Research Guidance & Education

Icahn School of Medicine at Mount Sinai

Guidance from Kris Alpi, Associate Dean of Libraries & Information Sciences and Allison Gottlieb, Sponsored Programs Eduation and Communications Director, GCO

Can you help me keep the leaves from falling off my Knowledge Tree below?

https://www.vecteezy.com/free-vector/owl

Dear Research Community,

In this communication we are focusing on Element 4 of the 6 Elements of the NIH Data Management and Sharing Plan. Element 4 is: Data Preservation, Access, and Associated Timelines. In many of these sections you may be tempted to write "Information is available on request." or "Information and/or data will be published with the supplemental files of the article." This will not meet the requirements; please follow our guidance instead.

https://www.vecteezy.com/free-vector/nature

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).

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.

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.

DMPTool Screenshot: Element 4A: Data Preservation, Access, and Associated Timelines > A. Repository where scientific data and metadata will be archived

There are various ways in which to disseminate, preserve, and make scientific data discoverable. In this section, you should describe where and when the scientific data associated with your research will be made available. The primary way to satisfy this requirement is to put your scientific data into a repository, which will support preservation of that data and provide long term access. In some cases, a repository may be specified by the funding Institute or Center; if a repository is not specified, then you have discretion in selecting the repository for your scientific data. When identifying a repository, consider how your scientific data will be made discoverable through the platform. For example, the repository's ability to provide persistent identifiers (e.g., DOIs, handles, ARKs) for your scientific data is a good starting point to ensure consistent access. If there are restrictions as to the manner or length of time in which the scientific data can be preserved and/or accessed, be clear as to what those restrictions are in this section.

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**)

B *I*	☰ ˅ ☰ ˅	🔗	⊞˅

[text area]

Save

Example answer

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.

Our comments: In addition to the name of the repository, please provide the supporting organization name to assist in recognizing similarly named repositories. Journals frequently change their policies and practices about supplemental files so this should not be your repository strategy. The requirement for free and easy access to the data means that you will need to retain your author rights to deposit the files elsewhere if you plan on publishing data as supplemental files in journals that require a subscription. More effective to deposit the data and then refer to the data citation in the repository in the article's data availability statement. Many repositories and journals require you to have an ORCID ID to uniquely identify you as a researcher when you publish your data. Sign up at ORCID.org.

NIH Guidance with Added Mount Sinai Comments

NIH provided additional information to assist in selecting suitable repositories: NOT-OD-21-016.

First priority is if the FOA or Institute specifies a repository, in which case that repository must be used. The next priority is approved Open Domain-Specific Data Sharing Repositories. If neither of those considerations fit, they offer other potentially suitable options: attaching data files up to 2 GB as supplementary material to author manuscripts submitted to PubMed Central with no embargo, approved generalist repositories, or an organization's institutional repository if you have collaborators at institutions with institutional

repositories. *Mount Sinai does not have an institutional repository for all research data at this time.*

[Genomic data has further guidance and considerations](#) to address in the Plan.

DMPTool Screenshot: Element 4B: Data Preservation, Access, and Associated Timelines > B. How scientific data will be findable and identifiable
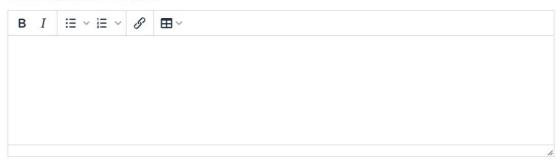
How scientific data will be findable and identifiable: Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

B	*I*	☰ ∨	☰ ∨	🔗	⊞ ∨

Save

Example answer

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].

Our comments: This section is specific to the repository you choose. Typically, the licenses will be one of the Creative Commons licenses—for example, the Dryad example uses a [Creative Commons Zero](#) *(CC0) waiver crafted specifically to facilitate the discovery, re-use, and citation of data. Note that the Example answer "OR through a request process" does not mean an investigator statement that the data will be available upon request, it is meant to indicate that your repository has restricted-use data request processes and what process you will use for data that requires these protections. For an example of these, see* [Restricted-Use Data Management at ICPSR (umich.edu)](#)*.*

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.

DMPTool Screenshot: Element 4C: Data Preservation, Access, and Associated Timelines > C. When and how long the scientific data will be made available.

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.

| **B** *I* | ☰ ˅ ☲ ˅ | 🔗 | ⊞ ˅ |

Example Answer Based on Mount Sinai Policy:

Data will be made available at the time of associated publication(s) or by the end of the performance period if unpublished. Per Mount Sinai policy, the data will be available for at least six years post-publication. [NOTE: If your repository provides perpetual access, please indicate that here as well.]

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. 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.

Genomic data has further guidance on release expectations and timelines.

Thus far, the Library and the GCO have provided guidance about creating your DMS plan, preparing the budget and developing Elements 1 – 4 of the plan. As a reminder, these communications will be made available online and you are welcome to contact Allison Gottlieb if you need any past communications. The next communication will be on Element 5: Access, Distribution, or Reuse Considerations; and Element 6 Oversight of Data Management and Sharing.

Please let us in the Library and the GCO know any questions you may have.

Sincerely,

Allison Gottlieb

Allison Gottlieb, M.S. | Director, Sponsored Programs Education and Communications | Grants and Contracts Office

and

Kris Alpi
MLS, MPH, PhD, FMLA, AHIP I Associate Dean of Libraries & Information Sciences