

# Directorate for Engineering Data Management Plans Guidance for Principal Investigators

updated: November 2018

The Directorate for Engineering (ENG) supports research covering a broad spectrum of communities of investigators, and each community has its own best practices. ENG is aware of the need to provide flexibility to programs, principal investigators (PIs), and reviewers in assessing the quality of individual Data Management Plans (DMPs) from various communities. Therefore, guidance has evolved to accommodate changing community standards and expectations. ENG relies on the merit review process to determine which DMPs best serve each community.

The following guidance is to assist ENG investigators, reviewers and Program Officers in developing and evaluating effective, complete, and competitive DMPs. It is important to recognize that while all DMPs should address the five categories of information as specified in the 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. DMPs are subject to peer review. Please contact your specific Program Officer if you have any questions related to DMPs in the program context.

## **PAPPG and NSF-WIDE REQUIREMENTS**

All proposals must include a supplementary document of no more than two pages labeled “Data Management Plan,” as described in [PAPPG Chapter II.C.2.j](#). The DMP is NOT part of the 15-page Project Description. ***Proposals that do not include a Data Management Plan will be returned without review.***

You may request funds to cover costs of publication, page charges, or preparation of data as a direct cost in your budget proposal, which is evaluated as part of the merit review process. Any costs associated with implementing the DMP should be explained in the Budget Justification.

Some NSF Program Solicitations may contain specific and/or additional instructions that deviate from this guidance and/or provide exceptions to the two-page limit. Instructions in the solicitation take precedence over this guidance. Please check solicitations carefully for this information.

## **DATA MANAGEMENT PLAN (DMP) CONTENT**

The DMP should clearly articulate how the investigators plan to manage and disseminate both the physical and digital data generated by the project, taking advantage of emerging information

technologies and cyberinfrastructure. **The plan must include sufficient detail for evaluation of its appropriateness and feasibility during merit review.** DMPs often include existing practices in the principal investigator's laboratory and the larger research community.

In an effort to assist the ENG community in developing effective DMPs, the five essential components of the DMP are listed below along with numerous examples of the types of questions 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 PI(s) plan to address each of the five essential components listed here:

1. Products of Research:

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? The DMP should describe the expected types of data to be retained, managed, and shared, and the plans for doing so. What descriptions of the metadata are needed to make the actual data products useful and reproducible for the general researcher? For collaborative proposals, 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.

2. Data Formats and Standards:

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)? 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? 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.

3. Dissemination, Access and Sharing of Data:

What specific dissemination approaches will be used to make data available and accessible to others, including any pertinent metadata needed to interpret the data? In this case, "available and accessible" refers to data that can be found and obtained without a personal request to the PI, for example by download from a public repository. 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? For software or code developed as part of the project, include a description of how users can access the code (e.g., licensing, open source) and specific details of the hosting, distribution and dissemination plans. 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? 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? “Data sharing” refers to the release of data in response to a specific request from an interested party. 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? Research centers and major partnerships with industry or other user communities should also address how data are to be shared and managed with partners, center members, and other major stakeholders; publication delay policies (if applicable) should be clearly stated.

4. Re-Use, Re-Distribution and Production of Derivatives:

What are your policies regarding the use of data provided via general access or sharing? For data to be 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. 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 with others should also be clearly articulated. For example, if you plan to provide data and images on your website, will the website contain disclaimers or condition regarding the use of the data in other publications or products?

5. Archiving of Data:

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? 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.

## **POST-AWARD MANAGEMENT**

After an award is made, the PI(s) must manage their data as described in the DMP and will be monitored primarily through the normal Annual and Final Report process and through evaluation of subsequent proposals. The NSF Guidance on [Technical Reporting Requirements](#) states that Annual and Final Reports should describe actions taken during the reporting period to bring a proposal’s Data Management Plan to completion. While these reports are a critical mechanism for communication between the PI and the award’s managing Program Officer, PI(s) are

encouraged to reach out to the cognizant Program Officer about changes to or difficulty implementing the proposed DMP.

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 (where these exist) as a long-lasting reference to digital resources. 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.

Subsequent Proposals submitted by the PI(s) to NSF will be partially evaluated on the implementation of the DMPs from previous awards. As described in the [PAPPG Part I Section II.C.2.d.iii](#), data management outcomes must be reported in the section “Results from Prior NSF Support” of future proposals.

## **RESOURCES**

There are many sites 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 may provide guidance for the community.

- Non-governmental organizations are now offering resources and training.
- The US Geological Survey, while intended for USGS researchers, has a wealth of training and best practices materials: <http://www.usgs.gov/datamanagement/index.php>
- Repository registry (and search) service: <http://www.re3data.org/>
- Finally, many university libraries now provide resource guides on data management planning and best practices; some provide direct support for DMP development.