

From: research.listserv@mssm.edu

Sent: Tuesday, December 27, 2022 10:12 AM

last updated 3/21/23

Subject: [GCO-Library]

Advice for Element 1 Data Type, Element 2: Related Tools, Software and/or Code: and Element 3: Standards in Your NIH DMSP



Research Guidance & Education

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

What Data is in Your Knowledge Tree?

What Tools Are You Using to Plant Your Tree?

What Are Your Standards for Planting?

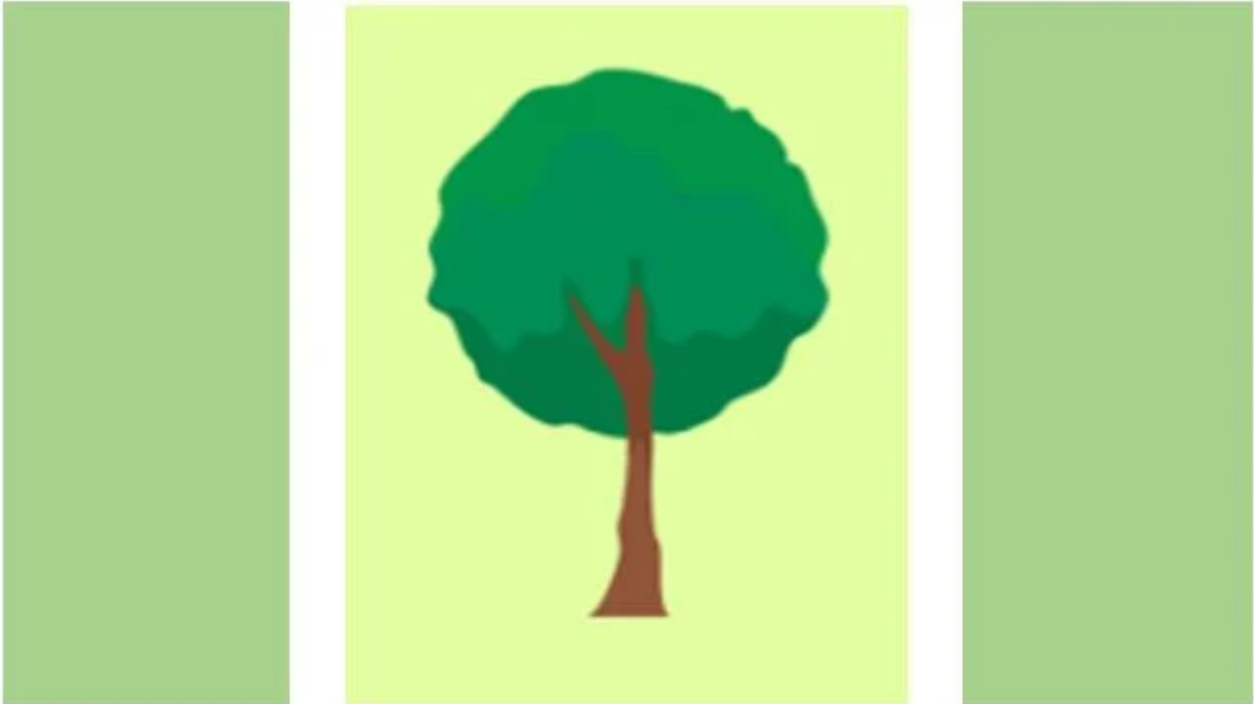


<https://www.vecteezy.com/free-vector/owl>

Dear Research Community,

In this communication we are focusing on Elements 1-3 of the 6 elements of the [NIH Data Management and Sharing Plan](#), namely Element 1: Data Type; Element 2: Related Tools, Software and/or Code; and Element 3: Standards.

Planting the Knowledge Tree



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

Let us help you plant that sturdy knowledge tree for your Data Management and Sharing Plan (DMSP) by applying the advice and instructions we provide here.

For each of the 3 elements, we are sharing a screenshot with sample responses from the free [DMPTool](#), which allows you to build your plan in the format required by NIH. The DMPTool provides a click-through wizard for creating your DMSP that complies with funder requirements. It is super easy to use and we strongly encourage you to use it. Remember: Your final plan should not be more than two pages in many circumstances. Please see the NIH FAQ about page limitation at the end of this communication. updated 3/21/23

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.

Our Comments: This should parallel what you described in your budget justification. Please see our communication about the budget justification for more information.

DMPTool Guidance: Research projects vary widely in the types of data produced. In this section, you will describe the categories, amounts, and

— Data Type (0 / 3)

Briefly describe the scientific data to be managed, preserved, and shared.

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)

B *I* [List Icons] [Link Icon] [Table Icon]

Save

Example answer

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.

Our Comments: For human subjects research, this should be in agreement with your IRB proposal, content and protection of human subjects plans. Click [here](#) for more information.

DMPTool 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.

DMPTool Screenshot: Element 1 Data Type > B. Data Preserved and Shared

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.

Our Comments: Consult the [RDA Metadata Standards catalog](#) or search to see if there is a metadata standard applicable to your subject area's research data.

DMPTool Screenshot: Element 1 Data Type > C. Metadata & Other

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.

B I [List Icon] [List Icon] [Link Icon] [Table Icon]

Save

Example answer
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, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

Our Comments: Where there are multiple approaches or tool options, focusing on what you will use while noting there are others is fine. E.g., you use a licensed tool like SAS rather than a free tool like R.

For proprietary or licensed tools and software when you cannot provide the tool, indicating the version, settings selected, transformations made, or scripting facilitates reproducibility and reuse. This can be included in a readme file deposited with shared data.

DMPTool Guidance:

Tool(s) and software should be identified; 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.

Screenshots DMPTool: Element 2 Related Tools, Software and/or Code

- Proprietary - The specification of the data encoding format is not released or 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.*

Screenshots DMPTool: 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.

B I [List Icon] [List Icon] [Link Icon] [Table Icon]

Save

Example answer

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.
 These tools will be shared openly via ____.

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.

Our Comments: Please be sure to search the [NIH Common Data Elements Repository](#) to see if there are standards you should apply to your data since NIH Program Officers will be familiar with this resource. All data with human subject participants should consider the ["Person" elements](#). For example, if coding for Age, apply the standard responses for [Age Units](#) that distinguish between Don't Know, Prefer not to answer, and Unknown.





If there are no common data standards, be sure to indicate that.

DMPTool Guidance:

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. However, for researchers within a community where more formal standards have not been well established, it can also be interpreted more broadly to refer to the adoption of the same (or similar) data management-related activities or strategies by different researchers and across different projects.

It is possible that your work will employ multiple formal standards or a mix of formal standards and other data management strategies. You should be as specific as possible when describing the standards used for each type of data included in your proposal.

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

B I    

Save

Example answer

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 _____ [common data elements, standardized survey instruments, etc] 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.

To recap, we've provided guidance about creating your DMS plan and preparing the budget. These communications will be made available online and you are welcome to contact Allison Gottlieb if you need any past communications. Please be on the lookout in the next few weeks for communications on Element 4: Data Preservation, Access, and Associated Timelines; 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

Excerpt from NIH's [Frequently Asked Questions](#)

11. Are applications for awards that support a variety of activities or generate many data types limited to a DMS Plan of two pages or less?

No. Some research projects may propose a larger number of activities or generate a larger number of data types, which may necessitate more detail in a DMS Plan. NIH provides an optional DMS Plan [format page](#), but use of this format page is not required. NIH will accept DMS Plans longer than the recommended two pages as long as Plans provide sufficient information for NIH program staff to assess the Plan. Please see supplemental information "[Elements of an NIH Data Management and Sharing Plan](#)" for more information.