

Tips for Writing a Data Management and Sharing (DMS) Plan

Researchers: Use the following tips to create a DMS Plan for your NIH-funded research projects. Refer to the detailed instructions in the application guide for developing your plan as well as to additional guidance on <https://sharing.nih.gov/>.

Overall Policy Reminders:

- Under the [2023 NIH DMS Policy](#), NIH requires all applicants proposing research that will generate scientific data to prepare a DMS Plan that describes how the scientific data will be managed and shared.
- The DMS Policy expects that researchers maximize the appropriate sharing of scientific data generated in the project by making data accessible to the broader research community or public through established repositories as soon as possible, and no later than the time of an associated publication, or the end of the award/support period, whichever comes first.
- DMS Plans may be made publicly available and should not include proprietary or private information.
- Applications subject to NIH's Genomic Data Sharing (GDS) Policy should also address GDS-specific considerations in the DMS Plan (described in [NOT-OD-22-198](#)).

As you review these tips, keep the following **definitions** in mind:

- **Scientific Data:** 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.
- **Data Sharing:** The act of making scientific data available for use by others (e.g., larger research community, institutions, public), such as via an established repository.

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.

Tips:

- For each data type, describe the source (e.g., human, other species or model organism, cells, sample types, administrative systems, machines), formats, and other descriptive details.
- Use any appropriate metrics to describe the projected scale of the data. If data will be collected from participants, subjects, or samples, consider describing the amount of data in terms of the projected number of participants, subjects, or samples. Other metrics may include the number of files and/or their anticipated volume (i.e., gigabytes, etc.).
- For genetic and genomic data types, provide sufficient detail to support [whether the project is subject to the GDS policy](#).

Policy Reminder: Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be).

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.

Tip: Describe which scientific data will be shared for each data type described in 1A. For any scientific data that will not be shared, provide a clear legal, ethical, or technical justification.

Policy Reminder: NIH expects that in drafting Plans, researchers will maximize the appropriate sharing of scientific data, acknowledging certain factors (i.e., legal, ethical, or technical) that may affect the extent to which scientific data are preserved and shared. Specific [data types and formats are expected for projects subject to the GDS Policy](#).

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.

Tips:

- Metadata and documentation associated with a dataset allow users to understand how the data were collected and processed, and how to interpret the data for reuse.
 - [Metadata](#) is information that enables discovery, reuse, and citation of the dataset. Metadata are collected and shared using a structure (schema) appropriate to, and ideally widely used across, a research community.
 - Examples of dataset-associated [documentation](#) include experiment protocols, data dictionaries, code books, study procedures, study visit timeline, questionnaire(s) and scoring instructions, and reference ranges for laboratory results.
- The appropriate metadata and associated documentation for a dataset will vary by scientific area, study design, the type of data collected, and characteristics of the dataset. To determine which metadata and associated documentation to share, look to your data repository's requirements, your research community's best practices, and other information that is critical to understand your data (e.g., experimental metadata).

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.

Tip: Describe how the relevant tools, software, or code can be accessed (e.g., [open-source license](#), available for a fee in the marketplace, available only from the research team).

Policy Reminder: The description of tools, software, and/or code in the DMS Plan should be specific to what is needed to "access or manipulate shared scientific data to support replication or reuse." For applications proposing to develop and share software or tools, use the "Resource Sharing Plan" section of the application rather than in the DMS Plan.

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.

Tips:

- [Data standards](#) specify how data and related materials should be stored, organized, and described, and improve data usability and interoperability. In the context of research data, standards typically refer to formats, schemas, vocabularies, ontologies, common data elements, and common data models that define the description and organization of data.
- To identify standards to use for collecting, managing, and sharing data, look to the data repository where the data will be submitted, standards that have been adopted by a specific scientific community, or data standards that may be specified by the NIH, for example in a Notice of Funding Opportunity (NOFO). Review [NICHD-relevant data standards](#).

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](#).

Tip: First identify data repositories that are specific to your research domain or data type by checking the [NICHD Data Repository Finder](#) or NLM's list of [NIH-Supported Data Sharing Resources](#); these repositories benefit from the expertise and established best practices of a given scientific community. If no appropriate domain or data type specific repository is available, consider "[generalist repositories](#)" that accept data regardless of data type, format, content, or disciplinary focus.

Policy Reminder: The DMS Policy strongly encourages the use of established repositories. Repositories used for sharing scientific data should make data accessible to the larger research community or public and adhere to other desirable characteristics described in [Selecting a Repository for Data Resulting from NIH-Supported Research \(NOT-OD-21-016\)](#).

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.

Tip: Find out if your selected data repository generates persistent unique identifiers for datasets, and which type. Data repositories typically create digital object identifiers (DOIs) or accession

numbers as a component of their data release 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.

Tips:

- Describe the timeline for data submission and/or sharing in terms of milestones, events, or relative dates (e.g., “during the third year of the project” or “4 months before publication or 6 months before the end of the award period, whichever is sooner”).
- Plan to submit data to a repository several months prior to expected public release (i.e., “data accessibility”) to leave sufficient time for the data repository to process the data and work with your data team to address any quality assurance/quality control questions. Do not wait until the end of the award period to submit all the data.
- Plan to share your data for as long as possible within the repository’s data retention policies.

Policy Reminder: Genomic data must adhere to the [GDS Policy timelines](#), which generally for human data expects submitting data 3 months after data generation and releasing no later than 6 months after submission.

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.

Tip: For the most part, it is not appropriate for investigators sharing data to place limits on the research questions or methods other investigators might pursue with the data. It is also not appropriate for the investigator who produced the data to control access to the data or require co-authorship as a condition for accessing the data.

Policy Reminder: Acceptable limitations on sharing and reuse are generally those based on ethical factors, federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements. The source and scope of these limitations must be explicitly described and justified in the DMS Plan. While certain agreements (e.g., those related to the development of commercial products) are consistent with NIH goals, agreements should not unnecessarily prohibit or restrict data sharing.

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).

Tip: Describe the access types that the repository will use to share scientific data with the public, which will vary depending on the data repository and the data sensitivity. Common approaches include:

- **Controlled access** typically requires established processes for verifying appropriate use of shared data, such as requiring verification of requestor identity, committee approval of proposed research use, and signing a data use agreement to access protected data. This approach is commonly used for sharing sensitive human data, even if de-identified (described in: [NOT-OD-22-213](#)).
- **Enclave** is controlled access where data cannot be downloaded or removed from a specific environment.
- **Registration required** access is open to all, but users need to register and sign into the resource to access the data.
- **Open access** has no restrictions or registration required to access the data.

Policy Reminder: To adhere to the Policy’s definition of data sharing, the approach must make the data accessible to the broad research community or public and not just a limited group of investigators.

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).

Tips:

- Describe the controls in place to protect privacy, including the deidentification standard used, such as:
 - [Health Information Portability and Accountability Act \(HIPAA\) Safe Harbor](#): removal of all 18 identifiers enumerated in the HIPAA Privacy Rule
 - HIPAA Expert Determination
 - Federal Policy for the Protection of Human Subjects, known as the Common Rule, whereby the definition of de-identified means the “identity of the human subjects cannot readily be ascertained”
- The GDS Policy currently expects adherence to the Common Rule and HIPAA Safe Harbor definitions of de-identification (described in [NOT-OD-14-124](#)).
- Sharing data via controlled access in a repository is also an approach to protecting sensitive human data and can be addressed in section 5B, described earlier.

Policy Reminders:

- NIH promotes the responsible sharing of scientific data consistent with protecting research participant privacy.
- NIH strongly encourages researchers to plan for how data management and sharing will be addressed in the informed consent process, including communicating with prospective participants how their scientific data are expected to be used and shared.
- NIH recommends scientific data be de-identified to the greatest extent that maintains sufficient scientific utility. More guidance is described in [NOT-OD-22-213](#) including resources for de-identifying various data types.

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).

Tips:

- List the title, roles, and responsibilities of the contact Principal Investigator and/or any other individuals on the project team who will be responsible for oversight of data management and sharing.
- Describe the frequency of major oversight activities planned throughout the project period (e.g., progress will be reviewed at monthly meetings).

Secondary Analysis Projects

Policy Reminders:

- Scientific data can result from secondary research, but researchers are not expected to share the existing, shared primary data used to conduct the secondary research.
- Researchers are, however, expected to maximize appropriate sharing of any new, derived data generated as a result of their research.

Tips:

- To align with the Policy's expectation to enable validation and replication of research findings, Plans for secondary analysis projects should describe:
 - The primary data used in the project and where (e.g., through a controlled access repository or administrative data source) and how (e.g., after signing a data use agreement) other researchers can access that data.
 - When, where, and how the documentation for processing and analyzing that data will be shared.
 - When, where, and how derived data, including aggregate or summary data, will be shared.

DMS Plan Budget and Costs

Tips:

- DMS activities are inherent components of the research process. Researchers should ensure they have requested sufficient funds to perform the activities that are described in their DMS Plan throughout the project, and not just at the end.
- Costs may be offset by collaborations with institutional librarians or investigators on other NIH awards, such as data coordinating centers.
- Guidance on [Allowable Costs for DMS](#) can help to determine which costs associated with data curation, preservation, and management are allowable costs and should be factored into your DMS budget justification.

Policy Reminder: In addition to a DMS Plan attachment, researchers should provide a brief summary of the DMS Plan (type and amount of shared scientific data and name of repository) and description of DMS costs within the budget justification attachment. Unlike the DMS Plan attachment, this summary will be accessible to peer reviewers, who will comment on whether the budget is reasonable; these comments will not impact score.

Budget Form options include:

- Modular budgets, which use the [Additional Narrative Justification](#) attachment of the PHS 398 Modular Budget Form.
- Detailed budget forms. DMS costs must be included with other costs in the appropriate cost categories (e.g., personnel, equipment, supplies, other expenses), following standard R&R Budget Form instructions.