



EScience in Action

Examination of Federal Data Management Plan Guidelines

Jennifer L. Thoegersen

University of Nebraska-Lincoln, Lincoln, NE, USA

Abstract

Data management plans as expectations of the grant proposal process are still fairly novel, and the expected format and content of these plans is still evolving. The objective of this research is to gain a greater understanding of the expected content for data management plans submitted as part of grant proposals to federal funding agencies. This paper examines federal funding agencies' data management plan guidelines in relation to the broad elements of data management identified by the Interagency Working Group on Digital Data: Description, Impact, Content & Format, Protection, Preservation, Access, and Transfer of Responsibility. Specifically, statements in agencies' guidelines were categorized into the most applicable category (or categories). The representation of each category within each agency's guidelines was addressed, and the statements falling in each category were analyzed. Some categories, including Access and Preservation, were represented in all or nearly all of the guidelines examined. Other categories — Impact and Transfer of Responsibility — were rarely addressed. The expectations for data management plans are evolving and will likely continue to evolve as more agencies require them.

Correspondence: Jennifer L. Thoegersen: jthoegersen2@unl.edu

Keywords: federally funded research, data management plans, research data



All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Introduction

In 2012, the Association of European Research Libraries (LIBER) working group on E-Science / Research Data Management provided a list of recommendations for how libraries could become involved in data management at academic institutions (Christensen-Dalsgaard et al. 2012). The first recommendation suggests offering “research data management support, including data management plans for grant applications” (3). In order to offer this support, libraries must cultivate adequate knowledge in relevant areas, therefore, many resources are being developed to assist information professionals in building their understanding of data management. From the seven-module New England Collaborative Data Management Curriculum (NECDMC) to Purdue’s Data Curation Profiles, these resources provide a wealth of information, including an overview of data management topics, suggestions on conducting consultations, research project case studies, and sample data management plans (DMPs). While guidance is provided on the general content of DMPs, emphasis is also placed on the uniqueness of each plan, as the content and structure is dependent on the research project, the particular solicitation, and the agency awarding the grant. For a librarian consulting on DMPs, the project and solicitation may be instance-specific, but certain funding agencies will likely be encountered frequently. Familiarity with DMP guidelines for federal funding agencies, including a rich understanding of the data management elements emphasized by each agency, can assist information professionals when providing DMP support.

Analysis of DMP requirements has been conducted by the Digital Curation Centre (DCC) for U.K. research funders, and Dietrich et al. used a modified version of the DCC’s rubric to analyze the DMP requirements of 10 federal funders (Digital Curation Centre 2015; Dietrich et al. 2012).

In contrast, this paper categorizes the content of DMP guidelines into the aspects of data management identified by the Interagency Working Group on Digital Data (IWGDD), which consisted of representatives from over two-dozen federal agencies. By examining DMP guidelines based on the IWGDD’s categories, this paper highlights the changing expectations of DMPs at the federal level.

Background

The DMP requirement for federal grant proposals has been gaining momentum for over a decade. The National Institutes of Health (NIH) began requiring a Data Sharing Plan in 2003 for applicants seeking \$500,000 or more in annual research funding (National Institute of Health 2003). The motivation behind this decision was “to expedite the translation of research results into knowledge, products, and procedures to improve human health” (National Institutes of Health 2003).

In 2009, the IWGDD was tasked to develop a plan for a framework for federal agencies to promote data preservation and sharing and published a report recommending agencies encourage data management planning for projects (Interagency Working Group of Digital Data 2009). The recommendation includes seven suggested aspects of data management that may be addressed in DMPs (the IWGDD’s descriptions for each aspect are presented in Table 1):

- Description
- Impact
- Content and format
- Access
- Protection
- Preservation

Table 1: Data Management Plan Elements recommended by the Interagency Working Group on Digital Data (Interagency Working Group of Digital Data 2009)

Element	Description
Description	Brief, high-level description of the digital scientific data to be produced.
Impact	Discussion of possible impact of the data within the immediate field, in other fields, and any broader, societal impact. Indicate how the data management plan will maximize the value of the data.
Content and Format	Statement of Plans for data and metadata content and format, including description of documentation plans and rationale for selection of appropriate standard. Existing, accepted standards should be used where possible.
Protection	Statement of plans, where appropriate and necessary, for protection of privacy, confidentiality, security, intellectual property, and other rights.
Access	Description of plans for providing access to data. This should include a description and rationale for any restrictions on who may access the data under what conditions and a timeline for providing access. This should also include a description of the resources and capabilities (equipment, connections, systems, expertise, etc.) needed to meet anticipated requests, including those needed for access locally, nationally and internationally.
Preservation	Description of plans for preserving data in accessible form. Plans should include a timeline proposing how long the data are to be preserved, outlining any changes in access anticipated during the preservation timeline, and documenting the resources and capabilities (e.g., equipment, connections, systems, expertise) needed to meet the preservation goals. Where data will be preserved beyond the duration of direct project funding, a description of other funding sources or institutional commitments necessary to achieve the long-term preservation and access goals should be provided.
Transfer of Responsibility	Description of plans for changes in preservation and access responsibility. Where responsibility for continuing documentation, annotation, curation, access, and preservation (or its counterparts, de-accessioning or disposal) will move from one entity or institution to another during the anticipated data life cycle, plans for managing the exchange and documentation of the necessary commitments and agreements should be provided.

In 2011, the National Science Foundation (NSF) implemented a DMP requirement for all grant proposals, and in 2013 the White House Office of Science and Technology Policy (OSTP) mandated that all agencies receiving \$100 million or more in research and development funds submit policies for data sharing in order to increase the availability of the products of federally funded research (Holdren2013). Over the past four years, the number of federal agencies with DMP guidelines for grant proposals has been increasing, most recently with the Department of Energy requiring DMPs as of October 2014 (DOE Office of Science 2014).

Methodology

As an initial step, funders with DMP guidelines were identified. While there are federal websites that identify funding agencies (e.g. CFDA¹, Grants.gov²), they do not delineate which ones require DMPs. The DMPTool³, a service of the University of California Curation Center, maintains a list of agencies with DMP requirements. This list was used to identify the funding agencies (and, for NSF, divisions and directorates) with specific guidelines. The funders whose guidelines were examined — along with their abbreviations — are outlined in Table 2.

Each funder's website was visited to obtain their DMP guidelines. One complication encountered was the plethora of DMP advice, suggestions, and guidance dispersed in multiple locations throughout several of the funders' sites. Due to the difficulty of determining which information should be included or excluded, and to maintain a consistent approach across agencies, only the main DMP guidelines were examined. These guidelines generally were presented in a list or description of elements that should be addressed in a DMP.

Once the guidelines were collected, they were broken into statements based on topic. These statements were then categorized into broad aspects of data management. Preliminary categories were derived from the DMP elements recommended by the Interagency Working Group on Digital Data in 2009, which are presented in Table 1. Statements that did not fall into any of these categories were placed in an "other" category. Once all guidelines had been examined, statements in the "other" category were grouped together iteratively. Statements that addressed multiple categories were placed in each applicable category.

¹ https://www.cfda.gov/?_so_list_aatda5df66d9f809a53b1563b95a6dce8e3=250

² <http://www.grants.gov/web/grants/applicants/applicant-resources/agencies-providing-grants.html>

³ <https://dmptool.org>

Table 2: List of federal agencies included in study

Code	Agency
APSF	Alfred P. Sloan Foundation
DOE	Department of Energy
GBM	Gordon and Betty Moore Foundation
GoMRI	Gulf of Mexico Research Initiative
IES	U.S. Department of Education, Institute of Education Sciences
IMLS	Institute of Museum and Library Services
JFSP	Joint Fire Science Program
NIH	National Institutes of Health
NEH-ODH	National Endowment for the Humanities, Office of Digital Humanities
NSF-AGS	National Science Foundation, Atmospheric and Geospace Sciences
NSF-AST	National Science Foundation, Astronomical Sciences
NSF-BIO	National Science Foundation, Biological Sciences
NSF-CHE	National Science Foundation, Chemistry Division
NSF-CISE	National Science Foundation, Computer & Information Science and Engineering
NSF-DMR	National Science Foundation, Division of Materials Research
NSF-EAR	National Science Foundation, Division of Earth Sciences
NSF-EHR	National Science Foundation, Directorate for Education & Human Resources
NSF-ENG	National Science Foundation, Directorate for Engineering
NSF-GEN	National Science Foundation
NSF-PHY	National Science Foundation, Division of Physics
NSF-SBE	National Science Foundation, Directorate for Social, Behavioral, and Economic Sciences
USGS	U.S. Geological Survey

Results

Table 3 presents the 22 DMP guidelines examined and indicates which guidelines included at least one pertinent statement for a given category. The results of the analysis are explored by content category.

Description

The majority (73%) of the guidelines requested a description of the data that would be produced. Most of these guidelines emphasized what data would be produced, while two (IMLS and NSF-AGS) highlighted the quantity of data. Many of guidelines listed potential data types; for instance, NSF-AST and NSF-CHE tailored potential data types to the astronomy and chemistry disciplines, respectively. However, most references to data types were a variation of the NSF-GEN statement requesting “the types of data, samples physical collections, software, curriculum materials, and other materials to be produced in the course of the project” (National Science Foundation 2014, II-20).

Impact

The impact element, which focuses on the prospective value of the data, has been largely disregarded by federal agencies’ DMP guidelines thus far, only being addressed by two guidelines. The IMLS and NIH guidelines mention potential impact, but only in specific instances. The NIH suggest applicants address significance specifically “if support is being sought to develop a large database that will serve as an important resource for the scientific community” (National Institutes of Health 2003). IMLS requires the applicant address novelty only if the project involves developing digital tools.

Access

All 22 guidelines include at least one statement related to access. This is only category represented in every guideline examined, and was by far the most extensively addressed. Access topics covered by guidelines included timeliness of sharing, what data will be shared, how data will be shared, the format of shared data, and the policies/licenses that will be implemented for reuse. Two agencies — APSF, GBM — request information on how sharing might be affected by the use of pre-existing data.

Content and format

Eighteen (82%) of the guidelines address aspects of the content and format category, including an indication of file formats, metadata standards, and intended documentation practices. Some agencies provide specific exemplar formats (e.g. IMLS, NSF-AGS, NSF-AST) and metadata standards (e.g. IMLS, JFSP). The JFSP guidelines explicitly mandate particular standards for spatial data sets.

Protection

The IWGDD’s protection element is concerned with any rights or security issues that need to be addressed during and after the research project. This includes how researchers will resolve

Table 3: Categories represented in agency data management plan guidelines

	Description	Impact	Content and Format	Protection	Access	Preservation	Transfer of Responsibility	Other
APSF	X		X	X	X	X		X
DOE				X	X	X		X
GBM	X		X	X	X	X		X
GoMRI	X		X		X	X		X
IES			X	X	X			X
IMLS	X	X	X	X	X	X		X
JFSP	X		X	X	X	X		X
NIH		X	X	X	X			X
NEH	X		X	X	X	X		X
NSF-AGS	X		X	X	X	X		X
NSF-AST	X		X	X	X	X		
NSF-BIO	X		X	X	X	X	X	X
NSF-CHE	X		X	X	X	X		
NSF-CISE	X		X	X	X	X	X	X
NSF-DMR					X	X		X
NSF-EAR					X	X		X
NSF-EHR	X		X	X	X	X		X
NSF-ENG	X			X	X	X	X	X
NSF-GEN	X		X	X	X	X		
NSF-PHY					X	X		X
NSF-SBE	X		X	X	X	X	X	X
USGS	X		X		X	X		X

the conflict between the requirement to share data with the necessity to protect privacy and comply with applicable regulations. Most of the guidelines (77%) included language related to protection, especially in connection with sharing data after project completion. Guidelines for five of the agencies (GoMRI, NSF-DMR, NSF-EAR, NSF-PHY, USGS) did not specifically address protection, possibly because the research funded by these agencies is less likely to have privacy or security concerns, especially when compared to research funded by, for example, the NIH or DOE.

Preservation

Preservation, which relates to preserving data following the completion of the research project, was addressed by all agencies except the IES and NIH. Preservation issues included what types of data will be preserved, how long data will be retained, who will maintain preserved data, and what resources, facilities and infrastructure will be utilized for preservation. Three agencies highlight validation of research as the main purpose for preservation. The DOE guidelines ask how researchers “will enable validation of results,” and the NSF-DMR and the NSF-PHY guidelines ask how the researchers’ plans for preservation will ensure their ability to “respond to a question about a published result” (DOE Office of Science 2014; NSF Division of Physics 2010, 2).

Transfer of Responsibility

Transfer of responsibility is addressed only by four NSF Directorates — BIO, CISE, ENG, and SBE — all of which use similar wording. However, while the IWGDD category focuses on anticipated transfer of responsibility through the normal data life cycle, these directorates’ guidelines concentrate on the contingencies in place should a project lead “leave the institution or project” (NSF Directorate for Computer & Information Sciences & Engineering 2015).

Other

There were many statements that did not fall into any of the IWGDD’s initial categories. As much as possible, these statements were grouped iteratively into additional categories and are presented in Table 4. These categories covered topics including: explaining how data will be managed in a collaborative environment, addressing costs associated with data management and preservation, outlining how data will be managed throughout the course of the research project, stating whether pre-existing data will be used and how it will be handled, explaining what quality control will be used, and identifying the roles and responsibilities for managing data.

Discussion

Of the seven categories identified by the IWGDD, five appear to be solidifying as necessary components for a DMP: Description, Content & Format, Protection, Access, and Preservation. The fact that Access alone is addressed by every guideline is unsurprising, as public availability of federally funded research has been a major driving force behind DMP requirements.

However, there is clearly a gap between the IWGDD’s designated content categories and the

expectations of the DMP guidelines. The remaining two categories, Impact and Transfer of Responsibility, have largely been ignored. This omission may suggest that these categories are considered less relevant to the underlying purpose of the DMP requirement. In addition, most of the guidelines analyzed included statements that did not fall into any of the IWGDD categories, demonstrating that DMPs are taking on a broader scope of content than originally anticipated by federal agencies.

Table 4: List of iterative groups generated from uncategorized statements

Group		Exemplar statements
Collaboration	NEH, NSF-DMR, NSF-EAR, NSF-EHR, NSF-ENG, NSF-PHY, NSF-SBE	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. (NSF-SBE)
Cost	NIH, NSF-CISE, NSF-ENG, NSF-SBE, USGS	Applicants may request funds in their application for data sharing. If funds are being sought, the applicant should address the financial issues in the budget and budget justification sections. (NIH)
Management	APSF, NEH, NSF-CISE, NSF-EHR, NSF-SBE	What tools, platforms, and processes will be used to manage project assets as they are created and used through the grant period? (APSF)
Pre-existing data	APSF, GBM, JFSP, IMLS	Does this project involve organization or analysis of pre-existing data, and what are the data sharing arrangements for these data? (GBM)
Quality Control	GBM, IMLS, JFSP	Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products). (IMLS)
Roles	APSF, GBM, GoMRI, IES, NSF-BIO, NSF-CISE, SF-ENG, NSF-SBE	Describe the roles and responsibilities of all parties with respect to the management of the data. (NSF-BIO)

Conclusion

The requirement for DMPs to be submitted with federal grant proposals is still a fairly novel concept, and the content expectations for these plans are still evolving. Current DMP guidelines provided by federal funding agencies largely reflect many of the categories suggested by the IWGDD. However, the categories and guidelines' content diverge in several ways, and there are significant variations in the content expectations of the various federal funding agencies. The only category that was addressed by the guidelines of all of the agencies was Access — likely reflecting the underlying emphasis at the federal level for publicly available data — followed by Preservation (addressed by all but NIH). In addition, even when guidelines included requirements for the same general content category, the specific statements varied greatly between guidelines. A more in-depth look at the statement level will be necessary to gain a fuller understanding of expectations for DMPs for federal grant proposals.

Disclosure

The author reports no conflicts of interest.

References

- Christensen-Dalsgaard, Birte, Marc van den Berg, Rob Grim, Wolfram Horstmann, Dafne Jansen, Tom Pollard, and Annikki Roos. 2012. "Ten Recommendations for Libraries to Get Started with Research Data Management." *Association of European Research Libraries*. Last modified April 7. http://libereurope.eu/wp-content/uploads/The_research_data_group_2012_v7_final.pdf
- Dietrich, Dianne, Trisha Adamus, Alison Miner, and Gail Steinhart. 2012. "De-Mystifying the Data Management Requirements of Research Funders." *Issues in Science and Technology Librarianship* 70. <http://dx.doi.org/10.5062/F44M92G2>
- Digital Curation Centre. 2015. "Overview of Funders' Data Policies." Accessed June 22. <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>
- Holdren, John P. 2013. "Increasing Access to the Results of Federally Funded Scientific Research." *White House Office of Science and Technology Policy*. Last modified February 22. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- Interagency Working Group of Digital Data. 2009. "Harnessing the Power of Digital Data for Science and Society." Last modified January 14. https://www.nitrd.gov/About/Harnessing_Power_Web.pdf
- National Institutes of Health. 2003. "NIH Data Sharing Policy and Implementation Guidance." Last modified March 5. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
- National Science Foundation. 2014. "Grant Proposal Guide, December 2014." Last modified November 20. http://www.nsf.gov/publications/pub_summ.jsp?ods_key=pgp15001&org=NSF
- NSF Directorate for Computer & Information Sciences & Engineering. 2015. "Data Management Guidance for CISE Proposals and Awards." Last modified March 15. http://www.nsf.gov/cise/cise_dmp.jsp
- NSF Division of Physics. 2010. "Advice to PIs on Data Management Plans." Last modified November 30. <http://www.nsf.gov/bfa/dias/policy/dmpdocs/phy.pdf>
- U.S. Department of Energy: Office of Science. 2014. "Statement on Digital Data Management." Last modified July 28. <http://science.energy.gov/funding-opportunities/digital-data-management>