



Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach

Lisa Johnston, Meghan Lafferty, Beth Petsan

University of Minnesota, Minneapolis, MN, USA

Abstract

This article describes the curriculum, implementation, and results of the research data management training offered by the University of Minnesota (UMN) Libraries. The UMN Libraries have offered the workshop titled, "Creating a Data Management Plan for Your Grant Application," to more than 300 researchers and faculty since late 2010. With University partnerships, this training satisfies the requirement for the continuing

education component to maintain PI eligibility. Based on workshop feedback, the authors conclude that academic libraries can provide support to researchers with federal mandates to share their research data by providing timely, discussion-based training and resources on how to create a data management plan. The unanticipated benefits for library staff education and professional development on this topic are explored.

Introduction

To understand the research behaviors, information resources, and service needs of scientists, the University of Minnesota Libraries conducted a user-needs study on the process of research (University of Minnesota Libraries, 2007). A key finding was the large unmet need to help manage research data. The UMN Libraries identified this as an opportunity to take a lead role in responding to researchers' concerns. The appropriateness of that role was further bolstered by comments from faculty such as, "There are probably better ways [to organize research data]. If there were a workshop on organization and file management, I would go. The Libraries do this so well." (University of Minnesota Libraries, 2007).

In 2009, the University's Office of Information Technology commissioned a survey

to explore data management needs in greater depth, and found clear evidence of an education gap among researchers (Johnston, 2010). This survey of 780 researchers found that over a quarter had lost important data due to the lack of backup, and 72% used unsecure media such as CDs, DVDs, and flash drives to back up data. While 92% of respondents from the study stated that they share their data, primarily within their own group or on campus (51% and 18%, respectively), few were aware of other on-campus options. As one researcher phrased it in the survey's published data, "If infrastructure exists for sharing data, the knowledge has not been imparted on me" (University of Minnesota President's Emerging Leaders Program, 2009). University of Minnesota researchers received \$769 million in sponsored funding during the 2011 fiscal year (Mulcahy, 2011). The substantial amount of sponsored awards and the growing list of

Correspondence to Lisa Johnston: ljohnsto@umn.edu

Keywords: data management plans (DMP), curriculum, data literacy

funding agencies mandating data management and sharing plans in grant applications (<http://lib.umn.edu/datamanagement/funding>) offered the UMN Libraries a prime opportunity to educate researchers.

Background

The University of Minnesota Libraries began developing educational and outreach programs to support research data management several years ago. In 2007, prompted by the call to action in the Association of Research Libraries' "Agenda for Developing E-Science in Research Libraries," the UMN Libraries formed the E-Science and Data Services Collaborative (EDSC) to assess the landscape and recommend priorities for action (Lougee, 2007; Johnston and Hanson, 2010). The group surveyed user-needs reports from other large research-intensive institutions in the US including Cornell University, University of North Carolina, Purdue University, and University of Washington as well as a federal government interagency working group (Steinhart, 2007; Granatino, 2008; Mullins, 2007; Fox et al., 2009; Furlani et al. 2009). The EDSC also studied the experience of creating an Australian national data repository and efforts to define library roles and responsibilities in the United Kingdom (ANDS, 2007; Lyon, 2007). Two major results of the work of the EDSC were "Managing Your Data," a web site addressing practical data management problems like file naming, data storage, and backup options on campus located at <http://www.lib.umn.edu/datamanagement>, and "Introduction to Data Management for Scientists and Engineers," a drop-in workshop modeled on the MIT Libraries Data 101 workshop (MIT Libraries, 2009).

Despite these successful early efforts, the libraries recognized the need to expand its reach and engage researchers in discussions about their unique issues. The impetus for change came with the 2010 National Science Foundation (NSF) announcement that all new NSF grant proposals must in-

clude a data management plan (DMP) to be considered for review after January 2011 (National Science Foundation, 2010). Based on prior user-needs assessments and training efforts, the libraries were well positioned to build upon that data management expertise and developed the following educational approach.

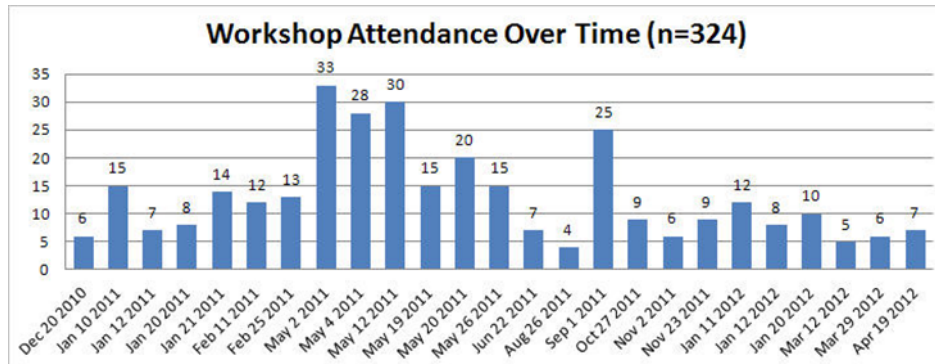
Training Researchers on "Creating a Data Management Plan"

The workshop, "Creating a Data Management Plan for Your Grant Application," was developed in the fall of 2010. Because the campus is large with more than 50,000 students, faculty, and staff, training needed to be scalable and amenable to the needs of a broad range of disciplines. The one and a half hour-long workshop incorporated each participant's needs and experience and used an active learning approach emphasizing sharing and group discussion (see Table 1). Co-taught by two librarians, a research services librarian and a subject librarian, the workshop consisted of an introduction followed by five sections based on the five questions (see Table 1) suggested by the NSF guidelines for writing a DMP listed in their grant proposal guide (NSF, 2012). Each section combined lecture, led by the research librarian, and an activity, facilitated by the subject librarian, who often had an existing relationship with the participants. This arrangement alleviated the subject librarian from the burden of needing expertise on all aspects of data curation, while allowing the librarian team to offer a more customized workshop to the participants by means of their disciplinary knowledge. The workshop typically met in a space that facilitated discussion (e.g., a room with a round table). At the start of the session, attendees received two handouts, an overview of and FAQ about the NSF requirement, and a checklist of questions to address when writing a data management plan (the workshop slides used and the handouts are available for reuse and sharing at

Table 1: Outline of the “Creating a Data Management” Active Learning Session

Run Time (min)	Activity/Section (Based on NSF guidelines)	Attendee Instructions	Instructor Tips
0-15	Introduction to Data Management Plans		Introduce yourself
16-30	What type of Data or file formats do you produce in your work/ research?	<p>Introduce yourself to your neighbor and describe the types of data you produce.</p> <p>Use the first section of the "Questions to consider" handout to give you ideas, like rate of growth to determine storage.</p>	After discussion period, go around the room and have people introduce themselves and describe their research data. Instructor: Ask follow-up questions as appropriate.
31-45	What standards will you use for data documentation and metadata format?	Imagine that I am about to enter your lab or department as a new grad student, what type of documentation about your files and data collection would I need to get up to speed. Are there metadata standards? If not, how would you make this process more standard so anyone in your field could understand? Discuss this possibility with your partner.	Facilitate two or three groups to report back (volunteers, or select).
46-60	How will you release your data for access?	<p>Break into three groups and each group give us 1-2 examples of steps that one would take to protect the following types of data:</p> <ol style="list-style-type: none"> 1) Privacy/Confidential 2) Security (physical or digital) 3) Intellectual Property rights. 	Facilitate each group to report back and respond with anything that may be missing.
61-75	How will you share your data?	Staying in the same three groups, each group give us one pro and one con for the following data sharing options: <ol style="list-style-type: none"> 1) Post online 2) Publish in journal 3) Made available on request. 	Facilitate each group to report back and respond with examples of data sharing examples that may be missing.
76-90	How will you archive the data for preservation and long-term access?	Thinking back to your graduate degree, would you be able to access the data from your dissertation today? Why or why not? How might this be different for students going forward?	Ask for one or two examples from the group as a whole.

Figure 1: Workshop Attendance from December 2010 to April 2012



<http://www.lib.umn.edu/datamanagement/workshops/dataplan>). Following the workshop, participants received the lecture slides and electronic versions of the handouts along with two additional key elements:

- A University-specific sample DMP, created with a UMN researcher, which included the five sections of the workshop. The sample gave attendees a model with boilerplate language to use in describing their data.
- An invitation to meet with the research services librarian for a future DMP consultation.

Results and Survey Feedback

The libraries have offered DMP workshops and consultations since December 2010. The “Creating a DMP” session has reached 325 faculty and researchers (see Figure 1) through 26 sessions with attendance varying over the academic year. Eight of the workshops were customized sessions requested by groups including:

- Center for Transportation Studies Scholars
- College of Education and Human Development
- Department of Biomedical Engineering
- Department of Computer Science and Engineering

- Department of Electrical and Computer Engineering
- Department of Neurology
- Department of Physics and Astronomy
- Sigma Xi Minnesota Chapter

Promoting the workshop on campus was crucial to its success. The session was initially offered as a walk-in library workshop open to all University affiliates and promoted through library blogs, printed posters, and direct emails to select departments. After some walk-in success, departments known to seek NSF and NIH funding were targeted for workshops at faculty meetings or department retreats which paired the research services librarian with the department’s subject liaisons. This arrangement gave liaisons the opportunity to be more involved in discussions about data management in their subject areas.

Attendees received Responsible Conduct of Research continuing education credit, an additional selling point for faculty members since they must renew the credit every three years. The libraries partnered with the Office of the Vice President of Research to allow researchers to earn the credit through the workshop.

A majority of participants in the general sessions were faculty (see Figure 2) from engineering and agricultural, biological, health, and physical sciences (see Figure 3). Fewer

Figure 2: Workshop Attendees by Academic Role

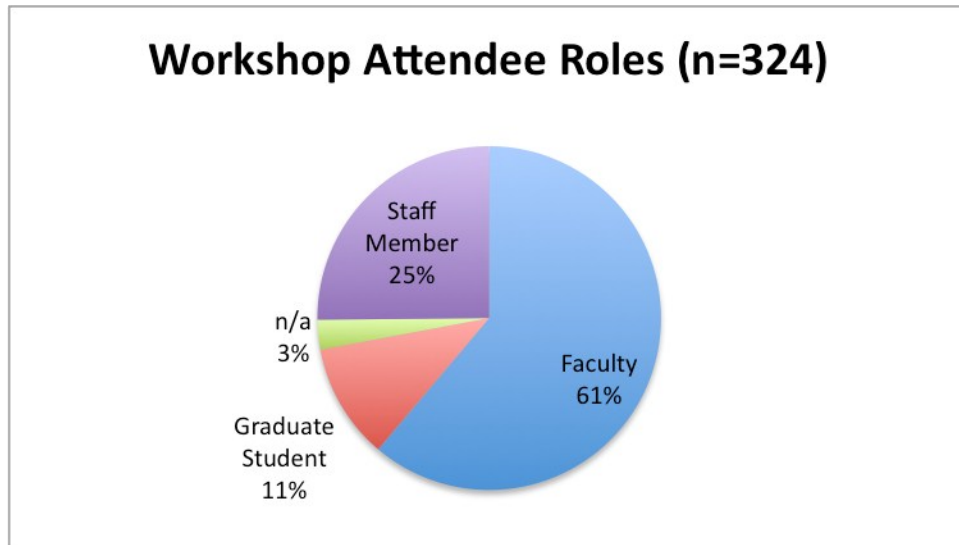
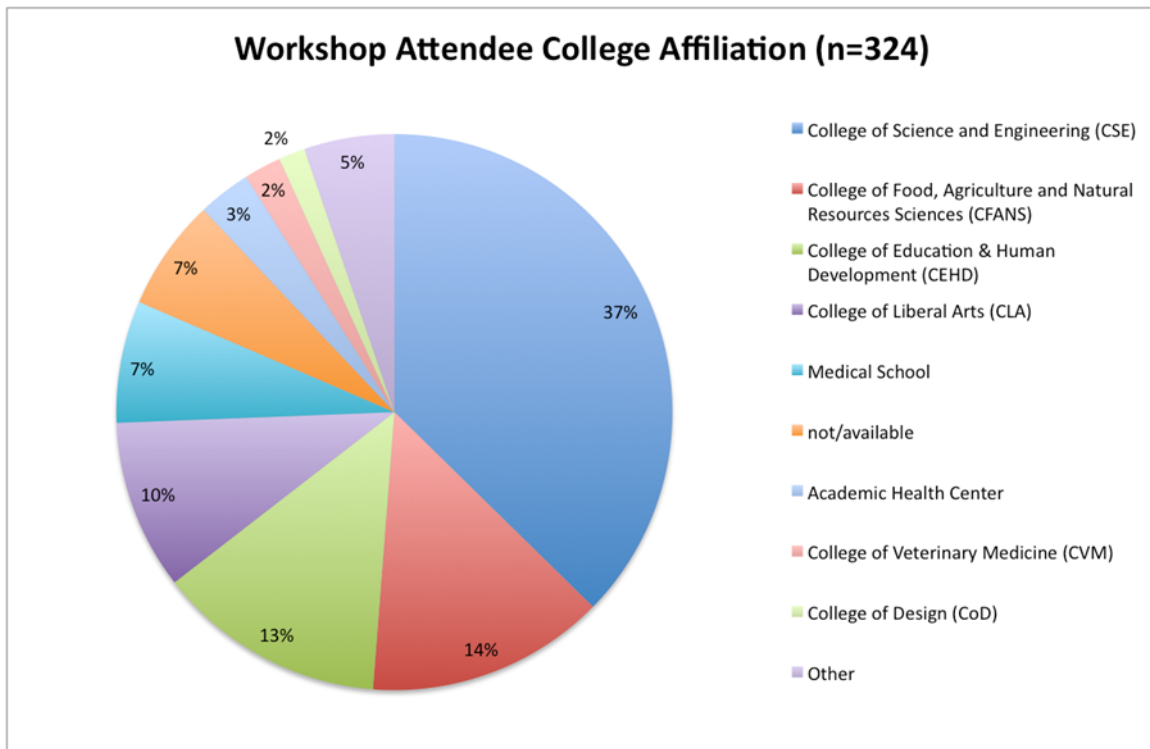


Figure 3: Workshop Attendees by College Affiliation, listed in decreasing order by college.



participants were from the social sciences, arts, or humanities, but their needs were no less apparent. A notable number of staff members in grant-writing support roles also attended.

Informal feedback from workshop attendees has been very positive:

- “Very knowledgeable presenters. Great breadth of coverage and general enough to help everyone see that it pertained to them.”
- “Format of talking to neighbors was good. Much better than listening to lecture for 90 minutes.”
- “Thanks the class was very valuable!”
- “Great workshop this morning. You raised issues that I'd not really thought about before. Thank you!”

The 178 “walk-in” participants (as opposed to department-based sessions) received a library workshop evaluation email survey (see Appendix A). Of the 62 responses received, the comments were overwhelmingly positive. Results, for example, when asked, “What did you find most useful about the workshop?” are as follows:

- 21% mentioned the handout (the checklist of DMP questions).
- 44% found the content in general most useful (e.g., “Topic was very useful, given new NSF data management requirements. I particularly liked that the workshop covered different options and gave some guidelines about how to tell what options might be most appropriate given a particular research domain.”).
- 31% found sharing and discussion with other participants most useful.
- Other useful aspects, with less than 10% of responses, included the slides, related library services, and “knowledgeable presenters.”

The survey also indicated that 92% agreed or strongly agreed that the “workshop content was delivered in a clear manner.” The

answers to the survey questions, “How might we improve the workshop?” and “Please provide any additional comments or suggestions” provided more critique. For example, even though 44% of respondents commented that the group discussions were the most useful aspect of the workshop, a small number of respondents, 11%, wanted less time for group discussion. When sessions had either less active participants or too few participants for effective group activities, the instructors had a few strategies. First, they were prepared to share information intended to promote discussion like stories of personal interactions with researchers or news stories on the subject. Second, the instructor would ask the entire group to work through activities together which allowed for more individualized attention.

An additional 15% of participants wanted more in-depth information. One comment summed up these sentiments: “More detailed info from organizers and (much) less time in small groups talking about issues.” Other improvements suggested included:

- Examples of the libraries’ role in developing data management plans.
- Case studies using real research projects to walk attendees through the process of creating a sample DMP.
- A session specific to grant-writing support staff.

Based on attendee feedback, the libraries planned to increase promotion of its free DMP consultation services and highlight its involvement with DMPTool (<https://dmp.cdlib.org>), recently developed by the California Digital Library and partners. Formal participation with this tool allows university-affiliated researchers to receive customized advice and links to further guidance from their institution’s library.

In addition to the workshops, recipients took advantage of free grant application DMP-writing consultation services. The research services librarian consulted with approxi-

mately a dozen researchers, and the (unsolicited) feedback has been very positive. Comments included:

- “We successfully submitted the NSF proposal (my first)! Thank you again for taking the time to review it. Your feedback and Layne [Johnson]’s were helpful and greatly appreciated.”
- “Thanks so much for your help over last few days. This is one less thing for me to learn how to do next go round.”

Conclusions

The success of this research training in creating data management plans was due to several factors. The team-teaching approach is scalable to multiple disciplines, and the active-learning components allow for addressing a variety of data management needs. An unanticipated benefit was enabling liaison librarians to engage with faculty in their departments in a new way. To date, 12 library staff members have co-taught the training sessions and many more have attended, thereby increasing their overall awareness of data management issues across the campus community. This training would be useful at other academic libraries, large and small, for library staff professional development as well that of researchers.

Next steps in data management training include a more in-depth study of how graduate students are taught data management skills. The UMN Libraries are participating in an IMLS grant led by Purdue University (<http://wiki.lib.purdue.edu/display/ste/>) called The Data Literacy Project to define a model for “data literacies” based on the work of Carlson, et al. (2011) in graduate education. The curriculum, to be offered via web-based modules to civil engineering students in the fall of 2012, has benefitted from the lessons learned in the data management plan workshops. The learning activities of the new online curriculum are primarily based on the five sections of the DMP workshop. A future goal for discipline-based data literacies train-

ing is to create easily adaptable web-based learning modules that can be customized for different disciplines.

In summary, the DMP workshops had an overwhelmingly positive impact on the libraries’ position on campus to respond to research data management needs. The successful workshops provided us with a wealth of knowledge and spurred new interactions between faculty and librarians around research. The resulting conversations allowed the libraries to better understand researchers’ needs and, in turn, offer solutions and advice in areas of expertise such as digital object curation and appropriate sharing and attribution. This emerging area of research support has not only led to new connections with faculty and researchers, but has also enriched relationships with campus units like the Office of the Vice President for Research.

References

- ANDS Technical Working Group. *Towards the Australian Data Commons: A proposal for an Australian National Data Service*. Accessed June 29, 2012, <http://www.pfc.org.au/twiki/pub/Main/Data/TowardstheAustralianDataCommons.pdf>.
- Carlson, Jacob, Michael Fosmire, C. C. Miller, and Megan Sapp Nelson. “Determining Data Information Literacy Needs: A Study of Students and Research Faculty.” *portal: Libraries and the Academy* 11,2 (2011): 629-657. doi: 10.1353/pla.2011.0022.
- Fox, Louis, Cara Lane, Ed Lazowksa, Janice Fournier, Greg Koester, and William Washington. “Information Technologies for eScience: A Preliminary Report from the University of Washington.” *Research Bulletin* 2009, 6 (2009): 1-12. Accessed June 29, 2012, <http://net.educause.edu/ir/library/pdf/ERB0906.pdf>.
- Furlani, Cita, Chris Greer, and Chuck Romine. *Interagency Working Group on Digital*

Data. Accessed June 29, 2012, http://sites.nationalacademies.org/xpedio/groups/pqasite/documents/webpage/pqa_047703.pdf.

Granatino, Christopher M. "Expanding the Grid: Collaborative Opportunities in e-Science." A Master's Paper for the M.S. in L.S degree, University of North Carolina at Chapel Hill, 2008. Accessed June 29, 2012, <http://ils.unc.edu/MSpapers/3436.pdf>.

Johnston, Lisa, and Cody Hanson. "e-Science at the University of Minnesota: a collaborative approach." Presentation at the International Association of Scientific and Technological University Libraries, 31st Annual Conference, West Lafayette, IN, June 20-24, 2010. Accessed June 29, 2012, <http://docs.lib.purdue.edu/iatul2010/conf/day2/3>.

Johnston, Lisa. "User-needs assessment of the research cyberinfrastructure for the 21st century." Presentation at the International Association of Scientific and Technological University Libraries, 31st Annual Conference, West Lafayette, IN, June 20-24, 2010. Accessed June 29, 2012, <http://docs.lib.purdue.edu/iatul2010/conf/day1/5>.

Lougee, Wendy, Sayeed Choudhury, Anna Gold, et al. "Agenda for Developing E-Science in Research Libraries." Final Report, Association of Research Libraries, 2007. Accessed June 29, 2012, http://www.arl.org/bm~doc/ARL_EScience_final.pdf.

Lyon, Liz. "Dealing with Data; Roles, Rights, Responsibilities and Relationships." Consultancy Report, UKOLN, University of Bath. 2007. Accessed June 29, 2012, http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf.

MIT Libraries. "Managing Your Data 101" [ppt]. 2009. Accessed June 29, 2012, <http://libraries.mit.edu/guides/subjects/data-management>.

Mulcahy, Tim. "2011 Research Report." Uni-

versity of Minnesota, 2011. Accessed June 29, 2012, <http://www.research.umn.edu/documents/2011Report.pdf>.

Mullins, James L. "Enabling International Access to Scientific Data Sets: Creation of the Distributed Data Curation Center (D2C2)." Presentation at the International Association of Technological University Libraries, 38th Annual Conference, Stockholm, Sweden, June 11-14, 2007. Accessed June 29, 2012, http://www.iatul.org/doclibrary/public/Conf_Proceedings/2007/Mullins_J_full.pdf.

National Science Foundation. "Chapter II - Proposal Preparation Instructions." *Grant Proposal Guide (GPG)*. Last modified January 2011. Accessed June 29, 2012, http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp.

National Science Foundation. "Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans." *Press Release 10-077*. Last modified May 10, 2010. Accessed June 29, 2012, http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928.

Steinhart, Gail, John Saylor, Paul Albert, et al. "Digital Research Data Curation: Overview of Issues, Current Activities, and Opportunities for the Cornell University Library." Working Paper, 2012. Accessed June 29, 2012, <http://hdl.handle.net/1813/10903>.

University of Minnesota Libraries. "Managing Your Data." Last modified 2012. Accessed June 29, 2012, <http://www.lib.umn.edu/datamanagement>.

University of Minnesota Libraries. "Understanding research behaviors, information resources, and service needs of scientists at the University of Minnesota." Report, 2007. Accessed June 29, 2012, <http://purl.umn.edu/5546>.

University of Minnesota: President's Emerging Leaders Program. "Implementing Re-

search Cyberinfrastructure for the 21st Century.” Final Report, 2009. Accessed June 29, 2012, <http://purl.umn.edu/91772>.

Disclosure: The authors report no conflicts of interest.

All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a Creative Commons Attribution-Noncommercial-Share Alike License <http://creativecommons.org/licenses/by-nc-sa/3.0/>

ISSN 2161-3974

Appendix A: Library Workshop Evaluation Survey

Timestamp *(collected automatically)*

Name of workshop you attended: *(fill-in the blank)*

What was the location of the workshop? *(drop-down)*

What was the date of the workshop? *(fill-in the blank)*

Workshop content was delivered in a clear manner. *(select one)*

- Yes, I strongly agree
- Yes, I agree
- No, I disagree
- Neutral, unsure
- No, I strongly disagree

What did you find most useful about the workshop? *(fill-in the blank)*

How might we improve the workshop? *(fill-in the blank)*

Please provide any additional comments or suggestions. *(fill-in the blank)*

Please tell us who you are. *(Select multiple)*

- Faculty/Instructor
- Staff
- Researcher
- Graduate Student
- Undergraduate Student

Would you like to receive periodic emails from the Libraries about upcoming workshops and events? *(provide email)*