FAQ for the 2023 NIH Policy

This Frequently Asked Questions (FAQs) document includes Harvard-focused answers based on current NIH guidance. They are intended to help clarify the implementation of the NIH Policy for Data Management and Sharing at Harvard University and will be updated on an ongoing basis. 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.

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. 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.

The Data Safety Application is an internal system for routing and recording Harvard researchers' data security plans. Researchers are only required to submit a request in the Data Safety Application if their scope of work involves Sensitive Research Data (Harvard Data Security Level (DSL) 3 or above), a Data Use Agreement, or as otherwise described in the Research Data Security Policy. If the proposal is successful, and Sensitive Research Data is implicated, the Principal Investigator must submit a request in the Safety Application, and upload the DMSP as part of the request, so the cognizant security reviewer and other research team members are aware of the immediate and long-term expectations for the data. The Research Support site has additional information on the Data Safety review processes. Don't hesitate to reach out to your School Security Officer with specific questions.

6) How do I determine the security level of my data?

The researcher managing the overarching project is responsible for assessing the sensitivity of the data throughout the project's lifecycle. This includes being aware of associated internal and external requirements and limitations that may apply to the management of the data, either per Harvard policy, IRB documentation, contractual obligations, or regulatory mandate. If the research team has any questions pertaining to the applicable DSL, they should speak to their local Information Security Officer. If at any point during the project's period of performance the data is determined to be Sensitive Research Data, the research team must immediately submit a request for security review in the Data Safety Application.

Harvard researchers frequently deal with sensitive information that relates to human subjects and other research areas. Examples can include proprietary information, personally identifiable information, and data that is subject to confidentiality requirements or domestic regulations. Most of these types of information will be categorized as DSL 3. However, certain personally identifiable data that could directly impact individuals' safety or financial standing, as well as certain regulated data (e.g. GDPR, CMMC, NIST 800-171), information with national security or export control implications, and medical information, is usually categorized as DSL 4 data. Harvard researchers must submit any such projects for Security Review by a local information security reviewer in the Data Safety Application.

7) Where can I store my data?

A majority of the University's tools and resources are consistent with at least DSL 3 standards, as well as many external regulations. These Harvard-managed tools are protected by contractual restrictions and security measures not available for consumer tools. Utilizing Harvard resources, as opposed to vendor or home-grown solutions will help ensure compliance with internal and external requirements, and expedite Data Safety and IRB reviews. A great place to start is with the Harvard Information Security Collaboration Tools Matrix or Research Support at Harvard for Active Research.

'Short-term storage' is typically described as the length of a research process from initiation to final publication. Your research data must be stored in a place that is secure and accessible while it's actively

being collected and analyzed. Again, refer to appropriate University tools and resources for short-term storage and/or file sharing services.

Long-term storage and preservation seek to ensure that research data will be available to those who seek it (e.g., your sponsors, the public, and other researchers) in a persistent and accessible format for the specific period of time outlined by your funder and parent institution. As the designated steward of research data, faculty are required to retain research records for a period of no fewer than seven (7) years after the end of a research project or activity. Refer to OVPR's Retention and Maintenance of Research Records and Data.

Permanent retention or archiving refers to the ongoing migration of electronic formats and storage costs, as well as care, maintenance, and access services for the records in perpetuity. The Records Management Services, a department of the Harvard University Archives, provides guidance to University staff, faculty, and administrators on how to understand their responsibilities for stewarding and managing their records.

8) 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. Learn more about data standards from the Harvard Biomedical Data Management Website and Research Support.

9) 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.

10) 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?
    - Consider Harvard Dataverse or other Harvard-suggested options

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 Harvard Biomedical Data Management guide to data repositories or the NIH guidance on selecting a data repository.

11) 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.

12) 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.

13) 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.

14) 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.

15) Who owns the research data I produce and what are my responsibilities regarding its management?

The University's Research Data Ownership Policy states that Harvard University asserts ownership over research data for all projects conducted at the University, under the auspices of the University, or with University resources. This enables the University to respond to inquiries from funders and third parties, as well as appropriately protect the data, data subjects, and researchers. Principal Investigators (PIs) and other researchers are stewards and custodians of research data. Therefore, the PI's responsibilities with respect to research data include:

- Ensuring proper management and retention of research data in accordance with the Guidance on Retention and Records of Research Records and Data
- Establishing and maintaining appropriate procedures for the protection of research data
- Ensuring compliance with program requirements
- Maintaining confidentiality of research data
- Maintaining appropriate data use agreements for the sharing of research data
- Complying with applicable federal, state, and local laws and regulations

Permanent retention or archiving refers to the ongoing migration of electronic formats and storage costs, as well as care, maintenance, and access services for the records in perpetuity. The Records Management Services, a department of the Harvard University Archives, provides guidance to University staff, faculty, and administrators on how to understand their responsibilities for stewarding and managing their records.

16) Where can I get help at Harvard?

The Harvard Library's Research Data Management Program is happy to help with selecting a repository, choosing a metadata standard, or answering questions about the policy. We can also refer you to other relevant groups on campus. See the Data Services Network for the appropriate library contact. School research administration offices are a great resource for questions related to NIH policy and requirements. Find your Office for Sponsored Programs Pre-Award or Research Finance contact.