



Data Management Training for Graduate Students at a Large Research University

Jessica Adamick, Rebecca C. Reznik-Zellen, Matt Sheridan

University of Massachusetts Amherst, Amherst, MA, USA

Abstract

This article describes UMass Amherst Libraries data management workshops and online resources developed for graduate students. Although students respond favorably to general “Data Management Basics” workshops, they offer suggestions for improve-

ment and request discipline-specific examples, tools, and resources to augment the general information presented. In response, the Libraries’ Data Working Group aims to develop both broad-based, discipline agnostic workshops as well as on-demand, discipline-specific workshops.

Introduction

Federal funding agency data management plan requirements—specifically the National Science Foundation (NSF), National Institutes of Health, and National Endowment for the Humanities—have a direct impact on research teams. In addition to principal investigators, graduate students that participate in sponsored research with data management plan requirements are affected; at a minimum, they need to be aware of and adhere to their principal investigator’s plan for the effective management, storage, and sharing of research data. University libraries are working to support graduate students who work with data through data management curriculum development, as evidenced by recent work at the University of Minnesota, University of Massachusetts Medical School, and Worcester Polytechnic Institute, among other institutions (University of Massachusetts Medical School Lamar Soutter Library and Worcester Polytechnic Institute George C. Gordon Library 2012; Johnston, Lafferty,

and Petsan 2012).

The University of Massachusetts Amherst (UMass Amherst) is classified as a Carnegie Foundation Research University with Very High research activity, and is a top-50 recipient of NSF funding. At an institution like UMass Amherst, data management education is essential due to the high research rate occurring across the institution and the significant portion of research funded through agencies with data management plan requirements. Library-developed data management education practices for graduate students are emerging, but these activities are relatively new to mainstream science librarianship, and should be further documented and explored. For example, Carlson et. al.’s study of faculty and graduate student data information literacy needs describes gaps in graduate student knowledge of basic data management skills and recommends data information literacy as a natural extension of the jurisdiction of librarian instruction (Carlson 2011). Insight gained from a case

Correspondence to Jessica Adamick: jadamick@library.umass.edu

Keywords: data management, data management training, graduate students, university libraries

study exploration of graduate-level data management education will contribute to the advancement of library-developed data management curricula. A case study research method was chosen for this paper in order to contribute to an “evidence base for professional applications” (Zucker 2009).

The UMass Amherst Libraries are working to support both faculty and graduate students responsible for managing research data. A Data Working Group (DWG) was formed in the library to address this growing need, and is charged with creating meaningful resources on data management for the University community. The DWG provides data management plan consulting services, online resources for data management best practices and local support, and data management workshops for faculty and graduate students.

The DWG conducted a graduate student focus group in October 2010 on data management, which demonstrated a need for greater data education; participants were engaged with the issue and wanted to know more about best practices and local resources. The focus group’s conversation revealed that the students present were responsible for the collection, documentation, and management of data for their research projects. None of the students present reported formal training on this topic through their departments or research groups. In response, the DWG began a series of educational workshops for graduate students, which are the basis of this case study.

Methodology

This paper is a Descriptive and Evaluative Case Study (Mariano 1993); the DWG’s graduate student data management workshops are described and evaluated. Student participation, expectations, and post-workshop feedback are analyzed to determine the effectiveness of the program and to identify prospects and strategies for data management education. There are multiple

units of analysis: survey responses, student demographics, and observational and field notes from question and answer sessions. The units were analyzed inductively, where “themes and categories emerge from the data through the researcher’s careful examination and constant comparison” (Zhang and Wildemuth 2009).

The DWG held four data management workshops for graduate students during the 2011-2012 academic year at UMass Amherst. UMass Amherst is a public research and land-grant university in Amherst, MA, and the flagship of the University of Massachusetts system, with 28,084 undergraduate and graduate students and 1,121 full-time instructional faculty. Workshops titled “Data Management Basics” aimed to provide students with a broader context for core elements of data management such as effective data storage options, sharing and reuse policies, metadata, ethical and legal considerations, and preservation of data. The workshops identified external data management tools as well as campus-based resources for data management. Instructors included DWG members, with librarian representatives from the following areas: scholarly communication, systems, science reference, social sciences research services, and special collections and archives.

The workshops were formatted as no-cost 90-minute sessions; all four workshops consisted of a 45-minute presentation by DWG members followed by a question-and-answer session. At the beginning of the first three workshops, index cards were circulated to the attendees, who were asked to write down questions about data management and what they hoped to learn from the workshop. For the fourth workshop, students were asked to submit these questions, which were addressed during the question and answer session of the workshops, with their RSVPs. Evaluations were collected at the conclusion of each workshop.

The workshops were advertised to the grad-

Table 1: What do you hope to learn in this workshop?

What students hope to learn about data management, general categories
Organization, management, tracking of data sets
Data storage and backup
Search and retrieval of data
Data description and documentation
Data manipulation
Data collection, data gathering
Compliance with federal and university requirements
Dealing with large amounts of data
Tools and resources

uate student body through the Graduate School, faculty contacts, the UMass Amherst campus-wide events calendar, and the Libraries' Data Management web pages. Two of the workshops were open to graduate students of all disciplines and two of the workshops were targeted toward major discipline groups: Science and Engineering disciplines and Social Sciences and Humanities disciplines. While the format of the workshops remained fairly consistent, the content was modified iteratively, based on workshop evaluation feedback, to better target the needs of each group. In conjunction with the first Data Management Basics workshop, the DWG created a Data Management LibGuide as a complementary resource specifically for graduate students. The LibGuide includes much of the information covered in the workshop presentations, including context and best practices for data management, links to local and third-party tools, and information on research ethics and data citation (UMass Amherst Libraries 2012).

Results

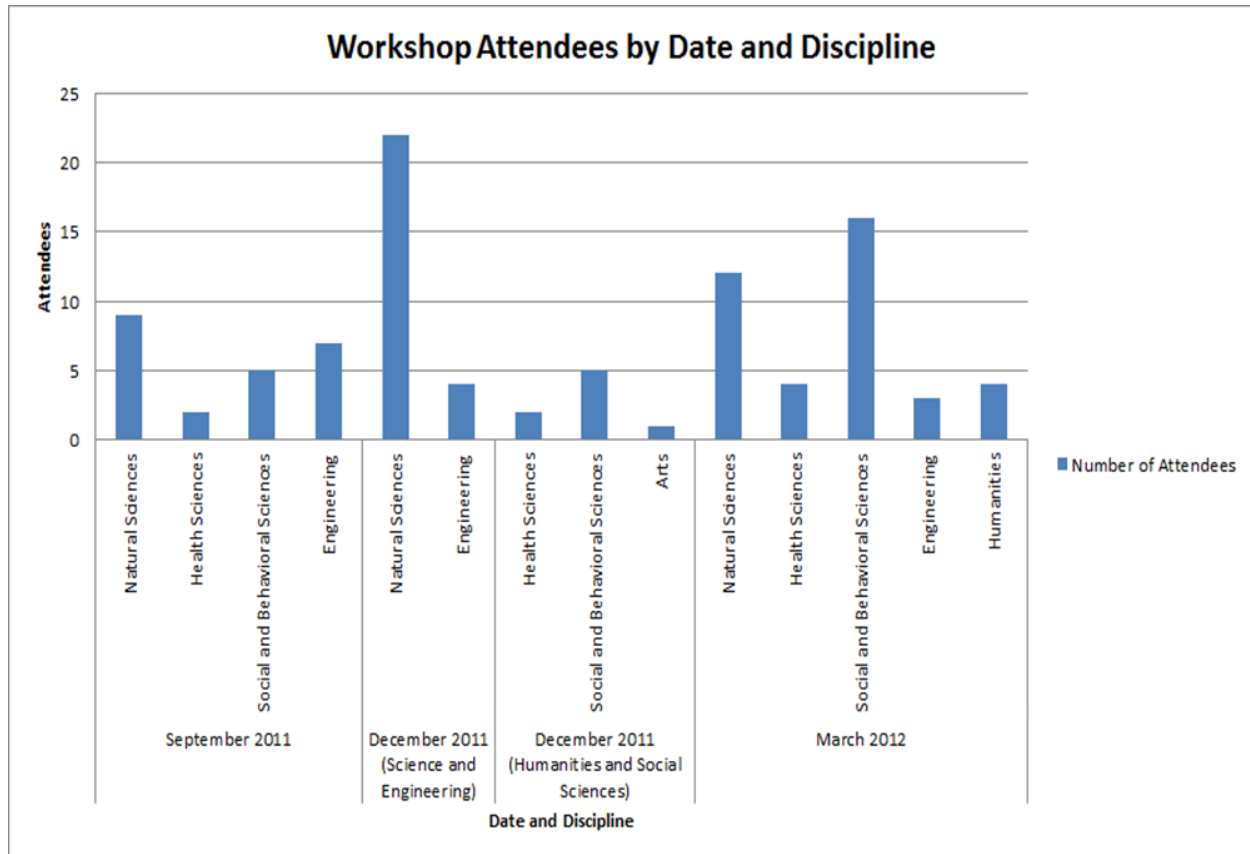
Pre-workshop responses

Pre-workshop participant responses on data management and what the participants hoped to learn were grouped into several general categories. Common responses are represented in Table 1. In general, students are overwhelmed by the amount and variety of data they encounter in their research and express a desire to learn about effective techniques and tools to stay organized. Data storage and organization were predominant themes in pre-workshop feedback, where data sharing and access were not mentioned prior to instruction.

Graduate Student Participation and Disciplinary Information

Twenty-three students attended a September 2011 workshop marketed to all graduate students; 10 students attended a December 2011 workshop for Social Sciences and Humanities graduate students; 27 students at-

Figure 1: Workshop Attendees by Date and Discipline



tended a December 2011 workshop for Science and Engineering graduate students; and 42 students attended a March 2012 workshop marketed to all graduate students.

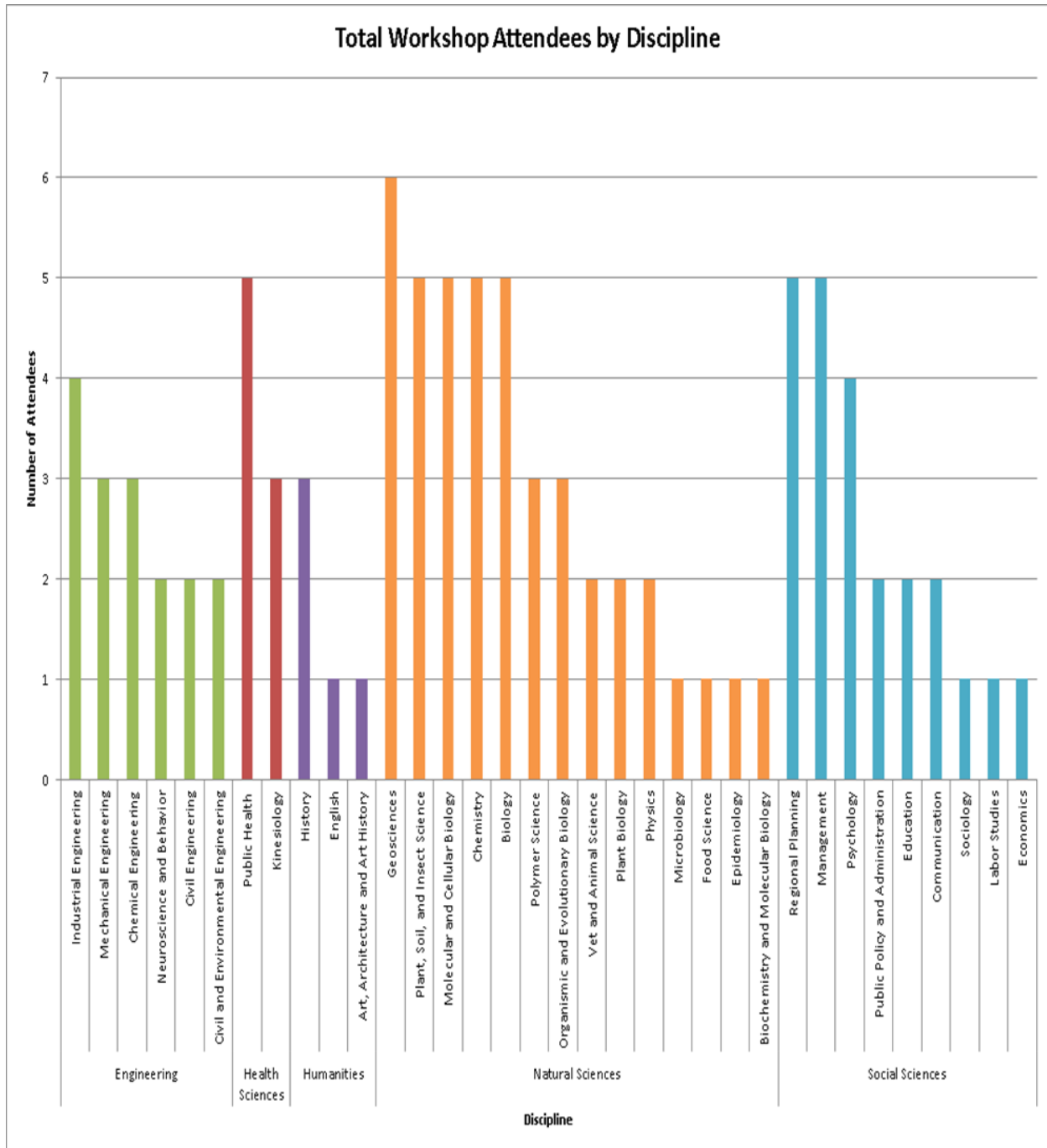
There was diverse disciplinary representation at the workshops. At all workshops, more students from the Natural Sciences attended than those from Engineering. While the turnout for the December 2011 workshop for Social Sciences and Humanities students was the lowest, Social and Behavioral Sciences and Humanities represented over 50% of the attendees at the March 2012 workshop, which was marketed to all graduate students. Students from the Health Sciences attended both general workshops, but none attended the session for Sciences and Engineering. Figure 1 represents workshop attendees by date and discipline.

Across all workshops, Geosciences was the most represented discipline with six attendees. Regional Planning; Public Health; Plant, Insect, and Social Sciences; Molecular and Cellular Biology; Management; Chemistry; and Biology were also well represented with five attendees. Psychology and Industrial Engineering each brought four attendees, while Polymer Science, Organismic and Evolutionary Biology, Mechanical Engineering, Kinesiology, History, and Chemical Engineering brought three attendees. The number of total workshop attendees by discipline is represented in Figure 2.

Post-workshop Feedback

After each workshop, participants were encouraged to fill out a brief evaluation. The survey questions were: What did you learn? What questions do you still have? How can

Figure 2: Total Workshop Attendees by Discipline



the workshop be improved?

The responses illustrated that attendees walked away with a general overview of data management practices. For example, respondents wrote:

[I learned] “basic data management procedures; resources available through university/public domain for managing data.”

“It was a useful overview of data management across disciplines. It also provided resources for further understanding and investigation.”

Table 2: What did you learn?

What did you learn?
Importance of data management practices
General information about organization and description
General information about backup and storage
File naming
Metadata
Awareness of repositories
Citation management tools
Backup
Funder policies

“Reminder about and introduction for terms and useful tools.”

Few respondents reported a concrete take-away:

“Back up data in at least 3 places. Make better file naming conventions.”

“RefWorks: you can take your database even after you graduate.”

Themes from the responses are presented in Table 2.

Many attendees reported that they didn’t have a clear understanding of metadata and its role in data management practice. Respondents wrote:

“What would the metadata file actually look like for a large database of xxxx?”

“How to go about adding metadata to files.”

“I still don’t really understand metadata.”

Many attendees wanted information about concrete resources they could use in data management. For example, respondents wrote:

“I guess what I’m thinking is about backup strategies. For instance, how to back up terabytes of data.”

“I would like to learn more about tools to help with versioning, mapping data files, and backup.”

“If there are free resources; a guide and what resources (eg: cloud) department @ UMass have (sic).”

Attendees also wanted information about general online resources where they could learn more about data management. Themes from the responses are presented in Table 3.

A majority of the suggestions for improving the workshop were based around providing more specific information, whether it was on tools, practices, or discipline-based differences in data management. For example:

Table 3: What questions do you still have?

What questions do you still have?
More information needed about sharing data with colleagues
Options for data backup
General online resources
Lack of understanding of metadata

Table 4: How could we improve this workshop

How could we improve this workshop?
Distribute slides in advance
Give examples from different fields
Give more explanation about metadata
Give several workshops on different data management topics
Present more specific examples
Partner with departments to promote the workshop

“Be more specific about actual methods. We never used (sic) these systems so it is hard to interpret what you are saying.”

“Give more specific examples, maybe choose a single field with good data management practices and show a data management plan.”

“I thought it would be much more specific about strategies for organizing Excel files (for example).”

“Show us large files and talk about what makes a good vs. poorly organized data set.”

Themes from the responses are presented in Table 4.

Discussion

Although students attend these workshops with general expectations about learning to organize, collect, describe, and manipulate data (Table 1), the DWG found that they are really looking for this information delivered in discipline-specific ways. Targeting large groups and broad categories of students makes preparing and delivering a successful workshop a challenge due to the students' need for examples, demonstrations, and tools specific to their disciplines. In multi-disciplinary groups, students reported that examples given were too generic to be meaningful. For example, the term metadata was confusing for many students. This concept may have been delivered more ef-

fectively in a discipline-focused setting where students could see metadata examples in their own field. In large groups, the range of experience with technologies varies widely. For those with a solid technical background, general best practices are unsatisfying, where they may be sufficient for those with little technical experience.

Based on the student feedback received, the DWG sees use in both broad-based data management overviews and tailored, discipline-specific workshops. Introductions to data management may be most beneficial to incoming graduate students before they become engrossed in their research. Given the feedback the DWG has received on past workshops, a more fully developed data management curriculum, targeted to and modified for specific disciplines, would provide a more effective mechanism for engaging graduate students about research data management at UMass Amherst. Going forward, the DWG plans to review and adapt existing approaches such as the University of Massachusetts Medical School and Worcester Polytechnic Institute Libraries' Frameworks for a Data Management Curriculum for undergraduate and graduate students. This framework is not discipline specific, but utilizes examples from scientific disciplines to illustrate the concepts covered.

In addition to continuing general best practice workshops, the DWG is exploring delivering discipline-specific workshops upon request for students who desire data management best practices in their disciplinary context. For example, when planning a workshop through the Digital Humanities Initiative at UMass, the DWG sent out a questionnaire to participants to directly inform the workshop content. The questionnaire aimed to discover the participants' role in data management, their need for collaboration, their familiarity with common data management tools, and their data management needs.

Conclusion

The University of Massachusetts Amherst Libraries Data Working Group conducted four workshops for graduate students during the 2011-2012 academic year. Although the workshops were well-received by attendees, evaluations indicate that more discipline-specific information is desired. Based on the feedback received before and after these workshops, the DWG plans to give both broad-based, discipline agnostic workshops as well as on-demand, discipline-specific workshops. Targeting workshops to narrow disciplines requires a greater investment of DWG members' time, but may be a more effective way to engage the range of graduate students with research data management training needs.

References

- Carlson, Jacob, Michael Fosmire, C.C. Miller, and Meghan Sapp Nelson. "Determining Data Information Literacy Needs: A Study of Students and Research Faculty." *portal: Libraries and the Academy* 11, no. 2 (2011): 629-657, <http://dx.doi.org/10.1353/pla.2011.0022>
- Johnston, Lisa, Meghan Lafferty, and Beth Petsan. "Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach." *Journal of eScience Librarianship* 1, no. 2 (2012): 79-87, <http://dx.doi.org/10.7191/jeslib.2012.1012>
- Mariano, Carla. "Case Study: The Method." In *Nursing Research: A Qualitative Perspective*, 2nd edition, edited by Patricia Munhall and Carolyn Oiler Boyd, 311-337. New York: National League for Nursing Press, 1993.
- UMass Amherst Libraries. "Good Practice for Data Management." Last modified July 21, 2012. <http://guides.library.umass.edu/content.php?pid=258091&sid=2130729>
- UMass Amherst Libraries Data Working Group. "Data Working Group Report." Am-

herst: UMass Amherst Libraries, 2010.
<http://www.library.umass.edu/assets/aboutus/attachments/Report2010-Final.pdf>.

University of Massachusetts Medical School, Lamar Soutter Library and Worcester Polytechnic Institute George C. Gordon Library. "Frameworks for a Data Management Curriculum." Worcester: University of Massachusetts Medical School Lamar Soutter Library and Worcester Polytechnic Institute George C. Gordon Library, 2012. http://library.umassmed.edu/data_management_frameworks.pdf

Zucker, Donna. "How to Do Case Study Research." In *Teaching Research Methods in the Humanities and Social Sciences*, edited by Mark Gardner, Barbara Kawulich and Claire Wagner. Surry, UK: Ashgate Publishing, 2009.

Funding Statement

Funding for this project comes from the National Science Foundation through grant number CCMI-1025020. Any opinions, findings, conclusions, or recommendations expressed here are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

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