NIH Office of Intramural Research (OIR)

Intramural Data Management and Sharing Plan Template

Intramural NIH research that involves the generation of scientific data is subject to the 2023 NIH Policy for Data Management and Sharing. The policy requires submission of a Data Management and Sharing (DMS) Plan, and compliance with the approved plan.

For all ongoing and new intramural research associated with a ZIA (and not included in a clinical protocol) that will be conducted on or after January 25, 2023, the investigator/project lead must submit a DMS plan prior to January 25, 2023. After that date, new and revised plans can be submitted throughout the year but must be included and approved as part of the annual reports process.

For research associated with a clinical protocol submitted for IC initial scientific review on or after January 25, 2023, a DMS plan must be submitted along with other protocol materials. For prior protocols, a DMS plan must be submitted as part of the quadrennial review.

If the proposed research will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in the DMS Plan. The DMS plan incorporates the Data Sharing Plan elements required under the 2015 intramural Human Data Sharing Policy. The Intramural DMS plan template aligns with the recommended template developed by NIH for the extramural research community, available here.

Additional guidance on the DMS policy is available at sharing.nih.gov and in the OIR Sourcebook. Contact sharing@nih.mail.gov for questions about NIH Sharing Policies including the new Data Management and Sharing Policy. The NIH Library offers one-on-one and group consultations, as well as courses and additional services. The DMP Tool, a service of the University of California, provides additional guidance and sample language. Some of the DMP Tool example answers are provided. These are provided as examples only and it is optional for investigators to use this language.

Requested Information

Investigator/Project Lead:

Project / Protocol Title(s):

ZIA Number(s):

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.

NIH 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.*

Example Answer (from DMP Tool):
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.

NIH expects that researchers will take steps to maximize scientific data sharing, but recognizes that certain factors (i.e., ethical, legal, or technical) may necessitate limiting sharing to some extent. Foreseeable limitations should be described. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data

Example Answer (from DMP Tool):
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 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. Indicate if none.

Example Answer (from DMP Tool):
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. If so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed. Indicate if none.

Example Answer (from DMP Tool):
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.
• 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].

Element 3: Data Standards

For DMS plans associated with clinical protocols, answer Element 3A and 3B. Other plans should answer only Element 3B.

Data standards refer to community-accepted approaches to methods of organizing, documenting, and formatting data to aid in data aggregation, sharing and reuse, and could refer to common data elements (CDEs), data dictionaries, data models, vocabulary, etc. Metadata standards (e.g., DataCite, DublinCore) are also data standards. Determining which standards are relevant to specific data may depend on the data repository where the data will be submitted as well as the standards adopted by the specific scientific community and/or by specific NIH ICs. Refer to: https://fairsharing.org/ that makes over 1,600 standards searchable.

A. Data standards for clinical protocols – Common Data Elements (CDEs):
Describe what Common Data Elements (CDEs) will be used. Justify if CDEs are not used.

CDEs allow data to be collected in the same way across multiple research studies and are defined unambiguously in both human and machine-computable terms. NIH has established a "Common Data Element (CDE) Resource Portal" (http://cde.nih.gov/) to help investigators identify CDEs when collecting data, particularly for clinical research, and when developing protocols and case report forms.

B. **Data standards for all plans:**
State what additional common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources; 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 proposed research. Plans not associated with a clinical protocol should describe the use of Common Data Elements here, if applicable. Indicate if no consensus standards exist.

A standard specifies how data and related materials should be stored, organized, and described. For 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. Some repositories require the use of specific standards.

Example Answer (from DMP Tool):
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 _____ [indicate standard] 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. Note that submission of a study to ClinicalTrials.gov meets the requirements of FDAAA but does not fulfill the requirements of the Data Management and Sharing Policy.

For additional guidance see Selecting a Data Repository and Repositories for Sharing Scientific Data. For some programs and types of data, NIH and/or IC policy may designate specific data repositories to be used. 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. Primary consideration should be given to data repositories that are discipline or data-type specific to support effective data discovery and reuse. If no appropriate discipline or data-type specific repository is available, researchers should consider other options, including generalist repositories and PubMed Central, which can accept small datasets (2 GB or less) along with the submitted manuscript.

Example Answer (from DMP Tool):
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.

Example Answer (from DMP Tool):
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, and if known, for how long data will be available. Scientific data must be made available no later than the time of an associated publication (when the publication first appears, either online or in print), or by the end of the project/protocol, whichever comes first.

Example Answer (from DMP Tool):
Data will be made available as soon as possible or at the time of associated publication, whichever comes first.

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).

Element 6: Oversight of Data Management and Sharing

Compliance with this plan will be monitored and managed by the Scientific Director (or designee). (No additional information is required).

Element 7: Other Elements (if applicable)

Include IC-specific or other additional information here. Indicate if not applicable.