

Resources to Support Faculty Writing Data Management Plans: Lessons Learned from an Engineering Pilot

Natsuko H. Nicholls
University of Michigan

Sara M. Samuel
University of Michigan

Leena N. Lalwani
University of Michigan

Paul F. Grochowski
University of Michigan

Jennifer A. Green
University of Michigan

Abstract

Recent years have seen a growing emphasis on the need for improved management of research data. Academic libraries have begun to articulate the conceptual foundations, roles, and responsibilities involved in data management planning and implementation. This paper provides an overview of the Engineering data support pilot at the University of Michigan Library as part of developing new data services and infrastructure. Through this pilot project, a team of librarians had an opportunity to identify areas where the library can play a role in assisting researchers with data management, and has put forth proposals for immediate steps that the library can take in this regard. The paper summarizes key findings from a faculty survey and discusses lessons learned from an analysis of data management plans from accepted NSF proposals. A key feature of this Engineering pilot project was to ensure that these study results will provide a foundation for librarians to educate and assist researchers with managing their data throughout the research lifecycle.

Received 13 January 2014 | *Accepted* 26 February 2014

Correspondence should be addressed to Natsuko Nicholls, 240E Clark Library, Hatcher South, Ann Arbor, MI, 48109-1190. Email: hayashin@umich.edu

An earlier version of this paper was presented at the 9th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



Introduction

The research community is quickly recognizing the value of data as an asset along with its potential to be reused or repurposed. The growing importance of research driven by large-scale accumulation of data highlights the need for more effective management of this new intellectual asset (Halbert, 2013). Another important attribute is that federal agencies, particularly the National Science Foundation (NSF), require a data management plan (DMP) to be submitted as part of research grant proposals. Researchers are now expected to make their data more understandable, discoverable, accessible (yet secure) and usable for others. To support increasing pressures of efficiencies and compliance, academic libraries, as trusted providers of information technology and information management expertise, are encouraged to establish data support services and resources.

Over the last few years, many academic libraries have begun to articulate the conceptual foundations, roles and responsibilities involved in research data management. The strategic plans developed by practitioners, librarians, IT technologists, and scholars have created a body of literature that compares and contrasts use cases and various organizational characteristics, which in turn have influenced the development decisions, outcomes achieved, and planned future deployments of research data services (Beitz et al., 2013; CLIR, 2013; Fearon et al., 2013; Jones et al., 2013; Pryor et al., 2013; University of Edinburgh, 2012). By reviewing existing literature, our environmental scan of peer institutions from a comparative perspective demonstrates that libraries are taking different approaches and are at different stages of data service development and implementation (Akers et al., 2014; Kouper et al., 2013; Raboin et al., 2012; Zilinski et al., 2013)¹.

While none of the processes and methods of data service development are the same, all libraries seem to share a common understanding of the emerging data landscape and goals. Librarians are being offered the opportunity to further embed themselves within the research infrastructure so that libraries can continue supporting researchers and new research paradigms in the most efficient and effective way possible. Thus, academic librarians have been increasingly engaging in assisting individuals with research data management through best practices, training, and services that address the complex and still emerging issues of data sharing and preservation (Halbert, 2013). Although none of these efforts are new, it is only recently that most academic libraries have begun to fully prepare for the critically important new activities of managing research data with library-offered services that support all phases of the research and data life cycles.

At the University of Michigan, the Library's Research Division has taken the lead in data service design and development while assessing needs and evaluating ways to organize the service. As part of developing support for research data management, a group of librarians have taken a pilot approach aimed at the College of Engineering (CoE) because of their high use of and creation of data. CoE received over \$38 million from NSF and their total research expenditure was over \$200 million for 2013. In May

¹ The environmental scan, conducted by Sferdean, Nicholls and Green in 2012, identified key players and stakeholders in the landscape of data management, preservation and curation at the University of Michigan. The external environmental scan comparing ten peer institutions illustrated the distinguishing features and different ways of establishing data services. These scans built the contextual foundation for this paper, and referred to Kouper et al., (2013), Zilinski et al., (2013), and Akers et al., (2014).

2013, three Engineering librarians and one CLIR/DLF Data Curation Postdoctoral Fellow were charged by the Director of Research Data Services to update the existing data management plan template document and to create pilot-based DMP consultation services for Engineering. These services are defined as the provision of a web-based NSF DMP guide and resources (e.g. LibGuide), in-person consultations for writing a DMP, workshops about data management planning, and overall promotion of the importance of data management as an integral part of research.

The project team membership excelled in meeting project goals. Subject librarians with strong domain knowledge and established ties to CoE (with approximately 500 faculty and researchers) are vital to promote partnerships and pilot projects. A postdoctoral data fellow with research orientation and current knowledge of data management and curation helped the team contextualize the pilot in a larger, more comprehensive effort of developing research data support infrastructure and services across research and data life cycles. This research data management project is the result of communication, planning and collaboration with various stakeholders across campus units, and is critical to the development of new services that most effectively meet the unique needs of the Michigan campus research community.

Approach and Method

A series of steps has been taken since the launch of the Engineering data support pilot. A specific approach was to assess the potential for research support improvement and make evidence-based suggestions for establishing data management support and resources.

Assessment for Improvement: Faculty Survey

In the summer of 2013, an email invitation was sent to CoE faculty to voluntarily complete a short online survey. The survey contained five questions and was administered using Qualtrics software. The survey was intentionally brief – it was not intended to be research-oriented, but rather to raise awareness and gather information. The survey functioned as an outreach tool to raise researchers' awareness of the library's developing support for research data management and as an assessment tool to learn about Engineering faculty's familiarity with and experience of writing a DMP as part of NSF grant proposals. Most importantly, the faculty survey was distributed to gather feedback on possible improvements to the previous NSF Engineering Data Management Plan Document developed by Jake Glenn, a science librarian, in 2011. The survey provided a link² to download and view this document. The survey results and the response to the faculty feedback are discussed in a later section.

Evidence-Based Approach: DMP Review and Analysis

To accurately gauge researchers' interest in and solicit evidence-based suggestions for potential data management services that are best suited for Engineering, a review was conducted of DMPs written by Engineering faculty whose NSF proposals had been granted. 104 DMPs were acquired from CoE administration, representing grant

² NSF ENG Data Management Plan Template for the University of Michigan College of Engineering: <http://hdl.handle.net/2027.42/86586>

proposals that were awarded between January 2012 and June 2013. The project team examined and compared each DMP to the NSF DMP requirements and the Engineering Directorate to determine how well each section of the DMP met the specifications. All DMP documents, which ranged from a half page to two pages, were carefully reviewed, evaluated and coded. For coding, categorical distinctions were defined by the characteristics of DMPs. In particular, different metrics (e.g. levels of detail from high to low; quality of DMP from good to poor) and categories were used, including:

- **Roles and responsibilities for data management;**
- **Types of data produced:** Description of data to be collected (e.g. content, type, format, volume, etc.);
- **Data and metadata standards:** Standards and methodologies for data collection and management;
- **Policies for access and sharing:** Ethics and intellectual property (e.g. restrictions on data sharing such as embargoes and confidentiality);
- **Policies for reuse and redistribution:** Plans for data sharing and access (e.g. how, when, to whom);
- **Data storage options;**
- **Plans for archiving and preservation:** Strategy for long term preservation, specifically noting if researchers mention Deep Blue, our institutional repository.

Note that these categories align with the requirements of the NSF Engineering Directorate³:

1. Roles and responsibilities,
2. Period of data retention,
3. Expected data,
4. Data formats and dissemination,
5. Data storage and preservation of access.

Findings and Implications

Assessment for Improvement: Faculty Survey

The faculty survey showed that the majority (60%) of the faculty who responded to the survey were at least moderately familiar with the requirements for a Data Management Plan, and approximately three-quarters (72%) had written a DMP (see Figures 1 and 2).

³ Data Management for NSF Engineering Directorate: http://nsf.gov/eng/general/ENG_DMP_Policy.pdf

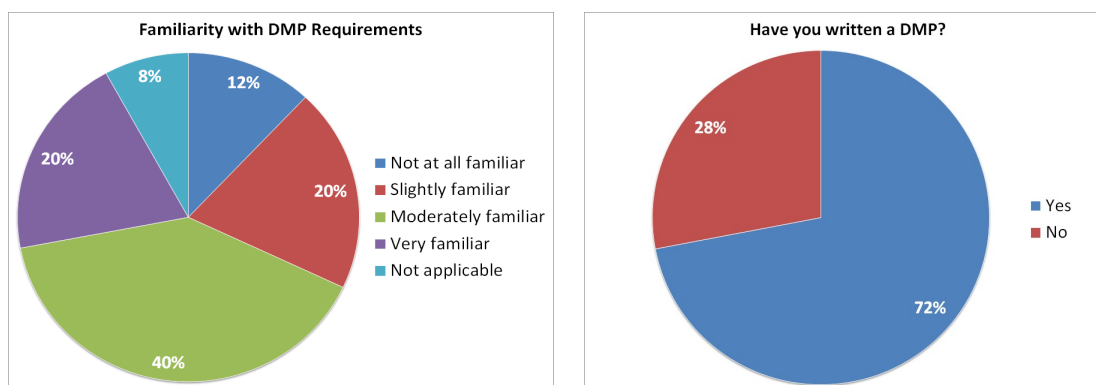


Figure 1. Familiarity with DMP requirements.

Figure 2. Experience of writing a DMP.

However, most were not familiar with the 2011 document that the library had provided – they did not know about it and did not use it (see Figure 3).

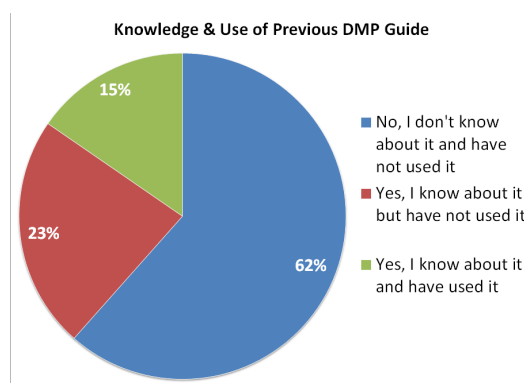


Figure 3. Knowledge and use of previous DMP guide.

The survey provided a link to the existing document, which could be downloaded and viewed. Several faculty took the opportunity to provide feedback about it in the open-ended questions. They noted that the document was too long, too detailed, and inflexible due to being in PDF format. Further, one faculty member commented that they are interested in receiving library assistance that goes beyond writing a DMP, i.e., the provision of secure storage for long term data preservation.

‘[T]he document is very detailed – useful for reading but not for creating the DMP.’

‘I just learned about it with this email, so I will try to use it for next time. It seems a bit long, but I guess once I get through my first DMP it will be easier to use for future times.’

‘I think the document is great, especially since examples of good statements and statements that need further work are provided. I have not really used it because I have been comfortable putting together my data management plans.’

‘What we need is a secure place to store data that is easy.’

In addition to gathering comments about the existing document, the survey also asked for comments on what resources the faculty have used to help them write data management plans. Multiple faculty members reported consulting a colleague’s DMP to assist them in writing their data management plan, which indicates the usefulness of providing a guide that includes boilerplate language for selected sections of a DMP.

Survey results and comments showed that the previous DMP document needed both simplification and increased visibility. In response, the team updated the document and reformatted it into the LibGuides⁴ format, which led to a more simplified layout and a presence on the library’s website. Instead of a long, 30-page document, users now find a tabbed web page that allows them to quickly browse to the sections of the guide that offer assistance with the elements that should be addressed in a DMP.

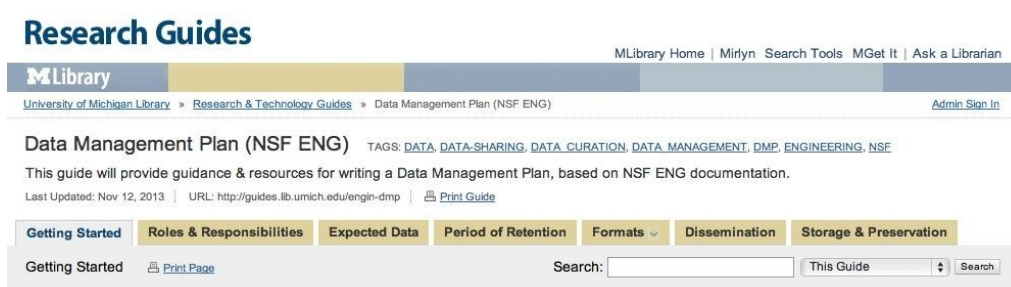


Figure 4. New Library research guide on DMP.⁵

With the DMP guide completed, the team contacted faculty and asked their opinions. One enthusiastic response showed appreciation for the new guide:

‘[The] material is very useful, good examples, and the links to other resources and related [topics are] very powerful. Reading the material, I would ask myself: “but what are the repository resources at U-M?” and then, here the link is. Terrific... I wish to thank you and the rest of the team who put this document together.’

Findings from DMP Analysis

The analysis of actual DMPs written by Engineering faculty whose NSF proposals had been granted confirmed some of the results of the survey: Faculty produced acceptable Data Management Plans, but many did not fully satisfy all the categories that were developed based on the elements specified by the NSF. The analysis exposed a need for assistance and education regarding the NSF DMP guidelines. The team determined that creating a service to evaluate DMPs before they are submitted would be useful, and the analysis could be a means of showing faculty that our services can help improve their DMPs. While researchers clearly value the use of data for their research, some do not yet recognize the full value of maintaining records and information about that data, which is the primary objective of a DMP. It is important to note that the creation of data

⁴ LibGuides is a proprietary application for presenting library information to users in standard formats: <http://springshare.com/libguides/>

⁵ MLibrary Research Guide, Data Management Plan: <http://guides.lib.umich.edu/engin-dmp>

and maintenance of records about the data are not at odds with each other, but are complementary approaches toward creating a sustainable, preservable data ecosystem.

Before discussing the findings in depth, it should be noted that of 104 DMPs being reviewed, nine percent represent cases in which the primary investigators provide little information in their DMPs except for noting that “[t]his is a workshop proposal” and/or “[n]o data will be generated.”⁶ In the analysis, these cases were kept and treated as “N/A (not applicable).” Figures 5 and 6 demonstrate the distribution of 104 DMPs by level of detail, measured by the amount of information provided in DMPs, versus the quality of DMPs, measured by the extent to which the DMP met the NSF DMP requirements. The two charts demonstrate that a high amount of information does not necessarily mean that all NSF requirements are met.

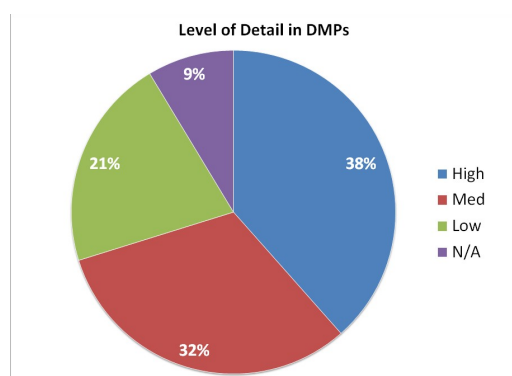


Figure 5. Level of detail in DMPs.

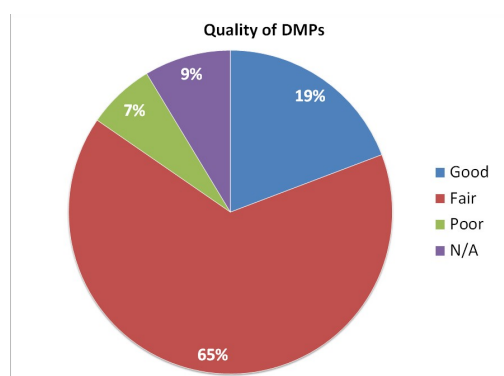


Figure 6. Quality of DMPs.

One of the most important findings was that the DMP requirement in the area of ‘expected data’ was well understood and fulfilled by Engineering faculty. They were able to provide detailed information in DMPs about types of expected data (e.g. experimental, numeric, simulation and computational data, etc.) and formats of expected data (e.g. text/docs, image files, video files, computer codes, algorithm and the files generated by computer programs, including Excel, Matlab, Gaussian, LabVIEW, Python, etc.).

On the other hand, findings indicate that there are areas of misunderstanding or omission. In one third of the DMPs, researchers did not name specific individuals or state their roles in managing specific types of data, and did not consider changes to roles and responsibilities that would occur should a PI or co-PI leave the institution. Likewise, 30% remained unclear about the period of data retention (see Figures 7 and 8).

In our review of DMPs, it was noted that there was ambiguity and/or confusion related to the volume of expected data, the difference between storage and long term preservation, data formats better suited for preservation, and reasons for not sharing data. Only ~20% of DMPs stated how much data the research would generate and at what rate. This could be due to researchers not knowing the final volume of the data they will be working with, or an expectation that the data volume may be so small that there is no need to consider required storage space prior to the research being done. Overall, an important lesson learned was to connect these findings to practical advice for researchers and be prepared to answer data management questions such as:

⁶ NSF requires researchers to include a DMP even for a project that is not expected to generate data or samples. However, according to the NSF guidance it is acceptable for PIs to state in the DMP that the project is not anticipated to generate data or samples that require management and/or sharing.

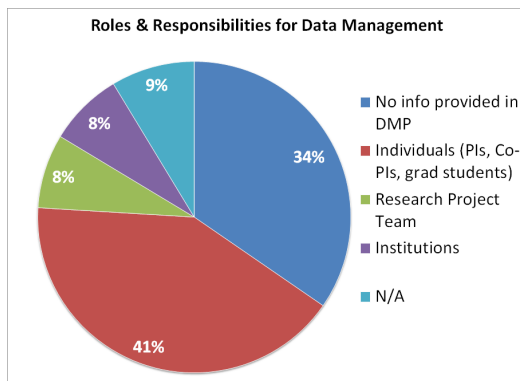


Figure 7. Roles and responsibilities.

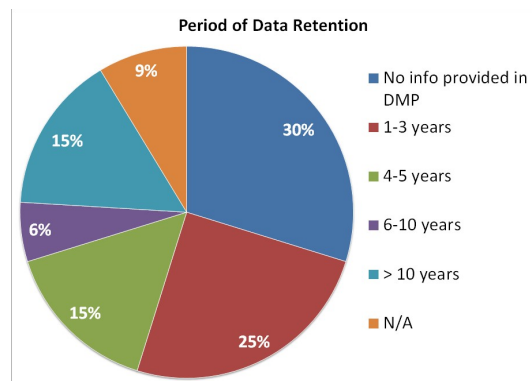


Figure 8. Data retention period.

- **Period of data retention:** Does it differ by raw, processed and compiled data?
- **Data storage and preservation:** Do all data (raw, processed, and compiled data) need to be saved? Does one need multiple locations for storage and preservation? Do some storage options (e.g. a project-based public website; a course management system) serve as a location for secure data storage and/or preservation?
- **Data dissemination and sharing:** Is publication a sufficient means of data sharing?
- **Repository:** In what way does a subject data repository better serve one's research needs than an institutional repository for depositing data?

Although some of the DMPs could be improved, many good statements were found in the DMPs that satisfied elements of the NSF guidelines:

‘Should any of co-PIs leave the University of Michigan, [name] will take the responsibility for the storage and access of data directly acquired by the leaving co-PI. Should the lead-PI leave the University of Michigan, the grant would likely be transferred. If not, [name] or [name] will assume the leadership of the project and responsibility for data storage and access.’

‘The proposed research is expected to generate data on near-field thermal conductance (data files ~1 KB for each measurement) and surface characterization data for the suspended island and other microfabricated devices (~1 MB each). This data will be stored as computer files. A total storage demand of <5 GB is anticipated over the three years of the program, based on ~5000 near-field conductance measurements and ~1000 surface characterization experiments.’

‘We will retain data in the form for which the University of Michigan’s long term data repository, Deep Blue, offers the highest level of support (level 1 support). For images and image renderings, the format will be .tiff; for confocal microscopy coordinate files, the format will be .txt.; for

data points appearing in tables and graphs the format will be .txt. The format for metadata will be .pdf, except for image processing tools, whose source code will be retained as .txt.’

Implications

There are two main implications from this work:

1. More research and education about data management plans and data management is needed.
2. More marketing and outreach is needed to educate faculty and graduate students about the NSF requirements and to raise their awareness of the helpful resources and services that the library offers.

DMP analysis shows that more education is needed to help researchers write better DMPs. This team’s immediate plans are for greater outreach to CoE researchers and graduate students.⁷ Librarians have met and discussed this work with the Engineering Faculty Library Advisory Committee, will offer open faculty and graduate student workshops in collaboration with CoE administration, and will seek additional opportunities to speak in engineering departmental faculty meetings about the new DMP guide and new data services.

Conclusion and Future Work

Based on the results of this pilot project, this team concludes that subject-based data services are best evolved from needs- and project-based services to more systematic, program-based services. The partnership between users (i.e. faculty and researchers) and subject and liaison librarians is important in assessing faculty’s data needs to create services worth transforming from initiative into sustained services. Initiatives allow exploration into the users’ ever-growing needs and better insight into what librarians can support given our available resources and infrastructure. Future work will focus on the goal of educating faculty and graduate students about data management, which includes writing DMPs. Long term success can be measured by conducting faculty and graduate student surveys and analyzing future DMPs, or by being part of the grant application process prior to submission. The team suggest these steps to take in the near future toward the final goal:

- Share these study results with campus partners (starting with CoE) to inform their specific research and data support initiatives and foster further collaborations with the library;
- Share these study results with librarians to increase their understanding of researchers’ data practices and needs with an eye towards effective and comprehensive data management in libraries;

⁷ Most recently, to reach out to graduate students, librarians interviewed 121 presenters at the 2013 Engineering Graduate Student Research Symposium. The interviews gave insights into graduate students’ familiarity with data management practices and DMP requirements. Librarians also introduced the students to the DMP guide. DMP information will now be mentioned in new graduate student orientation sessions, and DMP workshops are planned for the next semester.

- Use study results to guide the development of new services, particularly consultation services and infrastructure, to extend the library's support for the creation, analysis, storage, preservation and sharing of research data in the sciences, social sciences, and humanities;
- Develop workshops for faculty and graduate students about best practices for writing DMPs;
- Create boilerplate language for certain sections of the DMP that faculty could easily use in their proposals;
- Promote new library services for a data management tool. For example, DMPTool 2 when it is released in Spring 2014;
- Explore opportunities for librarians to further embed themselves in the research infrastructure, such as becoming part of the grant application process;
- Identify the range of barriers to effective research data management at scale.

Recent years have seen a growing emphasis on the need for improved management of research data. Through this pilot project, this team of librarians has identified areas where the library can play a role in assisting researchers with data management, and has put forth proposals for immediate steps that the library can take in this regard.

Acknowledgments

Authors wish to thank Linda Forsyth, Rachael Ristau, Elizabeth Wagner and Dr. Alec Gallimore; the College of Engineering, who contributed time to compile DMPs from NSF proposals and review an earlier draft of this paper; Elaine Westbrook and Jake Glenn; and the University of Michigan Library, who provided guidance to the project. Authors also thank anonymous Engineering faculty who took the time to take a survey to share their feedback.

References

- Akers, K., Sferdean, F., Nicholls, N., & Green, J. (2014). *Building support for research data management: Biographies of eight research universities*. A submitted manuscript.
- Beitz, A., Dharmawardena, K., & Searle, S. (2013). *Monash University research data management strategy and strategic plan 2012-2015*. Retrieved from <https://confluence-vre.its.monash.edu.au/display/rdmstrategy>
- Council on Library and Information Resources. (2013). *Research data management: Principles, practices, and prospects* (CLIR Publication No. 160). Washington, DC: Author. Retrieved from <http://www.clir.org/pubs/reports/pub160/pub160.pdf>

- Fearon, D., Gunia, B., Lake, S., Pralle, B., & Sallans, A. (2013). *Research data management services* (SPEC Kit 334). Washington, DC: Association of Research Libraries. Retrieved from <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/>
- Glenn, J. (2011). *NSF ENG data management plan template for the University of Michigan College of Engineering*. Retrieved from <http://hdl.handle.net/2027.42/86586>
- Glenn, J., Nicholls, N., & Sferdean, F. (2013). *Resources and services throughout the data lifecycle: Data management planning*. Paper presented at Cyberinfrastructure (CI) Days, Ann Arbor, MI.
- Halbert, M. (2013). Prospects for Research Data Management. In *Research data management: Principles, practices, and prospects* (CLIR Publication No. 160, pp. 1–15). Washington, DC: CLIR. Retrieved from <http://www.clir.org/pubs/reports/pub160/pub160.pdf>
- Inter-university Consortium for Political and Social Research. (2013). *Guidelines and resources for OSTP data access plans*. Retrieved from <http://www.slideshare.net/ICPSR/guidelines-for-ostp-data-access-plans>
- Jones, S., Pryor, G., & Whyte, A. (2013). *How to develop research data management services: A guide for HEIs*. Edinburgh, UK: Digital Curation Centre. Retrieved from <http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services>
- Kouper, I., Akers, K., Nicholls, N., & Sferdean, F. (2013). *A roadmap for data services*. Paper presented at the Joint Conference on Digital Libraries, Indianapolis, IN. doi:10.1145/2467696.2467763
- Pryor, G., Jones, S., & Whyte, A. (Eds.). (2013). *Delivering research data management services: Fundamentals of good practice*. London, UK: Facet Publishing.
- Raboin, R., Reznik-Zellen, R.C., & Salo, D. (2012). Forging new service paths: Institutional approaches to providing research data management services. *Journal of eScience Librarianship* 1(3): 134-147. doi:10.7191/jeslib.2012.1021
- Ray, J. (Ed.). (2014). *Research data management: Practical strategies for information professionals*. West Lafayette, IN: Purdue University Press.
- Sallans, A. (2012). DMP online and DMPTool: Different strategies towards a shared goal. *International Journal of Digital Curation* 7(2), 123–129. doi:10.2218/ijdc.v7i2.235
- University of Edinburgh. (2012). *Research data management (RDM) roadmap, August 2012 – January 2014*. Retrieved from <http://www.ed.ac.uk/schools-departments/information-services/about/strategy-planning/rdm-roadmap>
- Zilinski, L., Chan-Park, C., Dasler, R., & Nicholls, N. (2013). *Carpe data: Data curation services at four different institutions*. Paper presented at the Digital Library Federation Forum, Austin, TX. Retrieved from http://docs.lib.purdue.edu/lib_fspres/39/