**NSF Data Management Plan**

This reference sheet is designed to help you develop the 2 page Data Management Plan that must accompany your NSF proposal. The NSF review panel will consider your DMP during the merit review process as they assess the intellectual merit and broader impacts of your application.

We recommend using the bold numbered headings below, which have been taken directly from the NSF Proposal and Awards Policies and Procedures Guide (PAPPG). It is not necessary to use the underlined subheadings; these are just to help focus your attention on major topics/ issues to cover.

**IMPORTANT**: Some NSF Directorates provide specific guidance and/or impose additional Directorate-specific DMP guidelines. Details vary. You are responsible for determining whether the Directorate to which you are applying has such requirements and, if so, for ensuring that they are appropriately addressed. This link from the [University of Michigan with a summary of requirements by Directorate](http://www.lib.umich.edu/research-data-management-and-publishing-support/nsf-data-management-plans#contents) provides a concise overview and further links to the relevant NSF documentation. If in doubt, please consult your Program Officer.

**1. Data Description**

The proposed project will generate the following data *(give a short description of data to be gathered – nature, scope, and scale, amount if known and content of the data)*. *What types of data will you be creating or capturing (experimental measures, observational/qualitative, model simulation, processed, etc). How will you capture or create the data? If you will be using existing data, state that fact and explain where you obtained it.*

The project will also result in the production of the following materials: *briefly describe samples, physical collections, software, curricular materials, and other inventions/materials to be produced)*

Stewardship: *Who will act as the responsible steward for the data throughout the data life cycle? (Typically the PI oversees the collection and management of research data throughout the project period. If there is more than one PI, describe division of responsibilities.) Specify how you will ensure that the data meet quality assurance standards.*

Security: *How and where will you store copies of your research files to ensure their safety? How many copies will you maintain and how will you keep them synchronized?*

*How will you ensure that the data are secure? (This refers both to security for digital information and to security measures involved in protecting direct identifiers or links to direct identifiers during collection, cleaning, and editing of raw data. Processed data may or may not contain disclosure risk and should be secured in keeping with the level of disclosure risk inherent in the data. Secure work and storage environments may include access restrictions (e.g., passwords), encryption, power supply backup, and virus and intruder protection.)*

**2. Standards for Data and Metadata Format and Content**

Data Format**:** *Which file formats will you use for your data?* *Why have you chosen these particular standards/approaches (e.g., accepted standards in your discipline; staff expertise, Open Source, widespread usage; flexibility in display; preservation-readiness).* *(To the extent possible, preservation formats should be platform-independent and non-proprietary to ensure that they will be usable in the future.) (If existing standards are absent or inadequate, you will need to state that, along with how you propose to deal with this situation.)*

Data Organization: *If the organization of the data is in some way atypical, describe how the data will be managed during the project (with information about version control, naming conventions, etc. For example, some data collections are dynamically changing and version control is central to how the data will be used and understood by the scientific community.*)

Metadata: Metadata are often the only form of communication between the secondary analyst and the data producer, so they must be comprehensive and provide all of the needed information for accurate analysis.  *What types of metadata (contextual details) will you produce to support the data? How will you create or capture these details? What form will the metadata take? Will a metadata standard be used?* (structured or tagged metadata are often optimal, e.g., XML format of the Data Documentation Initiative standard.) *(If existing standards are absent or inadequate, you will need to state that, along with how you propose to deal with this situation.)*

**3. Policies for Access and Sharing**

Data Access: *How will you make the data available?* *Usually data access and sharing involves one or more of the following options: (1) self-dissemination through a dedicated web site, and/or (2) use of institutional repositories (e.g., DIVA) and/or (3) depositing the data in domain repositories. Briefly mention resources, equipment, systems, expertise — refer reader to your Facilities and Resources/Research Environment section for more details.*

*When will you make the data available? (Give details of any embargo periods for political/commercial/patent reasons.) Does the original data collector/creator/PI retain the right to use the data before opening it up to wider use?*

*What is the process for gaining access to the data? Will access be chargeable?*

Protection of Privacy: *Are there ethical and privacy issues? If so, how will these be resolved? (e.g., anonymisation of data, institutional ethics committees, formal consent agreements.)*

*Does the project require IRB approval? Are there legal constraints (e.g., HIPAA) on sharing data? How will you manage disclosure risk in the data to be shared and archived? Will your data be free of direct and indirect identifiers? If not, how will you share your restricted data? Will special terms of use be required? How will you handle informed consent with respect to communicating to respondents that the information they provide will remain confidential when data are shared or made available for secondary analysis?*

Confidentiality: *Privileged or confidential information should be released only in a form that protects the privacy of individuals and subjects involved. (General adjustments and, where essential, exceptions to this sharing expectation may be specified by the Program or Division/Office to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interests of investigators. A grantee or investigator also may request a particular adjustment or exception from the NSF Program Officer—if special arrangements of that nature have been made, describe.)*

Intellectual Property: *Is the dataset covered by copyright? If so, who owns the copyright to the data and other intellectual property/information created by the project? Will any copyrighted material (e.g., instruments or scales) be used? If so, how will the project obtain permission to use the materials and disseminate them?*

*(The PI’s university is usually considered to be the copyright holder for data the PI generates. Many archives do not ask for a transfer of copyright but instead just request permission to preserve and distribute the data. Copyright could also come into play if copyrighted instruments are used to collect data.)*

*How will the dataset be licensed if rights exist? (e.g., restrictions or delays on data sharing needed to protect intellectual property, copyright, or patentable data). Will copyright be transferred to another organization for data distribution and archiving?*

Legal Requirements**:** *Indicate whether any legal requirements apply to sharing your data. (e.g., data covered by HIPAA, proprietary data, and data collected through the use of copyrighted data collection instruments). Specify in detail how these issues have been taken into account in your data sharing plan.*

**4. Policies/Provisions for Re-Use, Redistribution, and Production of Derivatives**

*Which bodies or groups are likely to be interested in the data? What and who are the intended or forseeable uses/users of the data? Will any permission restrictions need to be placed on the data? Are there any reasons not to share or re-use the data? (e.g., ethical, non-disclosure, etc)*

**5. Plans for Archiving Data, Samples, and Other Research Products, and for Preservation of Access to Them**

Long-Term Archiving: *What is your long-term strategy for maintaining, curating, and archiving the data (e.g., How will you ensure that data are preserved for the long term? Will you use an archive/repository/central database/data center in which the data will ultimately be deposited for long-term archiving and preservation of access? How will the costs for creating data and documentation suitable for archiving be paid for?) You can request funds in your budget for this specific purpose if necessary.*

Preservation of Access: *What procedures do you (or your intended long-term data storage facility) have in place for preservation and backup? (important because digital data need to be actively managed over time to ensure that they will be available and usable as technologies change.) What transformations will be necessary in order to prepare the data for preservation/data sharing? (e.g., data cleaning; anonymization when appropriate). What metadata/documentation will be submitted along with the data or created on deposit/transformation in order to ensure that the data is reusable? What related information will be deposited? (reference reports, research papers, fonts, the original bid proposal, etc)*

Selection and Retention Periods:*How long will/should data be kept beyond the life of the project? Indicate how data will be selected for archiving, how long the data will be held, and what your plans are for eventual transition or termination of the data collection in the future. (Not all data need to be preserved in perpetuity, so thinking through the proper retention period for the data is important, in particular when there are reasons the data will not be preserved permanently.)*

**[REMINDER: YOUR FINAL DMP MAY NOT EXCEED 2 PAGES]**