



DataWorks! Data Management Plan (DMP) Challenge Evaluation Rubric

The rubric is based on the [Belmont Forum](#) and [DART Rubric](#) and has been adjusted to align with the 2023 NIH DMP guidance ([NOT-OD-21-014](#)).

Section 1: Types of data produced

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).

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

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.

Performance Criteria		Performance Level		
		Complete/detailed	Addressed issue, but incomplete	Did not address
1.1	<i>Describes what types of scientific data will be generated and/or used in the research</i>	Clearly defines data type(s).	Some details about data types are included, but missing details or wouldn't be well understood by someone outside of the project	No information about data types is included; fails to adequately describe data types.
1.2	<i>Describes which data will be preserved and shared, based on ethical, legal, and technical factors</i>	Clearly defines which data will be preserved and shared, and underlying rationale.	Missing some details regarding how data will be preserved and shared	Does not include information regarding how the data will be preserved and shared
1.3	<i>Identifies relevant other data, and any associated documentation that will be made accessible</i>	Clearly describes other relevant data and associated documentation	Missing some details regarding documentation so that data wouldn't be well understood by someone outside of the project	No information about data documentation.
1.4	<i>Identifies how much data (volume) will be produced</i>	Expected scale of data (GB, TB, etc.) is clearly specified.	Expected scale of data is vaguely specified.	Expected scale of data is not specified.



Section 2: 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.

		Performance Level		
Performance Criteria		Complete/detailed	Addressed issue, but incomplete	Did not address
2.1	<i>Describes what specialized or licensed software or tools are needed to access or manipulate data generated and/or used in the research</i>	Clearly defines what software or and tools are needed to access and manipulate data, and specifies which are proprietary, open source, and/or custom created by the researcher(s).	Some details about software or tools are included, but DMP is missing details or wouldn't be well understood by someone outside of the project	No information about software or tools is included; fails to adequately describe software or tools.
2.2	<i>Describes whether custom-created code or in-house software or tools are needed or will be created to access or manipulate data generated and/or used in the research and if/how this will be made available</i>	Clearly defines whether custom code or tools will be needed to access and manipulate data, and if/how they will be made available.	Missing some details about whether custom code or tools will be needed or created.	No information about custom code or tools is included.



Section 3: Standards for data and metadata

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). 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.

Performance Criteria		Performance Level		
		Complete/detailed	Addressed issue, but incomplete	Did not address
3.1	<i>Identifies metadata standards and/or metadata formats that will be used for the proposed project</i>	The metadata standard that will be followed is clearly stated and described. If no disciplinary standard exists, a project-specific approach is clearly described.	The metadata standard that will be followed is vaguely stated. If no disciplinary standard exists, a project-specific approach is vaguely described.	The metadata standard that will be followed is not stated and no project-specific approach is described.
3.2	<i>Describes data formats created or used during project</i>	Clearly describes data format standard(s) for the data.	Describes some but not all data formats, or data format standards for the data. Where standards do not exist, does not propose how this will be addressed.	Does not include information about data format standards.
3.3	<i>Identifies data formats that will be used for storing data</i>	Clearly describes data formats that will be used for storing data and explains rationale or complicating factors.	Only partially describes data formats that will be used for storing data and/or the rationale or complicating factors.	Does not describe data formats that will be used for storing data and does not explain rationale or complicating factors.
3.4	<i>If the proposed project includes the use of unusual data formats, the plan discusses the proposed solution for converting data into more accessible formats</i>	Explains how the data will be converted to a more accessible format or otherwise made available to interested parties. In general, solutions and remedies should be provided.	Vaguely explain[s] how the data may be converted to a more accessible format or otherwise made available to interested parties.	Does not explain how the data will be converted to a more accessible format or otherwise made available to interested parties.



Section 4: 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](#). 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.

Performance Criteria		Performance Level		
		Complete/detailed	Addressed issue, but incomplete	Did not address
4.1	<i>Provides details on where the data will be made publicly available</i>	Clearly specifies where the data will be made available to people outside of the project. Aligned with FOA requirements, as needed.	Verifies that the data will be made available outside of the project but does not identify a specific repository.	Does not specify where the data will be made available outside of the project.
4.2	<i>How the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.</i>	Clearly specifies findability of the data by describing how a unique and persistent identifier will be obtained for the data.	Describes how a landing page URL for the data file will be created, but no identifier.	Does not specify data findability parameters.
4.3	<i>When the scientific data will be made available to other users (i.e., the research community, institutions, and/or the broader public) and for how long.</i>	Clearly specifies when the data will be made available to people outside of the project.	Verifies that the data will be made available outside of the project but does not identify timing.	Does not specify when the data will be made available outside of the project.



Section 5: 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).

Performance Criteria		Performance Level		
		Complete/detailed	Addressed issue, but incomplete	Did not address
5.1	<i>Provides details for access to scientific data derived from patient data (if any).</i>	Clearly specifies restrictions imposed by federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements, and any other considerations that may limit the extent of data sharing.	Verifies that the data are derived from patient data and must be controlled but does not specify controls.	Does not specify whether there are patient-derived data.
5.2	<i>Describes what protections will be put into place to protect privacy or confidentiality of human research subjects, including vulnerable populations (if applicable)</i>	Clearly describe the actions that will be taken to address the sharing of sensitive data and demonstrate an appropriate balance of protecting sensitive data and sharing non sensitive data.	Actions that will be taken to address the sharing of sensitive data are vaguely described.	Actions that will be taken to address the sharing of sensitive data are not described.
5.3	<i>Describes what intellectual property rights to the data and supporting materials will be given to the public and which will be retained by project personnel (if any)</i>	Clearly defines the IP rights the public (or designated group) has in accessing the data and the rights retained by project personnel (if any).	Vaguely defines the IP rights the public (or designated group) has in accessing the data or that are retained by project personnel.	Does not address IP rights for the public, intended audiences or personnel in the research group.
5.4	<i>Describes security measures that will be in place to protect the data from unauthorized access</i>	Clearly describes the security measures that will be put into place to prevent authorized access to the data.	Vaguely describes the security measures that will be put into place to prevent unauthorized access to the data.	Does not describe the security measures that will be put into place to prevent unauthorized access to the data.
5.5	<i>If there are factors that limit the ability to share data, e.g. proprietary nature or commercialization of the data</i>	Clearly defines the population to whom the data will be made available, as well as any conditions on access, a justification for its limited release.	Vaguely discusses who will have access to the data or conditions on access.	Does not state who will be able to gain access to the data.



Section 6: Oversight of Data Management and Sharing

This section addresses titles and roles overseeing data management and sharing, within the investigator team or as key personnel, allowable costs required to perform the types of data management and sharing activities specified in the plan, as well as oversight protocols. 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. See [NOT-OD-21-015](#).

Performance Criteria		Performance Level		
		Complete/detailed	Addressed issue, but incomplete	Did not address
6.1	<i>Provides details of titles and roles of project personnel overseeing data management and sharing, within the investigator team or as key personnel.</i>	Clearly specifies who is responsible for execution and oversight of data management and sharing plan.	Verifies that there is a Data Management and Sharing team but does not provide names or titles.	Does not specify Data Management and Sharing team.
6.2	<i>Provides details of personnel, in line with allowable cost guidance for Data Management and Sharing</i>	Clearly specifies personnel required to perform the types of data management and sharing activities.	Vaguely describes personnel categories required to perform the types of data management and sharing activities.	Does not describe personnel for data management and sharing activities.
6.3	<i>Provides details of Data Management and Sharing oversight practices throughout the research project workflow</i>	Clearly specifies how compliance with the Plan will be monitored and managed, including frequency of oversight, and by whom (e.g., titles, roles).	Vaguely describes how compliance with the Plan will be monitored and managed. Leaves out data management and description processes during key stages of the project	Does not describe how compliance with the Plan will be monitored and managed.