
DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on [sharing.nih.gov](https://www.nih.gov/data-management/data-sharing). The Plan is recommended not to exceed two pages. **Text in italics should be deleted.** There is no “form page” for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Element 1: Data Type

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

Summarize the types and estimated amount of scientific data expected to be generated in the project.

Sample language:

This project will produce _____ [Data type, e.g., imaging, sequencing, experimental measurements] data generated/obtained from _____ [e.g., instrument, method, survey, experiment, data repository]. Data will be collected from ____ [number] of research participants/specimens/experiments, generating ____ [number] datasets totaling approximately ____ [amount of data] in size. The following data files will be used or produced in the course of the project: _____ [list input data files, intermediate files, and final, post-processed files]. Raw data will be transformed by _____ [analysis, method] and the subsequent processed dataset used for statistical analysis. To protect research participant identities, _____ [e.g., individual, aggregated, summarized] data will be made available for sharing.

B. Scientific data that will be preserved and shared, and the rationale for doing so:

Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

Sample language:

Based on _____ [ethical, legal, technical] considerations, the following data produced in the course of the project will be preserved and shared: _____ [list] **OR** All data produced in the course of the project will be preserved and shared.

C. Metadata, other relevant data, and associated documentation:

Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

Sample language:

To facilitate interpretation of the data, _____ [e.g., metadata, documentation, protocols, data collection instruments] will be shared and associated with the relevant datasets.

Element 2: Related Tools, Software and/or Code:

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

Sample language:

If no specialized tools are needed to access or manipulate the data:

_____ [Data type - Imaging data, survey data, etc] data will be made available in _____ [csv, txt, dicom, etc] format and will not require the use of specialized tools to be accessed or manipulated.

If specialized tools are needed to access or manipulate the data:

_____ [Data type] data will be made available in _____ format, which requires the use of specialized tools, such as _____ [include list of tools] to be accessed and manipulated.

If applicable, specify how needed tools can be accessed, (e.g., open source and freely available, generally available for a fee in the marketplace, available only from the research team) and, if known, whether such tools are likely to remain available for as long as the scientific data remain available.

Sample language:

- The _____ tool, which can be used to _____ is available free of charge through _____ [source name]

- The _____ tool, which can be used to _____ is available for a fee of _____ through _____ [source name].
- Custom tools to _____ will be/have been developed by the research team.
 - Requests for these tools should be directed to _____ [include details of members of the research team].
 - These tools will be shared openly via _____.

Element 3: Standards:

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

Sample language:

To facilitate their efficient use, all of our data and materials will be structured and described using the following standards:

If there are formal data standards for some/all of the data:

Whenever possible, we will use _____ [common data elements, standardized survey instruments, etc] to structure and organize our data.

Our _____ data will be structured and described using the _____ standard, which has been widely adopted in the _____ community. [Add additional information about this standard, if applicable - e.g. implementation in data repositories, utility in combining/reusing datasets]

If there are not formal standards:

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

Data will be stored in common and open formats, such as _____ for our _____ data. Information needed to make use of this data [e.g. the meaning of variable names, codes, information about missing data, other metadata etc] will be recorded in _____ [data dictionaries/codebooks] that will be accessible to the research team and will subsequently be shared alongside final datasets.

Information about our research process, including the details of our analysis pipeline will be maintained contemporaneously, using _____ [lab notebooks, protocols, etc]. This information will be accessible to all members of the research team and will be shared alongside our data.

Element 4: Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:

Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see [Selecting a Data Repository](#).

Sample language:

All dataset(s) that can be shared will be deposited in _____ [Add appropriate NIH-supported data repositories] OR _____ [Add appropriate subject or disease repositories]

Sample Language for Dryad Data Repository

Dataset(s) resulting from this research will be shared via the generalist repository Dryad, which provides metadata, persistent identifiers (i.e., DOIs), and long-term access. Dryad is the institutional data repository supported by the University of California and all data is shared under a CC0 waiver, which makes the dataset(s) publicly available. Data will be made available as soon as possible or at the time of associated publication. Dryad datasets are backed up to Merritt, the UC's CoreTrustSeal-certified digital repository, for long-term storage and accessibility. Procedures in place to ensure dataset preservation include storage of data files in multiple geographic locations, regular audits for fixity and authenticity, and succession plans in the event of repository closure.

B. How scientific data will be findable and identifiable:

Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Sample language:

The _____ [Insert repository name] provides metadata, persistent identifiers (i.e., insert whether DOI, handles, other), and long-term access. This repository is supported by _____ [Insert funder/organization] and dataset(s) are available under a _____ [Insert license information] **OR** through a request process

_____ *[Insert information about request process].*

C. When and how long the scientific data will be made available:

Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Sample language:

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

Element 5: Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See [Frequently Asked Questions](#) for examples of justifiable reasons for limiting sharing of data.

B. Whether access to scientific data will be controlled:

State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

C. Protections for privacy, rights, and confidentiality of human research participants:

If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

Sample language:

For researchers working with human subjects data

In order to ensure participant consent for data sharing, IRB paperwork and informed consent documents will include language describing plans for data management and sharing data, describing the motivation for sharing, and explaining that personal identifying information will be removed.

To protect participant privacy and confidentiality, shared data will be de-identified using the _____ method. [Describe de-identification method, noting any other applicable laws or policies such as HIPAA].

For researchers selecting controlled access repositories

Given the sensitive nature of the dataset, de-identified human subjects data will be made available in _____ data repository, which restricts access to the data to qualified investigators with an appropriate research question who sign a data use agreement. [Describe data repository access methods and security measures].

Element 6: Oversight of Data Management and Sharing:

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

Sample language:

The following individuals [or just the position titles if unknown] will be responsible for data collection, management, storage, retention, and dissemination of project data, including updating and revising the Data Management and Sharing Plan when necessary.

- Name, Position Title, Host Institution, ORCID, email

Sample Language for budgeting requirements

This project includes the following costs associated with data management and sharing.

For data curation and the development of related documentation, the project is requesting \$_____. These funds will allow us to prepare data for sharing including de-identification of data, the incorporation of metadata to ensure discoverability and the data transfer process to _____ repository for preservation and access. An additional cost of \$ _____ is required to cover data deposit fees for _____ repository, which will cover _____ years of hosting.