UNTAPPED POTENTIAL: A CRITICAL ANALYSIS OF THE UTILITY OF DATA MANAGEMENT PLANS IN FACILITATING DATA SHARING

Jake Carlson
University at Buffalo

ABSTRACT

Many funding agencies require researchers to include a data management plan with their grant applications explaining how they intend to make the data generated from the research publicly accessible. University administration and campus service providers could potentially leverage the content of data management plans to facilitate compliance and reduce the burden on researchers. A case study at the University of Michigan demonstrates the promise of using data management plans as a communications and information sharing tool and the barriers in doing so. I apply the results of a content analysis to develop a series of recommendations to funding agencies, university administration, and campus service providers to improve the utility of data management plans in supporting data sharing and compliance.

Keywords:

Data Management Plans, Data Sharing, Funding Agency Requirements, Data Curation, Institutional Data Policy and Practices

INTRODUCTION

Over the past decade, federal and private funding agencies have instituted a requirement that grant applications include a data management plan (DMP). The details of the requirement differ between agencies but the primary objective behind the DMP requirement is to make data from funded research available to others outside of project personnel.

Ideally, DMPs provide a description of the research team's approaches and practices they will use in developing and sharing the data from their project. Given their content, DMPs could be quite useful as a shared, centralized reference document between the researchers developing the data and the data curators who will eventually take on the responsibility of making the data accessible to others once the data are submitted to a repository (Chodocki et al., 2020). Having access to the DMP would also help local data service providers, such as libraries or IT, reach out to researchers who may benefit from having support as they develop their data. In addition, as data sharing requirements take root, research administrators will need to document the steps taken by researchers to ensure that they successfully made their data publicly available. Given the function of the DMP, it could be used by research administrators as the means to track the progress of researchers in meeting their obligations in complying with agency mandates.

In practice however, DMPs are rarely used in these ways. In part, this is due to DMPs being largely inaccessible beyond the researchers who submitted the proposal and the grant agency reviewing them (Miksa et al., 2019). Although awarded grants are considered public documents, universities do not make them readily available to

anyone outside of those personnel who are responsible for administering them. In addition, funding agencies often require all grant documents to be pasted together and submitted as a single document, making it quite difficult to extract and route specific pieces of a proposal, like a DMP, to a particular stakeholder. Furthermore, DMPs exist as static documents. The information they contain is representative of the researchers' thinking and intent at the start of their project. Although researchers can update their DMPs locally, these updates are not usually attached or associated with the grant itself.

Beyond considerations of their accessibility, it is important to consider the purpose of the DMP as structured by the funding agencies that require them. Funding agencies currently use DMPs to aid review panels in evaluating the proposal, not to communicate information to data repositories or local service providers who offer support for sharing data. Review panels are generally comprised of experts in relevant fields of study, who may not have knowledge or skills in curating or sharing data themselves. Moreover, it does not appear that reviewers consider the quality of DMPs in their determination of whether or not to award funding (Hudson-Vitale & Moulaison-Sandy, 2019; Mischo et al., 2014). Nor does it appear that having a DMP attached to a data set has an impact on whether or not researchers will share their data in ways that are accessible to the public (Van Tuyl & Whitmire, 2016). The overall lack of consideration on how DMPs could be shared and used outside of a grant review has created a situation in which a tool that was ostensibly developed to encourage researchers to share their data may in fact be a barrier.

When DMPs are studied, the focus of the analysis tends to be more on the perceived quality of a DMP (whether it meets the stated guidelines of the funding agency), than on the utility of the DMP itself (i.e., does it provide useful, actionable information that could be used to provide services and support). In this article, I analyze the content of DMPs written as a component of grants awarded to the University of Michigan as the lead institution over a period of twelve months to explore the potential utility of this content to data curators and local service providers. I then use this analysis to make recommendations for how DMPs could be modified to make them more useful as a means of communicating important information to data service providers, research administrators and data repositories. This in turn would help DMPs to serve their intended function of enabling public access to high quality data sets produced with federal tax money.

LITERATURE REVIEW

The idea for a data management plan requirement grew out of a desire to support the emergence of data-intensive research in the early 2000s. Several terms were coined to try to define this phenomenon and its potential: cyberinfrastructure, eScience, the Fourth Paradigm, Data Science, etc. Scholarly societies, federal agencies, and other research organizations formed working groups to articulate what investments in human and technical infrastructure would be needed to realize the potential of data intensive research. A key component in many of these reports was the recognition that data intensive research requires ready access to high quality, well described data sets and that processes would need to be developed to ensure that data would be widely available. For example, the 2009 report from the Interagency Working Group on Digital Data (IWGDD) stated: "*We envision a digital scientific data universe in which data creation, collection, documentation, analysis, preservation, and dissemination can be appropriately, reliably, and readily managed... To pursue this strategy, we recommend that... agencies promote a data management planning process for projects that generate preservation data*" (IWGDD, 2009, p. 1-2).

The National Science Foundation (NSF) was the first agency in the United States to require all applications for funding include a data management plan (DMP). The DMP, as defined by the NSF, is "*a document of no more than two pages and should include information about the types of data to be generated in the project, the standards to be used in formatting and contextualizing the data (metadata), the policies for accessing the data, the policies for reusing the data, and the plans for archiving the data*" (2020). As they are a part of the submitted grant application, DMPs are

evaluated as "an integral part of the proposal" (NSF, 2020). The NSF's announcement, and the subsequent policy memorandum from the Office of Science and Technology Policy in 2013 (also known as the Holdern Memo), led to many other federal agencies and private funding agencies in the US adopting data sharing policies and DMP requirements of their own.

The DMP is still evolving as a tool and funding agencies continue to make changes to their data sharing requirements. The NIH recently unveiled a new Data Management and Sharing Plan that went into effect in January 2023 for any proposal that will generate research data (NIH, 2020). Although their new DMP requirement is similar to the NSF's in some ways, including the two page limit and focus on community standards, several new elements are introduced. The NIH requirement asks researchers to include information about any tools or code used to generate the data, an explanation on how compliance with the plan will be monitored and managed, and a description of any considerations for access, distribution, or reuse of potentially sensitive data. The NIH also encourages researchers to update their DMP over the course of the award to reflect any changes in the management or plans for sharing the data.

However, the DMP requirements of funding agencies do not appear to have had the desired effect of increasing access to research data thus far. Several studies completed soon after funding agencies instituted their DMP requirements demonstrated the difficulties that researchers were having in trying to respond. Many researchers, even those with experience in sharing data informally with colleagues, were not accustomed to making their data publicly accessible to anyone at that time and so the new DMP requirement was out of step with their practices (Imker, 2017). The DMP requirement itself was confusing and raised many questions for researchers. Were they supposed to share all their data or just the data that directly supported their findings? How should they respond to the requirement to use "community standards", if such standards did not yet exist? (Steinhart et al., 2012).

Outside of increasing access, data librarians and others have studied DMPs as a means to learn more about the strategies and activities researchers employ in managing and sharing their data. Studies of DMPs do show some utility in understanding variability between different disciplines, but overall DMPs lack consistency and often do not even contain the information that funding agencies require. Many researchers do not appear to understand fully what funding agencies are asking for, or do not see the utility of spending time and effort in developing plans to share their data (Bishoff & Johnson, 2015; Parham et al., 2016). Therefore, DMPs, as currently structured, appear to have limited value as a means to understand researcher's approaches to sharing data as well as the corresponding practices. These types of analyses also reveal a need to push beyond the boundaries of DMPs as currently defined by funding agencies to identify and encourage approaches that would produce improved results. Despite their name, DMPs generally emphasize how post-publication data will eventually be shared rather than how the data will be collected, processed, and managed as it is being developed (Williams et al., 2017). The DMP policies as currently defined by funding agencies do not address data preservation, formats, documentation, and metadata clearly or sufficiently enough to support the data sharing vision articulated by the IWGDD (Dietrich et al., 2012).

Data curators consider what policies and practices are needed to assist researchers and their staff in developing a data set that has enduring value for themselves and others over time. For example, ICPSR, a data repository for the social sciences, recommends several additional elements including statements on the format of the data, determining who will have particular responsibilities for managing and sharing data, the intended audience for the data, and any Ethics and Privacy issues (Inter-university Consortium for Political & Social Research [ICPSR], 2012). The Digital Curation Center, an agency providing expert advice and practical help on storing, managing, sharing, and preserving data, developed their own checklist for a DMP as a means of codifying best practice. They also recommend including statements detailing who will have responsibilities for ensuring that data are well managed and shared appropriately, explicitly defining which data will be retained or shared, and addressing any ethical or

legal issues present in the data in the DMP (DCC, 2013). Studies have shown that when data curators work closely with research staff in developing a DMP, the resulting DMP is more comprehensive and more likely to be ingrained in the researcher's practices and workflows (Burnette et. al., 2016; Karimova et al., 2021).

An emerging approach to addressing the shortcomings of the DMP is to make them machine actionable. Most funding agencies with a DMP requirement require researchers to write up their DMP as a static document and then embed it with other components of the funding proposal. However, this approach makes it difficult for anyone to access the DMP once the award has been made, including its authors, and therefore unlikely that the DMP will be used as intended. In contrast, machine actionable data management plans (maDMPs) would be submitted as a digital document enabling information to be easily read and automatically extracted as needed. Given their accessibility and flexible structure, maDMPs could easily be adjusted and updated as needed to reflect current thinking and practice in managing, sharing, and curating the data as the data are being developed. They could also more easily incorporate relevant standards and best practices, such as the application of controlled vocabularies, to facilitate communication across stakeholders. Finally, maDMPs could be assigned a unique and persistent identifier, such as a DOI, to connect it definitively with the researchers associated with the data and the other outputs associated with the research (Sims et al., 2017; Stodden et al., 2019). The NSF believes that maDMPs are a promising direction in addressing the current shortcomings of DMPs and have encouraged their development and implementation (NSF, 2019).

METHODOLOGY

In my previous position, I oversaw a suite of services offered by the University of Michigan Library to support researchers in managing, sharing, and preserving their data (https://www.lib.umich.edu/research-and-scholarship/data-services). These services include reviewing DMPs, consulting and training sessions on preparing data for sharing, and operating Deep Blue Data, a data repository for sharing and preserving data generated at U-M (https://deepblue.lib.umich.edu/data). I was interested in in learning more about the data management and sharing practices and needs of the U-M community, as well as obtaining advance notice of data deposits that will be submitted to Deep Blue Data. Although researchers who intend to deposit their data into Deep Blue Data are encouraged to contact the library early in the lifecycle of their data, this did not happen often. I asked U-M's Office of Research for access to U-M's grants management system, which they granted in January 2020.

For this project, I reviewed the DMPs from proposals to federal agencies that were awarded to U-M as the lead institution for twelve months, from March 2020 to February 2021. I located DMPs by opening the narrative description of the proposal and using the find command available in the Google Chrome browser to search for the terms "data management", "data sharing", or "resource sharing". If unsuccessful, I scrolled to the bottom of the proposal narrative and then slowly scrolled up the document looking for a DMP. If I was still unsuccessful, I looked at other files attached to the record of the grant to see if the DMP is included as a separate document. If none of these approaches led to the discovery of a DMP, I listed the award as not having DMP available. Although descriptions of data, as well as data management and sharing activities, are included sometimes in the body of the narrative or in other documents outside of a DMP, I did not review or capture this information. My process for identifying the DMPs for this study is presented visually in Figure 1.

Figure 1: Process for Identifying DMPs

Once acquired, I then reviewed the DMPs to identify the following key pieces of information:

1. The types and formats of the data to be generated,
2. Any indications and descriptions of the metadata and documentation to be provided,
3. Any mention of intellectual property (IP) concerns or constrictions on making the data available to others,
4. Statements on how and when the data will be shared with others outside of the project, and
5. The expected duration of preservation needed for the data.

Every federal funding agency has a different set of requirements for researchers to follow when it comes to sharing the unique resources generated from funded projects. Funding agencies even have different names for their requirements: data management plan, data sharing plan, resource sharing plan, etc. In this paper, I refer to these plans as DMPs no matter how individual funding agencies refer to them. As my intent was to study the content of these plans rather than to ascertain the level of compliance to the directives of funding agencies, I reviewed each DMP for the same set of variables. I found wide variability in the coverage and depth of the information provided, as expected. However, I also found that many DMPs did not provide required information, as well as more than a few DMPs that provided more thorough and in-depth information than was asked for by the funding agency. I collected all this information from the DMP regardless of the repository or data sharing method mentioned by the PI.

The top funding agencies awarding grants to researchers at the University of Michigan from March 2020 to February 2021 were the National Institutes of Health (NIH), the National Science Foundation (NSF), the Department of Defense (DoD), and the National Aeronautics and Space Administration (NASA). I grouped information from DMPs of grants awarded to U-M by other agencies together and present them in the category of "Other". Table 1 displays this information according to the awarding agency. In this period, there were 744 grants from federal agencies where U-M was the lead institution. Of these, I was able to locate data management plans, or the equivalent, for 476 of them in U-M's grants management system. Not being able to locate a DMP does not necessarily mean that a DMP does not exist for the award, only that I was unable to locate one within the U-M grants system. There may in fact be good reasons why a DMP would not be included in the materials, such as a DMP was not required for the particular funding program, or the award was supplementary to one already given. Nevertheless, the absence of DMPs from awards is potentially problematic. Without a DMP, the institution does not have an easy means of understanding what the research team has promised to the funding agency in making data from the project available.

Fifteen of these DMPs indicated that no data would be generated, leaving 461 DMPs for analysis.

Table 1: Awards Made to U-M with and without DMPs by Funding Agency

Agency	Total # of Awards	Awards with a DMP	% of Awards with a DMP	Awards not generating data	Total # of Awards from DMPs Generating Data
NIH	384	236	61%	3	233
NSF	199	183	92%	10	173
DoD	51	12	24%	1	11
NASA	44	23	52%	1	22
Other	66	22	33%	0	22
Grand Total	744	476	64%	15	461

RESULTS

The Types and Formats of Data

I reviewed each DMP and attempted to determine the nature of the data that research teams would be generating in the awarded project. To aid my analysis, I developed a broad categorization scheme based on the patterns that emerged. The data categories and definitions I developed are as follows:

Administrative – Data that pertain to administering or evaluating research or teaching programs.

Clinical – Data that pertain to the direct observation and treatment of patients rather than theoretical or laboratory studies.

Code – Data where the inputs and outputs were primarily or entirely comprised of instructions for computers to follow.

Experimental – Data that result from experiments conducted in labs or other controlled environments.

Genomic – Data derived from or that pertain to the DNA, RNA, proteins or other genetic elements of humans, animals, or other organisms.

Observational – Data developed through observing people or phenomena. This would include data gathered through surveys, interviews, and other interactions with people as well as data gathered using sensors and other instruments.

Physical Specimens – Data comprised of physical objects such as human or animal tissues, rocks, plants, etc.

Secondary – Data that were originally developed through prior research or for a different purpose than the project described in the DMP.

Simulation / Model – Data that were created for, or resulted from, a computer simulation or a model of a particular phenomenon.

Occasionally I found that the data described in the DMP fit into more than one of the defined data categories. When this happened, I assigned multiple categories to the DMP. Table 2 shows the breakdown of data category by funding agency.

Table 2: Data Category Assigned to the Data Described in the DMP by Awarding Agency

Data Category	NIH n=233	NSF n=173	DoD n=11	NASA n=22	Other n=22	Total N=461
Administrative	0	2	0	0	0	2
Clinical	27	2	0	0	1	30
Code	6	25	0	1	0	32
Experimental	18	51	6	1	8	84
Genomic	98	6	4	0	0	108
Observational	16	72	0	4	7	99
Physical Specimens	1	9	0	0	1	11
Secondary	16	15	0	5	2	38
Model or Simulation	9	42	3	16	12	82
Data Not Listed	63	5	1	0	0	69
No Data will be Generated	3	10	1	1	0	15

Note: Columns do not add up as a DMP may have included statements that fit into more than one of the listed categories.

In reviewing DMPs, I also identified any mention of the format of the data that would be developed over the course of the project. This information was not often included in DMPs, even in DMPs submitted to the NSF, which asks for information about data formats directly. On average only a third or so of DMPs included any mention of the format of the data:

- Eighty-four of 233 DMPs to the National Institutes of Health (NIH) included some information about the format of the data (36%).
- Fifty-six of 173 DMPs to the National Science Foundation (NSF) included some information about the format of the data (32%).
- Four of the 11 DMPs to the Department of Defense (DoD) included some information about the format of the data (36%).
- Five of the 22 DMPs to the National Aeronautics and Space Administration (NASA) included some information about the format of the data (22%).
- Seven of the 22 DMPs to Other Federal Agencies included some information about the format of the data (32%).

Metadata and Documentation

In reviewing DMPs, I sought to identify any indication that metadata and documentation would be generated (Table 3). When metadata or documentation was mentioned in the DMP, I noted if it was mentioned in passing or if some detail was provided. "Mentioned" is defined as the researcher providing a general or broad statement indicating that metadata or documentation would be generated, and "detailed" as the researcher providing a description of at least some of the content of the metadata or documentation to be gathered for the data over the course of the project.

Table 3: References to Metadata and Documentation in DMPs by Agency

	NIH n=233	NSF n=173	DoD n=11	NASA n=22	Other n=22	Total N=461
Metadata						
Mentioned	14 (6%)	23 (13%)	2 (18%)	4 (18%)	2 (9%)	45 (10%)
Detailed	3 (1%)	41 (24%)	2 (18%)	3 (14%)	4 (18%)	53 (11%)
Total	17 (7%)	64 (37%)	4 (36%)	7 (32%)	6 (27%)	98 (21%)
Documentation						
Mentioned	33 (14%)	43 (25%)	2 (18%)	5 (23%)	3 (14%)	86 (19%)
Detailed	12 (5%)	40 (23%)	2 (18%)	1 (5%)	7 (32%)	62 (13%)
Total	45 (19%)	83 (48%)	4 (36%)	6 (27%)	10 (45%)	148 (32%)

Forty-four DMPs in this study, or nine percent, included information about both the metadata and documentation to be generated for the data.

I also noted if a metadata standard or if a specific type of documentation was listed in the DMP, regardless of how much detail about content of the metadata or documentation was included. The specific type of documentation to be used was mentioned much more often in DMPs than specific metadata standards. One hundred and thirteen DMPs, or 24% listed a specific type of documentation that they would develop for the data. The most popular types of documentation were: Lab Notebooks (mentioned in 28 DMPs), Read me files (20), and Codebooks (11). Only 19 DMPs, or four percent, listed a specific metadata standard that they would employ. Of the standards listed, Dublin Core was mentioned three times and the Data Documentation Initiative (DDI) standard was mentioned twice.

How Will the Data Be Shared?

Most researchers gave some kind of indication of how they are planning to share their data. Only 38 of them made a statement that they are not planning to share (Table 4).

Table 4: Methods of Sharing Data Listed in DMP

	NIH n=233	NSF n=173	DoD n=11	NASA n=22	Other n=22	Total N=461
Repository – Domain	103 (44%)	44 (25%)	6 (55%)	4 (18%)	5 (23%)	162 (35%)
Repository – General	7 (3%)	11 (5%)	0 (0%)	2 (9%)	1 (5%)	21 (5%)
Repository – Institutional	12 (5%)	55 (32%)	4 (36%)	12 (55%)	13 (59%)	96 (21%)
Repository – Not Specified	19 (8%)	2 (1%)	1 (9%)	1 (5%)	2 (9%)	25 (5%)
Repository – Total	141 (61%)	112 (65%)	11 (100%)	19 (86%)	20 (91%)	304 (66%)
Researcher Hosted Website or GitHub	30 (13%)	68 (39%)	2 (18%)	7 (32%)	9 (41%)	116 (25%)
In Presentations or Publications	75 (32%)	41 (24%)	4 (36%)	6 (27%)	8 (36%)	134 (29%)
On Request	52 (22%)	47 (27%)	5 (45%)	2 (9%)	4 (18%)	110 (24%)
Some or All of the Data will not be shared	3 (1%)	8 (5%)	0 (0%)	0 (0%)	0 (0%)	11 (2%)

Note: Columns do not add up as some DMPs included more than one method of sharing data.

Repositories were listed as the primary means by which researchers would share their data. Roughly two thirds of DMPs from NIH and NSF grants mentioned a repository of some kind to share their data. Repositories were listed at even higher rates in DMPs from the DoD, NASA, and other agencies. I made note as to whether the repository mentioned in the DMP was a domain repository, a general repository, or an institutional repository.

Domain repositories, those that host a particular type of data or serve a specific field of research, were listed most often, particularly in DMPs to the NIH and the DoD. As the NIH supports multiple data repositories that are widely known and used by researchers in the medical and life science fields, this result is not surprising.

Generalist repositories are repositories that accept a variety of data types and formats, that were not associated or tied to a particular field or subject, and that were not associated exclusively with a single institution. Repositories that met these criteria included Figshare, Dryad and the Open Science Framework (OSF). Researchers did not list generalist repositories in their DMPs as often as domain or institutional repositories.

Institutional repositories serve as an archive for collecting, disseminating, and preserving the intellectual output (research data in this case) of a specific institution. Researchers did mention institutional data repositories frequently in their DMPs. Deep Blue Data, the University of Michigan Library's data repository, accounted for 85 of the 96 mentions of an institutional repository. Institutional data repositories were not as much of a factor for researchers applying for NIH funding, but for those submitting to the NSF, institutional repositories were selected more often than domain repositories. Institutional repositories were also a popular choice for researchers seeking funding from NASA. This may be due to the NSF and NASA serving a number of research communities that have not yet developed robust and open repositories of their own yet. It is important to note that institutions without

an institutional data repository, or without an organization like the library that provides services for research data, would likely see different results in their DMPs.

Researchers also listed means of sharing data outside of repositories in their DMPs. Roughly a third of DMPs, regardless of the funding agency, included some mention of sharing their data through presentations given at conferences or through publishing the results of their research. It was often difficult to discern what the researcher meant when stating that data sharing would take place through their presentations or publications. Some researchers appeared to believe that the tables, charts, and graphics summarizing the data in their presentations and publications would be sufficient. Others alluded to sharing the data behind their figures as separate supplemental files. Many simply did not provide enough information for me to determine what data would be shared through presentations and publications and to what extent.

Two additional methods of sharing data were regularly mentioned in DMPs: the researcher hosting and disseminating their data themselves or sharing the data when requested to do so by a person outside of the original research team. Each of these methods appeared in roughly a quarter of all DMPs. These methods allow the researcher a greater degree of control over what data are shared, when, and to whom. However, studies have shown that researchers do not always follow up on the promises made in the DMP to deliver their data to individuals who request it (Krawczyk & Reuben, 2021; Tedersoo, 2021). In addition, researchers who reported hosting the data themselves generally did not describe the steps they would take to ensure ongoing access to the data or the duration of access to the data they would ensure in their DMPs.

Many DMPs listed more than a single method of making their data available to others. For example, some made a distinction as to how they would share their data before and after publication of the results, such as including statements that the PI would share the data on request prior to publishing the results and then deposit the data into a repository as a part of the publication process. Other researchers referred to using more than one type of repository as a means to share their data.

Some researchers listed just a single method for sharing their data in their DMP. The number of times a particular method of sharing data was listed in the DMP as the only method to be used is presented in Table 5.

Table 5: DMPs Listing a Single Method of Data Sharing

	Total # of Mentions in DMPs N=461	Mentioned as Only Method N=461
Repository – Domain	162	68 (15%)
Repository – General	21	3 (1%)
Repository – Institutional	96	35 (8%)
Repository – Not Specified	25	8 (2%)
Researcher Hosted Website or GitHub	116	28 (6%)
In Presentations or Publications	134	39 (8%)
On Request	110	32 (7%)
Some or All of the Data will not be shared	11	6 (1%)

When Will the Data be Shared?

Researchers mentioned the timing of the expected release of the data in half or less than half of the DMPs, regardless of the funding agency. When included, statements on when the data would be shared clustered around two events: the lifespan of the grant or project, or the publication of the results. As shown in Table 6, the publication of the results was the primary event triggering the release of the data from the project, with most researchers stating that they would share their data upon acceptance, upon publication, or after publication of the results.

Table 6: Expected Timing of When the Data Would Be Made Publicly Available

	NIH n=233	NSF n=173	DoD n=11	NASA n=22	Other n=22	Total N=461
DMPs mentioning when data will be shared	109 (47%)	77 (45%)	4 (36%)	11 (50%)	8 (36%)	209 (45%)
Before or at Project's End	5	4	0	0	1	10
After Project Ends	19	15	0	1	2	37
Before Publication	6	4	0	1	1	12
On Acceptance or on Publication	49	18	2	6	2	77
After Publication	32	25	1	1	4	63
Other Time Frame	10	19	1	3	0	33

Note: The numbers in the column do not add up to the total listed in the top row as some of the DMPs listed more than one time frame for sharing components of their data.

Intellectual Property Concerns or Other Restrictions on Making the Data Available

In their DMPs, researchers often raised stipulations around sharing their data. These stipulations included provisions to ensure that the researcher and the University of Michigan would retain the rights necessary to file patents or otherwise retain the benefits derived from the outcomes of their research. Other considerations

centered on following federal and university policies, such as using a Material Transfer Agreement as the means to share data with others. Researchers also mentioned developing Data Sharing Agreements as a means for the researcher to weigh the merits of a request in deciding whether or not to share it with the requester and as a means of setting and enforcing terms and conditions for making the data available.

Table 7 displays the different IP or other concerns researchers had about sharing their data and in some cases, what steps they intend to take to address these concerns. Most often the steps described centered on adhering to established U-M research policies or practices, asserting their rights or the rights of the University over the data as a part of the research, or asserting more control over the sharing of their data than making it publicly accessible would normally permit.

Table 7: IP or Other Restrictions on Data Sharing Listed in DMPs

	NIH n=233	NSF n=173	DoD n=11	NASA n=22	Other n=22	Total N=461
# of DMPs that included statements on IP or conditions for sharing	151 (64%)	117 (64%)	9 (75%)	1 (5%)	13 (59%)	291 (63%)
Retaining Patents, Tech Transfer or Commercialization Rights	58 (25%)	27 (16%)	6 (67%)	0	2 (9%)	93 (20%)
U-M Policies	21 (9%)	30 (17%)	0	0	4 (18%)	55 (12%)
Policies of Other Institutions	29 (12%)	14 (8%)	1 (11%)	0	4 (18%)	48 (10%)
Limitations on Commercial Use	19 (8%)	10 (6%)	0	0	0	29 (6%)
Material Transfer Agreement or Simple Letter Required	76 (33%)	4 (2%)	2 (22%)	0	1 (5%)	83 (18%)
Data Sharing / Use Agreement, or Approval of PI	34 (15%)	21 (12%)	0	0	0	55 (12%)
PI or 3rd parties retain ownership (including statements that part or all of the data will not be shared)	20 (9%)	21 (12%)	4 (44%)	0	3 (14%)	48 (10%)
Use of Licenses	9 (4%)	26 (15%)	1 (11%)	1 (5%)	2 (9%)	39 (8%)
Prohibits Re-distribution	8 (3%)	10 (6%)	0	0	0	18 (4%)
Prohibits Creating Derivatives	4 (2%)	8 (5%)	0	0	0	12 (3%)
Requires Attribution of PI or Funding Agency	11 (5%)	30 (17%)	1 (11%)	0	4 (18%)	46 (10%)
Copyright or Publisher Policies Limiting Sharing	1 (0%)	13 (8%)	1 (11%)	0	1 (5%)	16 (3%)
Rights of Research Participants Stated	3 (1%)	2 (1%)	0	0	0	5 (1%)
No Restrictions on Sharing or Use of the Data Stated Explicitly	11 (5%)	39 (23%)	0	0	3 (14%)	53 (11%)

Note: Columns do not add up as a DMP may have included statements that fit into more than one of the listed categories.

Preserving Access to the Data

Researchers did not regularly include information about how long they would preserve access to their data. This is particularly true for DMPs developed for NIH grants, but even when preserving access to the data is explicitly included in funding agency guidance, as it is for the NSF, many researchers did not mention it in their DMPs.

Table 8: Preservation of Access

	NIH n=233	NSF n=173	DoD n=11	NASA n=22	Other n=22	Total N=461
# of DMPs that included a preservation of access statement	10 (4%)	84 (49%)	5 (45%)	3 (14%)	12 (55%)	114 (25%)
3 years or less	2 (2%)	26 (15%)	1 (9%)	0	2 (9%)	31 (7%)
Between 4 to 9 years	5 (2%)	28 (16%)	2 (18%)	0	3 (14%)	35 (8%)
10 years	2 (2%)	14 (8%)	2 (18%)	2 (9%)	4 (18%)	24 (5%)
More than 10 years, defined	0	2 (1%)	0	0	0	2 (0%)
Indefinitely or Undefined	2 (2%)	23 (13%)	0	1 (5%)	4 (18%)	30 (7%)

Note: Columns do not add up as a DMP may have included statements that fit into more than one of the listed categories.

The average duration of data preservation is rather short. The majority of these DMPs stated that the data only needed to be preserved for 10 years or less. This aligns with earlier findings that few researchers think about the long-term preservation of their data (Jahnke & Asher, 2012).

DISCUSSION

This analysis suggests that DMPs, as they are currently defined and implemented by funding agencies, are not as useful a tool for communicating detailed information about the anticipated data output to future data curators and stewards of those data as they could be. The literature review suggests that this finding is not limited to the University of Michigan, but likely holds true for DMPs more broadly. Nevertheless, the need for a means to communicate information about the data across stakeholders and the potential for DMPs to serve this purpose is great enough that funding agencies should not abandon them. Instead, funding agencies should better define, clarify, and support the role of the DMP and push to make DMPs from award grants more widely available to data curators and other service providers. Revisiting the content, accessibility, and the utility of DMPs as a tool to support data curation would realign them with the vision of the Interagency Working Group on Digital Data as a means to support ready access to high-quality research data.

In order for DMPs to fulfill this role, funding agencies would need to make significant adjustments in both the structure and content of the DMP. Based on the results of my analysis, I would make the following recommendations for improving the utility of DMPs:

Data

Despite their name, researchers did not always include much information about the data they expected to generate in their data management plans. In some ways, this is understandable as the DMP is included as a part of the larger grant proposal document. If the researcher describes their data as a part of their research narrative, it may feel redundant to provide this information again later in the same document. In addition, researchers may not have a complete understanding as to what information about their data should be provided as funding agency guidelines are often vague and researchers may not be used to sharing their data.

One way to address this would be to change the format of the DMP away from being a solely narrative based document to include a list of short answer questions, or asking researchers to select the best answer from a list of options. Asking for this information directly and making it easier to respond in the data management plan would likely increase the chances that researchers would provide this needed information. There are a number of data management planning tools, such as the DMPTool and DMPOnline, which employ a structured, form-based approach in assisting researchers in developing their DMPs. It is time for the funding agencies to follow suit and adopt an approach that would increase the likelihood of generating more useful and actionable information from researchers.

This should include asking researchers to provide the following information about their data in their DMPs, at minimum:

- *Data Type* – Different types of data entail different considerations for management, sharing and preservation. By identifying what types of data will be generated, even at a high level, data curators could provide relevant guidance on best practices and connect researchers to the services they might need in developing their data.
- *Formats* – Knowing the file formats researchers intend to use would help data curators consider issues around the use of proprietary formats, the structure and organization of the data, and what connections may be needed between the different components of the data.
- *Expected Data Size and the Number of Files* – Having a sense of how much data to expect and how many files will be included can help data curators plan for the eventual deposit of the data or help the research make alternative arrangements if their repository is not able to accommodate the data set.

It may be difficult for the researcher to feel like they could provide accurate information on some of the elements of their data set given that they create their data management plans before they generate data. However, even having a broad understanding of the likely characteristics of the data can provide valuable information to the data curator and help identify issues before they become larger problems. The new NIH data management and sharing plan and guidance that went into effect in January 2023 is a step in the right direction, asking researchers to describe their data in ways "*that address the type and amount/size of scientific data expected to be collected and used in the project*" (NIH, 2020). Ideally, researchers would treat the DMP as a living document and update information about the data set as it comes more into focus.

Metadata and Documentation

A critical element for ensuring that the data generated in a project will be findable, understandable, trusted and used by other is the amount and quality of metadata and documentation that accompany the data set. Many DMPs, however, did not include any information about metadata or documentation at all. Others simply mentioned that they would generate metadata or documentation without providing much, if any, detail about the nature of what they would provide.

A better approach for DMPs would be to frame the requirement around two questions. First, what information about your data would someone else need to understand it, trust it, and make use of it? Second, how will you document this information and provide it with your data set? The first question could be asked as an open-ended narrative and the second question could provide a list of preset options for the researcher to select from (lab notebook, readme file, codebook, etc.) with a write-in option. Questions about metadata and documentation standards would still be included as a part of the DMP. Reframing how documentation and metadata are presented to the researcher in DMP guidance could help prompt researchers to think more critically in their responses. Sharing more detailed information about the researchers' ideas and plans for communicating information about their data with data curators could help jump start discussions about how to capture and present contextual information about the data effectively.

Methods of Sharing Data

Although they varied in detail, the vast majority of the U-M DMPs included a statement on how researchers intend to share their data with others. Data repositories were listed by more than half of the DMPs, which may indicate that repositories are becoming more integrated into the research process. What remains unclear is researchers' actual understanding of the services and support provided by data repositories. In the follow up discussions that I have had with researchers who listed Deep Blue Data as their designated repository for their data, it was evident that some researchers had listed Deep Blue Data without really knowing what it was, how to use it, or how to prepare their data for deposit. The follow up conversations with researchers about Deep Blue Data and what services we offer have been invaluable.

Currently, DMPs are not readily available as they are incorporated into a larger grant proposal when they are submitted to a funding agency. This makes it very difficult for data curators and repository managers to access useful information or to communicate our services to researchers who need them. Machine actionable data management plans (maDMPs) are a promising development towards making the content of DMPs more accessible and actionable and there is some indication that funding agencies are considering their adoption (NSF, 2019). However, simply endorsing a maDMP protocol would not address the communication barriers between researcher and curators. In order to work effectively at scale, maDMP systems will have to consider how to connect to existing institutional systems and practices as seamlessly as possible. More importantly, for an institution to adopt a maDMP, it must demonstrate that it will add value to an institution's research program sufficient to justify the costs that the institution will incur in its adoption. Applying for grant funding, in addition to tracking the work and ensuring that researcher and institutional commitments are met over the course of the award, is time and labor intensive for an Office of Research. Personnel in the Office of Research are understandably wary of introducing new steps or systems that may slow down or add "extra work" to their processes. In addition, institutional grants management systems are generally closed to all but those who are actively engaged in carrying out the research and the grants manager assigned to manage the award. A maDMP system could potentially make more information available to data curators and local service providers, but without a culture change to allow institutional service providers and potentially non-institutional actors (such as data repositories) access to this information, the impact of the maDMP would be sharply curtailed.

Although it is not always included in the DMP, an indication of who will be responsible for managing the data and preparing it to be shared who be invaluable for data curators and should be listed. Inevitably, data curators will have questions about the data set and the needs of researchers who are submitting the data to the repository. Being able to ask these questions to the people who are directly responsible for developing the data set would save a lot of time and be more efficient. Having contact information would also enable the data curator to connect with the research team to share needed information about the repository, communicate the services offered, and initiate a relationship so that preparation and curation work can start early in the data life cycle.

Intellectual Property

The variety and depth of intellectual property concerns of researchers, as revealed in the analysis of U-M DMPs, is an area for further exploration. Some researchers face potentially competing interests between having to share their data and protecting their ability to commercialize the results of their research. These researchers often invoked university policies (U-M or other universities) in their DMPs as a means to assert some degree of authority and control over sharing the data, while still adhering to the requirements made by funding agencies. Statements such as "...any Intellectual Property Developed Within This Proposal Will Be Administered By The University Of Michigan", were common in the DMPs. The outputs of grant-funded projects have long been important revenue generators for universities and researchers, and university policies are the foundation for the systems and practices developed to ensure that this revenue comes to fruition. Referring to university policies may provide researchers with some justification for delaying or denying sharing some or all of the data they generate. It is not clear how federal agencies or researchers' home institutions interpret these statements, or what the role of the university should be in enforcing data sharing requirements.

It is also apparent from studying U-M's DMP that some researchers may not be completely comfortable with the loss of control over the interpretation and use of the data that comes with making it publicly available. Researchers included a variety of statements that indicated a desire to retain some degree of control over the data or to place conditions on what others can do with it. Many of these assertions were fairly modest, such as requiring individuals who make use of the data to cite project personnel or the funding agency who supported the research. Some researchers included explicit statements prohibiting specific uses or actions with their data, such as creating derivatives, redistributing the data, or using it for commercial purposes. Other researchers went further and stated that they would require a data sharing agreement or the review and approval of the principal investigator before making their data available to others. A smaller minority of researchers included statements in their DMPs that they or third parties retained ownership over the data and would limit access to part or all of the data to be used or generated in the project. There are undoubtedly situations in which limitations on the access and use of the data are necessary and I did not have sufficient information to determine what constituted justifiable reasons for limiting or denying others access to the data. However, from the results of the analysis, it does appear that acceptable norms and expectations on how much and which elements of the data being generated must be shared have not yet been fully developed by funding agencies or universities, particularly in situations where there is a likelihood that the outputs of research will be patentable or commercialized.

Researchers take their cues on sharing their data from their scholarly communities and peers, but they are not the only influences on the cultures and norms of research. University policy plays a large role in determining acceptable research practice and influencing behavior in sharing their data. Many U-M researchers referred to university policy in their DMPs. However, their statements often did not go beyond making a generic observation that the university has authority over how data sharing will (or will not) take place. Researchers provided few, if any, details on what they expected the university's role to be in supporting data sharing. This may be because the researchers themselves may not be aware of or fully understand their university's policy as it pertains to data, assuming their university has a data sharing policy at all. The University of Michigan did have an Institutional Data Resource Management Policy (U-M, 2008) when I conducted my analysis of DMPs, but it did not address managing, sharing, or preserving research data specifically. Instead, the policy listed research data together with administrative, clinical, and educational data, despite the different purposes, contexts, and uses of each. The lack of a policy on research data is not unusual. As of 2015, out of 206 universities, only 90 had some kind of policy specifically addressing research data at the university level (Briney et al., 2015). As a part of its Research Data Stewardship Initiative launched in 2022, U-M did develop a new policy that directly addresses research data which will go into effect in 2014 (U-M, 2023).

Institutions provide the resources and support needed to carry out research and they too have an impact on research practice through policies and provisioning. In addition to the library, researchers mentioned a variety of other institutional offices and service providers in their DMPs, such as Information Technology Services, the Office of Research, and the University's Tech Transfer Office. Although these offices are likely aware of each other and the services that they provide, at least at a high level, their relationships are likely to be informal and personality based. Furthermore, interviews with Research Integrity Officers indicated some uncertainties and confusion around who at the university is responsible for overseeing DMPs and ensuring that the data are shared as required by the funding agency (Bishop, et al., 2021). The lack of clarity around roles and responsibilities in supporting data sharing work makes it harder for researchers to get the support they need and complicates the process of demonstrating institutional compliance with federal mandates. Data curators, research administrators, and researchers themselves would benefit from stronger, more defined relationships with more visible connections between departments.

One way to develop and communicate norms and expectations around data sharing would be for university administration and other institutional stakeholders to create a shared institutional strategic plan for research data. Such a plan would clearly and concisely describe the university's policies for making research data publicly accessible and list the units on campus providing services and support for doing so. A well-defined and publicized strategic plan, if stakeholders developed, agreed to, and maintained it collectively, could have a real impact in reducing the burden on researchers from having to figure out how to share their data on their own.

Preserving Access to the Data

Guidance from funding agencies does not always require statements from researchers on how they intend to archive or preserve their data to ensure long-term access, but such information is critical for data curators to have in working with data. Specifically, the intended duration of preservation needed for the data is important for curators to understand so that they know when to weed out data in their collection, as data may lose their value over time. The costs of data preservation are such that it is unrealistic to expect that all data would be preserved indefinitely. The researchers generating the data will likely have a better sense of when the data no longer merit the effort and resources it will take to preserve them than the data curator and so ideally researchers would communicate this information in their DMP. The DMP could serve as a starting point for the researcher and data curators to determine what actions and expenditures would be reasonable to undertake to ensure that the data are accessible and usable over time.

Unfortunately, only 25% of the reviewed DMPs included a statement about the length of time the researcher expected the data to be preserved. The majority of those who listed a timeframe for the preservation of their data stated that their data only needed preservation for 10 years or less. It is not entirely clear why researchers listed relatively short timeframes for preserving their data, but researchers may be equating data preservation with data retention. Data retention is more focused on record keeping for regulatory obligations than data preservation, which is centered on ensuring the long-term access and usability of the data. In addition, researchers may not be considering the full ramifications of what not having access to their data would mean for their research impact. Without considering how long their data may have value to others and themselves, researchers may incorrectly assume that their data will be preserved "indefinitely". Asking researchers to consider the value of their data over time directly in their DMPs, and asking them some specific short answer questions in addition to open-ended narratives, would help identify where discussions may need to happen to set reasonable expectations.

Table 9: Summary of Key Findings

Element of the DMP	Finding	Recommendations
Types and Format of the Data	Limited or no information about the data to be generated making it difficult to know if data sharing requirements were addressed adequately	Replace the free text narrative with a more structured approach, such as a checklist or short answer questionnaire Explicitly ask for information about the data type, format and expected size of the data
Metadata and Documentation	Insufficient detail about the metadata or documentation to be provided which limits the utility of the data	Reframe DMP guidelines to center on the information needed to understand, trust and make use of the data
Methods of Sharing Data	Lack of understanding of the services provided by repositories and how to use them making the deposit process more difficult	Reframe the DMP as a vehicle for communications between researchers and repositories
Intellectual Property	Uncertainty about loss of control over the data and the potential consequences of making it publicly accessible	Clarify questions of ownership and responsibilities with research data in University policies Develop a strategic plan for supporting data sharing with appropriate boundaries at the institutional level
Duration of Preservation	Frequently left out of DMPs and likely equated with data retention requirements	More discussion and guidance in determining the likely value of the data over time given the costs of its preservation

CONCLUSION

The current iterations of data management and sharing plan requirements by funding agencies are not adequate in communicating information from researchers to data curators, local service providers or administrators about making data publicly accessible to others. Many of the DMPs analyzed in this study provided insufficient information about the types of data to be produced, the metadata and documentation to make it discoverable and useable by others, how and when the data would be shared, intellectual property concerns and the expected preservation duration, to be useful. This was true even when the funding agency asked for this information directly.

This study was limited to DMPs from a single institution for a period of twelve months. As reported in the literature review, earlier studies performed at other institutions produced similar results. A larger study conducted across multiple institutions, or one conducted over a longer period may produce a more nuanced understanding of how researchers understand and respond to data sharing requirements. Furthermore, expectations around data sharing continue to evolve, as evidenced by the NIH's recent expansion of their data sharing policy and the release of the 2021 Nelson memo from the Office of Science and Technology Policy recommending federal agencies update their public access policies on research data. At the University of Michigan, these and other developments have led to the creation of a university wide Research Data Stewardship Initiative (RDSI) to raise awareness of data sharing requirements and to connect researchers with services to support them in making their data publicly accessible. The RDSI is led by U-M's Office of the Vice President for Research and membership includes representatives from the Library, Information and Technology Services, Regulatory Affairs, Research Integrity and Innovation Partnerships, among others. RDSI's efforts have included workshops and other educational programming for

researchers on how to navigate the data sharing requirements of funding agencies and develop actionable DMPs. The RDSI also led a successful effort to develop an institutional policy to articulate the university's expectations and guidance for the stewardship of research data more clearly. Given the attention that data sharing has received at U-M recently, a content analysis study on DMPs might yield different findings if it were done today.

If data sharing mandates are to succeed, then current practices and structures to facilitate the data sharing process need to be reconsidered. Although they are flawed as currently designed, DMPs have tremendous potential to serve as a powerful communications tool between stakeholders in support of sharing data. Within an institution, DMPs could be used as a centralized source of information to create a shared understanding of the data sharing commitments made to the funding agency and to connect researchers to the support they would need to fulfill them. Applying the DMP in this manner would help clarify administrative roles and responsibilities and promote solid working relationships across campus. With the DMP as the centerpiece of a documentation and tracking process, institutions could collect and analyze data from institutional stakeholders regarding the effectiveness of the strategies and approaches taken to support data sharing mandates. Reframing the structure of DMPs and making them into more accessible, extensible, and actionable documents would better enable them to fulfill their original intended function of facilitating widespread public access to high-quality data sets.

Author's Note

Jake Carlson
University at Buffalo
University Libraries
Associate University Librarian for Research, Collections and Outreach

Correspondence concerning this article should be addressed to Jake Carlson, Associate University Librarian for Research, Collections and Outreach, University at Buffalo, Buffalo, NY 14260, jakecarl@buffalo.edu

REFERENCES

Bishoff, C., & Johnston L. (2015). Approaches to data sharing: An analysis of NSF data management plans from a large research university. *Journal of Librarianship and Scholarly Communication, 3*(2), eP1231. https://doi.org/10.7710/2162-3309.123

Bishop, W. B., Nobles, R., & Collier, H. (2021). Research Integrity Officers' responsibilities and perspectives on data management plan compliance and evaluation. *Journal of Research Administration, 52*(1), 76-101. https://www.srainternational.org/viewdocument/spring-2021

Briney, K., Goben, A., & Zilinski, L. (2015). Do you have an institutional data policy? A review of the current landscape of library data services and institutional data policies. *Journal of Librarianship and Scholarly Communication, 3*(2), eP1232. http://dx.doi.org/10.7710/2162-3309.1232

Burnette, M. H., Williams, S. C., & Imker, H. J. (2016). From plan to action: Successful data management plan implementation in a multidisciplinary project. *Journal of eScience Librarianship, 5*(1), e1101. https://doi.org/10.7191/jeslib.2016.1101

Chodacki, J., Hudson-Vitale, C., Meyers, N., Muilenburg, J., Praetzellis, M., Redd, K., Ruttenberg, J., Steen, K., Cutcher-Gershenfeld, J., & Gould, M. (2020). Implementing effective data practices: Stakeholder recommendations for collaborative research support. Washington, DC: Association of Research Libraries. https://doi.org/10.29242/report.effectivedatapractices2020

DCC. (2013). *Checklist for a Data Management Plan: v. 4.0.* Edinburgh: Digital Curation Centre. https://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_ Checklist_2013.pdf

Dietrich, D., Adamus, T., Miner, A., & Steinhart, G. (2012). De-mystifying the data management requirements of research funders. *Issues in Science and Technology Librarianship, 70.* https://doi.org/10.5062/F44M92G2

Holdren, John (2013). *Increasing access to the results of federally funded scientific research.* Memorandum from the Office of Science Technology Policy. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Hudson-Vitale, C., & Moulaison-Sandy, H. (2019). Data management plans: A review. *DESIDOC Journal of Library and Information Technology, 39*(6), 322-328. https://doi.org/10.14429/djlit.39.6.15086

Imker, H. (2017). Overlooked and overrated data sharing: Why some scientists are confused and/or dismissive. In L. R. Johnston (Ed.), *Curating research data, volume one: Practical strategies for your digital repository* (pp. 127-150). Association of College and Research Libraries (ACRL). http://hdl.handle.net/2142/95024

Interagency Working Group on Digital Data. (2009). *Harnessing the power of digital data for science and society: Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council.* Networking and Information Technology Research and Development (NITRD). https://www.nitrd.gov/pubs/Report_on_Digital_Data_2009.pdf

Inter-university Consortium for Political & Social Research (ICPSR). (2012). *Guidelines for effective data management plans.* https://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf

Janke, L., & Asher, A. (2012). The problem of data. *CLIR Publication #154.* Council on Library and Information Resources. https://www.clir.org/pubs/reports/pub154/problem-of-data/

Karimova, Y., Ribeiro, C., & David, G. (2021). Institutional support for data management plans: Five case studies. In E. Garoufallou and M. A. Ovalle-Perandones (Eds.), *Metadata and semantic research. MTSR 2020. Communications in Computer and Information Science*, 1355. Springer. https://doi.org/10.1007/978-3-030-71903-6_29

Krawczyk, M., & Reuben, E. (2012). (Un)Available upon request: Field experiment on researchers' willingness to share supplementary materials. *Accountability in Research, 19*(3), 175-186. http://doi.org/10.1080/08989621.2012.678688.

Michener, W. K. (2015). Ten simple rules for creating a good data management plan. *PLoS Computational Biology, 11*(10), e1004525. https://doi.org/10.1371/journal.pcbi.10045252015

Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLoS Computational Biology, 15*(3), e1006750. https://doi.org/10.1371/journal.pcbi.1006750

Mischo, W. H., Schlembach, M. C., & O'Donnell, M. N. (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 Institutes of Health. (2020). *Supplemental information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan* (NOT-OD-21-014). National Institutes of Health. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-014.html

National Science Foundation. (2015). *Public access plan: Today's data, tomorrow's discoveries: Increasing access to the results of research funded by the National Science Foundation* (No. nsf15052; p. 6-7). National Science Foundation. https://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf

National Science Foundation. (2019). *Dear Colleagues Letter: Effective practices for data*. https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp

National Science Foundation. (2020). *Proposal & Award Policies & Procedures Guide, Chapter II proposal preparation instructions*. https://www.nsf.gov/pubs/policydocs/pappg20_1/pappg_2.jsp#IIC2j

Parham, S. W., Carlson, J., Hswe, P., Westra, B., & Whitmire, A. (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

Sims, S. R., & Jones, S. (2017). Next-generation data management plans: Global, machine-actionable, FAIR. *International Journal of Digital Curation, 12*(1) 36-45. https://doi:10.2218/ijdc.v12i1.513

Steinhart, G., Chen, E., Arguillas, F., Dietrich, D., & Kramer, S. (2012). Prepared to plan? A snapshot of researcher readiness to address data management planning requirements. *Journal of eScience Librarianship, 1*(2), e1008. https://doi.org/10.7191/jeslib.2012.1008

Stodden, V., Ferrini, V., Gabanyi, M., Lehnert, K., Morton, J., & Berman, H. (2019). Open access to research artifacts: Implementing the next generation data management plan. *Proceedings of the Association for Information Science and Technology, 56*, 481-485. https://doi.org/10.1002/pra2.51

Tedersoo, L., Küngas, R., Oras, E., Koster, K., Eenmaa, H., Leijen, A., Pedaste, M., Raju, M., Astapova, A., Lukner, H., & Kogermann, K. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8, 192. https://doi.org/10.1038/s41597-021-00981-0

University of Michigan. (2008). Institutional Data Resource Management Policy: Standard practice guides - University of Michigan. *UM Standard Practice Guides*. https://spg.umich.edu/policy/601.12

University of Michigan. (2023). Research Data Stewardship Policy: Standard practice guides - University of Michigan. *UM Standard Practice Guides*. https://spg.umich.edu/policy/303.06

Van Tuyl, S., & Whitmire A. L. (2016). Water, water, everywhere: Defining and assessing data sharing in academia. *PLoS ONE, 11*(2), e0147942. https://doi.org/10.1371/journal.pone.0147942

Williams, M., Bagwell, J., & Nahm Zozus, M. (2017). Data management plans: The missing perspective. *Journal of Biomedical Informatics, 71*, 130–142. https://doi.org/10.1016/j.jbi.2017.05.004