122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Paper ID #12072

Analyzing Data Management Plans: Where Librarians Can Make a Difference

Sara M. Samuel, University of Michigan

Sara is an engineering librarian at the University of Michigan University Library where she is a liaison to the Electrical Engineering and Computer Science Department and the Atmospheric, Oceanic and Space Sciences Department. Sara has a bachelor of arts from Hope College and a Master of Science in Information from the University of Michigan.

Mr. Paul F. Grochowski, University of Michigan

Paul Grochowski is an engineering librarian at the University of Michigan.

Ms. Leena N Lalwani, University of Michigan

Leena Lalwani is an Engineering librarian and the Coordinator for Engineering Collection at the Art, Architecture and Engineering Library (AAEL) at the University of Michigan. She is also the liaison Librarian for Biomedical Engineering, Chemical Engineering, Materials Science, Naval Architecture and Marine Engineering and Entrepreneurship. Leena has been a librarian at University of Michigan since 1995 in various ranks. Prior to joining University of Michigan, Leena has worked as Librarian at Gelman Sciences and American Tobacco Company. Leena has a M.L.S. degree from Catholic University of America and M.S. in Chemistry from the University of Mumbai.

Mr. Jake Carlson, University of Michigan Library

Jake Carlson is the research data services manager at the University of Michigan Library. In this role, he explores the application of the theories, principles, and practices of library science beyond the domain of traditional library work. In particular, Carlson seeks to increase the Library's capabilities and opportunities to provide services supporting data-related research. Much of his work is done through direct collaborations and partnerships with research faculty. Carlson is one of the architects of the Data Curation Profiles Toolkit (http://datacurationprofiles.org) developed by Purdue University and the University of Illinois at Urbana-Champaign and is the principal investigator of the Data Information Literacy project (http://datainfolit.org), a collaboration between Purdue University, Cornell University, the University of Minnesota, and the University of Oregon.

Analyzing Data Management Plans: Where Librarians Can Make a Difference

Abstract

Since January 18, 2011, any researcher applying to a National Science Foundation (NSF) grant must include a data management plan (DMP) in their proposal. Many librarians have responded to this mandate by establishing new data-related services. One potential area for engaging with researchers in the grant proposal process is offering a DMP review service. In preparation for offering a DMP review service, engineering librarians at the University of Michigan (U-M) reviewed twenty-nine DMPs that were part of successful NSF grant proposals accepted in early 2014. The librarians analyzed the DMPs using three different sets of criteria: two rubrics developed at U-M, and a rubric currently under development by the IMLS funded Data management plan as a Research Tool (DART) project. The librarians had access to this third set of criteria as its trial run.

Analysis of the DMPs shows that the overall quality of DMPs at U-M varies greatly. Some common weaknesses in the DMPs are: lack of roles and responsibilities; lack of metadata standards that will be used; and failure to mention intellectual property rights. Analysis of the DMPs also revealed gaps in the librarians' knowledge of DMP requirements. In addition to discussing the findings from this current set of analyses, overall DMP quality from this study is compared to DMP quality found in a similar analysis of engineering DMPs from 2013. Looking toward a future where the outcome of grant proposals may be more dependent on the quality of the DMP, this analysis gives the engineering librarians at U-M a foundation for creating a DMP service in the coming year, and can inform other librarians who wish to develop a similar service at their institution.

Introduction

In 2011, the National Science Foundation (NSF) began requiring researchers to include a data management plan (DMP) as a part of their submitted proposals for funding.[1] As defined by the NSF, a DMP should include:

- a description of the data being developed,
- the standards that will be employed in formatting and developing the content of the data and metadata produced,
- policies for accessing and sharing the data with others,
- statements on how the data may be re-used, re-distributed or used to produce derivatives, and
- how the data will be archived to ensure access in the long-term.[2]

Some of the directorates within the NSF have produced additional requirements or gone into more detail about what they expect to see in DMPs. The NSF Directorate for Engineering, for example, states that a DMP should "outline the rights and obligations of all parties as to their roles and responsibilities in the managing and retention of research data."[3]

Even before the requirements were enacted, librarians were discussing the impact of data management plans and how libraries could respond through offering consultation and other services.[4] Librarians recognized early on that developing and implementing DMPs would likely prove to be a challenge for most researchers, an assumption confirmed by the Cornell University Library in their investigation of faculty reactions to these requirements in 2012.[5] Delserone notes that "there is growing recognition, within the library profession and the scientific community as well, that our combined knowledge and skills may be valuable to this [data sharing] enterprise."[6] In response, many academic libraries launched consultation services to work with researchers to help them craft data management plans that both met the requirements of funding agencies and were tailored to the needs of the researchers. Librarians have also developed workshops and other educational programming to help researchers better understand what constitutes a strong DMP and to increase their capabilities to produce one.[7]

It has been four years since the data management plan requirement was instituted by the NSF. Do researchers now have a better understanding of the data management requirements and how to respond to them? In this paper, the content of twenty-nine DMPs is reviewed and the quality of these DMPs is compared to DMPs from a previous study in 2013. The aims of this research are to identify areas of growth in researchers' collective understanding of the DMP requirements as evidenced by the DMPs they have written, and to determine particular areas of need to inform how data services offered by the library may be improved.

Literature Review

Since the advent of the data management plan requirement from NSF, libraries and librarians have been actively exploring their place in the data management plan landscape: "Increasing attention to data management plans can be most readily attributed to their requirement by funders."[8] Dietrich et al. found that guidance for the data management plans from funding agencies ranges from being detailed to ambiguous, so there is an opportunity for librarians to help.[9] Federer writes that "[m]any researchers do not have formal training or expertise in data management, making it difficult for them to meet funder requirements for data sharing."[10] While examining an embedded medical informationist's involvement with a research team, Federer notes that the research team asked the informationist for assistance with writing a DMP.[10] At Colorado State University, librarians conducted focus groups with faculty and researchers where

"a few participants indicated that they would appreciate assistance creating DMPs."[11] One participant thought that "all researchers would benefit from experts 'helping us do it right the first time.'"[11] In 2012, Diekema et al. surveyed 253 faculty about data services and practices and found that 37.7% of the respondents would like help with writing DMPs.[12]

This is a growing area within libraries. Antell et al. surveyed science librarians in 2014 and found that 15.8% of respondents had the job duty described as "help researchers develop data management plans."[13] Additionally, 15.1% indicated that they "'promote, publicize, or advocate' the library's data management services."[13] When asked what skills are needed, 4.2% of the survey respondents indicated that "experience assisting with data management plans" would be useful, but only 2% thought they had those skills or were working on acquiring them.[13]

One method of providing support for researchers is to make DMP templates. Participants in the focus groups at Colorado State University expressed "awareness or prior use of, and appreciation for, the CSU Libraries' data-management plan templates."[11] The University of Illinois Library also provided DMP templates for specific NSF directorates.[14] Librarians at the University of Houston created an online form for researchers to use when writing a DMP.[15]

Similar to a template, DMPTool has been a useful resource that librarians have used and recommended to faculty. DMPTool is an online resource that provides templates and takes users through the creation of DMPs step-by-step. The University of California, Los Angeles spearheaded the development of DMPTool and it "has had significant use among the UC campuses."[16] Johnston et al. notes that feedback from the workshop attendees at the University of Minnesota encouraged the libraries "to increase promotion of its free DMP consultation services and highlight its involvement with DMPTool."[7] James Madison University also decided to join DMPTool to help their researchers navigate DMP requirements.[17] A similar resource, DMPOnline, was developed by the Digital Curation Centre (DCC) to support researchers in meeting the requirements of funding agencies primarily in the United Kingdom.[18]

Only a couple studies have been published that have evaluated the contents of DMPs. Mischo et al. did an analysis of storage venues and reuse mechanisms mentioned in 1,260 NSF DMPs from all subject areas at the University of Illinois. This study "found that there were no statistically significant differences between the specific storage venues and reuse mechanisms within funded and unfunded proposals."[14] Nicholls et al. looked at DMPs from funded NSF proposals from the University of Michigan's College of Engineering and did an analysis using the NSF Directorate for Engineering's guidance which included roles and responsibilities, period of data retention, expected data, data formats and dissemination, and data storage and preservation of access.[19] Both of these studies indicate a need for data education for both faculty and graduate students.

Consulting with faculty about their data management plans has been a leading initiative in developing data services in academic libraries, but this is by no means the only type of data service being offered or considered. Johns Hopkins University, for example, offers services to support data management and curation across the lifecycle of a data set, from consulting on DMPs to archiving the data using an institutional repository.[20] Purdue University has also developed an institutional repository that is dedicated solely to disseminating and curating research data[21] and have engaged in direct partnerships with faculty to address their distinct data needs.[22, 23] Several libraries are now offering data information literacy programming designed to teach students how to manage and curate the data they produce more effectively.[24, 25] As Akers reminds us, not all faculty will serve as principal investigators (PIs) on grants that require DMPs, and therefore libraries ought not to focus all of their efforts into data management planning.[26] Nevertheless, analyzing DMPs can be a good starting place in developing an understanding of the data needs of researchers. This understanding can then be applied directly towards improving DMP consultation services as well as identifying and informing other potential data services.

Looking forward, more research funders will likely require a DMP as part of a grant application, so writing DMPs will increasingly be part of a researcher's grant application process. In 2014, the U.S. Department of Energy announced that it would begin requiring DMPs on grant applications.[27] Halbert wrote about the NSF mandate as "...neither unprecedented nor an isolated intervention by one federal agency...for example, the National Endowment for the Humanities adopted a requirement for data management plans…"[28] Kozlowski reported on a panel presentation where speakers shared updates on agencies' responses to the Public Access to Research Results (PARR) memo.[29] Gherghina and Katsanidou wrote about data management requirements in Europe: "Since 2011, the Economic and Social Research Council, the main funding body of the United Kingdom, requires all funding proposals to include a data management plan…"[30] Diekema et al. studied the effect of funders mandating data management plans and found that the mandates caused changes in services at some sponsored programs offices, has impacted some researchers' data management practices, and found that survey respondents felt that "the role of central IT and the library has increased" with regard to helping with data.[12] As more funders are requiring DMPs, it is clear that the library can provide assistance.

Setting

This study takes place at the University of Michigan, a large research university with a strong library system. The university consists of nineteen schools and colleges, one of which is the College of Engineering. The College of Engineering has a large and diverse student and faculty body:

Tenure-Track and Tenure Faculty (Fall 2013)	381
Research Faculty (Fall 2013)	127
Masters & Doctoral Students (Fall 2014)	3,169
Undergraduate Students (Fall 2014)	6,204

The College of Engineering received over $162 million in federal funding in 2014, of which 20% ($33 million) was from the National Science Foundation.

The library at the University of Michigan has been building up support services around data for several years now. Spatial and Numeric Data Librarians were hired beginning in 2005 and have an established space for helping patrons find and use data.[31] CLIR (Council on Library and Information Resources) Data Fellows were on staff from 2012 - 2014 and conducted background research and helped lead the development of internal training for liaison librarians to learn more about data.[31] In 2014, a Research Data Services Manager was hired to more formally support the increasing data needs of researchers, and to support the liaison librarians whose faculty work with data.

In 2013, the authors conducted a rough review of data management plans[19] that were attached to accepted NSF grant applications to learn:

- more about the DMP requirements
- how well the researchers were addressing those requirements
- what the areas of need were

Now that data management plans have been required by the NSF for several years and the researchers have become accustomed to the requirement, it was time for another, more formal, review.

Method

This more formal review consisted of reviewing twenty-nine DMPs. These DMPs were acquired from the Office of Research in the College of Engineering and were from grant applications that were accepted by NSF between January and June 2014. The DMPs were written by researchers from a wide variety of disciplines within the College of Engineering.

The DMPs were evaluated using three different rubrics: the College of Engineering (CoE) Rubric, the University Rubric, and the Data management plan as a Research Tool (DART) Rubric.

CoE Rubric

The CoE Rubric was the first rubric used to evaluate the first set of DMPs from the College of Engineering. This rubric consists of eight criteria that were established based on the NSF Directorate for Engineering DMP guidance. To apply the CoE Rubric, librarians provided notes under each criterion to illustrate how the DMP met the criterion. If the DMP said nothing about a required criterion, it was coded as "no mention." In determining overall quality of a DMP, this rubric used a binary system. For each of eight criteria, the DMP was assigned 0 (did not meet, coded as "no mention") or 1 (met), based on the librarian's notes. These numbers were totaled, and overall DMP quality was determined on the scale: 0-2 (poor); 3-5 (fair); 6-8 (good). The eight criteria were: Roles and Responsibilities; Types and Formats; Storage; Size of Data; Documentation and Metadata; Dissemination/Provision for Re-use; Archiving and Preservation; and Retention. The CoE Rubric is included as Appendix A.

University Rubric

The University Rubric was developed by the CLIR Data Fellows for evaluating DMPs that were written by researchers from a different college within the university. The rubric consists of a thorough set of questions that cover all aspects of data collection and management that could potentially be included in a DMP, beyond even the requirements listed by NSF directorates. This rubric was used in addition to the CoE Rubric since it covers requirements beyond the Directorate for Engineering guidance and better facilitates comparisons between DMPs by using a very defined set of outputs. The majority of the criteria use a binary system for indicating if the piece of information is included in the DMP (For example, the responses to "Is the total amount of expected data specified?" are yes/no), but there are some that have more possibilities. For example, "Who will be responsible for data management?" has four potential answers: PI and/or co-PIs, Trainees (graduate students, postdocs or technicians), More than one of the above, and Not clear. The questions asked in the University Rubric are listed in Appendix B.

DART Rubric

The Data management plan as A Research Tool (DART) project[32] seeks to evaluate the content of data management plans as a means to inform the development of library services in managing and curating research data. With support from the Institute of Museum and Library Services (IMLS), librarians from Oregon State University, the University of Oregon, Penn State University, the George Institute of Technology and the University of Michigan are developing and testing a rubric for this purpose. In addition to the use of this rubric by individual librarians as a means of determining how closely a researcher has complied with the requirements issued by an NSF directorate, the rubric will enable standardized evaluations of DMPs across multiple institutions. Thus, the library community will have a tool that will produce meaningful

comparisons that could lay the groundwork for identifying common issues and creating best practices to address them.

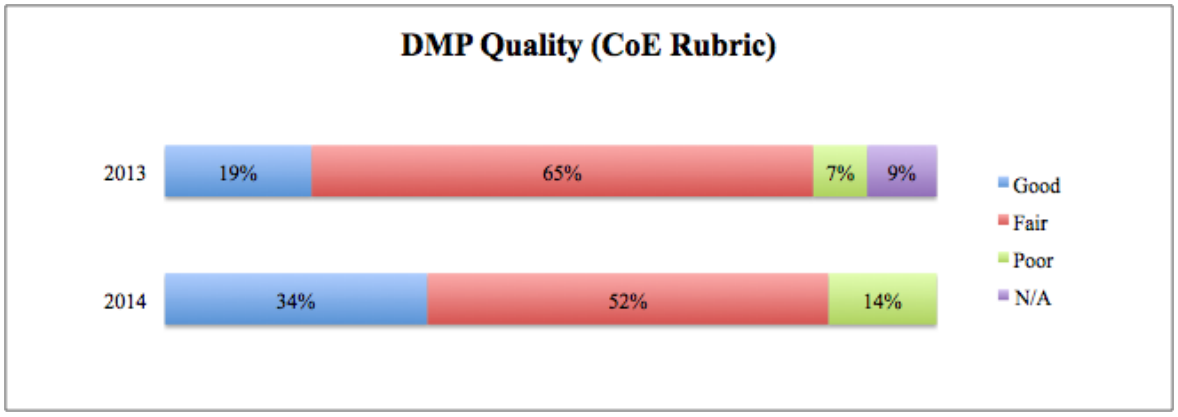
This study made use of an early iteration of the DART Rubric and served as a beta test of its effectiveness. The results produced by this study and the experiences of the librarians who made use of the DART Rubric were used to further the development of the rubric. It provided an alternate method of evaluating the DMPs that differed from the CoE Rubric and University Rubric by using a three-tiered method of evaluation.

The DART Rubric used in this study consisted of twelve items that matched common DMP requirements mentioned by all NSF directorates. Each item was rated on a three-point scale - low, medium, or high - depending on the level of detail provided. "Low" means there is no information fulfilling the requirement, "Medium" indicates there is little and/or vague information provided, and "High" indicates that the DMP clearly meets the requirement. The early iteration of the DART Rubric is available as Appendix C.

The librarians reviewed the DMPs using all three rubrics, each of which contributed to a more complete understanding of the content of the DMPs. The librarians reviewed DMPs that were written by researchers in their assigned liaison areas - two librarians reviewed eleven DMPs each, and the other librarian reviewed seven DMPs.

Results

One measure of the CoE Rubric, as discussed above, was the overall quality of the DMP. This rubric was originally used in an earlier study of a larger set of DMPs.[19] The authors of this study were interested in comparing overall quality of DMPs between the two studies to determine if there was any change in quality since the first study. Findings are shown below.
N/A refers to documents that stated a DMP was not needed due to the nature of the project.



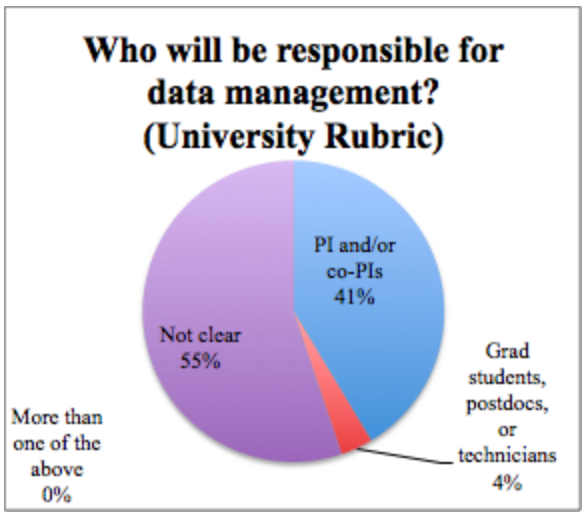
Looking at the overall quality of the DMPs, the DMPs reviewed for this study were slightly better than the DMPs reviewed in 2013. Although only an additional 2% of DMPs in this study were considered "good" or "fair," there was a marked increase in "good" DMPs, from 19% in the original analysis to 34% in this analysis. Only one DMP reviewed during this current analysis rated as high as it could on the CoE Rubric, scoring an 8 on the 0-8 scale.

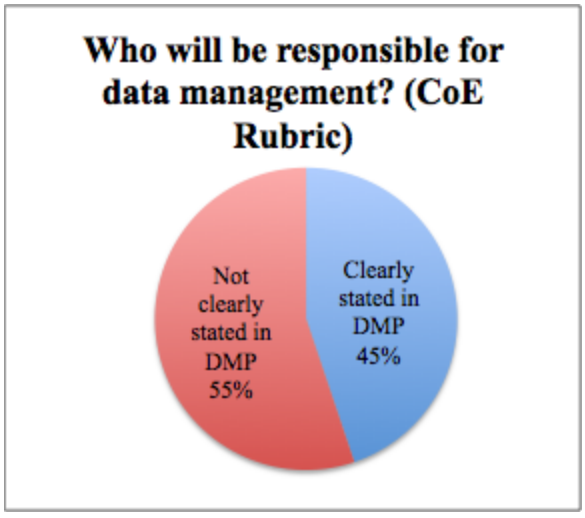
Although the DART Rubric did not include an overall quality score, information can be inferred by tallying the number of "High" and "Low" marks for each DMP. Only two DMPs had zero "Low" scores across all twelve areas, and only three more had just one "Low" score. None of the DMPs were rated "High" across all twelve categories.

The University Rubric did not include an overall quality score.

Roles & Responsibilities

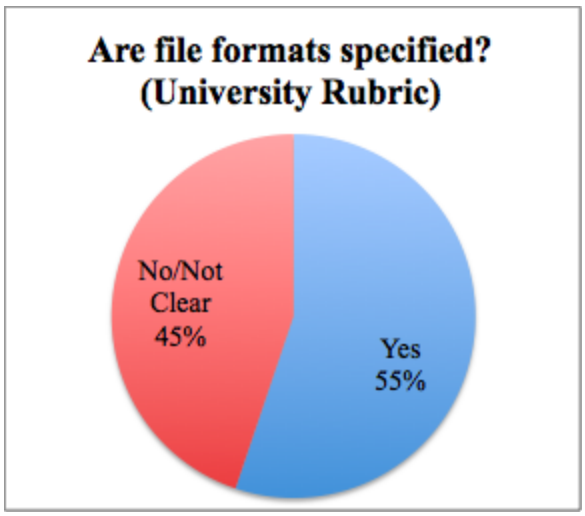
One of the first items looked for in the CoE Rubric and University Rubric was whether a specific person was listed as being responsible for the data produced by the research project. As expected, when a responsible person was mentioned it was usually the PI or co-PI, but it was also interesting that in 45% of the DMPs, it was not made clear who was in charge of the data.

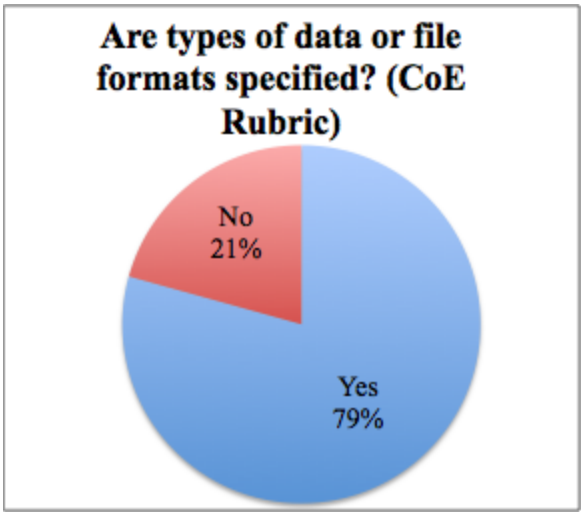




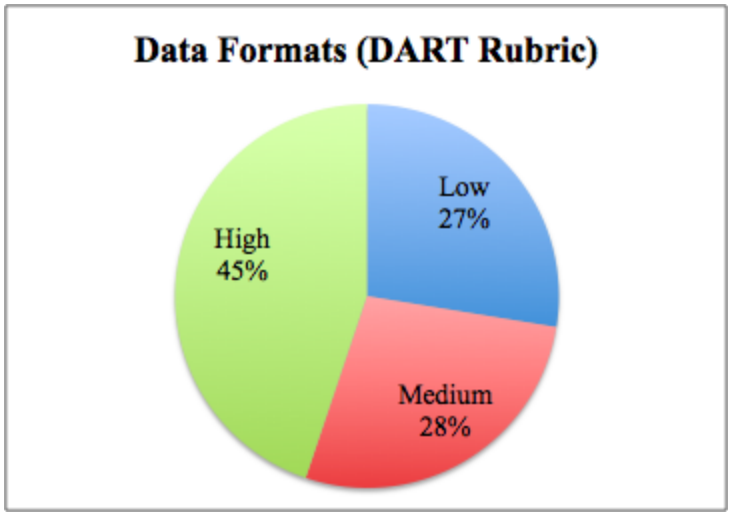
File Formats

Another area that was interesting to look at was whether or not specific file formats were listed in the DMP. File formats refers to specific programs or file extensions. Note that the CoE Rubric's criteria looked for both file formats and/or data types (referring to a mention of general types such as images, spreadsheets, software, etc.). In total, 55% of the DMPs specifically mentioned file formats or a program that was going to be used to capture data, while 79% met the broader category of mentioning either or both file formats and data types.





The DART Rubric helps paint another picture regarding data formats and how detailed the description was. This three-tiered evaluation helps capture a bit more detail than the University Rubric does with its two evaluation options, and it helps to discern that although some DMPs mention data formats, they are not as clear as they could be.



Low	No information
Medium	Little and/or vague information
High	Clearly meets requirement

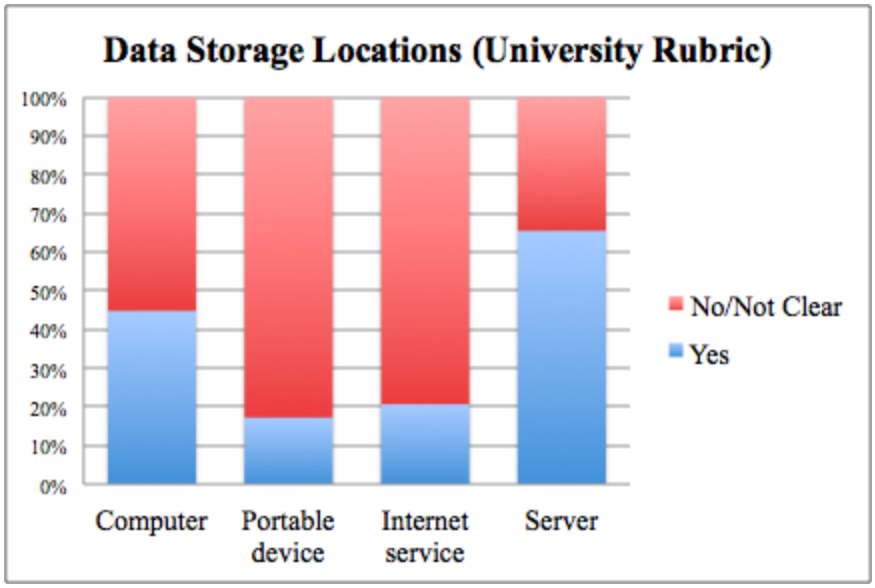
Amount of Data

Although NSF guidance does not require including an estimated amount of data, the authors consider the amount of data to be an important element to consider when drafting a DMP. The NSF guidance does require discussion of plans for preservation and storage of data, and one

might conclude that considering the amount of data is an important step to take before planning for data storage. Only 7% of the DMPs reviewed explicitly stated how much data the research project was expected to generate.

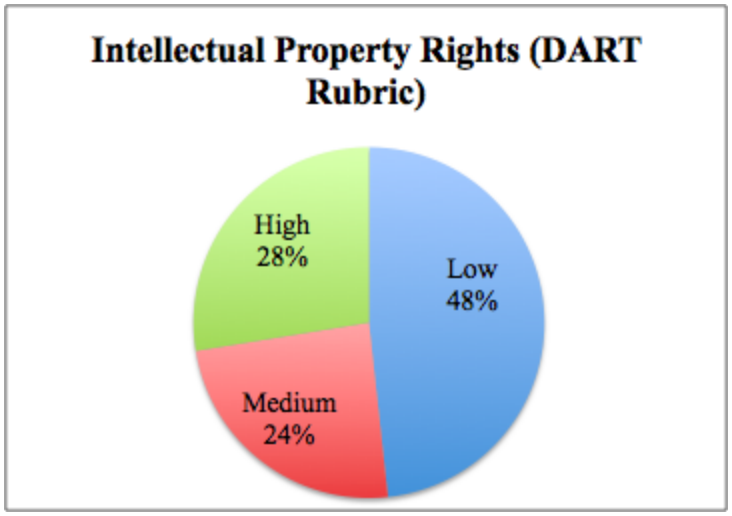
Data Storage

Another area of importance in a DMP is what storage options will be used for the data. The DMPs were reviewed to see if each one mentioned one or more of the storage option categories using the University Rubric. Note that there could be overlap if a DMP mentioned using both a computer and a server for data storage. Most of the DMPs, 19 out of 29, indicated using a department-run or university-run server. The next most popular option was the use of a laptop or desktop computer to store data. Only four of the DMPs did not have at least one storage option explicitly mentioned.



Intellectual Property

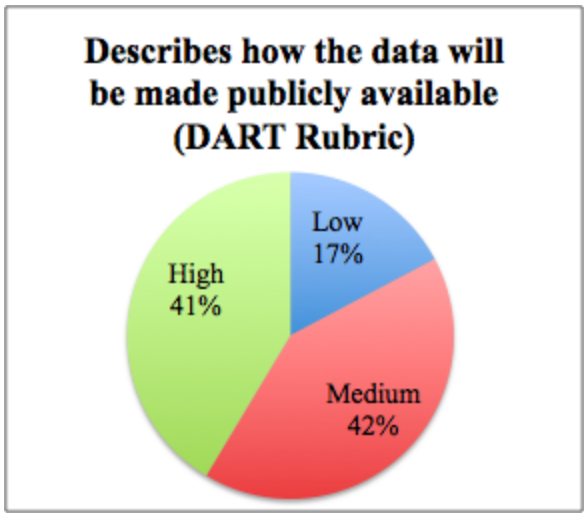
Intellectual property rights are important when sharing data. The DART Rubric asks if a DMP "Describes what intellectual property rights to the data and supporting materials will be given to the public and which will be retained by project personnel (if any)." Almost half of the DMPs neglected to mention any sort of rights to the data.

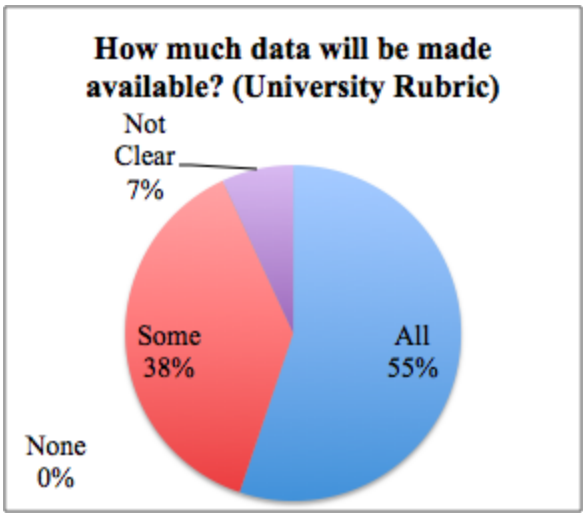


Low	No information
Medium	Little and/or vague information
High	Clearly meets requirement

Data Sharing

Although the general DMP guidance from NSF states that the DMP "should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results,"[24] there were some DMPs (17%) that did not adequately describe how the researchers were planning to make their data publicly available. However, it is encouraging to note that none of the DMPs stated that the data would not be available after the project, as was teased out by using the University Rubric.

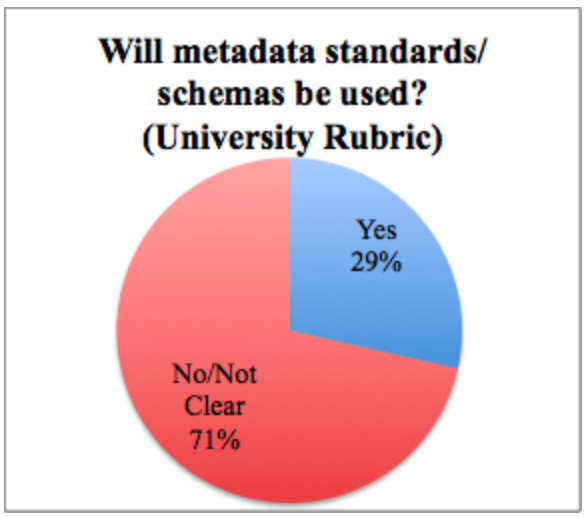


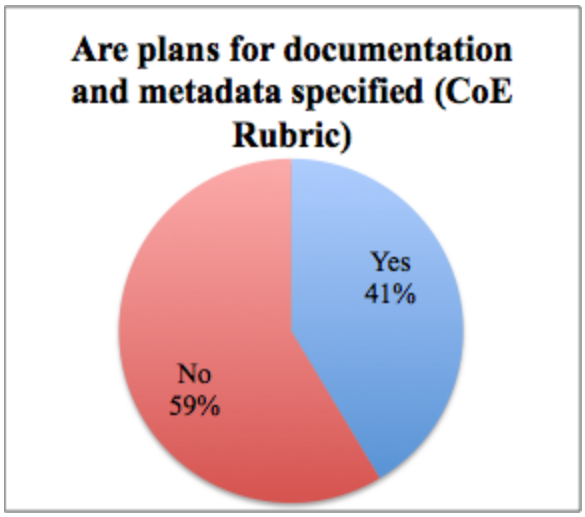


Low	No information
Medium	Little and/or vague information
High	Clearly meets requirement

Documentation and Metadata

If data are to be shared, future researchers need appropriate documentation and metadata to make use of the data. The different rubrics ask slightly different questions about documentation and metadata, with these results:





Discussion

It is encouraging to see that a majority of DMPs are listed as "fair" or better using the CoE Rubric, and that the quality of the DMPs has improved slightly over time. It is also good that researchers are thinking about where they will be storing their projects' data before the project is even funded. Most encouraging is that none of the DMPs stated that they would not be sharing the data from their project. One of the goals of the NSF is to encourage researchers to share their data, so this is a positive display of the culture around data sharing.

However, as mentioned previously, only one DMP rated as high as it could on the CoE Rubric, and none were rated "High" across all categories of the DART Rubric, so there is clearly room for improvement. Overall, most DMPs could easily be improved by including just a bit more detailed information about: who will be in charge of the data, what file formats will be produced or what programs will be used in generating the data, and the intellectual property rights of the resulting data.

The DMP can be more than just a document that satisfies an NSF proposal requirement. Three faculty at the University of Houston reported in a survey that they have a data management plan in place because "it was just good research practice to do so, that it helped them stay organized, and that data are simply too valuable not to manage."[15] A DMP can be used in the lab to help guide new researchers and graduate students, and it can give PIs a document that they can refer

to when questions arise about who is in charge or where data should be stored. Although naming a specific person to be in charge of the data is not required by all directorates of the NSF, it is good practice to do so, and it will help make sure that someone is designated in that role up front.

Not all NSF directorates require the DMP to state the volume of data a project is expected to produce, but it is a beneficial exercise for a PI to consider this before beginning the project. This information can help determine possible costs for storage and maintenance of the data, which can inform the project's overall budget.

Additionally, it is important that researchers clearly indicate their IP rights to the data. If the data are going to be made available, users of the data need to know how to credit the data creator and if they can alter the data in any way. Creative Commons[33] is a useful resource for this since their IP licenses provide specific details about how a work may be used.

Lack of metadata standards is also a weakness in many of these DMPs, but this may be more challenging to address in a DMP since researchers need to be familiar with metadata standards before they can write about them. Even without knowing about metadata standards, though, researchers could improve many of these DMPs if they commit to writing a document that describes the data.

Conclusion

Hamasu et al. report that at a discussion of twenty-two librarians, the consensus was that "training will be essential to apply their traditional knowledge and abilities in new ways"[34] and a useful training method is to have hands-on experience. Examining the contents of DMPs provided the authors the opportunity to better understand how faculty consider and approach planning for managing their data, and to identify areas of need where librarians could provide assistance. The librarians view the work done for this study as a part of their training in developing the library's data services.

A significant benefit of this study is that it gave the librarians the chance to become more acquainted with DMPs and the NSF's DMP requirements. Although the experience provided a lot of information, there are several areas that librarians could continue to study in order to increase their knowledge. One of those areas is the specific NSF directorates' DMP requirements. There are a number of different guidance documents, and although the librarians have a general understanding of what should be in a DMP, what each directorate wants in a DMP differs. The librarians wish to be familiar with the different requirements so that they can give accurate feedback to DMP authors.

There are three different paths for librarians to help researchers at U-M improve their DMPs: DMP workshops will continue to be offered, a DMP Review Service will be established, and librarians will begin working with departments to offer data education options.

A workshop was developed in early 2014 to help educate the researchers about the DMPs, and was successfully offered four times so far, reaching seventy researchers. DMP workshops will continue to be offered, but the content will be altered to specifically address the weak areas that were noted in this study. The workshops will continue to educate attendees about the NSF DMP requirements and will include information about why certain information, like intellectual property rights, is beneficial to include in a DMP.

A DMP Review Service pilot was launched in February 2015 to provide feedback to researchers on their NSF DMPs before the researchers submit their grant proposals. This service is a collaboration between the University Library, the Office of the Associate Dean for Research, and IT staff. A memorandum of understanding between the three service units has been developed in which each unit has agreed to participate in the service at some level. Researchers will send DMPs by email and the service units will share comments on the plans with the researchers. Part of this service includes providing template statements about storage options offered at U-M so researchers are accurately describing where their data will be stored. The DMP Review Service will be reviewed after six months, and it may be modified in the future to include more guidance as other funding agencies begin to require data management plans.

A final path that librarians at U-M will pursue is to begin working with individual departments to develop specialized data education for researchers. Liaison librarians will connect with faculty and researchers in their departments to gather information about what the needs are and get feedback about the options for offering data education, whether it is a one-shot workshop, a series of seminars, or integration with a for-credit course. Doing this at the departmental level will help ensure that what is developed will be relevant and the faculty who provide feedback will be more apt to encourage their graduate students to attend.

With these interventions comes a potential for positive impact on the DMPs that are written with librarian assistance, whether the DMP author attends a workshop or submits a DMP for review. The impact may not be realized for several years, so a follow-up study will be done in the future to determine if the librarian efforts are helping to noticeably improve the DMPs being written by researchers at U-M.

In conclusion, this study was useful for identifying areas of need for improving DMPs and informing data services. Time will tell if the planned interventions make a difference in the content and quality of DMPs.

References

[1] National Science Foundation. (n.d.). *Dissemination and Sharing of Research Results*. Retrieved from http://www.nsf.gov/bfa/dias/policy/dmp.jsp

[2] National Science Foundation. (2011). *Grant Proposal Guide, Chapter II - Proposal Preparation Instructions*. Retrieved from: http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp

[3] National Science Foundation. (n.d.). *Data Management for Engineering Directorate Proposals and Awards*. Retrieved from http://nsf.gov/eng/general/ENG_DMP_Policy.pdf

[4] Gabridge, T. (2009). The Last Mile: Liaison Roles in Curating Science and Engineering Research Data. *Research Library Issues*, *265*(August), 15–21. Retrieved from http://old.arl.org/bm~doc/rli-265-gabridge.pdf

[5] 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), Article 1. http://dx.doi.org/10.7191/jeslib.2012.1008

[6] Delserone, L. M. (2008). At the Watershed: Preparing for Research Data Management and Stewardship at the University of Minnesota Libraries. *Library Trends, 57*(2), 202-10. http://dx.doi.org/10.1353/lib.0.0032

[7] Johnston, L., Lafferty, M., & Petsan, B. (2012). Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach. *Journal of eScience Librarianship, 1*(2), Article 2. http://dx.doi.org/10.7191/jeslib.2012.1012

[8] Krier, L., & Strasser, C. A. (2014). *Data Management for Libraries*. Chicago, IL: American Library Association.

[9] 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*(Summer). Retrieved from http://www.istl.org/12-summer/refereed1.html

[10] Federer, L. (2013). The Librarian as Research Informationist: A Case Study. *Journal of the Medical Library Association, 101*(4), 298-302. http://dx.doi.org/10.3163/1536-5050.101.4.011

[11] McLure, M., Level, A. V., Cranston, C. L., Oehlerts, B., & Culbertson, M. (2014). Data Curation: A Study of Researcher Practices and Needs. *portal : Libraries and the Academy, 14*(2), 139-64. http://dx.doi.org/10.1353/pla.2014.0009

[12] Diekema, A. R., Wesolek, A., & Walters, C. D. (2014). The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories. *The Journal of Academic Librarianship, 40*(3-4), 322-331. http://dx.doi.org/10.1016/j.acalib.2014.04.010

[13] Antell, K., Foote, J. B., Turner, J., & Shults, B. (2014). Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. *College & Research Libraries, 75*(4), 557-74. http://dx.doi.org/10.5860/crl.75.4.557

[14] 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), Article 3. http://dx.doi.org/10.7191/jeslib.2014.1060

[15] Peters, C., & Dryden, A. R. (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. http://dx.doi.org/10.1080/0194262X.2011.626340

[16] Starr, J., Willett, P., Federer, L., Horning, C., & Bergstrom, M. L. (2012). A Collaborative Framework for Data Management Services: The Experience of the University of California. *Journal of eScience Librarianship, 1*(2), Article 7. http://dx.doi.org/10.7191/jeslib.2012.1014

[17] Shorish, Y. (2012). Data Curation Is for Everyone! The Case for Master's and Baccalaureate Institutional Engagement with Data Curation. *Journal of Web Librarianship, 6*(4), 263-73. http://dx.doi.org/10.1080/19322909.2012.729394

[18] Donnelly, M., Jones, S., and Pattenden-Fail, J. W. (2010). DMP Online: The Digital Curation Centre's Web-Based Tool for Creating, Maintaining and Exporting Data Management Plans. *International Journal of Digital Curation, 5*(1), 187-193. http://dx.doi.org/10.2218/ijdc.v5i1.152

[19] Nicholls, N. H., Samuel, S. M., Lalwani, L. N., Grochowski, P. F., Green, J. A. (2014). Resources to Support Faculty Writing Data Management Plans: Lessons Learned from an Engineering Pilot. *International Journal of Digital Curation, 9*(1), 242-252. http://dx.doi.org/10.2218/ijdc.v9i1.315

[20] Shen, Y., & Varvel, V. E. (2013). Developing Data Management Services at the Johns Hopkins University. *The Journal of Academic Librarianship, 39*(6), 552-557. http://dx.doi.org/10.1016/j.acalib.2013.06.002

[21] Witt, M. (2012). Co-designing, Co-developing, and Co-implementing an Institutional Data Repository Service. *Journal of Library Administration, 52*(2), 172-188. http://dx.doi.org/10.1080/01930826.2012.655607

[22] Garritano, J. R., & Carlson, J. (2009). A Subject Librarian's Guide to Collaborating on e-Science Projects. *Issues in Science and Technology Librarianship, 57*(Spring). http://dx.doi.org/10.5062/F42B8VZ3

[23] Carlson, J., & Stowell-Bracke, M. (2013). Data Management and Sharing from the Perspective of Graduate Students: An Examination of the Culture and Practice at the Water Quality Field Station. *portal: Libraries and the Academy, 13*(4), 343-361. http://dx.doi.org/10.1353/pla.2013.0034

[24] Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *portal: Libraries and the Academy, 11*(2), 629-57. http://dx.doi.org/10.1353/pla.2011.0022

[25] Carlson, J., & Johnston, L. (Eds.). (2015). *Data Information Literacy: Librarians, Data and the Education of a New Generation of Researchers*. West Lafayette, IN: Purdue University Press.

[26] Akers, K. G. (2014). Going Beyond Data Management Planning: Comprehensive Research Data Services. *College and Research Libraries News, 75*(8), 435-436. Retrieved from http://crln.acrl.org/content/75/8/435

[27] [anonymous]. (2014). Energy Department to Require Digital Management Plans. *Information Outlook, 18*(4), 6.

[28] Halbert, M. (2013). Prospects for Research Data Management. In *Research Data Management: Principles, Practices, and Prospects.* Retrieved from http://www.clir.org/pubs/reports/pub160/pub160.pdf

[29] Kozlowski, W. (2014). Funding Agency Responses to Federal Requirements for Public Access to Research Results. *Bulletin of the Association for Information Science and Technology, 40*(6), 26-30. Retrieved from http://www.asis.org/Bulletin/Aug-14/AugSep14_Kozlowski.pdf

[30] Gherghina, S., & Katsanidou, A. (2013). Data Availability in Political Science Journals. *European Political Science, 12*(3), 333-349. http://dx.doi.org/10.1057/eps.2013.8

[31] Akers, K. G., Sferdean, F. C., Nicholls, N. H., & Green, J. A. (2014). Building Support for Research Data Management: Biographies of Eight Research Universities. *International Journal of Digital Curation, 9*(2), 171-191. http://dx.doi.org/10.2218/ijdc.v9i2.327

[32] Oregon State University. (n.d.). *Data Management Plans as A Research Tool*. Retrieved from http://dmpresearch.library.oregonstate.edu/

[33] Creative Commons. (n.d.). Retrieved from https://creativecommons.org/

[34] Hamasu, C., Jones, B., & Kelly, B. (2012). Discussing "eScience and the Evolution of Library Services." *Journal of eScience Librarianship, 1*(2), Article 5. http://dx.doi.org/10.7191/jeslib.2012.1015

Appendix A

Items in the College of Engineering (CoE) Rubric rated on a 2-point scale (Yes - No). Criteria listed in italics are criteria considered when assessing overall quality of DMPs.

- *Roles and responsibilities*
- *Data types and formats*
- *Data storage op1*
- Data storage op2 (if indicated)
- Data storage op3 (if indicated)
- *Size of data*
- *Data documentation and metadata*
- *Data dissemination/provision for data re-use/re-distribution*
- *Data retention*
- *Archiving and preservation*
- Data sensitivity/access limitation/privacy protection, confidentiality or security concerns (if indicated)
- Intellectual property rights (if indicated)

Appendix B

Questions asked in the University Rubric:

1. Roles and responsibilities for data management
1.1. Where will the data be managed and/or stored during the project?
1.2. Are the individual(s) responsible for data management specifically named (or referred to "the PI" or "co-PI")?
1.3. In general, who will be responsible for data management?

2. Types of data
2.1. Will the project generate/analyze quantitative data (e.g., spreadsheets, GIS data)?
2.2. Will the project generate/analyze qualitative data (e.g., text)?
2.3. Will the project generate/analyze multimedia data (e.g., images, videos, audio, etc.)?
2.4. Will the project generate software/code/scripts?
2.5. Will the project generate physical materials (i.e., samples, cell lines, etc.)?
2.6. Are file formats specified (e.g., file extensions, names of data collection software)?
2.7. Is the total amount of expected data (in terms of MB, GB, TB, etc.) specified?
2.8. What is the total amount of expected data?
2.9. Is the rate of expected data generation (in terms of MB/GB/TB per day, etc.) specified?
2.10. What is the rate of expected data generation?

3. Metadata and data documentation
3.1. Will metadata standards/schemas be used?
3.2. Are other data documentation strategies mentioned (e.g., readme files, codebooks, lab notebooks)?

4. Short-term data storage
4.1. Will data be stored on laptop or desktop computers?
4.2 Will data be stored on portable storage devices (e.g. external hard drives, USB drives, CDs, DVDs)?
4.3 Will data be stored on the internet (e.g., Box, etc.)?
4.4 Will data be stored on an institutional server?
4.5. Is a method or location of data backup/duplication specified?

5. Long-term data accessibility
5.1. Will data/code be made accessible after the project?
5.2. Why will data/code not be made accessible?
5.3. Will data/code be made available upon request?

5.4. Will data/code be posted on researchers' website or available via a personally-curated database?
5.5. Will data/code be submitted to journals as supplemental material?
5.6. Will data/code be submitted to a repository/archive?
5.7. Will our institutional repository be used for data preservation?
5.8. Will data/code be embargoed before release?
5.9. How long will data/code be embargoed before release?
5.10. Is length of data/code retention specified?
5.11. How long will data/code be retained?

Appendix C

Items rated in the DART Rubric (beta version from August 2014) on a 3-point scale (Low - Medium - High):

1. Data Types: Describes what types of data will be captured or created

2.1 Metadata: Identifies metadata standards or formats that will used for the proposed project
2.2 Formats: Describes data formats created or used during project

3.1 When: Provides details on when the data will be made publicly available
3.2 How: Describes how the data will be made publicly available
3.3 Privacy: If the data are deemed to be of a "sensitive" nature, describes what protections will be put into place to protect privacy or confidentiality of research subjects
3.4 IP: Describes what intellectual property rights to the data and supporting materials will be given to the public and which will be retained by project personnel (if any)
3.5 Security: Describes security measures that will be in place to protect the data from unauthorized access

4.1 Use/Reuse: Describes the policies in place governing the use and reuse of the data
4.2 Redistribution: Describes the policies for redistribution of the data
4.3 Derivatives: Describes policies for building off of the data, such as through the creation of derivatives

5. Archived: Describes whether or not the data will be archived