DATA MANAGEMENT PLAN GUIDANCE

A Data Management Plan ("DMP") explains how data generated in the course of the research or work performed under an assistance award will be shared and preserved or, when justified, explains why data sharing or preservation is not possible or scientifically appropriate.

DMP Requirements

In order for a DMP to be considered acceptable, the DMP must address the following:

At a minimum, the DMP must describe how data sharing and preservation will enable validation of the results from the proposed work, or how results could be validated if data are not shared or preserved.

The DMP must provide a plan for making all research data displayed in publications resulting from the proposed work digitally accessible at the time of publication. This includes data that are displayed in charts, figures, images, etc. In addition, the underlying digital research data used to generate the displayed data should be made as accessible as possible in accordance with the principles stated above. This requirement could be met by including the data as supplementary information to the published article, or through other means. The published article should indicate how these data can be accessed.

The DMP should consult and reference available information about data management resources to be used in the course of the proposed research work. In particular, a DMP that explicitly or implicitly commits data management resources at a facility beyond what is conventionally made available to approved users should be accompanied by written approval from that facility. In determining the resources available for data management at DOE User Facilities, researchers should consult the published description of data management resources and practices at that facility and reference it in the DMP. Information about other DOE facilities can be found in the additional guidance from the sponsoring program.

The DMP must protect confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; avoid significant negative impact on innovation, and U.S. competitiveness; and otherwise be consistent with all laws (e.g., export control laws), and DOE regulations, orders, and policies.

Data Determination for a DMP

The Principal Investigator should determine which data should be the subject of the DMP and, in the DMP, propose which data should be shared and/or preserved in accordance with the DMP Requirements noted above.

For data that will be generated through the course of the proposed research/work, the Principal Investigator should indicate what types of data should be protected from immediate public disclosure by DOE (referred to as "protected data") and what types of data that DOE should be able to release immediately. Similarly, for data developed outside of the proposed research work at private expense that will be used in the course of the proposed research work, the Principal Investigator should indicate whether that type of data will be subject to public release

or kept confidential (referred to as “limited rights data”). Any use of limited rights data or labeling of data as “protected data” must be consistent with the DMP Requirements noted above.

Suggested Elements for a DMP

The following list of elements for a DMP provides suggestions regarding the data management planning process and the structure of the DMP:

Data Types and Sources: A brief, high-level description of the data to be generated or used through the course of the proposed research work and which of these are considered digital research data necessary to validate the research findings or results.

Content and Format: A statement of plans for data and metadata content and format including, where applicable, a description of documentation plans, annotation of relevant software, and the rationale for the selection of appropriate standards. Existing, accepted community standards should be used where possible. Where community standards are missing or inadequate, the DMP could propose alternate strategies for facilitating sharing, and should advise the sponsoring program of any need to develop or generalize standards.

Sharing and Preservation: A description of the plans for data sharing and preservation. This should include, when appropriate: the anticipated means for sharing and the rationale for any restrictions on who may access the data and under what conditions; a timeline for sharing and preservation that addresses both the minimum length of time the data will be available and any anticipated delay to data access after research findings are published; any special requirements for data sharing, for example, proprietary software needed to access or interpret data, applicable policies, provisions, and licenses for re-use and re-distribution, and for the production of derivatives, including guidance for how data and data products should be cited; any resources and capabilities (equipment, connections, systems, software, expertise, etc.) requested in the research proposal that are needed to meet the stated goals for sharing and preservation (this could reference the relevant section of the associated research proposal and budget request); and whether/where the data will be preserved after direct project funding ends and any plans for the transfer of responsibilities for sharing and preservation.

Protection: A statement of plans, where appropriate and necessary, to protect confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; and avoid significant negative impact on innovation, and U.S. competitiveness.

Rationale: A discussion of the rationale or justification for the proposed data management plan including, for example, the potential impact of the data within the immediate field and in other fields, and any broader societal impact.

Additional Guidance

In determining which data should be shared and preserved, researchers must consider the data needed to validate research findings as described in the DMP Requirements, and are encouraged to consider the potential benefits of their data to their own fields of research, fields other than their own, and society at large.

DMPs should reflect relevant standards and community best practices and make use of community accepted repositories whenever practicable.

Costs associated with the project description/scope of work and resources articulated in a DMP may be included in the proposed research budget as permitted by the applicable cost principles.

To improve the discoverability of and attribution for datasets created and used in the course of research, DOE encourages the citation of publicly available datasets within the reference section of publications, and the identification of datasets with persistent identifiers such as Digital Object Identifiers (DOIs). In most cases, DOE can provide DOIs free of charge for data resulting from DOE-funded research through its Office of Scientific and Technical Information (OSTI) DataID Service.

Definitions

Data Preservation: Data preservation means providing for the usability of data beyond the lifetime of the research activity that generated them.

Data Sharing: Data sharing means making data available to people other than those who have generated them. Examples of data sharing range from bilateral communications with colleagues, to providing free, unrestricted access to anyone through, for example, a web-based platform.

Digital Research Data: The term digital data encompasses a wide variety of information stored in digital form including: experimental, observational, and simulation data; codes, software and algorithms; text; numeric information; images; video; audio; and associated metadata. It also encompasses information in a variety of different forms including raw, processed, and analyzed data, published and archived data.

Research Data: The recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This 'recorded' material excludes physical objects (e.g., laboratory samples). Research data also do not include:

(A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and

(B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.”

Validate: In the context of DMPs, validate means to support, corroborate, verify, or otherwise determine the legitimacy of the research findings. Validation of research findings could be accomplished by reproducing the original experiment or analyses; comparing and contrasting the results against those of a new experiment or analyses; or by some other means.