OMB No. 0925-0001 and 0925-0002 *(Rev. 07/2022 Approved Through 01/31/2026)*

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. 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.

Commented [JB1]: For more information about how the Stanford DMP Service evaluates plans, see our rubric.

https://zenodo.org/record/7710001

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,

Commented [JB2]: The italicized text throughout this template should be deleted before the plan is submitted.

- This section can be written in a narrative format but a table is also acceptable if you are working with multiple types of data.
- When describing "amount" of data, it is acceptable to give a range for file sizes or number of samples.

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.

- **Preservation** in this context roughly refers to long term storage for a wide array of research materials and products (including datasets) that may or may not be made available to others.
- **Sharing** in this context roughly refers to making research materials and products available for use by others. This does not *necessarily* mean they're available to anyone for any purpose (i.e. "open" data sharing) - the DMS policy allows for restricted forms of data sharing.
- In terms of what data should be shared, it is whatever underlies research findings. Most typically being those that are described in research publications.
- Almost certainly more data will need to be preserved than shared. A researcher may preserve data in its raw form and then share a cleaned/processed/analyzed form, for example. An important element to this section is explaining why data that is going to be preserved should be preserved, why data that is going to be shared should be shared, etc.

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.

- **Metadata** in this context roughly just refers to information/materials related to the data that is needed to make use of the data. Common forms of "metadata" in this context include things like data dictionaries, codebooks, protocols, etc. Typically, these things can be submitted alongside permanent datasets in a repository.

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.

- Because the DMSP reflects proposed practices at the time it is written (i.e. before the start of the proposed project), it is unlikely that every detail about the use of specific software tools, custom code, etc will be known. This section does not need to include all these details. The focus should be on:
  - If particular types of data require the use of particular computational tools, what those tools are and how they can be accessed by other researchers.
  - Plans related to sharing custom code. At present this is recommended by not explicitly required.

- An important reminder here that HYPERLINKS SHOULD NOT BE EMBEDDED INTO DMSPs. If software is to be made available via a Github repository or similar, that fact should be mentioned (and the repository name should be given), but a URL is not appropriate.

- There is no expectation in the DMS policy that custom code or other software created as part of the analysis process be made available to others. We can encourage it, however.

- There is no expectation that the software tools mentioned in this section should be openly or freely available. If a piece of proprietary software is needed to make use of shared data, then that just needs to be noted in the plan.

- If the data is in a common or open (i.e. non-proprietary) file format and therefore does not need specialized software, that fact should be - pending more guidance from NIH - mentioned in the DMSP.

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.

- This section is, far and away, what we get the most questions about. In simple terms, a **data standard** can be thought of as an agreed upon set of rules about how data should be organized, described, etc. What NIH seems generally to be referring to is less particular file formats (e.g. fastq, dicom, etc) and more things like common data elements, data models, and other standardized practices for organizing and representing data (e.g. the BIDS standard for neuroimaging data).

- If a formal data standard does not exist for a particular type of data/metadata, then that also should be noted in the plan. Because standards are developed to aid in usability and interoperability, we recommend giving additional information about related practices and strategies as an alternative.
  - In this context, **interoperability** generally refers to the degree to which a dataset can be combined with other data, read and acted upon by software tools, and included research workflows.

- **Usability** is a less formal version of interoperability, loosely summed up by "Could another researcher open up these files and make use of them?"

  - In the absence of formal data standards, we've been recommending that researchers discuss how they'll be standardizing practices within the research group (shared protocols, use the same variable names across files, maintenance of documentation, etc).

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).

- The use of repositories is encouraged, but they may not be appropriate under all circumstance (i.e. the data is clinical data from Stanford Hospital).

- If the FOA mentions the use of a specific repository, we strongly recommend using that repository.
  - If there is a specialized repository for your data, we strongly recommend using that repiostory.
  - If you need a general repository for your data, we recommend Dryad (for low/moderate risk data) or Vivli (for clinical trial data).

Commented [JB3]: Contact the Stanford DMP Service for template language for these tools.

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.

- Most data repositories assign uploaded data a persistent identifier (typically a DOI or accession number). If sharing data in this way, simply include a statement here about how the persistent identifiers/URLs for all datasets shared in a repository will be cited in a relevant paper.

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.

- The NIH DMS Policy states that data should be made available at the time of publication or the end of the grant period, whichever comes first.

- NIH and Stanford both require that data be kept for *at least* three years following project closeout, but both entities encourage researchers to make scientific data available for as long as they anticipate it being useful for the larger research community, the broader public, and other stakeholders.

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.

- This is where researchers can list any factors that restrict their ability to share data.
  - For some research projects, especially those that don't work with data derived from human participants, there may be no such factors.

- The DMS policy is not an open data mandate. Restrictions can be indicated as long as they can be justified.

- Acceptable reasons for restricting data sharing:
  - Informed consent will not permit or will limit the scope or extent of sharing.
  - Existing consent prohibits sharing or limits the scope or extent of sharing.
  - Privacy or safety of research participants would be compromised.
  - Explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure.
  - Datasets cannot practically be digitized with reasonable efforts.

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).)

A limited number of repositories (all typically specialized, many funded by NIH themselves) allow for controlled access. If any of the following is true, a researcher should consider controlling access:
- There are explicit limits places on subsequent use (i.e. by laws, policies, informed consent, etc)
- The data could be considered sensitive.
- The data can not be deidentified sufficiently to mitigate the rise of re identification.

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).

NIH gives the following best practices for protecting participant privacy:
- Ensure appropriate deidentification.
- Establish data sharing and use agreements.
- Understand legal protections against disclosure and misuse.

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).

- Some of the sample DMSPs provided by NIH include a statement about an office on campus tasked with monitoring compliance with the contents of DMSPs. Stanford has no such office (neither does any other university as far as we know). So the responsibility ultimately falls to the PI on the grant.
- For grants written by a junior researcher who will be supervised by other PIs, we recommend that the researcher indicate that they will do everything “under the mentorship of Dr(s). [insert senior PIs name]”.

Commented [JB4]: Contact the Stanford DMP Service for standardized language for this section.