

NSF-PHY: Physics

National Science Foundation

Types of data produced

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

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

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).

Datasets need metadata to be usable. Think about what details (metadata) someone else would need to be able to use these files. For example, you may need a readme.txt file to explain variables, structure of the files, etc. Consider the following:

- What contextual details (metadata) are needed to make the data you capture or collect meaningful?
- 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?

Policies for access and sharing, and provisions for appropriate protection/privacy

Policies for access and sharing; Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

Explain how and when the data will become available. 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, address the means by which access will be restricted. Consider these questions:

- How will you make the data available?
- What resources are needed to access or use the data? Examples are software or equipment
- When will you make the data available?
- What is the process for gaining access to the data?
- 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?
- Who will hold the intellectual property rights to the data and how might this affect data access?

Policies and provisions for re-use, re-distribution

Policies and provisions for re-use, re-distribution, and the production of derivatives.

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. **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:

- 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

Plans for archiving data, samples, and other research products, and for preservation of access to them.

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?