Data Management for NSF SBE Directorate Proposals and Awards

OVERVIEW

The Data Management Plan (DMP) should be short, no more than two pages, and submitted as a supplementary document. The plan will need to address:

- *What data are generated by your research?*

- *What is your plan for managing the data during and after your project is completed, including are you making your data and/or metadata available to others and how, or why not if you are not making the data available?*

This document is meant to provide guidance for investigators within the Social, Behavioral, and Economic (SBE) Sciences as they develop their Data Management Plans. A thoughtful and thorough consideration of data management issues for the SBE sciences is available in a workshop report, *Public Access to NSF-Funded Research Data for the Social, Behavioral, and Economic Sciences*. NSF's comprehensive Public Access Plan, *Today's Data, Tomorrow's Discoveries: Increasing Access to the Results of Research Funded by the National Science Foundation*, was published in 2015.

"Data" are defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings. This includes original data, but also "metadata" (e.g., experimental protocols, code written for statistical analyses, etc.).

It is acknowledged that there are many variables governing what constitutes "data," and the management of data. Moreover, each area of science has its own culture regarding data. The data management plan will be evaluated as part of your proposal. Proposals must include sufficient information so that peer reviewers can assess both the data management plan and past performance. The plan should reflect best practices in your area of research and should be appropriate to the data you generate.

BACKGROUND

The Requirement: Include a Data Management Plan in Proposals

An appropriate data management plan is required as a supplementary document (maximum of two pages) for all full research proposals submitted. This plan is to be included in the Supplementary Documents section of the proposal and is not part of the 15-page limit for the Project Description. The NSF will not accept any proposal submitted that is lacking a DMP. Even if no data are to be produced, (e.g., the research is purely theoretical or is in support of a workshop where there will be no report), a DMP is required. In such a case, the DMP can simply state that no data will be produced.

The plan should describe how the PIs will manage and disseminate data generated by the project, and how will data and/or metadata be stored and made available to the public. The DMP will be considered by NSF and its reviewers during the proposal review process. Strategies and eventual compliance with the proposed DMP will be evaluated not only by proposal peer review, but also through project monitoring by NSF program officers, by Committees of Visitors, and by the National Science Board.

NSF is aware of the need to provide flexibility in assessment of data management plans. In developing a plan, researchers may want to consult with university officials as many universities have explicit data management policies. Some professional organizations also have recommended data management practices and refer PIs to repositories. NSF does not endorse the use of any specific repository.

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 should also consider changes to roles and responsibilities that will occur should a principal investigator or Co-PI leave the institution or project. 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, or other materials to be produced during the project. It should then describe the expected types of data to be retained.

The Federal government defines 'data' in OMB Circular A-110 (now 2 CFR, Ch. II, §215.36(d)(2)(i), and codified in 5 U.S.C. 552(a)(4)(A)) as:

> Research data is defined as 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.

PIs should use the opportunity of the DMP to give thought to matters such as:

- The types of data that their project might generate and eventually share with others, and under what conditions
- How data are to be managed and maintained until they are shared with others
- Factors that might impinge on their ability to manage data, e.g. legal and ethical restrictions on access to non-aggregated data
- The lowest level of aggregated data that PIs might share with others in the scientific community, given that community's norms on data

- The mechanism for sharing data and/or making them accessible to others
- Other types of information that should be maintained and shared regarding data, e.g. the way it was generated, analytical and procedural information, and the metadata

Period of data retention. SBE is committed to timely and rapid data distribution. However, it recognizes that types of data can vary widely and that acceptable norms also vary by scientific discipline. It is strongly committed, however, to the underlying principle of timely access, and applicants should address how this will be met in their DMP statement.

Data formats and dissemination. The DMP should describe data formats, media, and dissemination approaches that will be used to make data and metadata available to others. 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.

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. These can include third party facilities and repositories.

Additional possible data management requirements. More stringent data management requirements may be specified in NSF solicitations or result from local policies and best practices at the PI's home institution. Additional requirements will be specified in the program solicitation and award conditions. Principal Investigators to be supported by such programs must discuss how they will meet these additional requirements in their Data Management Plans.

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.

Annual Reports. Annual reports, required for all multi-year NSF awards, must provide information on the progress on data management and sharing of the research products. This information could include citations of relevant publications, conference proceedings, and descriptions of other types of data sharing and dissemination of results.

Final Project Reports. Final Project Reports are required for all NSF awards. The Final Project Report must discuss execution and any updating of the original DMP. This discussion should describe:
- Data produced during the award;
- Data to be retained after the award expires;
- Verification that data will be available for sharing;
- Discussion of community standards for data format;
- How data will be disseminated;
- The format that will be used to make data available to others, including any metadata; and
- The archival location of data.

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