Directorate of Mathematical and Physical Sciences

Division of Materials Research (DMR)

Updated Guidance to Principal Investigators on Data Management Plans, March 18, 2020

Data are a product or byproduct of most scientific research. As such, it must satisfy NSF policy as discussed first in the guidance below. 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 provides one example of 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 DMR community and forms the basis of DMR-specific guidance; a good DMP supports data provenance and assures that proper credit is ascribed to the creator of the data.

DATA MANAGEMENT PLAN: NSF POLICY REQUIREMENTS

According to NSF 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 (https://www.nsf.gov/bfa/dias/policy/dmp.jsp). The implementation of this policy requires that proposals to the National Science Foundation contain an at most two-page-long Data Management Plan (DMP) uploaded into the Supplementary Documentation section of the proposal, as described in Chapter II.C.2.j of the NSF Proposal and Award Policies and Procedures Guide (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).

DIVISION OF MATERIALS RESEARCH SPECIFIC GUIDANCE: OVERVIEW

The Division of Materials Research (DMR) recognizes the need for flexibility in developing Data Management Plans that are appropriate for the practices and needs of each of the diverse research areas under its purview. The Data Management Plan must be consistent with community expectations and best practices appropriate for the proposed research and education activities. DMR relies on the process of peer review to enable the broad materials community to determine the adequacy and responsiveness of a Data Management Plan.

Increasingly, modern materials research values and expects data in digital form that is findable, accessible, interoperable, reusable, and properly presented together with metadata. The metadata provide 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; see the more detailed comments on Data Management Plans in **DATA MANAGEMENT PLAN CONTENT** below.

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

A Data Management Plan that states that a detailed plan is not required can be valid provided that the assertion is accompanied by a clear and compelling project-specific justification as to why this is the case.

DATA MANAGEMENT PLAN CONTENT

The content of the Data Management Plan (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 findable, accessible, interoperable, and reusable, the discussion below will expand considerations for data and only briefly comment on other products. The DMP should include adequate project-specific detail and should convince the 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 that may be useful in developing an efficacious Data Management Plan.

For many projects, an effective DMP will respond appropriately to the specific elements identified in the PAPPG Chapter II.C.2.j:

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 Data Management Plan and the requirements to preserve, curate, and share the products that result from the project.***

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. **Does the data format facilitate further analysis through widely used software tools? Is it compliant with other instruments?** 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. The need for deviation from existing standards, or for development of new ones, should be justified and relevant plans should be adequately documented.

3. **Access to Data and Data Sharing Practices and Policies**: Data should generally be accessible ***without*** need for explicit or required requests from interested parties. Plans should be provided for enabling broad community access to data, including websites maintained by the research group and direct contributions to appropriate public databases or repositories. **Will data be registered and indexed to enable their discovery?** 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 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.

4. **Policies for Re-Use, Re-Distribution, and Production of Derivatives:** Describe your policies regarding the use of 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. The rights and obligations of those who access, use, and share your data.

5. **Archiving of Data, Samples, and Other Relevant Research Products:** Describe plans for archiving data, samples, and other relevant research products. **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?**

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 and final project reports. DMR 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, DMR 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.

[1]Note 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.

ADDITIONAL CONSIDERATIONS FOR CENTER AND FACILITY PROPOSALS

DMR-supported facilities, including Materials Innovation Platforms (MIPs), and Shared Experimental Facilities supported by the Materials Research Science and Engineering Centers (MRSECs) provide services to the community in the form of access to instrumentation which results in users creating data. The associated Data Management Plans of facilities and MRSECs should describe plans and policies concerning storage, curation and access of data, addressing both intramural and extramural research activities. When appropriate, the DMP should reference or include provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. **Data Management Plan guidance and requirements provided in solicitations and other proposal preparation or review instructions specific to center and facility proposals supersede the considerations presented here.** For example, the Data Management Plan for a Materials Innovation Platform (MIP) proposal should be a component of a broader Knowledge Sharing Plan, as MIPs not only share data, but also share tools, codes, samples, and know-how.

REPORTING

If an award is made, data-related activities and actions taken to execute the Data Management Plan (DMP) should be described in annual and final project reports, and through subsequent proposals. The NSF guidance on Technical Reporting Requirements states that project reports should describe actions taken during the reporting period to bring a proposal's data management plan to completion. The NSF project report template includes specific sections on the accomplishments and products of the research, including how the results have been disseminated to communities of interest. The annual and final project report 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. Final project reports should describe the implementation of the Data Management Plan and include any major changes from the original Data Management Plan.

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.

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.