Directorate of Mathematical and Physical Sciences: Advice to PIs on Data Management Plans

NSF has published a revised version of the NSF Proposal and Award Policies and Procedures Guide (PAPPG) (NSF 11-1) that will require, in all proposals submitted , or due, on or after January 18, 2011, a supplementary document of no more than two pages describing a Data Management Plan for the proposed research. Fastlane will not permit submission of a proposal that is missing the Data Management Plan. The Data Management Plan 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. This policy states:

a. Investigators are expected to promptly prepare and submit for publication, with authorship that accurately reflects the contributions of those involved, all significant findings from work conducted under NSF grants. Grantees are expected to permit and encourage such publication by those actually performing that work, unless a grantee intends to publish or disseminate such findings itself.

b. Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. 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 funding NSF Program or Division/Office for a particular field or discipline to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interest of investigators. A grantee or investigator also may request a particular adjustment or exception from the cognizant NSF Program Officer.

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

d. NSF normally allows grantees to retain principal legal rights to intellectual property developed under NSF grants to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community, to make results, data and collections available to other researchers.

e. NSF program management will implement these policies for dissemination and sharing of research results, in ways appropriate to field and circumstances, through the proposal review

process; through award negotiations and conditions; and through appropriate support and incentives for data cleanup, documentation, dissemination, storage and the like.

MPS-supported research covers a broad spectrum of communities of investigators, from individual investigators on experimental and theoretical topics to support for users at national and international facilities to large national and international collaborations of investigators involving tens or hundreds of individuals. Each Division within MPS has developed a set of information items to provide guidance to the communities served by that Division in preparing a Data Management Plan that will meet the goals of the NSF plan.

AST
CHE
DMR
DMS
PHY

MPS Divisions will rely heavily on the merit review process in this initial phase to determine those types of plan that best serve each community and update the information accordingly.

AST
Astronomy Division Proposals: Data Management Plan Guidelines

Background:

NSF has issued a revised version of the Proposal and Award Policies and Procedures Guide (PAPPG) which will require, in all proposals submitted on or after January 18, 2011, a supplementary document of no more than two (2) pages describing a Data Management Plan (hereafter DMP) for the proposed research. The goal is to provide clear, effective, and transparent implementation of the long-standing NSF policy which provides that: "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." Extra costs of data management or archiving are allowable within the proposal, but must be justified along with all other portions of the proposed budget.

In general, a DMP will contain descriptions of one or more aspects of data generation and handling. The plan should contain information on how the data will be managed but must not be used to circumvent the 15-page Project Description limitation by discussing other aspects of planned research. Proposals without a DMP will not be accepted by FastLane. In some cases, the DMP may contain a statement such as "a data management plan is not relevant to the proposed activities"; such statements should be explained by the proposer.

This short document provides guidelines that appear to be most helpful for AST proposers in developing the content of their DMP. We emphasize that these are guidelines that seem most appropriate for AST; the contents of the DMP are subject to the standard merit review process by panels and/or ad-hoc reviewers, and proposers should write the DMP with this merit review in mind.

The following five items are listed in the same order as those in the revised PAPPG. Examples are given to help guide the Astronomy community, but these guidelines are not intended to replace the guidance given in the PAPPG. In other words if there is conflict, the PAPPG will take precedence.

1. Products of the Research

Describe the types of data and products that will be generated in the research, such as images of astronomical objects, spectra, data tables, time series, theoretical formalisms, computational strategies, software, and curriculum materials.

2. Data Format

Describe the format in which the data or products are stored (e.g., ASCII, html, FITS, VO-compliant tables, XML files, etc.). Include a description of the metadata that will make the actual data products useful to the general researcher. Where data are stored in unusual or not generally accessible formats, explain how the data may be converted to a more accessible format or otherwise made available to interested parties. In general, solutions and remedies should be provided.

3. Access to Data and Data Sharing Practices and Policies

"Access to data" refers to data made accessible without explicit request from the interested party, for example those posted on a website or made available to a public database. Describe your plans, if any, for providing such general access to data, including websites maintained by your research group, and direct contributions to public databases. If maintenance of a web site or database is the direct responsibility of your group, provide information about the period of time the web site or data base is expected to be maintained. Note that data taken at national or private observatories may be accessible through public archives (perhaps after a standard proprietary period). Various forms of data (e.g., FITS images and tables, other data tables) also may be deposited with published articles in the AAS journals and other journals. Particular attention should be paid to data sets that are products of well-defined surveys. Also describe your practice or policies regarding the release of data for access, for example whether data are posted before or after formal publication.

"Data sharing" refers to the release of data in response to a specific request from an interested party. Describe your policies for data sharing, including where applicable provisions for protection of privacy, confidentiality, intellectual property, national security, or other rights or requirements.

4. Policies for Re-Use, Re-Distribution, and Production of Derivatives

Describe your policies regarding the use of data provided via general access or sharing. For example, if you plan to provide data and images on your website, will the website contain disclaimers, or conditions regarding the use of the data in other publications or products? If the data or products (e.g., images) are copyrighted (by a journal, for example), how will this be noted on the website?

5. **Archiving of Data**

Describe whether and how data will be archived and how preservation of access will be handled. If the data will be archived by a third party (e.g., national observatory or journal), please refer to their preservation plans if available.

CHE
Division of Chemistry Advice to PIs on Data Management Plans

NSF has issued a revised version of the Proposal and Award Policies and Procedures Guide (PAPPG) which will require, in all proposals submitted on or after January 18, 2011, a supplementary document of no more than two pages describing a Data Management plan for the proposed research (see Section VI-8, NSF-10-1). The goal is to provide clear, effective, and transparent implementation of NSF's long-standing policy on the dissemination and sharing of research results. Details of this policy are given at the following link: Award Administration Guide, Section VI.D.4.

The Division of Chemistry recognizes that in most cases, principal investigators will publish data (and relevant supplementary information) in peer-reviewed journal articles within a reasonable time, and that the chemistry research community maintains a significant number of databases that provide for access to data. Such disclosure of data meets the majority of needs for robust and open scientific discourse. The purpose of the Data Management Plan is to provide a means for highlighting the existing practices of the principal investigator's laboratory and larger research community, and to encourage innovations that, where appropriate and practical, take advantage of emerging information technologies and cyber-infrastructure.

In general, a Data Management plan will describe the nature of the data or products generated by the research, and how these will be managed with regard to access, sharing and archiving. The plan should not be used to circumvent the 15-page Project Description limitation in a proposal. The Data Management plan will be subject to peer review, and proposals without such plan will

be returned without review. In certain special cases, the Data Management plan may contain a statement such as "a data management plan is not relevant to the proposed activities." Such exemptions will be explicitly noted in the Program Description or solicitation. Finally, principal investigators and their institutions should be aware of the provisions of the Bayh-Dole Act (35 U.S.C. § 200-212, implemented under C.F.R. 401), which afford the institution certain rights to the intellectual property, inventions, and other products derived from federally funded research.

The following five items are listed in the same order as those in the guidelines of the revised PAPPG, and have been paraphrased to guide the Chemistry community. They are not intended to supplant the guidance given in the PAPPG.

1. **Products of the Research**

   Describe the types of data and products that will be generated in the research, for example numerical data on chemical systems such as spectra, diffraction patterns, physical properties, time-dependent information on chemical and physical processes, theoretical formalisms, computational strategies, final or intermediate numerical results from theoretical calculations, software, and curriculum materials.

2. **Data Format**

   Describe the format in which the data or products are stored (e.g., hardcopy notebook and/or instrument outputs, ASCII, html, jpeg or other formats). Where data are stored in unusual or not generally accessible formats, explain how the data may be converted to a more accessible format or otherwise made available to interested parties. You may also comment on the current or anticipated need for interested parties outside of your laboratory to access your primary data.

3. **Access to Data and Data Sharing Practices and Policies**

   "Access to data" refers to data made accessible without explicit request from the interested party, for example those posted on a website or made available to a public database. Describe your plans, if any, for providing such general access to data, including websites maintained by your research group, and direct contributions to public databases (e.g., the Protein Data Bank, Cambridge Crystallographic Data Centre, Inorganic Crystal Structure Database in Karlsruhe, Zeolite Structure Database). Also note if you submit your data in the form of tables, graphs, computer code or other format to the supplementary materials sections of peer-reviewed journals. Describe your practice or policies regarding the release of data for access, for example whether data are posted before or after formal publication. Finally, note as well any anticipated inclusion of your data into databases that mine the published literature (e.g., PubChem, NIST Chemistry WebBook).

"Data sharing" refers to the release of data in response to a specific request from an interested party. Describe your policies for data sharing, including where applicable provisions for protection of privacy, confidentiality, intellectual property, national security, or other rights or requirements.

4. Policies for Re-Use, Re-Distribution, and Production of Derivatives

Describe your policies regarding the use of data provided via general access or sharing. For example, if you plan to provide data and images on your website, will the website contain disclaimers, or conditions regarding the use of the data in other publications or products? Describe these disclaimers and/or terms of use.

5. Archiving of Data

Describe how data will be archived and how preservation of access will be handled. For example, will hardcopy notebooks, instrument outputs, and physical samples be stored in a location where there are safeguards against fire or water damage? Is there a plan to transfer digitized information to new storage media or devices as technological standards or practices change? Will there be an easily accessible index that documents where all archived data are stored and how they can be accessed?

We reiterate that the Data Management Plan requirement will apply to proposals submitted on or after January 18, 2011. Proposals submitted during the November 2010 CHE submission window are not subject to this requirement. However, those investigators who elect to include a Data Management Plan in November submissions will not be asked to remove them prior to the commencement of the review process.

DMR

Division of Materials Research Advice to PIs on Data Management Plans

The Division of Materials Research will rely on the process of peer review to allow the broad materials' community to identify best practices. Due to the diverse communities supported by DMR, the Division is not in a position to recommend a Division-specific single data sharing and archiving approach.

DMR PIs should include in their Data Management Plan those aspects of data retention and sharing that would allow them to respond to a question about a published result. Members of formal collaborations may refer to the collaboration's existing policies and practices. Guidance on what is to be included in a Data Management Plan can be found at the following link: Award Administration Guide, Section VI.D.4.

DMS

Division of Mathematical Sciences Advice to PIs on Data Management Plans

Proposals submitted or due to NSF on or after January 18, 2011 are required to provide plans for management and sharing of data produced by NSF-supported research. For many proposals to DMS, a statement that no data management plan is necessary will suffice, provided that a clear justification for this claim is given. All principal investigators should consult the *Proposal and Award Policies and Procedures Guide* for points to be addressed in a data management plan and its inclusion within a proposal.

Many of the proposals to DMS that require significant data management plans will be interdisciplinary submissions that involve the mathematical sciences and research topics, data, or collaborators from one or more other disciplines. In such cases DMS expects principal investigators to address the customary data practices of partner disciplines in their proposals' data management plans, and reviewers are likely to be asked to comment on the suitability of those plans from the perspectives of the relevant disciplines. Principal investigators may find it useful to consult statements on data management plans from other divisions and directorates of NSF in the preparation of interdisciplinary proposals.

PHY

The Physics Division is not in a position to recommend a Division-specific single data sharing and archiving approach applicable to the disparate communities supported through the Division. The Division will rely on the process of peer review to allow each of these communities to identify best practices.

Physics Division PIs should include in their Data Management Plan those aspects of data retention and sharing that would allow them to respond to a question about a published result. If there is no such data, this should be stated.

Members of formal collaborations may refer to the collaboration's existing policies and practices.