

Data Management for NSF Engineering Directorate Proposals and Awards

Goal

Provide for clear, effective, and transparent implementation of NSF policy for data management and dissemination:

- (1) By requiring inclusion of a Data Management Plan (DMP) in proposals submitted to the Engineering Directorate (ENG);
- (2) By stating data-management expectations to help principal investigators compose their plans and to help reviewers evaluate them.

Reasons

NSF policy is summarized in the NSF Proposal and Award Policies and Procedures Guide:

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data ... created or gathered in the course of work under NSF grants.

From http://www.nsf.gov/publications/pub_summ.jsp?ods_key=gpg, Section VI.D.4.b.

The full policy specifically requires prompt preparation and submission of publications, recognizes allowances for restricting release of privileged or proprietary information, encourages sharing of software and inventions, and recognizes intellectual property rights. Such dissemination of data is necessary for the community to stimulate new advances as quickly as possible and to allow prompt evaluation of the results by the scientific community.

The requirement: Include a Data Management Plan in proposals

An appropriate DMP must be provided as a supplementary document (maximum of two pages) for all research proposals submitted to ENG and its programs. [As a supplementary document, it is not part of the 15-page limit for proposal bodies.] Efficiency and effectiveness of the DMP will be considered by NSF and its reviewers during the proposal review process. Archiving of both physical and digital data must be addressed in the plan, but in the present document, particular attention is given to the new needs for managing digital information.

Proposals must include sufficient information that peer reviewers can assess both DMPs and past performance. Strategies and eventual compliance with this policy will be evaluated not only by proposal peer review but also through project monitoring by NSF program officers, by division and directorate Committees of Visitors, and by the National Science Board.

Data to be managed

Under the following definitions, all data must be included in the DMP that result fully or in part from activities supported by ENG.

What data are included? Research data are formally defined as “the recorded factual material commonly accepted in the scientific community as necessary to validate research findings” by the U.S. Office of Management and Budget (1999).

The basic level of digital data to be archived and made available includes (1) analyzed data and (2) the metadata that define how these data were generated. These are data that are or that should be published in theses, dissertations, refereed journal articles, supplemental data attachments for manuscripts, books and book chapters, and other print or electronic publication formats.

- Analyzed data are (but are not restricted to) digital information that would be published, including digital images, published tables, and tables of the numbers used for making published graphs.
- Necessary metadata are (but are not restricted to) descriptions or suitable citations of experiments, apparatuses, raw materials, computational codes, and computer-calculation input conditions.

What data are not included at the basic level? The Office of Management and Budget statement (1999) specifies that this definition does not include “preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues.” Raw data fall into this category as “preliminary analyses.”

Exceptions. Some proposals may involve proprietary or other restricted data. For example, projects having proprietary information that will eventually lead to commercialization, such as Engineering Research Center (ERC), Nanoscale Science and Technology Center (NSEC), Industry/University Cooperative Research Center (I/UCRC), Small Business Innovative Research (SBIR), Small Business Technology Transfer (STTR), and Grant Opportunities for Academic Liaison with Industry (GOALI) awards. In addition, membership agreements, contracts, involvement with other agencies, and similar obligations may place some restrictions on data sharing.

Any such data-management issues should be discussed as well as the conditions that might prevent or delay the sharing of data. The proposal’s DMP would address the distinction between released and restricted data and how they would be managed. Exceptions to the basic data-management policy should be discussed with the cognizant program officer before submission of such proposals.

A similar situation arises from the special circumstances that arise from requirements to protect human subjects. Data-sharing policies for awards that involve human subjects should recognize and address human-subjects protocols and the need to protect privacy and confidentiality ([see http://www.nsf.gov/bfa/dias/policy/](http://www.nsf.gov/bfa/dias/policy/)).

Additional possible data management.

NSF-supported large research facilities and other focused research programs may specify more stringent data-sharing and archiving procedures for research conducted using these facilities or under these programs. This policy speaks to the minimal DMP requirements for all ENG supported research. More stringent data management requirements may be specified in particular NSF solicitations or result from local policies and best practices at the PI’s home institution. Any such more-stringent requirements will be specified in the program solicitation and award conditions. Principal Investigators to be supported by such programs must discuss in their Data Management Plans how they will meet these additional requirements.

Contents of the Data Management Plan

The DMP should clearly articulate how “sharing of primary data” is to be implemented. It should outline the rights and obligations of all parties as to their roles and responsibilities in the

management and retention of research data. It must also consider changes to roles and responsibilities that will occur should a principal investigator or co-PI leave the institution. Any costs should be explained in the Budget Justification pages. Specific components are listed below.

Expected data. The DMP should describe the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project. It should then describe the expected types of data to be retained.

Period of data retention. The DMP should describe the period of data retention. Minimum data retention of research data is three years after conclusion of the award or three years after public release, whichever is later.

Public release of data should be at the earliest reasonable time. A reasonable standard of timeliness is to make the data accessible immediately after publication, where submission for publication is also expected to be timely.

Exceptions requiring longer retention periods may occur when data supports patents, when questions arise from inquiries or investigations with respect to research, or when a student is involved, requiring data to be retained a timely period after the degree is awarded. Research data that support patents should be retained for the entire term of the patent.

Longer retention periods may also be necessary when data represents a large collection that is widely useful to the research community. For example, special circumstances arise from the collection and analysis of large, longitudinal data sets that may require retention for more than three years. Project data-retention and data-sharing policies should account for these needs.

Data formats and dissemination. The DMP should describe the specific data formats, media, and dissemination approaches that will be used to make data available to others, including any metadata. Policies for public access and sharing should be described, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

Research centers and major partnerships with industry or other user communities must also address how data are to be shared and managed with partners, center members, and other major stakeholders. Publication delay policies (if applicable) must be clearly stated. Investigators are expected to submit significant findings for publication quickly that are consistent with the publication delay obligations of key partners, such as industrial members of a research center.

Data storage and preservation of access. The DMP should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data. In collaborative proposals or proposals involving sub-awards, the lead PI is responsible for assuring data storage and access.

Post-award monitoring

After an award is made, data management will be monitored primarily through the normal Annual and Final Report process and through evaluation of subsequent proposals.

Subsequent proposals. Data management must be reported in subsequent proposals by the PI and Co-PIs under “Results of prior NSF support.”

References and resources

- Council on Governmental Relations, Access to and Retention of Research Data: Rights and Responsibilities, March 2006. <http://206.151.87.67/docs/CompleteDRBooklet.htm>
- National Science Foundation, Proposal and Award Policies and Procedures Guide, http://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp
- Office of Management and Budget, Circular A-110, September 30, 1999. White House Website, OMB Home. <http://www.whitehouse.gov/omb/circulars/a110/a110.html>