Division of Chemistry
Advice to Principal Investigators on Data Management Plans
January 2, 2018

The Division of Chemistry (CHE) offers the following guidance for CHE investigators to consider in developing required Data Management Plans (DMPs) for their proposal submissions. This document is a supplement to the data management plan requirements summarized in the Proposal & Award Policies & Procedures Guide (PAPPG),[1] and has been developed to aid Principal Investigators (PIs) in Chemistry in developing effective, complete, and competitive DMPs. It is important to recognize that while all DMPs should address the five categories of information as specified in the PAPPG, they should not be generic. Each DMP should be appropriate for the data, metadata, samples, software, curricula, documentation, publications, and other materials generated in the course of the proposed research. DMPs should reflect best practices and standards for the proposed research and types of data being generated, whether experimental, computational, or text-based. DMPs are subject to peer review. Please contact your CHE Program Officer if you have any questions related to DMPs in the program context.

For more information on the history of the DMP requirement, and NSF's expectations for the dissemination and sharing of research results, see the appendix.

PAPPG and NSF-WIDE REQUIREMENTS

All proposals must include a supplementary document of no more than two pages labeled "Data Management Plan," as described in PAPPG Chapter II.C.2.j. Any specific instructions and exceptions to the two-page limit will be found in specific Program Solicitations.

- A proposal without a supplementary DMP will not be accepted.[2]
- You may request funds to cover costs of publication, page charges, or preparation of data as a direct cost in your budget proposal, which is evaluated as part of the merit review process. Any costs associated with implementing the DMP should be explained in the Budget Justification.
- The DMP will be reviewed as an integral part of the proposal, considered under Intellectual Merit or Broader Impacts or both, as appropriate for the scientific community of relevance.
- A valid DMP may state that no detailed plan is needed as long as statement is accompanied by a clear justification.[3]

---

[1]This document refers to the PAPPG effective January 29, 2018, document number nsf18001. Please check if there have been significant changes and clarifications to the PAPPG since the publication of this document.
[2]https://www.nsf.gov/bfa/dias/policy/autocompliance.jsp
[3]In the case of a workshop or REU proposal, the DMP could discuss the management of data that may be generated as part of the proposed activity (e.g. participant lists, exit surveys, community reports).

- If proposers feel that the DMP cannot fit within the 2-page limit, they may also use part of the 15-page Project Description for additional data management information.
- A DMP that lacks detail and simply states "see project description" is not considered sufficient.

DATA MANAGEMENT PLAN CONTENT

CHE-supported research covers a broad spectrum of communities of investigators and each community has its own best practices. CHE is aware of the need to provide flexibility to reviewers and Programs in assessing the quality of individual DMPs. The standards for DMPs are evolving to accommodate changing standards and expectations. CHE relies on the merit review process to determine which DMPs best serve each community.

The DMP should clearly articulate how the investigators plan to manage and disseminate data generated by the project, taking advantage of emerging information technologies and cyberinfrastructure. **The plan must include sufficient detail for evaluation of its appropriateness and feasibility during merit review.** DMPs often include existing practices in the principal investigator's laboratory and the larger research community. CHE strongly encourages innovations that, where appropriate and practical, enable efficient and effective data sharing and management to stimulate and promote scientific advances.

The five essential components of the DMP are listed in the same order as in the PAPPG, with examples relevant to the Chemistry community:

1. **Products of the Research.** Describe the types of data (including metadata and annotations, primary or analyzed) and products that will be generated by the research, for example description of samples, numerical data on chemical systems such as spectra, chemical and physical properties, time-dependent information on chemical and physical processes, theoretical formalisms, experimental protocols, algorithm specifications, database schemas and data tables, data produced by simulations, and software. Data and products generated from Broader Impact activities, such as educational materials, participant information, tutorials and other web-based materials, as well as assessment results, should also be included in the DMP.

2. **Data Format.** Describe the format and media 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. In general, solutions and remedies to providing data in an accessible format should be provided with minimal added cost.

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 or software repositories (e.g., NMRShiftDB, the Protein Data Bank, Cambridge Crystallographic Data Centre, Inorganic Crystal Structure Database in Karlsruhe, Zeolite Structure Database, and Github). For software or code developed as part of the project, include a description of how users can access the code (e.g., licensing, open source) and specific details of the hosting, distribution and dissemination plans. Also describe your practice or policies regarding the release of data for access, for example whether data are posted before or after formal publication. Note as well any anticipated inclusion of your data in databases that mine the published literature (e.g., PubChem, NIST Chemistry WebBook). Consider using the Digital Object Identifiers (DOI) assignment mechanism not just for journal articles, but for suitably-archived, publishable data sets.

"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. Discussion on the compliance with the NSF's Public Access Policy is also encouraged.

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. Practices for appropriate protection of privacy, confidentiality, security, intellectual property, and other rights should be communicated. The rights and obligations of those who access, use, and share your data with others should be defined. 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?

5. **Archiving of Data**. Describe when the data should be archived, how data will be archived, and how preservation of access will be handled. Are there provisions for data backup? 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? What are the physical and cyber resources and facilities that will be used for data preservation and storage? Will there be an easily accessible index that documents where all archived data are stored and how they can be accessed? What are the roles and responsibilities of all parties with respect to the management and archiving of the data after the grant ends? How long will the data be maintained after the grant ends?

CHE-supported large research centers or other programs may specify more stringent data storage, sharing and archiving procedures for research conducted under their awards. Such requirements will be specified in the program solicitation and award conditions.

POST-AWARD MANAGEMENT
If an award is made, the PI must manage their data as described in the DMP and should

report these data-related activities in annual and final project reports, and through subsequent proposals. The [NSF guidance on Technical Reporting Requirements](#) states that annual and final reports should describe actions taken during the reporting period to bring a proposal's Data Management Plan to completion. These reports are a critical mechanism for communication between the PI and the award's managing Program Officer.

The NSF report format includes specific sections on the accomplishments and products of the research, including how the results have been disseminated to communities of interest. The project reports should include specific information such as identifier or accession numbers for data sets, metadata and data annotation, citations of relevant publications, conference proceedings, details of software hosting, and other types of data sharing and dissemination, and updated information on project mechanisms for data storage, protection, and backup. CHE encourages investigators to use persistent identifiers (where these exist) as a long-lasting reference to a digital resource. Publications from new awards resulting from proposals submitted after January 25, 2016 must be deposited in the NSF Public Access Repository (NSF-PAR). For more information, see NSF's [Public Access Initiative](#) and [Frequently Asked Questions (FAQs) for Public Access](#).

Final project reports should describe the implementation of the DMP and include any changes from the original DMP. Simply putting data in Supplementary Materials of a publication is not sufficient data management. The availability of the data should be advertised through a publicly accessible website and there should be adequate annotation provided, including what the data is and parameters used to generate it, to allow for reproducibility.

SUBSEQUENT PROPOSALS

DMP implementation will also be considered during review of subsequent proposals. As described in the [PAPPG Part I Section II.C.2.d.iii,](#) the following information pertaining to past data management must be provided in the section 'Results from Prior NSF Support:'

(e) Evidence of research products and their availability, including, but not limited to: data, publications, samples, physical collections, software, and models, as described in any Data Management Plan

DATA MANAGEMENT RESOURCES

There are many resources available to PIs that can provide assistance and information when planning and implementing a DMP. **Please note that inclusion of a resource in the list below is not intended as an endorsement by the NSF or the Division of Chemistry.**

- Many university and college libraries provide resource guides or e-library consulting   services to assist PIs in data management planning and best practices. These university data management groups can serve as a source of information for DMP topics such as data archiving and backup and open source distribution. For example:
    - Boston University [Data Services](#)

- o UC San Diego Library – [Research Data Curation Program](#)
- If you are unsure where to deposit your data, online registries of research data repositories exist. See [re3data.org](#) for an extensive, though not exhaustive, list.
- Professional societies will often also provide guidance for the community. The American Chemical Society has a position statement on [Ensuring Access to High-Quality Science](#).
- Numerous non-governmental organizations offer resources and training in developing DMPs. These can be quite helpful, even if the target scientific discipline is not Chemistry.   For example:
  - o [DataOne](#)
  - o [Software Carpentry](#)
  - o [DMPTool](#)
  - o [LTER Intranet](#)

APPENDIX - BACKGROUND

Beginning in January 2011, NSF implemented a data management plan requirement for all proposals, which is described in the [Proposal & Award Policies & Procedures Guide (PAPPG) Part I Section II.C.2.j](#). This requirement was created to aid in the dissemination, accessibility, and   preservation of data generated by NSF-funded research. The goal of a DMP should be to provide clear, effective, and transparent implementation of the NSF policy on [Dissemination and Sharing of Research Results](#) as described in the [PAPPG Part II Chapter XI.D.4](#) and below.

Dissemination and Sharing of Research Results

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 performing the 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 an 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 a way 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.

f. Each NSF grant contains, as part of the grant terms and conditions, an article implementing dissemination and sharing of research result.