes six sections, each with specific requirements associated with the Data. The submitted DMSP will be assessed by peers and NIH administrators, and will directly impact whether the proposal is accepted.

6) **What is a data or metadata standard? What standards are relevant to my research?**

Data standards specify how data and related materials should be stored, organized, and described. In the context of research data, the term typically refers to the use of specific and well-defined formats, schemas, vocabularies, and ontologies in the description and organization of data. However, for researchers within a community where more formal standards have not been well established, it can also be interpreted more broadly to refer to the adoption of the same (or similar) data management-related activities, conventions, or strategies by different researchers and across different projects.

7) **Am I expected to share all data generated during my research?**

No. Under the DMS Policy, researchers are expected to maximize the appropriate sharing of scientific data, which is defined as data commonly accepted in the scientific community as being of sufficient quality to validate and replicate the research findings. Not all data generated during NIH-supported research will constitute scientific data under the DMS Policy. See the NIH FAQ for more.

NIH Institutes, Centers, or Offices (ICOs), Notice of Funding Opportunities (NOFOs), funding opportunity announcements (FOAs), and other NIH policies (e.g., the Genomic Data Sharing Policy) may have additional expectations for what data should be shared.

Researchers are expected to maximize appropriate sharing of any new, derived data generated as a result of their research. Note that use of data obtained from repositories or other sources and derived data may be subject to limitations on sharing as a condition of access.

8) **What data repository should I use?**

Some programs, types of data, ICOs, or Funding Opportunity Announcements (FOAs) may require data deposition in particular data repositories, and "primary consideration should be given to data repositories that are discipline or data-type specific to support effective data discovery and reuse." NIH encourages the use of established repositories. To select a repository relevant to your data

consider:

- Is there a specific NIH repository named in the FOA?
- Is there a data repository specific to the data type(s) relevant to your research and your scientific discipline?
- Is there a data repository specified by the journal in which you are publishing or hope to publish?
- If there are no relevant discipline-specific repositories, is there a generalist data repository you can use?

For data generated from research for which no data repository is specified by NIH, researchers are encouraged to select a data repository that is appropriate for the data generated from the research project and is in accordance with the NIH Desirable Characteristics for All Data Repositories. To learn more, check out the NIH guidance on selecting a data repository.

There are also DMS options at Broad - Terra, DUOS and Broad's Data Access Committee

Terra, DUOS and Broad' Data Access Committee are three internal offerings that may be of value in helping researchers meet the terms of the DMS policy.

- **Terra - a platform for data storage, version control, and researcher analysis, and secure collaboration.** Terra meets all of the DMS requirements and is already approved as NIH-Supported Scientific Data Repository (Terra powers NHGRI AnVIL and NCI GDC).
- **DUOS - an application for registering controlled-access datasets with a data access committee, and making them available in a public catalog with persistent identifier (PID)** DUOS satisfies multiple DMS requirements, most specifically for controlled-access data.
- **Broad Data Access Committee (DAC)** - an additional service offered by Broad for datasets that use DUOS to broker access to data. In this case, a dedicated team from the Broad serves as the DAC for a given set of controlled access datasets. Broad's DAC can help ease the operational lift and liability for researchers in ensuring secondary data use access requests are approved and compliant with Data use Limitations.

Researchers who need a robust solution to meet the DMS are welcome to use any combination of Terra, DUOS and or Broad's DAC. The costs associated with use of these services depend on your project needs, but there are cost options ranging from mostly free / self-service to more formal engagements where one or more teams provide services to ensure the DMS requirements are met.

.

9) **Where can I find more info about these offerings?**

Information on Terra is located at terra.bio, and DUOS information is available at a high-level at duos.org, with more detailed user guidance on DUOS is located here. The DSP team will be converging all of this information with DMS specific guidance under one website in the coming weeks.

In the interim, attached is a DMS Plan template that includes language on Terra, DUOS and Broad's DAC. If you would like to use Terra / DUOS / Broad's DAC for satisfaction of the DMS, please consult the Broad Data Sciences Platform team prior to submission of any DMS Plan that involves Terra and or DUOS. The team can be consulted at dsp-pi@broadinstitute.org.

FAQ for the 2023 NIH Policy

10) Can I make my data available only upon request?

NIH expects that researchers will take steps to maximize scientific data sharing, but acknowledges that certain factors (i.e., ethical, legal, or technical) may necessitate limiting sharing, to some extent.

Foreseeable limitations (e.g. bounds of consent documentation, substantial risk to privacy of data subjects, restrictions imposed by regulations or contract), the proposed method for disseminating such data, and the rationale must be described in the DMSPs for the NIH to assess.

11) When do I need to make my data available?

NIH encourages scientific data to be shared **as soon as possible, and no later than at the time of an associated publication or the end of the performance period, whichever comes first.**

The time of an associated publication: Scientific data underlying peer-reviewed journal articles should be made accessible no later than the date on which the article is first made available in print or electronic format.

The end of the performance period: Scientific data underlying findings not disseminated through peer- reviewed journal articles should be shared by the end of the performance period unless the grant enters a no-cost extension. If a no cost extension is permitted, then the recipient should share the data by the end of the extended performance period. These scientific data may underlie unpublished key findings, developments, and conclusions; or findings documented within preprints, conference proceedings, or book chapters. For example, scientific data underlying null and negative findings are important to share even though these key findings are not always published. Researchers should be aware that some preprint servers may require the sharing of data upon preprint posting, and repositories storing data may similarly require public release of data upon preprint posting.

12) What happens if I do not comply with the NIH policy or make my data available as described in the DMSP?

- NIH Program Staff will be monitoring compliance with the policy during the funding period. “Noncompliance with Plans may result in the NIH ICO adding special Terms and Conditions of Award or terminating the award. If award recipients are not compliant with Plans at the end of the award, noncompliance may be factored into future funding decisions.” See more under Section VIII of the Final NIH Policy.

13) How do I get started writing my plan, is there a template?

- Yes, NIH has a template.
- A DMS template with language on Broad’s Terra, DUOS, and DAC offerings is also available.
- We also recommend using DMPTool which is used by many of our peer institutions. This tool walks you through each section of the NIH form, along with helpful guidance. We’ve created a step-by-step guide to help you get started.

14) **What data management and sharing costs can I include in my grant?**

Allowable costs can include those for:

- data curation and developing documentation (i.e., formatting data, de-identifying

data, preparing metadata, curating data for a data repository)
- data management systems (e.g., unique and specialized information infrastructure necessary to provide local management and preservation before depositing data in a repository)
- preserving data in data repositories (i.e., data deposit fees)

Read the NIH Supplemental Information on Allowable Costs for Data Management and Sharing.

15) Are there any tools to help plan my budget?

- DSP can provide a budget template and work with investigators individually to estimate costs for data management and sharing (dsp-pi@broadinstitute.org).
- More generally, Amazon Web Services (AWS) created a generic cost estimation tool that has been shared in the NIMH Data Archive (NDA) and may be of help to study investigators, available here.