Federation of American Societies for Experimental Biology (faseb.org): DataWorks! Data Management and Sharing Plan Challenge

Data Type

A general summary of the types and estimated amount of scientific data to be generated and/or used in the research. 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)

Guidance:

The final [NIH DMS Policy](#) 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."

Even those scientific data not used to support a publication are considered scientific data and within the final DMS Policy's scope. We understand that a lack of publication does not necessarily mean that the findings are null or negative; however, indicating that scientific data are defined independent of publication is sufficient to cover data underlying null or negative findings.

A description of which scientific data from the project will be preserved and shared.

Guidance:

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.

A brief listing of 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:

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.

Related Tools, Software and/or Code

An indication of whether specialized tools are needed to access or manipulate shared scientific data to support replication or reuse, and name(s) of the needed tool(s) and software.

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.

Standards

An indication of what standards will be applied to the scientific data and associated metadata (i.e., data formats, data dictionaries, data identifiers, definitions, unique identifiers, and other data documentation).

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.

Data Preservation, Access, and Associated Timelines

The name of the repository(ies) where scientific data and metadata arising from the project will be archived.

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
 For some programs and types of data, NIH and/or Institute, Center, Office (ICO) policy(ies) and Funding Opportunity Announcements (FOAs) identify particular data repositories (or sets of repositories) to be used to preserve and share data. For data generated from research subject to such policies or funded under such FOAs, researchers should use the designated data repository(ies).

For data generated from research for which no data repository is specified by NIH or the NIH ICO (as described above), researchers are encouraged to select a data repository that is appropriate for the data generated from the research project and is in accordance with the desired characteristics, taking into consideration the following guidance:
1. Primary consideration should be given to data repositories that are discipline or data-type specific to support effective data discovery and reuse. NIH makes a list of such data repositories available (see [Open Domain- Specific Data Sharing Repositories](#)).

2. If no appropriate discipline or data-type specific repository is available, researchers should consider a variety of other potentially suitable data sharing options:

1. Small datasets (up to 2 GB in size) may be included as supplementary material to accompany articles submitted to PubMed Central (see [PMC policies](#)).
2. Data repositories, including generalist repositories (see [Generalist Repositories)](#) or institutional repositories, that make data available to the larger research community, institutions, or the broader public.
3. Large datasets may benefit from cloud-based data repositories for data access, preservation, and sharing.

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

When the scientific data will be made available to other users (i.e., the larger research community, institutions, and/or the broader public) and for how long.

Access, Distribution, or Reuse Considerations

Describe any applicable factors affecting subsequent access, distribution, or reuse of scientific data related to whether access to scientific data derived from humans will be controlled (i.e., made available by a data repository only after approval).

Guidance:

- Informed consent (e.g., disease-specific limitations, particular communities' concerns).
- Privacy and confidentiality protections (i.e., de-identification, Certificates of Confidentiality, and other protective measures) consistent with applicable federal, Tribal, state, and local laws, regulations, and policies.
- Any restrictions imposed by federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements (e.g., with third party funders, with partners, with Health Insurance Portability and Accountability Act (HIPAA) covered entities that provide Protected Health Information under a data use agreement, through licensing limitations attached to materials needed to conduct the research).
- Any other considerations that may limit the extent of data sharing.

Oversight of Data Management and Sharing

Indicate how compliance with the Plan will be monitored and managed, frequency of oversight, and by whom (e.g., titles, roles).

Guidance:

This section should address titles and roles overseeing data management and sharing, within the investigator team or as key personnel.

Personnel costs required to perform the types of data management and sharing activities are

allowable. Examples of costs may include time and effort for data curation processes; local specialized infrastructure (only those not covered by institutional F&A costs); or fees for preserving and sharing data. Reasonable, allowable costs for management and sharing may be included in NIH budget requests. Funds for these activities must be spent during the performance period, even for scientific data and metadata preserved and shared beyond the award period. See NIH's [supplementary guidance on allowable costs for data management and sharing](#)