

NSF Data Management Plan Requirement

All proposals submitted to the agency on or after January 18, 2011, are required to include a supplementary document of no more than two pages titled **Data Management Plan**.

The supplement should address how the proposed research will meet NSF policy on the dissemination and sharing of research results. See the NSF Proposal and Award Policies and Procedures Guide, Chapter II.C.2.j for full policy implementation:

http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp.

As a supplementary document, the data management plan (DMP) is not included in the 15-page limit for proposal bodies. Fastlane will not permit submission of a proposal that is missing the DMP. Proposers who feel that the plan cannot fit within the supplement limit of two pages may use part of the 15-page Project Description for additional data management information; the plan may not be used to circumvent the 15-page Project Description limitation.

The DMP will be reviewed as part of the intellectual merit or broader impacts of the proposal, or both, as appropriate. The goal is to provide clear, effective, and transparent implementation of the long-standing NSF POLICY ON DISSEMINATION AND SHARING OF RESEARCH RESULTS, which may be found in the Award Administration Guide, Section VI.D.4:

http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4.

After an award is made, compliance with the data management plan will be monitored through the Annual and Final Report process and through evaluation of subsequent proposals. Data management activities must be reported in subsequent proposals by the PI and Co-PIs under "Results of prior NSF support."

Each directorate may also have specific guidelines that address unique data management issues within the respective community. **Be sure to look at the proposal details and main directorate and/or division website for addition guidelines:** <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>. If guidance specific to the program is not available, follow the general NSF DMP guidelines.

The following categories are to be covered:

1. **Types of data** to be produced
2. **Data & metadata standards** to be used
3. Provision & conditions for **data access and sharing**, including requirements for protection of appropriate privacy, confidentiality, security or intellectual property rights
4. Policies & provision for **re-use & re-distribution** of research data
5. Plans for **long-term preservation and access**

This document includes a series of questions for researchers to consider when developing a NSF data management plan, and a quick checklist of these items.

Thanks to the University of Virginia Scientific Data Consulting Group for sharing their *NSF Data Management Plan Template*.

NSF Data Management Plan Checklist

For more details regarding these categories, see discussion questions below.

I. Types of data

1. Description of research data to be produced
2. Estimated amount of data to be produced
3. Method of data creation or capture
4. Software or tools needed for data creation or capture

II. Data & metadata standards

5. File formats of research data
6. File naming convention for data
7. Reasons for using these file formats
8. Metadata standards for describing the data
9. Method of metadata creation or capture
10. Reasons for using this metadata standard

III. Data access & sharing (and protection)

11. Method for providing access to research data
12. Date & duration of data availability
13. To whom the data will be made available
14. Method for storing & backing up data during research
15. Reasons data should not be shared with others
16. Actions needed to make data sharable
17. Methods for restricting or limiting access to data

IV. Data re-use & re-distribution

18. Intended or potential audience for the research data
19. Future potential uses of the data
20. Reasons for restricting re-use or re-distribution of data
21. Rights restrictions placed on data
22. Duration of data availability for re-use

V. Data preservation

23. Duration research data should be kept after conclusion of research
24. Archive, repository or data center chosen to preserve data (if available)
25. Changes made to data to prepare for long-term preservation
26. Metadata needed for long-term preservation
27. Resources needed for long-term preservation
28. Preservation procedures
29. Transfer of responsibility for research data

NSF Data Management Plan Elements

I. Types of data

Brief, high-level description of the data to be produced.

1. What kind of data are you collecting or producing? (e.g., numerical data, image data, text sequences, modeling simulation, etc.)
2. What is the estimated amount of data to be generated? (in gigabytes, terabytes, etc.)
3. How will data be generated or collected? Do you need special software or tools to create or collect the data?
4. If using existing data, where did you get them? What is the relationship between new research data and the existing data?

II. Standards for data and metadata

Description of the standards to be used for data and metadata content. If no standards currently exist or are inadequate, such information should also be documented, along with proposed remedies.

1. What file formats will you use for your data? (e.g., XLS, SQL, XML, TIF, etc.). You may also want to include information about file naming standards.
2. Why do you plan to use these formats? (e.g., the formats are non-proprietary; the formats are required for use with a specific tool or software, etc.)
3. What metadata (i.e., contextual details documenting the data to make it meaningful) are you using to describe the data? For example, documentation regarding how research data are created or modified.
4. Which metadata standards will you use? What form will the metadata take? If no metadata standards exist for your type of data, state this fact.
5. How will you create or capture metadata details?
6. Why have you chosen particular metadata standards or methods for creating contextual documentation? (e.g., staff expertise, accepted domain-specific standards, widespread usage, etc.)

III. Policies for data access and sharing (including protection of non-sharable data)

Description of how data will be accessed during and after research, including protection of confidential or proprietary data.

1. How will you make research data available? What resources and capabilities are required? (e.g., web portal, database, streaming server, etc.)
2. When will you make your data available? Include any embargo periods for commercial reasons.
3. What is the process for someone to gain access to the data? Who will be able to access the data?
4. What are your plans for short-term storage & backup? Who will have access to data while research is on-going? How is data backed up? Who is charge of managing research data and insuring reliability?

Statement of plans, where appropriate and necessary, for protections of privacy, confidentiality, security, intellectual property and other rights:

1. Are there any ethical or privacy concerns with the data?
2. If yes, how will these issues be resolved? (e.g., data will be made anonymous; system authentication will be required to access data, etc.)
3. What have you done to comply with IRB Protocol?
4. Are the data covered by copyright? Who owns the data?
5. If rights exist, how will the data be licensed? (e.g., Creative Commons License). Include any restrictions on data due to protection of intellectual property.

IV. Policies and provisions for data re-use and re-distribution

1. Who will likely be interested in using your data?
2. What are intended or possible future uses of the data?
3. Are there any reasons to not allow re-use and re-distribution of the data?
4. Will any permission restrictions need to be placed on data re-use?
5. How long will the data be accessible to others?

V. Plans for archiving and providing long-term access to data

Description of long-term strategy for maintaining, curating and preserving data in accessible form.

1. How long should your data be kept beyond the life of the project?
2. Have you identified an archive /repository (such as Georgia Tech's [SMARTech](#)) /central database /data center for depositing and preserving your data?
3. What transformations will be necessary to prepare data for long-term preservation? (e.g., data normalization, data anonymization, etc.)
4. What metadata or other documentation will be submitted with the data to make them accessible and reusable? How will future users make sense of the data?
5. If data will be preserved beyond project duration, what funding sources or institutional commitments will be necessary?
6. What procedures does your intended long-term data storage facility have in place for preservation (e.g., file migration, continued documentation) and backup?
7. If responsibility for continuing documentation, annotation, curation, access, and preservation (or its counterparts, de-accessioning or disposal) of data will move from one entity or institution, plans for managing the exchange and documentation of the necessary commitments and agreements should be provided.