Data Management for NSF EHR Directorate

Proposals and Awards

The National Science Foundation has released a new requirement for proposal submissions regarding the management of data generated using NSF support. Dissemination of data is necessary for the community to stimulate new advances as quickly as possible and to allow prompt evaluation of the results. All proposals submitted for a deadline or target date of January 15, 2011 or later must include a data management plan (DMP). As summarized in the NSF Proposal and Award Policies and Procedures Guide:

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 (Section VI. D. 4. b). See the NSF Grant Proposal Guide (GPG) Chapter II.C.2.j for a description of the full policy implementation. The full policy recognizes conditions under which restricting release of privileged or proprietary information would be appropriate, encourages sharing of software and inventions, and recognizes intellectual property rights.

The Federal government defines "data" in OMB Circular A-110 as:

- "... the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This 'recorded' material excludes physical objects (e.g., laboratory samples). Research data also do not include:
 - (A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and
 - (B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study."

An appropriate data management plan is required as a supplementary document (maximum of two pages) for all proposals submitted. This plan is to be included in the Supplementary Documents section of the proposal and is not part of the 15-page limit for the Project Description. The NSF will not evaluate any proposal that lacks a DMP for competitions that have a due date or target date on or after January 18, 2011. Even if no data are produced, e.g., the research is purely theoretical or is in support of a workshop, a DMP is required. In such cases, the DMP can simply state, with justification, that no data will be produced. Any costs for data management should be explained in the Budget Justification. The DMP should reflect best practices in the PI's area of research and should be appropriate to the data generated. For proposals submitted to EHR, the plan should address two main questions:

What data are generated by your project?

What is your plan for managing the data?

According to the NSF Proposal and Award Policies and Procedures Guide (see <u>AAG Chapter VI.D.4</u>), the Data Management Plan may include:

- 1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- 2. 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);
- policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- 4. policies and provisions for re-use, re-distribution, and the production of derivatives; and
- 5. plans for archiving data, samples, and other research products, and for preservation of access to them.

Data on EHR projects involving human subjects should be made available to the public subject to constraints imposed by IRB decisions. Other data, such as software, publications, and curricula, should be made available subject to intellectual property rights. Any data collection required by the program announcement should be incorporated into the proposal's DMP. For example, the management of assessment, evaluation, or monitoring data required for all projects within a given program should be addressed in the data management plan. Some universities and professional organizations have explicit data management policies that PIs might wish to consult for guidance in developing a DMP. For example, the Inter-University Consortium for Political and Social Research has guidelines and useful examples (http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/index.jsp).

PIs should use the opportunity of the DMP to give thought to matters such as:

- The types of data that their project might generate and eventually share with others, and under what conditions
- How data are to be managed and maintained until they are shared with others
- Factors that might impinge on their ability to manage data, e.g. legal and ethical restrictions on access to non-aggregated data
- The lowest level of aggregated data that PIs might share with others in the scientific community, given that community's norms on data
- The mechanism for sharing data and/or making them accessible to others
- Other types of information that should be maintained and shared regarding data, e.g. the way it was generated, analytical and procedural information, and the metadata

Period of data retention.

EHR is committed to timely and rapid data distribution. However, it recognizes that types of data can vary widely and that acceptable norms also vary by scientific discipline. It is strongly

committed, however, to the underlying principle of timely access, and applicants should address how this will be met in their DMP statement.

Data formats and dissemination.

The DMP should describe data formats, media, and dissemination approaches that will be used to make data and metadata available to others. Policies for public access and sharing should be described, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. Research centers and major partnerships with industry or other user communities must also address how data are to be shared and managed with partners, center members, and other major stakeholders.

Data storage and preservation of access.

The DMP should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data. These can include third party facilities and repositories.

Additional possible data management requirements.

More stringent data management requirements may be specified in particular NSF solicitations or result from local policies and best practices at the PI's home institution. Additional requirements will be specified in the program solicitation and award conditions. Principal Investigators to be supported by such programs must discuss how they will meet these additional requirements in their Data Management Plans.

Post-Award Monitoring

After an award is made, data management will be monitored primarily through the normal Annual and Final Report process and through evaluation of subsequent proposals.

Annual Reports.

Annual reports, required for all multi-year NSF awards, must provide information on the progress on data management and sharing of the research products. This information could include citations of relevant publications, conference proceedings, and descriptions of other types of data sharing and dissemination of results.

Final Project Reports.

Final Project Reports are required for all NSF awards. The Final Project Report must discuss execution and any updating of the original DMP. This discussion should describe:

- Data produced during the award;
- Data to be retained after the award expires;
- Verification that data will be available for sharing;
- Discussion of community standards for data format;

- How data will be disseminated;
- The format that will be used to make data available to others, including any metadata; and
- The archival location of data.

Subsequent proposals.

Data management must be reported in subsequent proposals by the PI and Co-PIs under "Results of prior NSF support."

Examples

Some examples specific to EHR follow. These are only examples and every proposal will have its own data management plan that may differ substantially from these examples.

- 1. A proposal for a workshop that will result in a workshop report.
 - a. The DMP could consist of a statement to the effect that a workshop report will be produced and disseminated, e.g., via a website, publication in a journal, or other means.
- 2. A proposal for developing a new undergraduate course in chemistry during which software will be developed to help students visualize chemical reactions.
 - a. The DMP could address how the software will be released (open source, through a commercial license, within the cloud) and how the curriculum materials will be released (in a textbook, on the web, etc.). If data on student learning is also collected, the DMP should include information on how the data will be released (anonymized data, refereed article) and on the constraints for sharing the data to protect student privacy.
- 3. A proposal in an education program that requires all projects to report on the graduation rate of participants.
 - a. The DMP, in addition to addressing other data produced by the project, should state that the graduation rate data will be collected and reported to the appropriate collection site as required by the program announcement.
- 4. A proposal in a program that requires that projects collect data that will be used for project-level evaluation (e.g., survey data).
 - a. The DMP, in addition to addressing other data in the project, should identify the project-level evaluation data to be collected.
- 5. An education research proposal that employs a quasi-experimental approach to testing intervention hypotheses.
 - a. The DMP should identify the data collected, how it will be analyzed, and how the analyzed data will be disseminated. The methodologies used in the quasi-experiment should be part of the research plan and not part of the DMP.
- 6. A proposal in which video data on teaching or instruction will be collected.
 - a. The data management plan should address the form in which the data would be released, e.g., in journal publications or as data files released to interested researchers, and conditions for release of annotated videos for research purposes. It may be that the videos cannot be released because of IRB privacy restrictions.

References

NSF publications, including the Proposal and Award Policies and Procedures Guide and the Grant Proposals Guide can be found at

http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_index.jsp

Office of Management and Budget, Circular A-110, September 30, 1999. White House Website, OMB Home. http://www.whitehouse.gov/omb/circulars/a110/a110.