

2018-12-21

# Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices

Sara Mannheimer

*Montana State University-Bozeman*

**Corresponding Author(s)**

Sara Mannheimer, Montana State University, P.O. Box 173320, Bozeman, MT 59715;  
sara.mannheimer@montana.edu

Follow this and additional works at: <https://escholarship.umassmed.edu/jeslib>

 Part of the [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

## Recommended Citation

Mannheimer, Sara. 2018. "Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices." *Journal of eScience Librarianship* 7(3): e1155. <https://doi.org/10.7191/jeslib.2018.1155>

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Journal of eScience Librarianship by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

---

# Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices

## Keywords

data management plans, research data, data librarianship

## Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

## Rights and Permissions

Copyright Mannheimer © 2018

## Acknowledgments

This research was funded by the National Library of Medicine, National Institutes of Health under cooperative agreement number UG4LM012343 with University of Washington. Many thanks are also in order for the work of my outstanding student research assistant, Wangmo Tenzing. The Montana State University HELPS Lab scheduled and transcribed the PI interviews. Figure 3 was created by Laurie Rutemeyer at Montana State University Statistical Consulting and Research Services, which is supported by Institutional Development Awards (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under Awards P20GM103474, 5U54GM104944, U54GM115371, and 5P20GM104417.



## Full-Length Paper

### **Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices**

Sara Mannheimer

Montana State University, Bozeman, MT, USA

---

#### **Abstract**

Data Management Plans (DMPs) are often required for grant applications. But do strong DMPs lead to better data management and sharing practices? Several recent research projects in the Library and Information Science field have investigated data management planning and practice through DMP content analysis and data-management-related interviews. However, research hasn't yet shown how DMPs ultimately affect data management and data sharing practices during grant-funded research. The research described in this article contributes to the existing literature by examining the impact of DMPs on grant awards and on Principal Investigators' (PIs) data management and sharing practices. The results of this research suggest the following key takeaways: (1) Most PIs practice internal data management in order to prevent data loss, to facilitate sharing within the research team, and to seamlessly continue their research during personnel turnover; (2) PIs still have room to grow in understanding specialized concepts such as metadata and policies for use and reuse; (3) PIs may need guidance on practices that facilitate FAIR data, such as using metadata standards, assigning licenses to their data, and publishing in data repositories. Ultimately, the results of this research can inform academic library services and support stronger, more actionable DMPs.

---

**Correspondence:** Sara Mannheimer: [sara.mannheimer@montana.edu](mailto:sara.mannheimer@montana.edu)

**Keywords:** data management plans, research data, data librarianship

**Rights and Permissions:** Copyright Mannheimer © 2018

**Disclosures:** The substance of this article is based upon a lightning talk presentation at RDAP Summit 2018. Additional information at end of article.



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 and Background

In our data-centric era, research data have become valuable resources for encouraging research reproducibility, supporting information equity, accelerating science, and amplifying the impact of research (Vision 2010). Acknowledging the value of research data, federal funding agencies increasingly require that data management plans (DMPs) be included in grant proposals to support good data stewardship practices and to promote data sharing and reuse (Holdren 2013). In response to these trends, academic libraries are seeing increased demand to provide research data services (Tenopir et al. 2017; Bryant, Lavoie, and Malpas 2017; Tenopir, Sandusky, Allard, and Birch 2014). Common library research data services include education and training in data management best practices; consultations on writing DMPs for grants; and publication services for open datasets (Cox, Kennan, Lyon, and Pinfield 2017; Yoon and Schultz 2017). The ultimate goal of DMPs and library research data services is to support good data stewardship practices. Therefore, the DMP requirement from funding agencies is a commitment to the idea that strong data management planning can help ensure that published research data adheres to community standards like the FAIR Data Principles (Wilkinson et al., 2016). The FAIR principles support the idea that data must be carefully structured and described in order to be Findable, Accessible, Interoperable, and Reusable.

DMPs have been researched from various angles. A review of DMPs from the University of Illinois revealed that the proposed storage venue for datasets did not have a statistically significant effect on successful funding of National Science Foundation (NSF) proposals (Mischo, Schlembach, and O'Donnell 2014). A study at Georgia Tech found that faculty widely share and reuse language from previous DMPs when producing new DMPs (Parham and Doty 2012). The Data Management Plans as a Research Tool (DART) project has produced a rubric for analyzing DMPs (Whitmire, Carlson, Westra, Hswe, and Parham 2016), as well as an analysis of data management practices across domains (Parham, Carlson, Hswe, Westra, and Whitmire 2016). Researchers at the University of Houston conducted interviews with NIH and NSF grant recipients to illuminate data management needs on their campus (Peters and Dryden 2011). Most recently, Berman (2017) analyzed 35 DMPs and conducted six interviews in order to understand campus needs and inform development of future library research data services.

At Montana State University Library, research data services have been offered since 2014. These services include assistance with writing DMPs for grants. The purpose of this research is to build an evidence-based understanding of the impact of DMPs on grant-funded projects. This research also aims to inform library research data services, helping Principal Investigators (PIs) efficiently create useful, actionable DMPs that support FAIR research data.

## Research Questions and Methods

A 2013 memo from the White House Office of Science and Technology Policy (OSTP) (Holdren, 2013) suggests that publicly accessible research data can accelerate discovery and more broadly share the benefits of scientific research. The memo directs “each Federal agency with over \$100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government” (Holdren, 2013, p. 2). DMPs are one method for supporting public access to research data, through asking researchers to plan ahead for data management and data

sharing. NSF began requiring DMPs prior to the 2013 OSTP memo in January 2011. In theory, DMPs should help PIs build strong policies and procedures that support FAIR data. This research aims to understand how DMPs work in practice. Specifically, this research aims to determine the utility of DMPs by asking two key questions:

*Q1a. Are grant proposals with more complete/detailed DMPs more likely to be funded?*

*Q1b. Which sections within DMPs have the most and least complete/detailed information?*

*Q2. Does writing a DMP affect PIs' data management and data sharing practice for grant-funded projects?*

The research team consisted of the author and one student research assistant. To answer Q1a and Q1b, the author trained the student research assistant to analyze the content of 186 DMP documents from awarded and declined NSF grant proposals at Montana State University. To answer Q2, semi-structured interviews were conducted with 17 PIs who were selected from among the PIs whose DMPs came from awarded grants. These interviews were also used to provide additional insight to the results of the DMP content analysis when answering Q1a and Q1b.

### **DMP Content Analysis**

*Q1a. Are grant proposals with more complete/detailed DMPs more likely to be funded?*

*Q1b. Which sections within DMPs have the most and least complete/detailed information?*

This research was designated as exempt by the Montana State University Institutional Review Board. DMPs were collected through Montana State University's online approval system for all grant proposals. The online approval system only requires that a proposal abstract be uploaded to the system, but some PIs choose to upload their grant proposal in full. Therefore, not all entries in the online approval system included a DMP. Since NSF's DMP requirement went into effect on January 18, 2011 (NSF 2018), results were filtered to contain only NSF proposals from 2011 or later. Each proposal was then examined by hand, and full proposals were downloaded if they were present. Since 2011, 88 awarded full proposals and 110 declined full proposals have been uploaded to the online approval system. Therefore, the dataset consisted of 198 full proposals. The DMP was pulled from each full proposal, and the remainder of the proposal was discarded. Of the 198 DMPs, 11 reported that no data would be produced, which left a total of 187 DMPs that were suitable for analysis. All the PIs whose DMPs were included in the dataset were sent an email offering an opportunity to opt out of the research. One PI opted out of the research, producing a final dataset of 186 DMPs.

A Qualtrics survey available from the Data Management Plans as a Research Tool (DART) project was used to conduct content analysis the 186 DMP documents. The survey is based on the DART Rubric (Whitmire, Carlson, Westra, Hswe, and Parham 2016), which is organized according to DMP components outlined by NSF:

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);
3. 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;
5. plans for archiving data, samples, and other research products, and for preservation of access to them” (NSF 2017).

The DART rubric shortens each NSF component as follows:

1. types of data produced;
2. standards for data and metadata;
3. policies for access and sharing;
4. policies and provisions for re-use and redistribution; and
5. plans for data archiving and preservation of access.

For each section, the rubric suggests standards to evaluate whether each DMP:

1. provides complete/detailed information;
2. addresses the issue, but information was incomplete; or
3. does not address the issue.

For example, in the section “standards for data and metadata,” the rubric asks whether the DMP “identifies metadata standards and/or metadata formats that will be used for the proposed project.” To qualify as complete/detailed, the DMP must clearly state and describe a metadata standard that will be followed; if no disciplinary standard exists, the DMP should clearly describe a project-specific approach. See Figure 1 for more detail. See DART Rubric Guidance for complete information (Whitmire, Carlson, Westra, Hswe, and Parham 2017).

### **Semi-structured Interviews with Principal Investigators (PIs)**

*Q2. Does writing a DMP affect PIs’ data management and data sharing practice for grant-funded projects?*

Interview request emails were sent to all PIs with DMPs in the dataset that came from awarded grants. 17 PIs responded positively to the request to be interviewed. The semi-structured interviews examined whether PIs used DMPs during their grant-funded research process, whether they shared data from their grant project, and whether they considered writing the DMP to be a helpful exercise. The full interview instrument is available as Appendix A.

**Section 2: Standards for data and metadata**  
*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)*

Performance Criteria		Performance Level			Directorates w/this req.
		Complete/detailed	Addressed issue, but incomplete	Did not address	
General assessment criteria	2.1 Identifies <b>metadata standards and/or metadata formats</b> that will be used for the proposed project	The metadata standard that will be followed is clearly stated and described. If no disciplinary standard exists, a project-specific approach is clearly described. <i>"Data will be described using Darwin Core Archive metadata, and accompanied by readme.txt files providing information on field methods and procedures."</i>	The metadata standard that will be followed is vaguely stated. If no disciplinary standard exists, a project-specific approach is vaguely described. <i>"We will observe common data format and content protocols. For behavioral and the neuroimaging data, the participant log/response files include metadata embedded within the initial data collection. In addition, the behavioral and neuroimaging protocol/stimuli delivery files discussed above document the data and metadata format and content."</i>	The metadata standard that will be followed is not stated and no project-specific approach is described.	All
	2.2 Describes <b>data formats created or used during project</b>	Clearly describes file format standard(s) for the data. <i>"Data will be collected in plain text, Excel, SPSS, R, Matlab, Access DB, ESRI Shapefile, TIFF, JPEG, WAV, MP4, XML, HTML, or other software-specific file formats."</i>	Describes some but not all file formats, or file format standards for the data. Where standards do not exist, does not propose how this will be addressed. <i>"We produce data in the form of photon-counting events in a given time interval. The data is transferred to a laboratory server and stored. We also collect digital images of optical spectra from a spectrometer and store them on the server. We also collect digital samples of voltages monitoring various optical signals and store them on the server. The data will be available as raw data files."</i>	Does not include information about data format standards.	All
Directorate- or division-specific assessment criteria	2.3 Identifies <b>data formats that will be used for storing data</b>	Clearly describes data formats that will be used for storing data and explains rationale or complicating factors. <i>"NMR data will be saved in proprietary format to preserve embedded information, and converted to JCAMP files for ease of access and in case the proprietary systems fail or become unavailable."</i>	Only partially describes data formats that will be used for storing data and/or the rationale or complicating factors. <i>"Complex data analysis will be done using either Origin 8.0 or Matlab software and will be saved in a format that is easily transferable."</i>	Does not describe data formats that will be used for storing data and does not explain rationale or complicating factors.	GEO AGS, MPS AST, MPS CHE
	2.4 If the proposed project includes the use of <b>unusual data formats</b> , the plan discusses the proposed solution for converting data into <b>more accessible formats</b>	Explains how the data will be converted to a more accessible format or otherwise made available to interested parties. In general, solutions and remedies should be provided. <i>"Still Images from microscopes will be converted from proprietary formats to Open Microscopy Exchange format (OME-Tiff) for preservation and sharing."</i>	Vaguely explain[s] how the data may be converted to a more accessible format or otherwise made available to interested parties.	Does not explain how the data will be converted to a more accessible format or otherwise made available to interested parties.	MST AST, MPS CHE

Figure 1: DART Rubric example

The author conducted and recorded the semi-structured interviews, which were transcribed in full by the Montana State University Human Ecology Learning and Problem Solving (HELPS) Lab<sup>1</sup>. The author and the student research assistant conducted a conventional qualitative content analysis of the interview transcripts using an inductive coding approach, as outlined in Zhang and Wildemuth (2009). After reviewing the research questions, each coder individually identified chunks of text in the interview transcripts that represented key themes of the research. After coding each transcript, the two coders convened to compare and normalize themes, in accordance with the constant comparative method (Glaser and Strauss 1967). Through an iterative process, as the two coders continued to code interview transcripts, coding consistency was checked through assessment of intercoder agreement. If any disagreement occurred, the two coders discussed and resolved the disagreements, as suggested by Schilling (2006). Once sufficient coding consistency was achieved, the coding rules were then applied to all 17 interview transcripts.

Four key themes emerged from the interviews, and PI responses relating to these key themes were grouped into subthemes. See Table 1 for an overview of key themes and subthemes.

1 <http://helpslab.montana.edu>

**Table 1:** Thematic Analysis

Key Theme	Example Quotes	Subthemes
1. Resources used to write the DMP	<p>“You just copy and paste and add and subtract. I probably didn't write a word of that stuff.”</p> <p>“You have to know who to ask [for help].”</p>	<ol style="list-style-type: none"> <li>1. Colleagues and collaborators</li> <li>2. Reused own DMP language</li> <li>3. Library data services online resources</li> <li>4. NSF guidance</li> <li>5. Library consultation</li> <li>6. Institutional policies</li> <li>7. NSF reviewer feedback</li> </ol>
2. Utility of the DMP during the grant-funded project	<p>“I don't remember what I wrote in 2015.”</p> <p>“The DMP is putting fancier words to what we do very informally.”</p>	<ol style="list-style-type: none"> <li>1. DMP describes usual activities</li> <li>2. Activities outlined in DMP were updated later</li> <li>3. DMP informed policies for internal data sharing</li> <li>4. Didn't use DMP</li> </ol>
3. Plans for publicly sharing data	<p>“How is somebody able to just come in and look at this data—and we've spent so much time generating it and understanding it—and use it in a way that it should be used?”</p> <p>“It wouldn't be a bad idea to publish code, or make everything more transparent. I'm always a fan of it. But usually with so much stuff to do, it tends to fall by the wayside.”</p> <p>“I want to see more open access to science and data.”</p>	<ol style="list-style-type: none"> <li>1. Reasons to share data <ol style="list-style-type: none"> <li>a. Data sharing mandate</li> <li>b. Transparency</li> <li>c. Public good</li> </ol> </li> <li>2. Concerns about data sharing <ol style="list-style-type: none"> <li>a. Desire to control own data/research narrative</li> <li>b. Not sure what data is relevant/required to be shared</li> <li>c. Data is very large</li> <li>d. Confidentiality and identification concerns</li> <li>e. Sharing requires additional time and effort</li> </ol> </li> </ol>
4. Writing the DMP as a helpful or unhelpful exercise	<p>“DMPs are an important piece to try to instill a sense of importance on the data that is collected, as well as its legacy impact on science. And to train faculty and students on how to maintain that structure, so that ten years from now you could go back to that data and it could still be useful.”</p> <p>“I thought it was busy work because I do this anyway.”</p>	<ol style="list-style-type: none"> <li>1. Helpful <ol style="list-style-type: none"> <li>a. Useful exercise to reflect on what data would be produced, and general data management practices</li> <li>b. A good DMP makes the grant look better</li> <li>c. Valuable for accountability</li> <li>d. Helped formalize existing lab procedures</li> </ol> </li> <li>2. Unhelpful <ol style="list-style-type: none"> <li>a. Not helpful, but required</li> <li>b. Balancing time spent on DMP with good-enough quality</li> <li>c. Can feel like a bureaucratic impediment</li> </ol> </li> </ol>



## Results

### *Q1a. Are grant proposals with more complete/detailed DMPs more likely to be funded?*

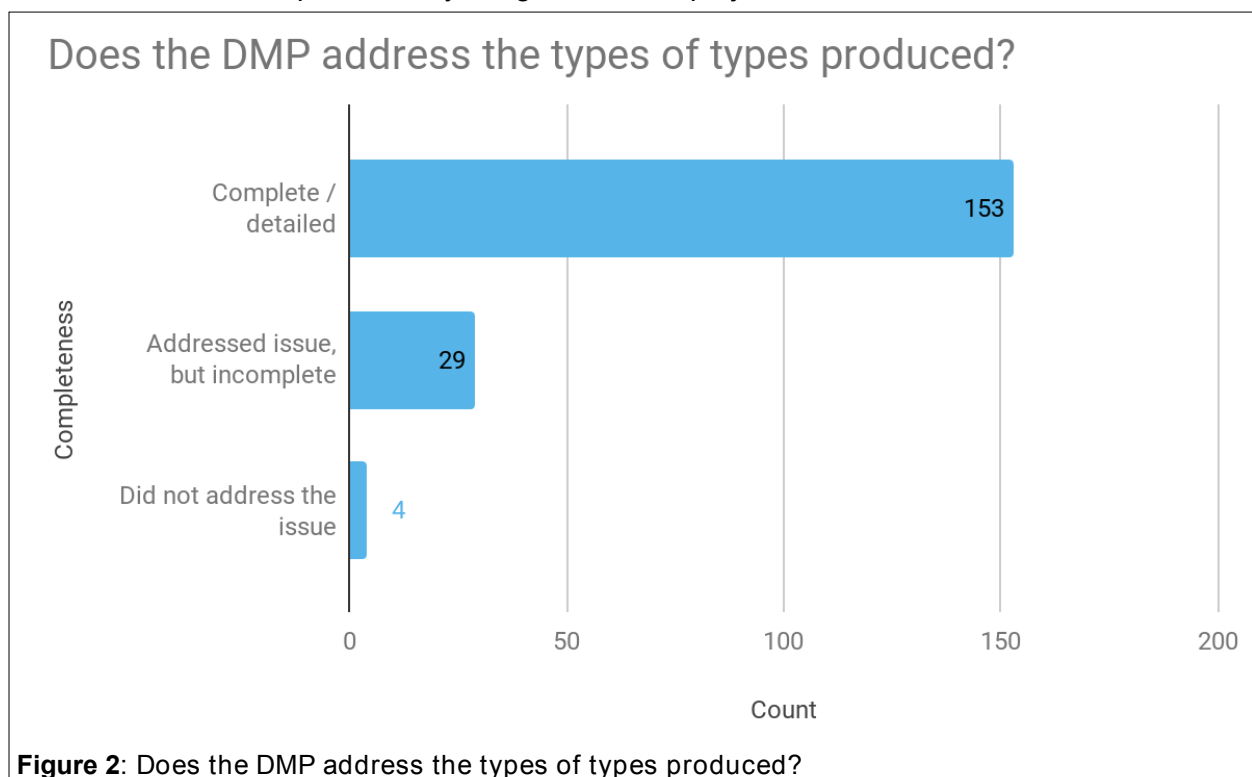
The difference in completeness between DMPs from awarded and declined DMPs was not statistically significant. This result is underscored by interview responses. Some interviewees had experience serving on grant review panels (n=2, 12%), and suggested that while the DMP was always reviewed, it was usually simply deemed either adequate or inadequate. Two PIs (n=2, 12%) suggested that the DMP is viewed as a supplementary document in the proposal—less important than the scientific research at the core of the proposal. While the DMP must be adequate to pass review, the completeness of a DMP does not appear to influence the success of a grant proposal.

### *Q1b. Which sections within DMPs have the most and least complete/detailed information?*

The DMP content analysis and subsequent PI interviews revealed some trends in the completeness/detailedness of the DMPs. Figures 2-6 show the completeness of DMPs in the five main sections of the DART rubric. Standards for “complete/detailed,” “addressed issue but incomplete,” and “did not address the issue” are defined by the DART Rubric (Whitmire, Carlson, Westra, Hswe, and Parham 2017).

### *Data Types*

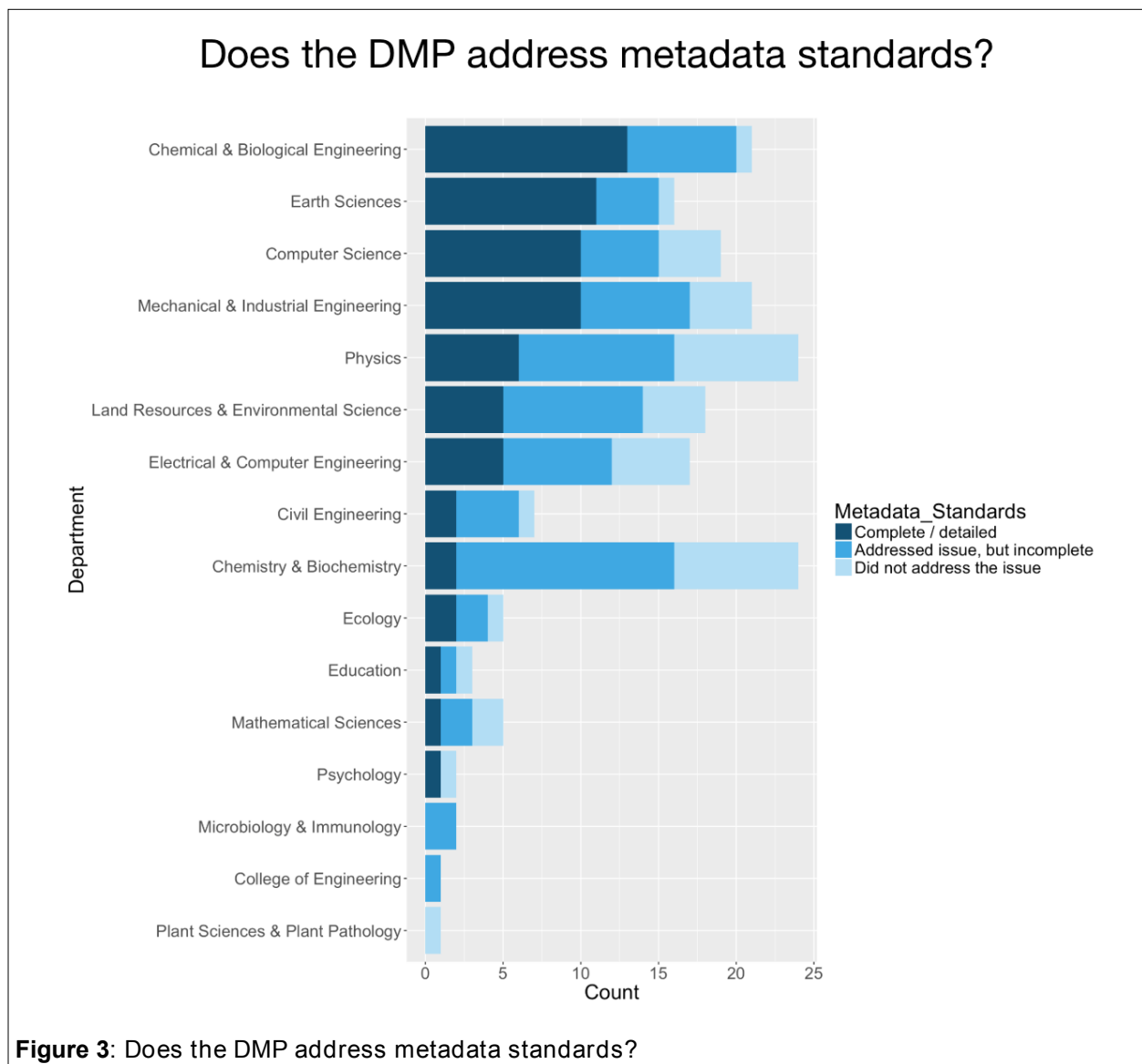
82% (n=153), of DMPs provided complete lists of data types. 16% (n=29) of DMPs had incomplete lists, and 2% (n=4) of DMPs didn’t address the type of data they would be collecting. Figure 2 shows the completeness/detailedness of DMPs when addressing the types of data that would be produced by the grant-funded project.



**Figure 2:** Does the DMP address the types of types produced?

*Metadata Standards*

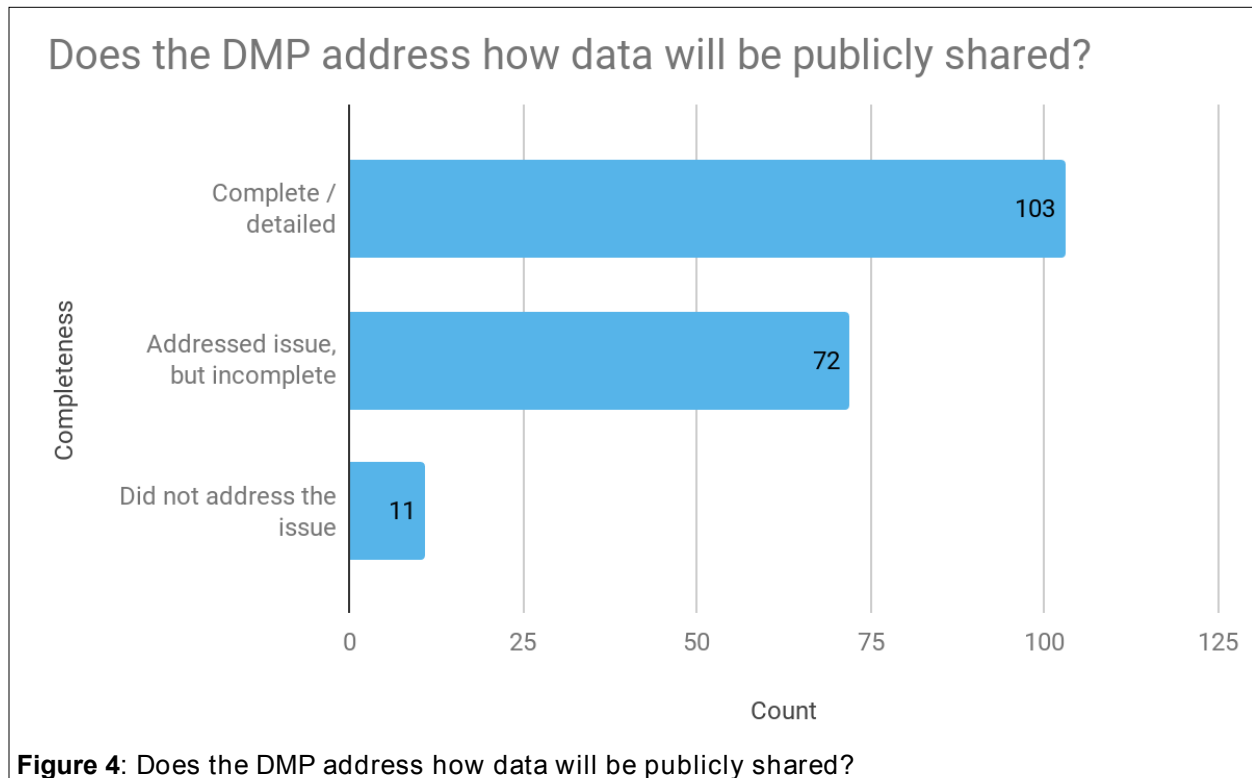
This research suggests that PIs would benefit from metadata guidance. 37% (n=69) of DMPs offered complete/detailed information about the metadata they planned to use in their grant-funded project, 40% (n= 75) addressed the issue, but incompletely, and 23% (n=42) did not address the issue. However, some departments' DMPs tended to provide more detailed metadata information than others (see Figure 3). Figure 3 shows results from DMPs organized by academic department at Montana State University; please note that some departments are represented at higher rates than others. Figure 3 illustrates that metadata standards information was more complete in DMPs from Chemical and Biological Engineering (62%, or 13 out of 21 DMPs, were complete/detailed), Computer Science (53%, or 10 out of 19 DMPs, were complete/detailed), and Earth Sciences (69%, or 11 out of 16 DMPs, were complete/detailed).



**Figure 3:** Does the DMP address metadata standards?

*How the Data will be Publicly Shared*

94% (n=175) of the DMPs in this analysis included plans for making data publicly accessible (See Figure 4). 55% (n=103) of DMPs had “complete/detailed” information about the plan to make data public, and 39% (n=72) of DMPs “addressed the issue, but incomplete.” This result suggests that PIs at Montana State University know that they are required to share their data publicly.



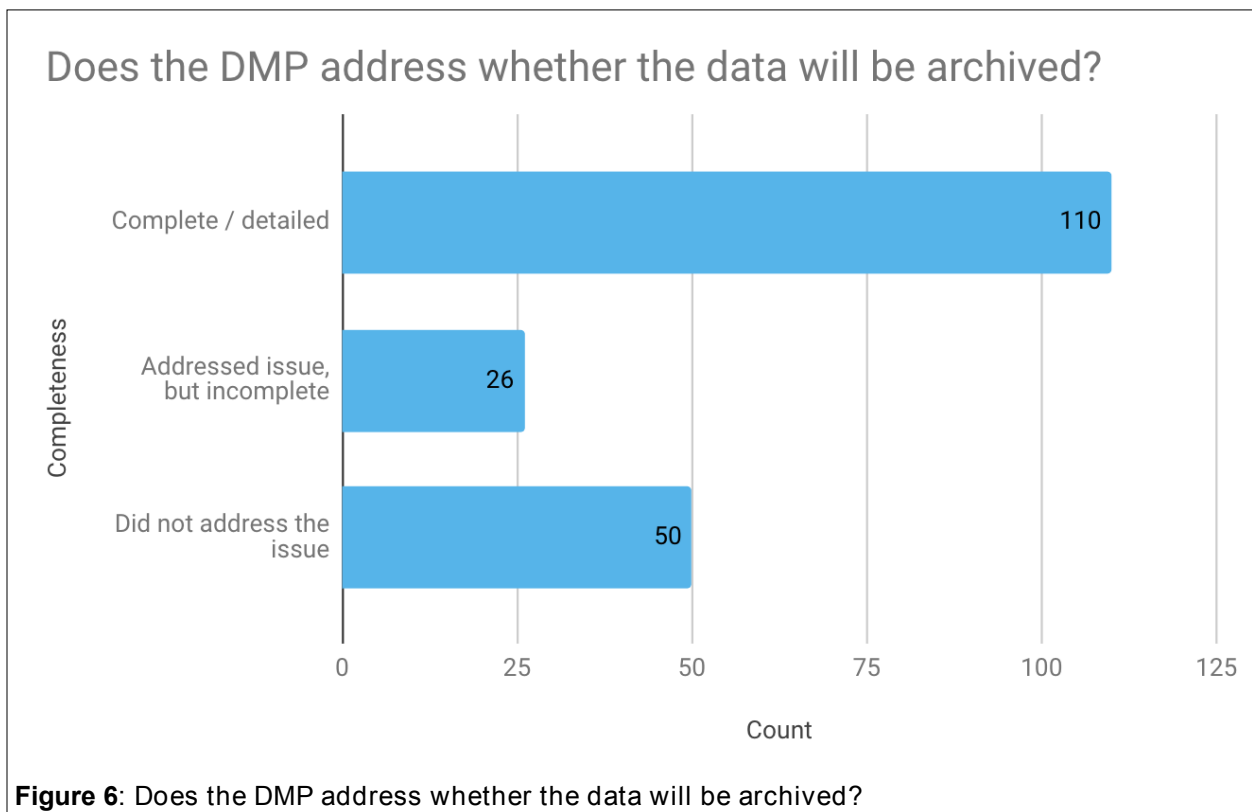
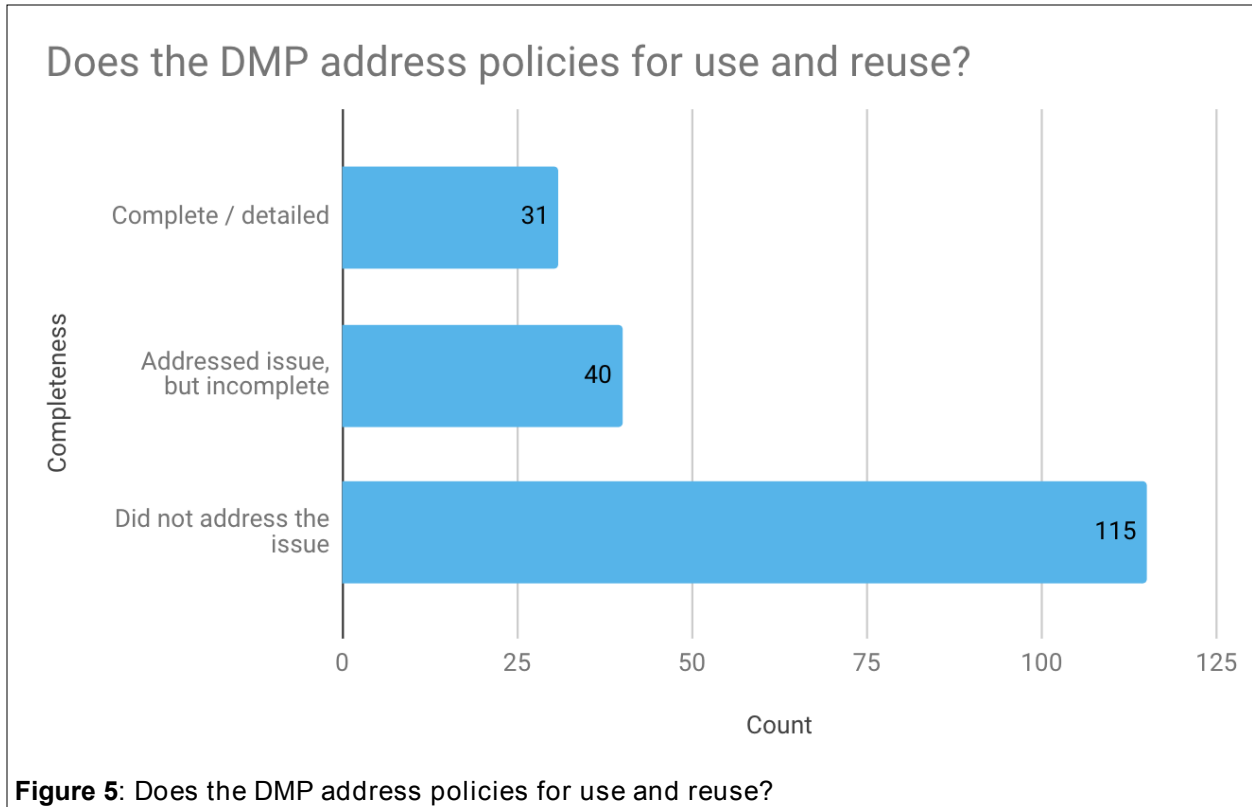
However, this research also suggests that many PIs do not use data repositories to facilitate long-term access and reuse. Most DMPs (56%, n=104) indicated that data would be shared via informal venues like a personal website or wiki, through a file sharing service like Dropbox or Google Drive, or upon request. Of the 186 DMPs analyzed, only 44% (n=82) proposed to share data via a repository or as a supplement to a manuscript.

*Policies for Use and Reuse*

The DMPs in this analysis also rarely provided policies for the use and reuse of data, as shown in Figure 5. 62% (n=115) of DMPs did not address the issue of policies for use and reuse. Such policies are a key part of FAIR data, enabling future users to actually use and reuse the data.

*Data Storage and Archiving*

Most DMPs showed that PIs understand how to store and archive their data over the long term (see Figure 6) with 73% (n=136) of DMPs either providing complete/detailed information, or addressing the issue incompletely.

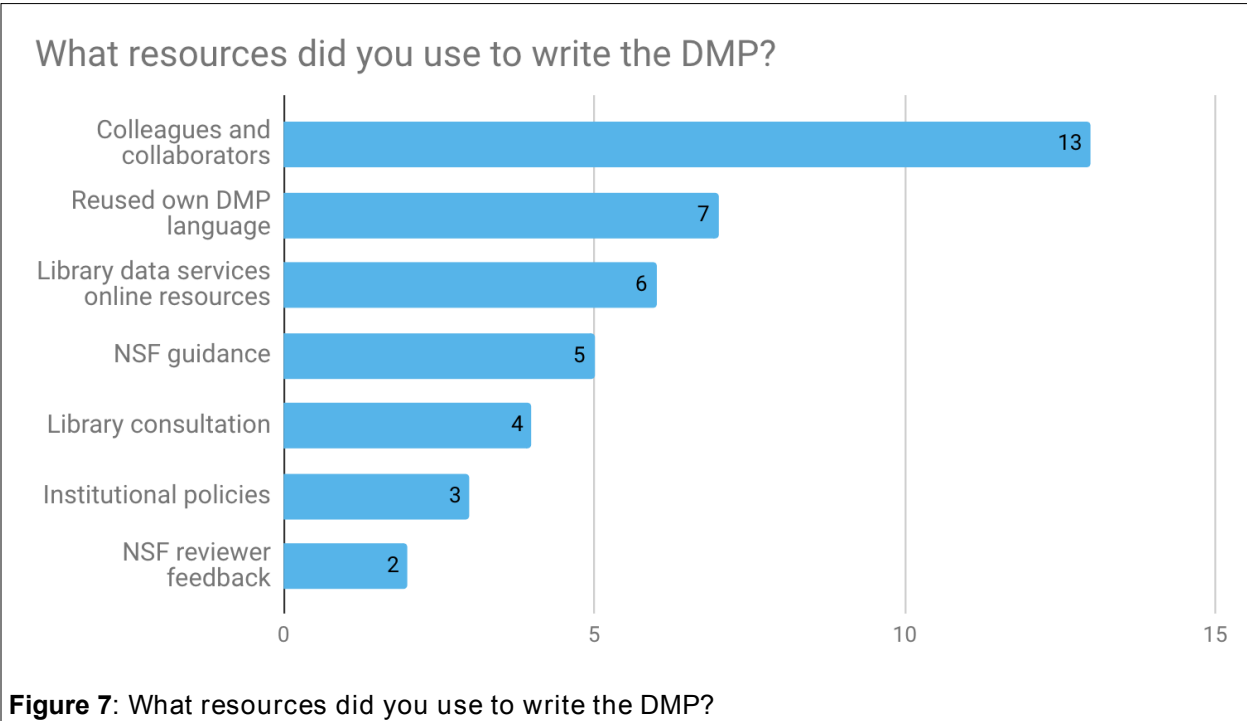


**Q2. Does writing a DMP affect PIs' data management and data sharing practice for grant-funded projects?**

Research question Q2 was answered using the semi-structured interviews with PIs. Four key themes emerged from the interviews, which are explained in more detail here. For more information about methods and themes, please see *Research Questions and Methods* and Table 1, above.

**1. Resources used to write the DMP**

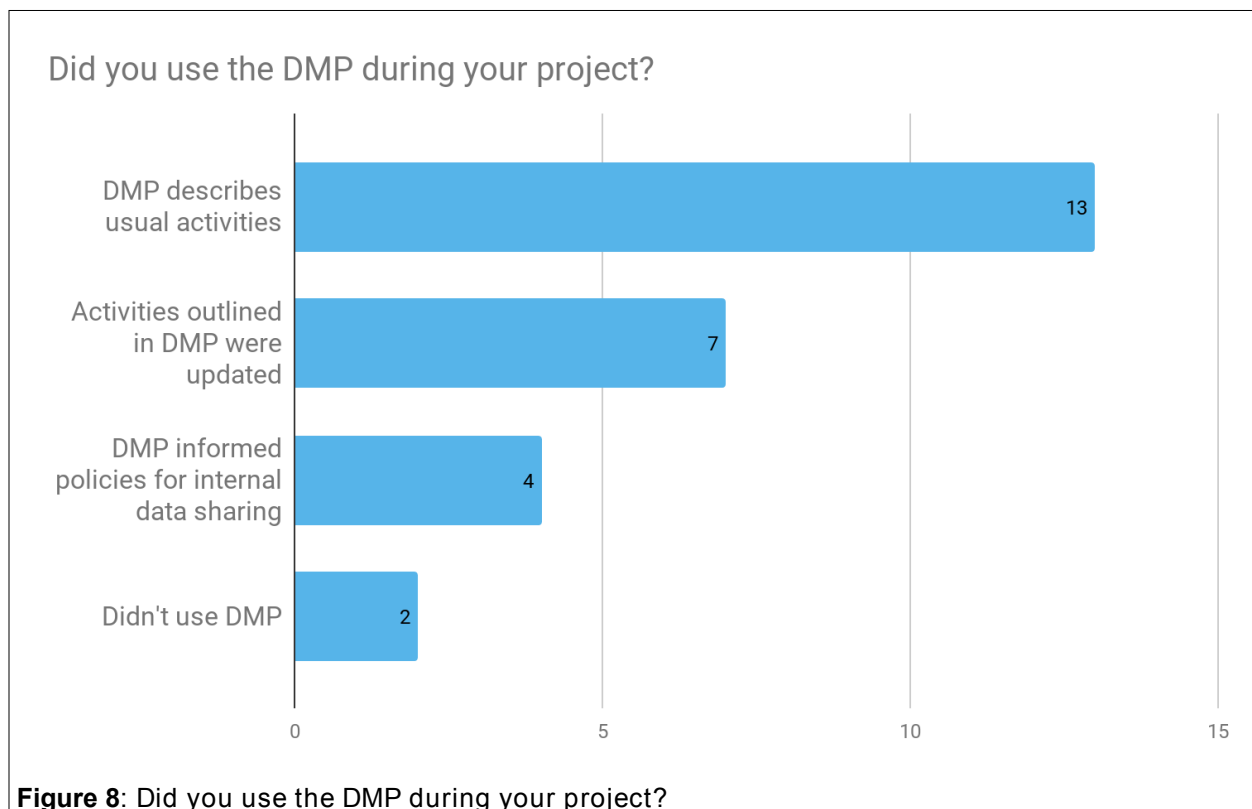
The PIs interviewed reported using various sources to piece together their DMP, including gaining insights from colleagues and collaborators (77%, n=13), reusing their own DMP language (41%, n=7), using online resources provided on the library data services website (35%, n=6), using guidance provided by NSF (29%, n=5), in-person library consultations (24%, n=4), institutional policies on data stewardship and data ownership (18%, n=3), and feedback from NSF reviewers (12%, n=2) (see Figure 7).



**2. Utility of the DMP during the grant-funded project**

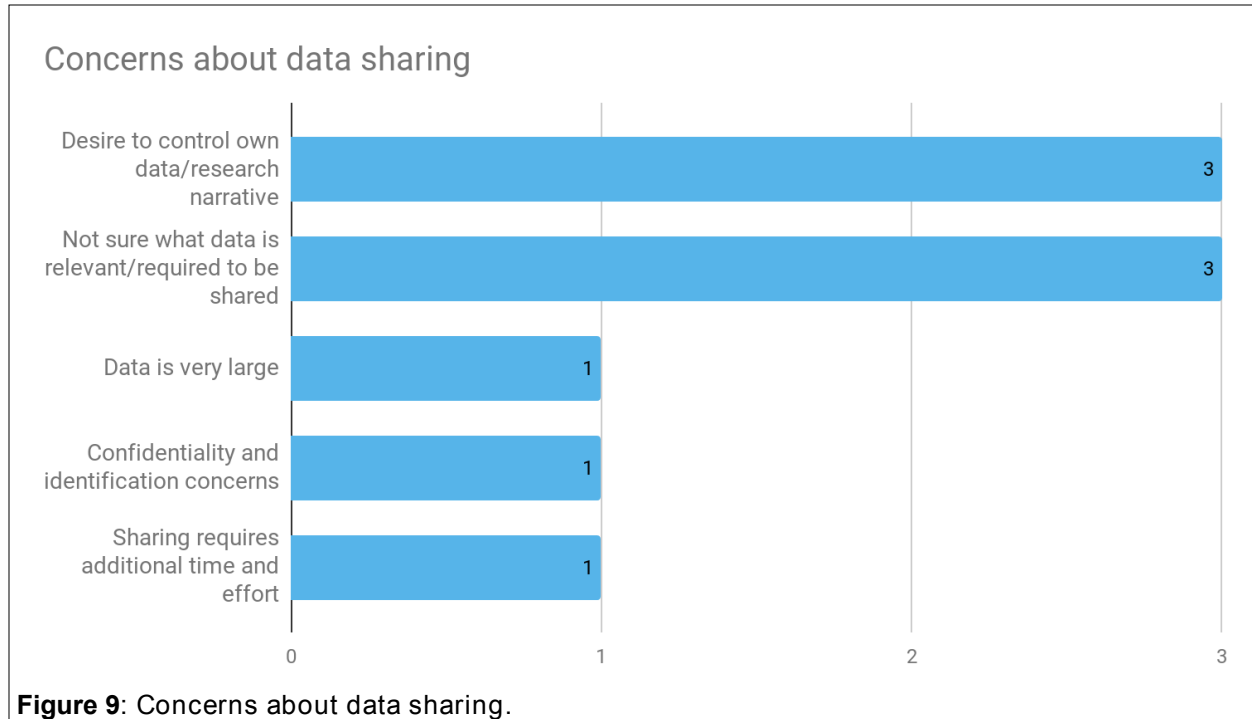
None of the PI interviewees used the DMP as a guiding document for data management in their lab. However, some (n=4, 24%) reported that the process of writing the data management plan did inform policies to facilitate sharing within their research team and to help with personnel transitions. Most (n=13, 76%) reported that while they generally followed the protocol outlined in the DMP, this was only because the DMP describes their usual activities and standard data management practices in their field. Some PIs (n=5, 29%) said that they adhered to the spirit of their DMP, but their actual data management practices were continually being updated to account for efficiency, changing best-practices, and new research strategies.

For example, one PI said that the DMP is "an earnest best effort to describe what you are going to do. It's not a contract, it's a grant. [NSF] wants good science. And if you're going to do good science a different way, it's like 'whatever is best for the science.'" Another PI said, "to the letter the DMP isn't exactly what we are doing, but this is not a change in scope so much as because we ended up using different models to answer the same questions." Some PI interviewees (n=2, 12%) reported that they used generic language in their DMP such as "we will comply with standards in the field," and noted that these standards are constantly changing. One PI joked that after writing the DMP to get the grant, they "look back once a year to say, 'oh, that was eye opening.'" Figure 8 illustrates the subthemes that emerged while discussing the use of DMPs during grant-funded projects.



### 3. Plans for publicly sharing data

Overall, the PI interviewees understood that data sharing is required by NSF. However, some PIs (n=3, 18%) reported that they generally aim to share as little data as possible to prove the experiment or published conclusions. As one PI said, "it seems like a big extra bit of work to get all that done, especially if no one is asking for it." Many interviewees (n=10, 59%) reported concrete plans for sharing their data; however, about a third of those interviewees (n=3, 18%) planned to share upon request, which is less sustainable than sharing using a data repository or supplementary material. Of the interviewees who offered reasons for sharing their data (n=5, 29%), some cited the NSF data sharing mandate (n=3, 18%), and some suggested that sharing supports replication and validation of research results (n=2, 12%). While the majority of interviewees (n=15, 88%) planned to share at least some of their data publicly, many (n=9, 53%) also voiced concern over data sharing requirements, as shown in Figure 9.



#### 4. Writing the DMP as a helpful or unhelpful exercise

Some PIs (n=5, 29%) reported that writing the DMP felt like yet another task to complete at the end of a long grant-proposal-writing process, and irrelevant to their actual data management practices. For example, one PI said, “I thought that [writing the DMP] was busy work, because I do this anyway. The only way for me to communicate what I am doing is by publishing, by saving the data, by sharing it with colleagues. And I do that anyway. I don’t need to write a data management plan to tell someone I do it.” However, most PIs (n=13, 76%) reported that the DMP-drafting process encouraged them to reflect on their own data management practices. Illustrative quotes include:

- The DMP ensured that “the data itself would be backed-up, so we wouldn’t lose it. We followed protocols established for data management for that [purpose].”
- The DMP was an opportunity to “spend a little time describing what we actually do.”
- “DMPs are an important piece to try to instill a sense of importance on the data that is collected, as well as its legacy impact on science. To train ... faculty and students on how to maintain that structure, so that ten years from now you could go back to that data and it could still be useful.”
- The importance of creating and “enforcing [data management policies] on a regular basis. [Writing the DMP] sort of clarified it in my head.”
- Considering that the “data management strategy [should] evolve with the advent of new storage capabilities.”
- “You have to kind of tailor for the discipline, so if I need a data management plan for an engineering project, it’s probably a lot different than for social psychology and human subjects.”

In addition, one PI reflected upon the bigger picture of data reuse, saying “I think that [writing a DMP is] incredibly important because I want to see more open access to science and data, and the usability of other people's data is important.”

## **Discussion and Recommendations for Practice**

By understanding which sections of the DMP are most and least complete/detailed, libraries can provide tailored resources that provide guidance to PIs in areas of the DMP that tend to be the least complete/detailed. This research indicates that library data management planning services should aim to streamline the DMP writing process, focusing on the following areas.

### *Understanding what PIs Already Know*

This analysis reveals that PIs have a strong grasp on what types of data will be produced over the course of their research (as illustrated in Figure 2, above). It also shows that PIs know how to document their data to facilitate work by a rotating research staff. This analysis also shows that the DMP writing process helps PIs consider data archiving and sharing requirements, and many of them include complete/detailed information in their DMPs describing archiving and publicly sharing data (as illustrated in Figures 4 and 6 above).

### *Specialized Concepts in the DMP*

PIs need help navigating the structure of the DMP, and understanding how concepts like metadata, data sharing, and data archiving match with the internal data management work that they are already doing. Two of the PI interviewees also suggested that a lack of understanding of key terms such as “metadata,” “archiving,” and “data licensing” was a barrier to writing better DMPs. Referring to the sections of the DMP on metadata and data storage, one PI said that their DMP had “a lot of wording about computers” and that their own “lack of computer literacy made [them] self-conscious.”

### *Facilitating FAIR Data*

PIs need more information about how to facilitate findability, accessibility, interoperability, and reuse for their data, including metadata, licensing, and selecting a data repository. This DMP analysis revealed that only 37% (n=69) of DMPs addressed the issue of metadata standards completely/in detail. This research additionally shows that metadata guidance may be especially useful if tailored to faculty in the departments whose metadata information was least complete in this analysis, such as Chemistry and Biochemistry (in which 92%, or 22 out of 24 DMPs either addressed the issue of metadata standards incompletely, or did not address the issue). Libraries should provide guidance on licenses such as Creative Commons that govern use and reuse of research data, since 62% (n=115) of DMPs in this analysis did not include policies for use and reuse.

### *Stronger, more Actionable DMPs*

These recommendations for practice aim to help data librarians create tailored DMP resources that respond to PI needs regarding data management planning and public sharing of FAIR data. Such tailored services support stronger, more actionable DMPs.



## Limitations

### *Sampling*

The DMPs and PI interviews were both convenience samples. DMP documents were selected according to which PIs had uploaded a full proposal to the online proposal approval system, producing a total of 198 DMPs to analyze. Of those 198 DMPs, 11 reported that no data would be produced, and one PI opted out of the research, producing a final dataset of 186 DMPs. The 17 PI interviewees were volunteers.

### *DMP Scoring*

A single student research assistant scored the DMP documents. The DART tool has been tested for intercoder consistency; the author also provided in-depth training and conducted spot checks on the completed rubric scoring. However, some inconsistencies may exist in the DMP scoring data.

## Conclusion

This analysis of DMP documents and semi-structured interviews with PIs produced the following answers to the research questions.

### *Q1a. Are grant proposals with more complete/detailed DMPs more likely to be funded?*

The difference in completeness between DMPs from awarded and declined DMPs was not statistically significant.

### *Q1b. Which sections within DMPs have the most and least complete/detailed information?*

The sections of the DMPs that had the most complete/detailed information were the sections on data types, data archiving and public data sharing (although most DMPs (56%, n=104) indicated that data would be shared via informal venues, rather than data repositories). DMPs tended to be least complete/detailed when discussing specialized ideas such as metadata standards and policies for use and reuse.

### *Q2. Does writing a DMP affect PIs' data management and data sharing practice for grant-funded projects?*

This research suggests that PIs generally value the DMP writing process as a moment of reflection about data management and sharing. For internal data management, this moment of reflection is simply an opportunity to refine the activities already being done in their labs, not a major learning experience; PIs continue using their already-established practices. For data sharing, the moment of reflection prompted by the DMP writing process did appear to help PIs consider how they would share their data.

The DMP content analysis and the subsequent PI interviews suggest a few key takeaways. (1) Most PIs practice internal data management in order to prevent data loss, to facilitate sharing within the research team, and to seamlessly continue their research during personnel turnover; (2) PIs still have room to grow in understanding specialized concepts such as metadata and

policies for use and reuse; (3) PIs may need guidance on practices that facilitate FAIR data, such as using metadata standards, assigning licenses to their data, and publishing in data repositories. The insights produced by this research can inform how library data services are delivered, in order to help PIs get more value out of the DMP writing process, and to support FAIR data practices from the early stages of a grant-funded project.

## Supplemental Content

### Appendix A

An online supplement to this article can be found at <http://dx.doi.org/10.7191/jeslib.2018.1155> under “Additional Files”.

## Data Availability

Data associated with this article are available from Zenodo at <https://doi.org/10.5281/zenodo.2432419>.

## Acknowledgements

This research was funded by the National Library of Medicine, National Institutes of Health under cooperative agreement number UG4LM012343 with University of Washington. Many thanks are also in order for the work of my outstanding student research assistant, Wangmo Tenzing. The Montana State University HELPS Lab scheduled and transcribed the PI interviews. Figure 3 was created by Laurie Rutemeyer at Montana State University Statistical Consulting and Research Services, which is supported by Institutional Development Awards (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under Awards P20GM103474, 5U54GM104944, U54GM115371, and 5P20GM104417. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

## Disclosures

The substance of this article is based upon a lightning talk presentation at RDAP Summit 2018: “Make the plan, work the plan: measuring the impact of DMPs on data management and data sharing practices” <https://osf.io/3n6cz>.

## References

- Berman, Elizabeth. 2017. “An Exploratory Sequential Mixed Methods Approach to Understanding Researchers’ Data Management Practices at UVM: Findings from the Qualitative Phase.” *Journal of eScience Librarianship* 6(1): e1097. <https://doi.org/10.7191/jeslib.2017.1097>
- Bryant, Rebecca, Brian Lavoie, and Constance Malpas. 2017. *The Realities of Research Data Management* (four-part series). Dublin: OCLC Research. <https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html>
- Cox, Andrew M., Mary Anne Kennan, Liz Lyon, and Stephen Pinfield. 2017. “Developments in research data management in academic libraries: Towards an understanding of research data service maturity.” *Journal of the Association for Information Science and Technology* 68(9): 2182-2200. <https://doi.org/10.1177/0961000613492542>
- Glaser, Barney G., and Anselm L. Strauss. 1967. *Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine. <https://www.taylorfrancis.com/books/9780203793206>

- Holdren, John P. 2013. *Increasing Access to the Results of Federally Funded Scientific Research*. Washington, DC: Office of Science and Technology Policy.  
[https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)
- Mannheimer, S. & Tenzing, W. (2018). Data From: Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices. *Zenodo*. <https://doi.org/10.5281/zenodo.2432419>
- Mischo, William, Mary Schlembach, and Megan O'Donnell. 2014. "An Analysis of Data Management Plans in University of Illinois National Science Foundation Grant Proposals." *Journal of eScience Librarianship* 3(1): e1060.  
<https://doi.org/10.7191/jeslib.2014.1060>
- National Science Foundation (NSF). 2018. *Dissemination and Sharing of Research Results*.  
<https://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- National Science Foundation (NSF). 2017. Chapter II.C.2.j. Special information and supplementary documentation. Proposal & Award Policies & Procedures Guide (PAPPG), NSF 17-1.  
[https://www.nsf.gov/pubs/policydocs/pappg17\\_1/pappg\\_2.jsp#IIC2j](https://www.nsf.gov/pubs/policydocs/pappg17_1/pappg_2.jsp#IIC2j)
- Parham, Susan Wells, and Chris Doty. 2012. "NSF DMP Content Analysis: What Are Researchers Saying?" *Bulletin of the American Society for Information Science and Technology* 39(1): 37-38.  
<https://doi.org/10.1002/bult.2012.1720390113>
- Parham, Susan Wells, Jake Carlson, Patricia Hswe, Brian Westra, and Amanda Whitmire. 2016. "Using Data Management Plans to Explore Variability in Research Data Management Practices Across Domains." *International Journal of Digital Curation* 11(1): 53-67. <https://doi.org/10.2218/ijdc.v11i1.423>
- Peters, Christie, and Anita Riley Dryden. 2011. "Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston." *Science & Technology Libraries* 30(4): 387-403.  
<https://doi.org/10.1080/0194262X.2011.626340>
- Schilling, Jan. 2006. "On the Pragmatics of Qualitative Assessment." *European Journal of Psychological Assessment* 22(1): 28-37. <https://doi.org/10.1027/1015-5759.22.1.28>
- Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. 2014. "Research Data Management Services in Academic Research Libraries and Perceptions of Librarians." *Library & Information Science Research* 36(2): 84-90.  
<https://doi.org/10.1016/j.lisr.2013.11.003>
- Tenopir, Carol, Sanna Talja, Wolfram Horstmann, Elina Late, Dane Hughes, Danielle Pollock, Birgit Schmidt, Lynn Baird, Robert J. Sandusky, and Suzie Allard. 2017. "Research Data Services in European Academic Research Libraries." *LIBER Quarterly* 27(1): 23-44. <https://doi.org/10.18352/lq.10180>
- Vision, Todd J. 2010. "Open Data and the Social Contract of Scientific Publishing." *BioScience* 60(5): 330-330.  
<https://doi.org/10.1525/bio.2010.60.5.2>
- Whitmire, Amanda, Jacob Carlson, Brian Westra, Patricia Hswe, and Susan W. Parham. 2016. "The DART Project: Using Data Management Plans as a Research Tool." *Open Science Framework*.  
<https://doi.org/10.17605/OSF.IO/KH2Y6>
- Whitmire, Amanda, Jacob Carlson, Brian Westra, Patricia Hswe, and Susan W. Parham. 2017. "Guidance for Using the DART Rubric to Assess NSF Data Management Plans." *Open Science Framework*. <https://osf.io/y9gze>
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3(March): 160018. <https://doi.org/10.1038/sdata.2016.18>
- Yoon, Ayoung, and Teresa Schultz. 2017. "Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites." *College & Research Libraries* 78(7): 920-933.  
<https://doi.org/10.5860/crl.78.7.920>

Zhang, Yan, and Barbara M. Wildemuth. 2009. "Qualitative Analysis of Content." In *Applications of Social Research Methods to Questions in Information and Library Science*, edited by Barbara M. Wildemuth, 308-319. Westport: Libraries Unlimited. <https://www.abc-clio.com/ABC-CLIOCorporate/product.aspx?pc=A4819P>