National Science Foundation (USA): NSF - generic

Types of data

Note the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project

Guidance:

Provide a description of the data you will collect or re-use, including the file types, dataset size, number of expected files or sets, and content. Data types could include text, spreadsheets, images, 3D models, software, audio files, video files, reports, surveys, patient records, etc. **Data that underlie the findings reported in a journal article or conference paper should be deposited in accordance with the policies of the publication and according to the procedures laid out in the DMP included in the proposal that led to the award on which the research is based.**

Consider the following:

- What data will be generated in the research?
- What data types will you be creating or capturing?
- How will you capture or create the data?
- If you will be using existing data, state this and include how you will obtain it.
- What is the relationship between the data you are collecting and any existing data?
- How will the data be processed?
- What quality assurance & quality control measures will you employ?

Data and metadata standards

Note 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).

Guidance:

Describe the format of your data and how it will be documented. Think about what details (metadata) someone else would need to be able to use these files. For example, you may need a "readme file" to explain variables, structure of the files, etc.

Metadata associated with the data should conform to community standards and the requirements of the host repository. NSF does not currently specify a single metadata standard. However, any acceptable minimum set of data elements would include the names of all authors, date of publication or release, and Universal Resource Locator (URL) or other persistent identifier, as required by Biographical Sketches in proposals (Section 3.2.3 and Grant Proposal Guide, Chapter II C.2.f.i (c)).

Consider these questions:

- Which file formats will you use for your data, and why?
- What form will the metadata describing/documenting your data take?
- How will you create or capture these details?
- Which metadata standards will you use and why have you chosen them? (e.g. accepted domain-local standards, widespread usage)
- What contextual details (metadata) are needed to make the data you capture or collect meaningful?

Policies for access and sharing

Note the policies for access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements

Guidance:

Explain how and when the data will become available. Will data be accessible on a web page, by email request, via open-access repository etc.? If there is an embargo period for sharing the data, make sure you provide details explaining this delay (e.g. publisher, political, commercial, patent reasons). And if the data is of a sensitive nature - human subject concerns, potential patentability, species/ecological endangerment concerns - that public access is inappropriate, address here the means by which granular control and access will be achieved (e.g. formal consent agreements; anonymizing data; restricted access, only available within a secure network).

Practices governing use of embargos and delayed data release vary widely across the research communities supported by NSF and should be discussed as part of the DMP. For large-scale projects that are supported primarily to generate data for community use, the timing of release will be part of the award terms and conditions and clearly stated in the public award abstracts. NSF recognizes that some classes of data, particularly those that relate to human subjects, education, personally identifiable information, national security, or proprietary interests, may be subject to restrictions. Such restrictions must be described in the DMP and changes addressed in annual and final reports.

Small Business Innovation Research (SBIR)/Small Business Technology Transfer (STTR) proposals and any other proposal may allow for exceptions for proprietary or otherwise restricted data, including but not limited to personally identifiable information, business confidential information, security, among other concerns outlined in section 4.a. of the OSTP memo. Any such data management issues as well as conditions that might affect, delay, or limit data sharing should be discussed in the DMP. Coordination with the Cognizant Program Officer prior to submitting the proposal is also advised.

Consider these questions:

- How will you make the data available? (Include resources needed to make the data available: equipment, systems, expertise, etc.)
- When will you make the data available?
- What is the process for gaining access to the data?
- Will access be chargeable?
- How long will the original data collector/creator/principal investigator retain the right to use the data before making them available for wider distribution?
- Are there any embargo periods for political/commercial/patent reasons? If so, give details.
- Are there ethical and privacy issues? If so, how will these be resolved?
- What have you done to comply with your obligations in your IRB Protocol?
- Who will hold the intellectual property rights to the data and how might this affect data access?

Policies for re-use, redistribution

Note the policies and provisions for re-use, re-distribution, and the production of derivatives

Guidance:

Explain how the policies outlined in the previous question can be applied to the re-use and re-distribution of your data. Identify who will be allowed to use your data, how they will be allowed to use your data and whether or not they will be allowed to disseminate your data. If you are planning on restricting access, use or dissemination of the data, you must explain in this section how you will codify and communicate these restrictions. Consider the following:

- Will any permission restrictions need to be placed on the data?
- Who is likely to be interested in the data?
- What and who are the intended or foreseeable uses the data?

Plans for archiving and preservation

Note the plans for archiving data, samples, and other research products, and for preservation of access to them

Guidance:

Provide a description of your long-term strategy for archiving and preserving the data you plan to generate/use. **Data that underlie the findings reported in a journal article or conference paper should be deposited in accordance with the policies of the publication and according to the procedures laid out in the DMP included in the proposal that led to the award on which the research is based. All**

data resulting from the research funded by the award, whether or not the data support a publication, should be deposited at the appropriate repository as explained in the DMP. Rarely does NSF expect that retention of all data that are streamed from an instrument or created in the course of an experiment or survey will be required. See your specific directorate or solicitation for details. There are several technical strategies for achieving long-term preservation including redundancy, dark archives, secure data centers, and so on. In general, good practice calls for duplicating the collection at a geographically distinct location and for regular monitoring and format migration, given exigencies of media degradation and format obsolescence.

Consider the following:

- What is the long-term strategy for maintaining, curating and archiving the data?
- Which archive/repository/database have you identified as a place to deposit data?
- What procedures does your intended long-term data storage facility have in place for preservation and backup?
- How long will/should data be kept beyond the life of the project?
- What data will be preserved for the long-term?
- What transformations will be necessary to prepare data for preservation / data sharing?
- What metadata/ documentation will be submitted alongside the data or created on deposit/ transformation in order to make the data reusable?
- What related information will be deposited?