

Data Management Guidance for CISE Proposals and Awards

March 15, 2015

(To provide feedback and comments on this Guidance, please email CISEdmp@nsf.gov (<mailto:CISEdmp@nsf.gov>).

The National Science Foundation's (NSF) Proposal and Award Policies and Procedures Guide (PAPPG), which is available at https://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp [<https://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp>](https://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp), contains up-to-date information about NSF's policy on data management and dissemination of the products of research. Specifically, NSF's Award and Administration Guide (AAG; https://www.nsf.gov/publications/pub_summ.jsp?ods_key=aag [<https://www.nsf.gov/publications/pub_summ.jsp?ods_key=aag>](https://www.nsf.gov/publications/pub_summ.jsp?ods_key=aag)) describes NSF's policy on data management and dissemination of the products of research (see AAG Chapter VI.D.4), and NSF's Grant Proposal Guide (GPG; https://www.nsf.gov/publications/pub_summ.jsp?ods_key=gpg [<https://www.nsf.gov/publications/pub_summ.jsp?ods_key=gpg>](https://www.nsf.gov/publications/pub_summ.jsp?ods_key=gpg)) specifies the requirements of data management plans that principal investigators (PIs) must include as supplementary documents in all proposals that they submit to NSF (see [GPG Chapter II.C.2.j](https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#IIC2j)) [<https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#IIC2j>](https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#IIC2j).

This document provides **guidance** for CISE investigators to consider in developing their required data management plans. CISE affirms its commitment to the advancement of science and the interests of the public by thoughtful consideration of plans for dissemination and sharing of data and research products.

Summary of CISE's Guidance on Data Management Plans:

Beginning in January 2011 (and following the required period of notification and comment), NSF implemented a data management plan requirement in the GPG ([GPG Chapter II.C.2.j](https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#IIC2j) [<https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#IIC2j>](https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#IIC2j)). **All proposals must include a data management plan (DMP); NSF will not evaluate any proposal that is lacking a DMP.** Even if no research data are to be produced (e.g., the proposed activity entails conducting a workshop), a DMP is required. In such cases, the DMP is expected to discuss the management of the data that may be generated as part of the proposed activity (e.g., participant lists, exit surveys, community reports, etc.).

The DMP should be no more than two pages and must be submitted as a supplementary document. The DMP does not count toward the 15-page limit (or any solicitation-specific page limits) specified for the Project Description. For proposals submitted to CISE programs (or to multi-disciplinary programs that are led by CISE), the DMP should address the following:

- The types of data, metadata, samples, physical collections, software, curriculum materials, and other materials to be collected and/or generated in the course of the project;
- The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- The physical and/or cyber resources and facilities (including those supplied by third parties) that will be used to store and preserve the data after the grant ends;
- The policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- The policies and provisions for re-use, re-distribution, and the production of derivatives;
- The plans for archiving data, samples, and other research products, and for preservation of access to them after the award ends; and
- The roles and responsibilities of all parties with respect to the management of the data (including contingency plans for the departure of key personnel from the project) after the grant ends.

The DMP will be evaluated as an integral part of each proposal during the merit review process. It must include sufficient information to enable reviewers to assess both the current plan and past performance. The DMP should reflect best practices in the relevant research community(ies) and be appropriate for the data to be generated as part of the proposed activities.

Definition and Policy:

As noted in the Code of Federal Regulations (2 CFR 215.36), "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 specimens)."

This definition includes not only original data but also "metadata" (e.g., experimental protocols, software code written for statistical or experimental analyses or for proofs-of-concept, etc.).

As summarized in the AAG Section VI.D.4:

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...

...Investigators and grantees are encouraged to share software and inventions created under an award or otherwise make them or their products widely available and usable.

This is a longstanding NSF policy on intellectual property and the sharing and dissemination of research results. The full policy also recognizes intellectual property rights and the need to restrict release of privileged information.

Additional Guidance for DMP Content:

CISE is aware of the need to provide flexibility in assessment of data management plans. There are many variables governing what constitutes "data" and its management, and each community within CISE has its own practices. CISE divisions will rely heavily on the merit review process to determine which plans best serve each community, and will continually revise this Guidance document accordingly.

The DMP should clearly articulate how the PI and co-PIs plan to manage and disseminate data generated by the project. The plan should outline the rights and obligations of all parties as to their roles and responsibilities in the management and retention of research data, and consider changes that would occur should a PI or co-PI leave the institution or project. It should describe how the research team plans to deposit data into any relevant and appropriate disciplinary repositories that are appropriately managed and that are likely to maintain the metadata necessary for future use and discovery. Any costs associated with implementing the DMP should be explained in the Budget Justification.

The DMP should describe the types of data, metadata, scripts used to generate the data or metadata, experimental results, samples, physical collections, software, curriculum materials, or other materials to be produced in the course of the project. The plan should then describe the types of data to be retained, managed, and shared, and the plans for doing so. The DMP should cover the following, as appropriate for the project:

- the period of time the data will be retained and shared;
- how data are to be managed, maintained, and disseminated;
- factors that limit the ability to manage and share data, e.g., legal and ethical restrictions on access to human subjects data;
- provisions for appropriate protection of privacy, confidentiality, security, and intellectual property;
- mechanisms and formats for storing data and making them accessible to others, which may include third party facilities and repositories; and
- other types of information that would be maintained and shared regarding data, e.g. the means by which it was generated, detailed analytical and procedural information required to reproduce experimental results, and other metadata.

Note that individual solicitations may specify additional DMP requirements. If guidance specific to a particular program is not available, then the requirements established in the GPG apply, and CISE PIs are encouraged to consider the additional guidance provided in this document.

NSF maintains an FAQ on DMP at <https://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp>
<<https://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp>>.

Review and Reporting:

As noted in the GPG, the DMP will be considered by NSF and by reviewers during the merit review process. Additionally, PIs are reminded that, if any PI or co-PI identified on a proposed project has received NSF funding (including any current funding) in the five years preceding the proposal submission, information on that (those) award(s) is required, irrespective of whether the support was directly related to the proposal. As part of this information, 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, must be provided. For additional information, PIs are encouraged to refer to [GPG Chapter II.C.2.d\(iii\)\(e\)](#).

<https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#II.C.2.diii>.

Annual project 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). NSF encourages investigators to employ persistent identifiers for all research products (where these exist) and citation practices common to the discipline.

Final project reports required for all NSF awards should describe the implementation of the DMP, including any changes from the original DMP, and should contain 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 and verification that they will be available for sharing;
- The format (including community standards) that will be used to make the data – including any metadata – available to others; and,
- Where the data generated by the project have been deposited/are being stored for long-term public access (see "Additional Guidance on Selecting or Evaluating a Repository" below).

Data management outcomes should be reported in subsequent proposals by the PI and Co-PIs under the heading "Results of Prior NSF support."

Additional Guidance on Selecting or Evaluating a Repository:

The following questions are intended to assist PIs and panel members to prepare Data Management Plans and to evaluate them during merit review, respectively. The questions are sequential, that is, if (1) applies, then the remaining questions are irrelevant unless (2) also applies or the PI chooses to deposit the data or software in multiple repositories. The more detailed questions, (4)-(6), apply if (1) and (2) do not.

1. Does the solicitation specify a repository for the data or software?
2. Does the PI's home institution have an institutional repository that mandates local deposit of the data/software?
3. Is there a discipline-relevant repository used by the research community either as the expected repository for data/software or as the expected repository for discovering and reusing

data/software?

4. Is the repository sustainable? And if not, are there contingency plans?
5. Does the repository require at least minimal identification and description sufficient to enable discovery, access, and retrieval? For purposes of data citation, NSF requires a persistent identifier and some level of metadata including acknowledgement of the creator/author and federal support.
6. Has the PI made any contingency plans in the event a designated repository becomes unavailable?

National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, USA Tel: (703) 292-5111, FIRS: (800) 877-8339 | TDD: (800) 281-8749