# National Institutes of Health (nih.gov): NIH-NIMH: The National Institute of Mental Health

## Data Type

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.*

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). 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).

*Example Answer*:

All example DMSPs can be found at the end of the NIMH data sharing page or in the sample plans section of the sharing.nih.gov site.

- For human clinical and/or MRI data, please refer to the clinical imaging example on the NIMH site, also known as Sample Plan A on the NIH sharing site.
- For human genomic data, consider the human genomics example on the NIMH site, also known as Sample Plan B on the NIH sharing site.
- For genomic data from a non-human source, consider the non-human genomics example on the NIMH site, also known as Sample Plan C on the NIH sharing site.
- For secondary data analysis, consider the secondary data example on the NIMH site, also known as Sample Plan D on the NIH sharing site.

*Guidance*:

**NIH Guidance**

The final 2023 NIH DMS Policy (NOT-OD-21-013) 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.

**NIMH Guidance**

Please check the 2023 NIMH DMS policy (NOT-MH-23-100) for NIMH-specific requirements in addition to the 2023 NIH DMS policy (NOT-OD-21-013).

NIMH expects data sharing plans to also include a description of the standard(s) and/or data dictionaries that will be used to describe the data set, as well as a proposed schedule to validate that the data are compliant with the data dictionary that is being used.

NIMH has certain requirements for these standards such as a set of common data elements (see NOT-MH-20-067 for non-HIV research; for HIV-related research, see NOT-MH-23-105).

The NDA provides an online Data Dictionary with a searchable interface to find data structures that awardees are expected to use for new data collection. The NDA Data Dictionary is updated as researchers extend existing data collection instruments or create new instruments. For more information, see the NDA NIMH Common Data Elements page.

**NIH Genomic Data Sharing (GDS) Policy Considerations**

Check if your research is subject to NIH GDS (Genomic Data Sharing) policy usingthis criteria and list those data and the levels of processing here.

Individual NIH Institutes and Centers (IC) may have additional expectations or requirements for genomic data sharing as well. NIMH expects genomic data to be deposited at the NIMH Data Archive (NDA) unless NIMH agrees to a different database.

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.*

*Guidance*:

**NIH Guidance**

Per the Policy, even those scientific data not used to support a publication are considered to be scientific data and to fall within the final DMS Policy's scope.

NIH does not anticipate that researchers will preserve and share all scientific data generated in a study. Researchers should decide which scientific data to preserve and share based on ethical, legal, and technical factors that may affect the extent to which scientific data are preserved and shared. Provide the rationale for these decisions.

**Additional Guidance from DMPTool**

If human subjects data will be collected and only de-identified subsets are to be shared, consider specific de-identification approaches that fit the population and purposes. Guidance on protecting privacy is at NOT-OD-22-213.

**NIH GDS Policy Considerations for NIMH**

If you are generating genomic data, follow specific sharing requirements (data submission and release expectation) under the NIH GDS policy (five levels of processing and associated expectations for data submission and release)

NIMH expects genomic data to be deposited at the NIMH Data Archive (NDA) unless NIMH agrees to a different database. All data associated with new projects at the NIMH Repository and Genomics Resource will also be deposited in the NDA. Awardees who are measuring human genomic data are required to register with the Database of Genotypes and Phenotypes (dbGaP).

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.

*Guidance*:

**NIH**

In addition to the documentation examples, consider metadata that will provide additional information intended to make scientific data interpretable and reusable (e.g., date, independent sample and variable construction and description, methodology, data provenance, data transformations, any intermediate or descriptive observational variables).

**NIMH Guidance**

Investigators should review the planning section of the NIMH Data Archive website, Use of the NDA includes the NDA data harmonization approach for metadata and documentation.

## 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.

*Guidance*:

**Additional Guidance from DMPTool**

Tool(s) and software should be identified; then plans should specify how the 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). When known, the longevity or period of time for which custom or proprietary tools will be available should be addressed.

In addition, file formats in which data are saved in a digital format can be divided into two general categories.

- Proprietary - The specification of the data encoding format is not released or is restricted in some way. Proprietary formats can only be easily opened and manipulated by particular software tools.
- Open - The specification of the data encoding format which can be used and implemented by anyone. Open formats can often be easily opened and manipulated by a large number of software tools.

*Example Answer*:

All example DMSPs can be found at the end of the NIMH data sharing page or in the sample plans section of the sharing.nih.gov site.

For human clinical and/or MRI data, please refer to the clinical imaging example on the NIMH site, also known as Sample Plan A on the NIH sharing site.

- For human genomic data, consider the human genomics example on the NIMH site, also known as Sample Plan B on the NIH sharing site.
- For genomic data from a non-human source, consider the non-human genomics example on the NIMH site, also known as Sample Plan C on the NIH sharing site.
- For secondary data analysis, consider the secondary data example on the NIMH site, also known as Sample Plan D on the NIH sharing site.

## Standards

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and describe how these data standards will be applied. If applicable, indicate that no consensus standards exist.

*Guidance*:

**NIH Guidance**

While many scientific fields have developed and adopted common data standards, others have not. In such cases, the Plan may indicate that no consensus data standards exist for the scientific data and metadata to be generated, preserved, and shared.

**NIMH Guidance**

Applicants are strongly encouraged to use clinical and phenotypic data collection instruments/data dictionaries that have already been defined rather than create new versions of those data dictionaries. There are several required data collection instruments, mostly related to demographic and sample information, that must be used by all researchers for data harmonization purposes except for HIV-related applications (https://nda.nih.gov/contribute/harmonization-standards.html).

The NIHM Data Archive (NDA) provides its own standards for metadata and data structures. The standards that a PI intends to use for compatibility with the NDA should be briefly described in this section.

**Additional Guidance from DMPTool**

A *standard* specifies how exactly 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.

Furthermore, if formal standards such as specific imaging or sequencing filetypes, descriptive metadata, collection formats, etc., beyond the NDA standards will be used, that should also be addressed in this Element.

*Example Answer*:

All example DMSPs can be found at the end of the NIMH data sharing page or in the sample plans section of the sharing.nih.gov site.

- For human clinical and/or MRI data, please refer to the clinical imaging example on the NIMH site, also known as Sample Plan A on the NIH sharing site.
- For human genomic data, consider the human genomics example on the NIMH site, also known as Sample Plan B on the NIH sharing site.
- For genomic data from a non-human source, consider the non-human genomics example on the NIMH site, also known as Sample Plan C on the NIH sharing site.
- For secondary data analysis, consider the secondary data example on the NIMH site, also known as Sample Plan D on the NIH sharing site.

## Data Preservation, Access, and Associated Timelines

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)

*Guidance*:

**NIH Guidance**

NIH has provided additional information to assist in selecting suitable repositories for scientific data resulting from funded research: NOT-OD-21-016. Selecting a Data Repository Page of the NIH sharing website.

**NIMH Guidance**

Research funded by the NIMH are required to deposit all raw and analyzed data (including, but not limited to, clinical, genomic, imaging, and phenotypic data) from studies involving human subjects into the NIMH Data Archive (NDA).

**NIMH Guidance on NIH GDS Policy Considerations**

NOT-MH-23-100 requires that the NIMH Data Archive (NDA) serve as the repository for genomic data funded by the NIMH unless the NIMH approves a different data repository during the negotiation of the terms and conditions of the award.

Awardees who are measuring human genomic data are required to register with the Database of Genotypes and Phenotypes (dbGaP Submission Process). After registration, all data (including, but not limited to, clinical, genomic, imaging, and phenotypic data) will be deposited in the NDA. A link to NDA will be added to the dbGaP registration. Aggregating the genomic data in a single cloud-based data archive will facilitate the re-analysis, replication, and additional analyses of these important data sets. Computational credits may be available to conduct these analyses in the cloud.

All data associated with new projects at the NIMH Repository and Genomics Resource will be deposited in the NDA. Appropriate data will then be transmitted to the NIMH Repository and Genomics Resource for quality control and to allow the research community to identify samples that are relevant to their research efforts.

**Additional Guidance from DMPTool**

The NIMH notice of data sharing policy does not provide specific guidance on access and preservation of non-human data for the data management and sharing policy. Consult the NIMH data sharing website or the trans-NIH repositories list for further possibilities.

*Example Answer*:

All example DMSPs can be found at the end of the NIMH data sharing page or in the sample plans section of the sharing.nih.gov site.

- For human clinical and/or MRI data, please refer to the clinical imaging example on the NIMH site, also known as Sample Plan A on the NIH sharing site.
- For human genomic data, consider the human genomics example on the NIMH site, also known as Sample Plan B on the NIH sharing site.
- For genomic data from a non-human source, consider the non-human genomics example on the NIMH site, also known as Sample Plan C on the NIH sharing site.

**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.

*Guidance*:

**NIH Guidance**

Unique Persistent Identifiers: The repository assigns datasets a citable, unique persistent identifier, such as a digital object identifier (DOI) or accession number, to support data discovery, reporting, and research assessment. The identifier points to a persistent landing page that remains accessible even if the dataset is de-accessioned or no longer available.

**Additional Guidance from DMPTool**

NIMH Data Archive data collections have DOIs to help make the data in them findable and identifiable.

**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.

*Guidance*:

**NIH Guidance**

NIH encourages scientific data be shared as soon as possible, and no later than time of an associated publication or end of the performance period, whichever comes first. The NIMH further specifies that data must be shared with the research community when papers using the data have been accepted for publication or at the end of the award period (including the first no cost extension). Researchers are encouraged to consider relevant requirements and expectations (e.g., data repository policies, award record retention requirements, journal policies) as guidance for the minimum time frame scientific data should be made available. NIH encourages researchers to make scientific data available for as long as they anticipate it being useful for the larger research community, institutions, and/or the broader public. Identify any differences in timelines for different subsets of scientific data to be shared.

**NIH GDS Policy Considerations**

Genomic data is subject to further guidance on release expectations and timelines.

**NIMH Guidance**

The general expectation is that data from NIMH-funded awards that involve human subjects will be submitted to NDA every 6 months throughout the duration of the award (typically January and July). Awardees will provide a Data Submission Agreement signed by the principal investigator and an institutional business official within 6 months of the notice of award. Although submission does not lead to release of the data, awardees are encouraged to share basic demographic and raw baseline data shortly after data submission to encourage collaborations in the research community.

In addition to regular submission of data associated with an award, awardees are expected to separately submit to NDA the specific data that was used for each resulting publication by creating an NDA Study.

## Access, Distribution, or Reuse Considerations

**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.

*Guidance*:

**NIH Guidance**

The DMS Policy acknowledges certain factors (i.e., legal, ethical, or technical) that may affect the extent to which scientific data are preserved and shared.

In addition, NOT-OD-22-213 addresses specific considerations about human subjects privacy.

**Additional Guidance from DMPTool**

This is the section to describe what legal, ethical, or technical issues may require limiting the sharing of your data. Examples may include ethical considerations such as IACUC restrictions on sharing videos or images of procedures. Other nonhuman data factors affecting distribution may include existing legal limits such as data licenses or use agreements or technical limits about the size or structure of the data.

For human data, the NIMH requires the application of privacy protections through the use of the NDA. It is likely to be sufficient in this subsection to say that human ethics require the use of the NIMH Data Archive for privacy.

**NIH GDS Policy Considerations**

Genomic data may have further considerations to address. The NIH now expects a single data sharing plan at the time of funding application to satisfy both the Genomic Data Sharing (GDS) Policy and the DMS Policy (per NOT-OD-22-198). How to access genomic data varies depending on which repository you selected. Please refer to the Accessing Genomic Data from NIH Repositories page on the NIH sharing site.

*Example Answer*:

All example DMSPs can be found at the end of the NIMH data sharing page or in the sample plans section of the sharing.nih.gov site.

- For human clinical and/or MRI data, please refer to the clinical imaging example on the NIMH

site, also known as Sample Plan A on the NIH sharing site.
- For human genomic data, consider the human genomics example on the NIMH site, also known as Sample Plan B on the NIH sharing site.
- For genomic data from a non-human source, consider the non-human genomics example on the NIMH site, also known as Sample Plan C on the NIH sharing site.
  For secondary data analysis, consider the secondary data example on the NIMH site, also known as Sample Plan D on the NIH sharing site.

**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).

*Guidance*:

The NDA uses controlled access. Summary information on the data shared in NDA will be available in the NDA Query Tool without the need for an NDA user account. To request access to record-level human subject data, researchers must submit a Data Access Request.

**Additional guidance from DMPTool**

Check the repository you intend to use to find out more about whether and how the repository supports controlled access.

**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).

*Guidance*:

**NIH Guidance**

Effective data stewardship and protection of human research participant (hereinafter "participant") privacy are achieved in tandem through responsible scientific data sharing practices. Accordingly, NIH has developed supplemental information to the DMS Policy to assist researchersin responsible data sharing by establishing 1) operational principles for protecting participants' privacy when sharing scientific data, 2) best practices for implementing these principles, and 3) points to consider for choosing whether to designate scientific data for controlled access.

**NIMH Guidance**

Applicants should also plan to collect the data neededto generate global unique identifiers (GUIDs) for each study subject. The GUID, or Global Unique Identifier, is used as an identifier for a research participant. The GUID provides a secure mechanism to link research participants within and across research project datasets in NDA.

Informed consent documents should describe how study data will be shared with NDA and the research community. The NDA has provided a plain-language description of the NDA as an example when creating informed consent language.
NIMH also expects these additional points to be considered in human subjects research.

## 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).

*Example Answer*:

All example DMSPs can be found at the end of the NIMH data sharing pageor in the sample plans section of the sharing.nih.gov site.

- For human clinical and/or MRI data, please refer to theclinical imaging example on the NIMH site, also known as Sample Plan A on the NIH sharing site.
- For human genomic data, consider the human genomics example on the NIMH site, also known as Sample Plan B on the NIH sharing site.
- For genomic data from a non-human source, consider the non-human genomics example on the NIMH site, also known as Sample Plan C on the NIH sharing site.
- For secondary data analysis, consider the secondary data example on the NIMH site, also known as Sample Plan D on the NIH sharing site.

*Guidance*:

**NIH Guidance**
This element refers to oversight by the funded institution, rather than by NIH. The DMS Policy does not create any expectations about who will be responsible for Plan oversight at the institution.

**Additional guidance from DMPTool**
Please confer with your Office of Sponsored Programs, Office of Research, etc., about any additional oversight considerations. Oversight of human data with the NIMH Data Archive may have specific factors to address. All NIMH applicants should consult local campus departments and policies about oversight.

## Validation schedule

The Data Management and Sharing Plan must propose a schedule to validate the quality of the data being uploaded so these data are compliant with the data dictionary or other standards that are being used.

*Guidance*:

**NIMH Guidance**

All NIMH DMS Plans must propose a schedule to validate the quality of the data being uploaded so these data are compliant with the data dictionary or other standards that are being used. In cases where a data dictionary has been defined in the NIMH Data Archive, the NDA Data Validation and Uploading Tool should be used to ensure your data files are harmonized to the NDA Data Dictionary. Compliance with the approved data management and sharing plan will become a term and condition in the Notice of Award and will be monitored by the NIMH throughout the duration of the award.