

Guidance for Data Management Plans for DMREF Proposals

The following guidance is provided to assist Designing Materials to Revolutionize and Engineer our Future (DMREF) investigators, reviewers, and Program Officers in developing and evaluating effective, complete, and competitive Data Management Plans (DMPs). It is important to recognize that while all DMPs should address the five categories of information as specified in the National Science Foundation's (NSFs) Proposal and Award Policies and Procedures Guide (PAPPG), they should not be generic. Each DMP should appropriately identify the data, metadata, samples, software, algorithms, curricula, documentation, publications, and other materials generated in the course of the proposed research. Moreover, the DMPs should describe how these materials will be disseminated, made accessible, and archived while incorporating the best practices and standards for the proposed research. DMREF relies on the merit review process to determine the potential for DMPs to serve the community.

Data are a product or byproduct of most scientific research. The ability to make data easily accessible in digital form enables a vision for how materials research can be done more efficiently and in ways that enable research to effectively build on past research. The Materials Genome Initiative (MGI) envisions how easily found, accessed, and reused digital data can accelerate the discovery of new materials and speed their incorporation into new products. More generally, data accessibility is a prerequisite for materials research at the desktop. This aspect is embraced by the broader materials community. An effective Data Management Plan (DMP) supports data provenance and assures that proper credit is ascribed to the creator of the data.

NSF POLICY REQUIREMENTS

According to [NSF's Policy](#), investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. The implementation of this policy requires that proposals to the NSF contain DMPs not exceeding two pages, uploaded into the Supplementary Documentation section of the proposal, as described in [Chapter II.C.2.j](#) of the NSF PAPPG. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results and products of the project ([Chapter XI.D.4](#)).

DMREF SPECIFIC GUIDANCE: OVERVIEW

The DMREF program recognizes the need for flexibility in developing DMPs that are appropriate for the practices and needs of each of the diverse research areas under its purview. The DMP must be consistent with community expectations and best practices appropriate for the proposed research and education activities. DMREF relies on the process of peer review to enable the broad materials community to determine the adequacy and responsiveness of a DMP.

Increasingly, modern materials research values and expects data in digital form that is Findable, Accessible, Interoperable, Reusable, (FAIR) and properly presented together with metadata. The metadata provides adequate information about the data to enable reproduction. Data available in this way accelerates materials research, enables and supports data intensive research, and may be reproduced and extended by other researchers. These expectations are reflected in the reviewing community.

Data management under an award is expected to be dynamic. Annual reports must discuss how the DMP was carried out and record changes made to that plan in the course of the project (see below).

DATA MANAGEMENT PLAN CONTENT

The content of the DMP provides the explanation of how the proposal complies with NSF policy and prevailing best practices on dissemination and sharing of the research and education products of the project. Because there is community interest in capturing research data in digital form and making it broadly available in a form that is FAIR, the discussion below will expand considerations for data and only briefly comment on other products. The DMP must include adequate project-specific detail for evaluation of its appropriateness and feasibility during merit review, thus convincing reviewers that it is consistent with the research and education data products produced by the specific project. [Dear Colleague Letter: Effective Practices for Data](#) highlights two effective data practices (use of persistent IDs for research data and use of DMP tools that create machine readable DMPs) that may be useful in developing an efficacious DMP.

In an effort to assist the DMREF community in developing effective DMPs, the five essential components of the DMP identified in the PAPPG [Chapter II.C.2.j](#) are listed below along with examples of the types of questions that PIs should consider when constructing their proposed DMPs. It is important to note that while it is not necessary to answer all of the specific sub-questions below, an effective DMP should clearly state how the PIs plan to address each of these components:

1. **Products of Research:** Describe the types of data and products to be produced during the project. Examples of data and products include: materials samples; characterization data; (meta)data that provides information on the data, e.g. synthesis conditions or community codes used; simulation data; and software. Data and other products generated from Broader Impact activities, such as education materials and assessment results, should also be included in the plan, together with Institutional Review Board (IRB) considerations and clearance, if applicable. This inventory should inform the scope of the DMP and the requirements to preserve, curate, and share the products that result from the project. The DMP should describe the roles and responsibilities of all parties with respect to the management of data (including contingency plans for the departure of key personnel from the project) both during and after the grant cycle.

Questions to be considered:

- What types of data (experimental, computational, or text-based), metadata, samples, physical collections, models, software, curriculum materials, and other materials will be collected and/or generated in the course of the project?
 - What descriptions of the metadata are needed to make the actual data products useful and reproducible for the general researcher?
2. **Data Format Standards:** Describe the format and media in which the data or products along with metadata are stored. The description should discuss the rationale for the format and to what extent it conforms to any existing standards, e.g. formats for image data, instrument outputs, and simulation data. Existing standards for data and metadata format and content should be used insofar as they facilitate the reuse of the data and its further processing. When existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies. In general, solutions and remedies to providing data in an accessible format should be offered with minimal added cost.

Questions to be considered:

- In what format and/or media will the data or products be stored (e.g., hardcopy notebook, and / or instrument outputs, ASCII, html, jpeg, or other formats?) Does the data format facilitate further analysis through widely used software tools? Is it compliant with other instruments?
 - Where data are stored in unusual or not generally accessible formats, how may the data be converted to more accessible formats or otherwise made available to interested parties?
3. **Access to Data and Data Sharing Practices and Policies:** Mechanisms for sharing data among DMREF team members should be addressed. Data should generally be accessible to interested external parties *without* need for explicit or required requests. Plans should be provided for enabling broad community access to data, including websites maintained by the research groups and direct contributions to appropriate public databases or repositories. Practices regarding the release of data for access should be described. For example, data and data products will be made available on completion of the project.¹ Persistent IDs, such as Digital Object Identifiers ([DOI](#)) can enable proper citation for suitably-archived, publishable data sets. A DOI is often automatically obtained when data are published in a major repository. Significant software or code developed as part of the project should be distributed open-source, and should include a description of how users can access the code, how to obtain documentation on how to use the code, and the conditions under which they can use and modify the code. A software license should be explicitly specified, if applicable.

Questions to be considered:

- What specific dissemination approaches will be used to make data available and accessible to others, including any pertinent metadata needed to interpret the data?
 - What plans, if any, are in place for providing access to data, including websites maintained by the research group and contributions to public databases/repositories?
 - If maintenance of a website or database is the direct responsibility of the research group, what is the period of time the website or database is expected to be maintained?
 - Will data be registered and indexed to enable their discovery?
 - What are the practices or policies regarding the release of data – for example, are they available before or after formal publication? What is the approximate duration of time that the data will be kept private?
 - What are the policies for data sharing, including, where applicable, provisions for protection of privacy, confidentiality, intellectual property, national security, or other rights or requirements?
4. **Policies for Re-Use, Re-Distribution, and Production of Derivatives:** For data deemed re-usable, it must be accompanied by any metadata needed to reproduce the data, e.g., the means by which it was generated, detailed analytical and procedural information required to reproduce experimental results, and other pertinent metadata. Describe the policies regarding the use of

¹Note that data should be disseminated in a timely matter to facilitate scientific progress. The PAPPG ([Chapter XI.D.4](#)) provides potentially helpful information on balancing dissemination and intellectual property.

data provided via general access or sharing, or specific licensing provisions, if applicable. Practices for appropriate protection of privacy, confidentiality, security, intellectual property, and other rights should be communicated. Describe the rights and obligations of those who access, use, and share your data.

Question to be considered:

- If you plan to provide data and images on a website, will the website contain disclaimers or condition regarding the use of the data in other publications or products?

5. **Archiving of Data, Samples, and Other Relevant Research Products:** Describe plans for archiving data, samples, and other relevant research products. If the data will be archived by a third party, please refer to their preservation plans (if available). Where no data or sample repository exists for collected data or samples, metadata should be prepared and made publicly available over the Internet and the PI should employ alternative strategies for complying with the general philosophy of sharing research products and data as described above.

Questions to be considered:

- How will the research products including data be preserved and stored?
- What measures will be taken to assure that they will be maintained after the grant ends?
- When and how will data be archived and how will access be preserved over time? For example, will hardcopy logs, instrument outputs, and physical samples be stored in a location where there are safeguards against fire or water damage?
- Is there a plan to transfer digitized information to new storage media or devices as technological standards or practices change?
- Will there be an easily accessible index that documents where all archived data are stored and how they can be accessed?

In the spirit of promoting an open digitally accessible materials research environment, a minimal strategy would be to make the data findable and accessible to the community in a form that links the data to adequate annotation, including what the data are and what parameters were used to generate them utilizing robust mechanisms. The latter could include well-maintained and sustained websites, digital libraries, repositories, and other data resources, that should be described in annual reports.

DMREF encourages investigators to use persistent identifiers (e.g., DOIs) as a long-lasting reference to a digital resource (see [DOI](#)) that can aid in making data findable and citable. Repositories often assign DOIs automatically when datasets are submitted. Publications from new awards resulting from proposals submitted after January 25, 2016 must be deposited in the NSF Public Access Repository (NSF-PAR). For more information, see NSF's [Public Access Initiative](#) and [Frequently Asked Questions \(FAQs\) for Public Access](#).

BUDGETARY CONSIDERATIONS

According to the PAPPG ([Chapter II.C.2.g.\(vi\).b](#)), "the proposal budget may request funds for the costs of documenting, preparing, publishing or otherwise making available to others the findings and products of the work conducted under the grant." The cleanup, documentation, storage and indexing of data and databases are among allowed items in the proposal budget (Line G). Infrastructure, human resources, and education may also be involved in an effective plan to manage data that is appropriate for the project. A compelling justification for any costs associated with implementing the Data Management Plan should

appear in the Budget Justification section of the proposal. Consistent with community expectations, DMREF encourages innovations that, where appropriate and practical, enable efficient and effective data curation, sharing, reuse, and management through cyberinfrastructure that operates under the principles that data should be findable, accessible, interoperable, and reusable. Data management strategies should use and leverage existing cyberinfrastructure and resources to the fullest extent practical.

REPORTING

If an award is made, data-related activities and actions taken to execute the DMP must be described in annual and final project reports, and through subsequent proposals. The [NSF guidance on Technical Reporting Requirements](#) states that reports should describe actions taken during the reporting period to bring a proposal's data management plan to completion.

Annual Reports required for all NSF multi-year awards should include information about progress made in data management and sharing of research products (e.g., identifier or accession numbers for data sets, citations of relevant publications, conference proceedings, and other types of data sharing and dissemination). These activities may be documented under Accomplishments, as Major Activities, Other Achievements, or in response to how the results have been disseminated, as appropriate. NSF encourages investigators to employ persistent identifiers for all research products as a long-lasting reference to digital resources. The NSF report template includes specific sections on the accomplishments and products of the research. The sections "How have the results been disseminated to communities of interest?", "Other Products", and "Websites" may be particularly helpful in discussing how data and software products have been disseminated to the community. URLs for archived metadata and data may be included in the section entitled "Products-Websites."

Final Reports should describe the implementation of the DMP and include any changes from the original DMP. The Final Report should clearly describe the following information:

- The data produced during the award period,
- The data that will be retained after the award expires,
- How the data will be disseminated along with verification that data will be accessible or made available for sharing,
- The format (including reference to any and all pertinent metadata) that will be used to make the data available and usable by others, and
- Where the data generated by the project have been deposited/are being stored for long term public access.

Final Reports must document compliance or explain why it did not occur. In cases where the Final Report is due before the required date of sample or data submission, the PI must report submission of metadata and plans for final submission. The PI should notify the cognizant Program Officer by e-mail after final data and/or sample submission has occurred, even if this is after the expiration date of the award.

Results from Prior NSF Support

A description of data and other products created or generated during the research supported by an NSF award must be included in the section 'Results from Prior NSF Support'. The following information should be provided and reflects on past data management, as discussed in the PAPPG [Chapter II.C.2.d.iii](#):

(e) evidence of research products and their availability, including, but not limited to: data, publications, samples, physical collections, software, and models, as described in any Data Management Plan;

In this way, data management and the products of the project are subject to the review process of future proposals through the evaluation of Results from Prior NSF Support.

ADDITIONAL RESOURCES

There are many resources available to PIs that provide specific guidance on data management practices. The following list is neither exhaustive nor intended to endorse these particular resources.

These provide an entry point for assistance:

- Journals and data repositories may have specific formatting and metadata requirements for data publishing or archival deposit.
- Professional and scholarly societies often provide guidance for the community.
- Numerous non-governmental organizations are now offering resources and training.
- Many university libraries provide resource guides on data management planning and best practices; some provide direct support for DMP development.

DISCLAIMER

The preceding guidelines are not intended to replace the guidance given in the PAPPG and solicitations. In any perceived conflict, the PAPPG or the solicitation will take precedence as appropriate for the proposal.