



Full-Length Paper

Required Data Management Training for Graduate Students in an Earth and Environmental Sciences Department

Bonnie L. Fong and Minglu Wang

Rutgers University – Newark, Newark, NJ, USA

Abstract

The increasing importance of data management in the sciences has led the Department of Earth & Environmental Sciences at a research intensive university to work closely with the Physical Sciences Librarian and Data Services Librarian on campus to provide mandatory training to its graduate students. Although integrating data management training into the graduate program curriculum may not be possible, there are still opportunities to ensure students learn such skills prior to graduating. This article describes the four approaches taken thus far — a seminar about basic data management during the department's weekly seminar series, creation of a Data Profile form that students were asked to complete, an interactive workshop during the department's annual retreat, and assistance with writing data management plans. Buy-in for requiring data management training was essential from both faculty and students and was possible because both groups understood the value of research data management skills. Also vital to the success of these approaches was how the subject specialist and data librarians leveraged their respective areas of expertise in a complementary fashion to address disciplinary as well as broader data-related concerns.

Correspondence: Bonnie L. Fong: bonnie.fong@rutgers.edu

Keywords: data management training, research data services, graduate students, library instruction, data information literacy, data management, academic libraries



All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Introduction

Growing interest in research data from funding agencies, publishers, and higher-education institutions have led researchers and the librarians who support them to contemplate current practices of data management and to seek methods for improving them. Data management plans (DMPs) are being written, and options for data sharing are being explored. Recognized for their expertise in information management and sharing, librarians are a logical group to help faculty and student researchers with similar needs regarding data.

Proper data management is necessary to make data sharing possible, and both are important aspects of efficient research. Good data management practices result in well-organized, well-documented, and well-preserved research data, allowing for ease in finding, accessing, and understanding of the data by current researchers as well as other interested parties at a later date. Research data that is freely shared provides others with the opportunity to re-use it, thus preventing duplication of effort, saving not just time and effort, but money as well.

This is one reason more and more funding agencies want to know how researchers are managing their data, and are beginning to require data sharing. Since 2003, the National Institutes of Health (NIH) has mandated data sharing for most grants receiving over \$500,000 a year (National Institutes of Health 2003). Since 2011, The National Science Foundation (NSF) has expected their grant proposals to include data management plans (National Science Foundation 2010). In 2013, the White House Office of Science and Technology Policy (OSTP) issued a memorandum stating that data collected through federally funded scientific research should be shared (Holdren 2013).

Another growing trend is the expectation of publishers in the sciences for authors to share their data. Some require data be submitted as supplementary information. Many do not host the data themselves, but instead direct authors to deposit their data in specified repositories (American Association for the Advancement of Science 2014; Dryad 2014; Macmillan Publishers Limited 2015; PLOS ONE 2014).

At research universities, faculty advisors may be the visionaries, grant writers, and authors of data management plans, but graduate students are the ones collecting much of the data and managing the daily operations in the laboratory. How they handle the data impacts its quality and usability (Carlson and Stowell-Bracke 2013), and yet, they may arrive at graduate school unprepared for such responsibility; therefore, it is important that they are properly trained in research data management. The frequent turnover in research groups as new students matriculate and other students graduate is another motivation for well-organized, properly backed up, and well-documented data. This ensures that the research group's data will be easy to find and access rather than become lost. Documentation helps other group members understand the data — a crucial part of data sharing. Given that today's graduate students will become tomorrow's scientists, it is vital they develop good data management habits early on in their research careers.

To support data management and data sharing, universities and colleges are providing training and consultation on the former and creating institutional repositories for the latter. Academic libraries are typically leading these initiatives (Tenopir, Birch, and Allard 2012). Increasingly, new Data Services or eScience Librarian positions are being created. Even positions with

more traditional job titles now have job descriptions that include mention of data. Typically, Data Services Librarians, Metadata Librarians, Subject Specialist Librarians, and/or Digital Library Services must work together as a team to address the complexity of managing and reserving research data in the many different disciplines and in the many possible formats (Antell et al. 2014; Starr et al. 2012).

In addition to teaching (traditional) information literacy skills, librarians are now teaching data information literacy skills, which “builds upon and reintegrates data, statistical, information, and science data literacy” (Carlson et al. 2011). Although audiences for this training might include faculty, graduate students, undergraduate students, and other librarians, this article focuses on graduate students. There are a multitude of ways that graduate students can be taught data information literacy skills — via a separate course (Carlson et al. 2011; Qin and D’Ignazio 2010), general or discipline-specific seminars or workshops (Adamick, Reznik-Zellen, and Sheridan 2012; Massachusetts Institute of Technology 2014; University of Minnesota 2014), and/or online tutorials (Hswe and Musser 2014; Piorun et al. 2012). However, due to the varying nature of research data collected and data management needs, one size does not fit all. While introductory material may be appropriate for a broad audience new to data management, students prefer more specific details directly relevant to their discipline of study (Adamick, Reznik-Zellen, and Sheridan 2012). Thus, training tailored to students in a specific department or even research group is likely to be most effective.

This article discusses how the authors (a Physical Sciences Librarian and Data Services Librarian) are meeting the data training needs of graduate students in one particular science department on their campus by bringing to the collaboration their individual areas of expertise. It describes the four approaches taken thus far to ensure students in the department will be data information literate prior to degree completion.

Background

On the research intensive Rutgers University – Newark campus, the Research Office and the John Cotton Dana Library pay close attention to the growing need for data management training. They began co-sponsoring research data management workshops for faculty, oftentimes including discussion about the NSF DMP requirement. It was the announcement for one of these workshops in fall 2011 that caught the eye of the Department of Earth & Environmental Sciences’ (DEES) graduate program director. Unable to attend, the director contacted the organizers for copies of distributed handouts. He was provided with the slides, but was also told that a separate workshop could be presented to his department — including graduate students — if there was interest. His response was very enthusiastic as members of the department were highly interested in improving their research groups’ research data management skills — especially in the areas of data files organization, documentation, and back-up. Thus began an ongoing, mutually edifying relationship between the authors and DEES regarding data management.

Four Approaches

Seminar Series

Like many science departments in research universities, DEES has a weekly seminar series

that graduate students are required to attend; faculty usually sit in. Typically, presentations are about current scientific research, with presenters ranging from scientists from around the world, to graduate students in DEES. Occasionally, guest speakers have included librarians and grant specialists.

The authors' first instance of data management training for the department's graduate students (and faculty) was during one of these 90-minute seminar series sessions in January 2012. Per the graduate program director's request, the session emphasized the basics such as the importance of data management and some tips on how to do it. Materials from the original faculty workshop were modified to address these needs and then tailored to researchers in the Earth and Environmental Sciences. The students (and faculty) in attendance were encouraged to contact the authors with any data-related questions they might have. They were also provided with a list of additional training opportunities available for free online by other institutions and disciplinary service organizations. For slides used and handouts provided during the session, see http://libguides.rutgers.edu/data_EES (Fong and Wang 2013).

Based on positive responses received from this session, the authors, graduate program director, and another interested DEES faculty member discussed additional opportunities for further data management training that would address the specific needs and concerns of both the faculty and graduate students in DEES. The authors also began offering data management workshops targeted at the general graduate student population on campus.

Data Profile Form: Understanding Current Data Management Practices & Concerns

In an effort to better understand the data needs of DEES, a Data Profile form was developed in February 2013 (see Appendix 1) to gather information about the type of research data DEES generates and/or collects (e.g., scope of data, types of data, data formats, and approximate size of files), to learn how the data is managed, and to find out what data management challenges the student researchers faced. This form was adapted from Purdue University's "Data Curation Profiles Toolkit" (2012) — with modifications made by the authors, graduate program director (who had since also become Chair of the department), and another DEES faculty member — in order to address local needs. Changes included shortening the form due to concerns that the length of the original form and some of the more advanced topics on the form might be overwhelming to the graduate students.

Graduate students in the department were expected to complete the forms about their current research projects. There were a few reasons why the students were asked to do this rather than the faculty. The students were the ones collecting much of the data in DEES, so their forms would provide a good overview about the research performed in DEES. It was also the hope that filling out the forms would encourage the students to reflect on their projects and how they manage their data. Additionally, there was great interest in knowing what the students' perspective was, what issues they were currently having and expecting to have in the future concerning research data management. The authors generated an example of a completed Data Profile form with fictitious project information that provided guidance for the students as they filled out the form. Although the authors offered to meet with students individually to go through the form and complete it together, only one student took advantage of this opportunity.

Altogether, 13 Data Profiles were collected and analyzed by the authors, then summarized for the DEES Chair and other faculty member. There were a few students who did not complete the form, but based on those who did, it seemed most of the research in DEES focuses on near-surface geophysics, with fewer students studying environmental geochemistry, atmospheric chemistry, and environmental science. Interestingly, the student studying environmental science had a faculty advisor in the related Department of Biological Sciences. As might be expected, students indicated they received internal (departmental or university) support for their research. Only one student's project was supported by a corporate sponsor. Generally, external funding came primarily from NSF or from other federal agencies. This is significant because both faculty and graduate students must adhere to any research data management and data sharing requirements of their funding sources.

Research data in DEES is quite diverse, with some gathered via laboratory work and others via field work. Some projects consist of image files, some use spreadsheets. Projects may result in small files (1 KB) or larger files (2 MB) and range from just a dozen different files to a few hundred files. Back-up frequency also varied quite a bit, from those who backed up daily to those who did so annually. Where data was being backed-up differed, too, with an assortment of devices being used, such as portable drives, desktop computers, Dropbox, or the department server.

Challenges noted by graduate students were similar to concerns expressed by faculty. Files organization — especially file directories — was most worrisome. The quantity of data files collected overwhelmed a few graduate students. Others pointed to a need for standardized file naming conventions and hoped that would also solve versioning confusion. Faculty agreed that standardization was needed, indicating they were sometimes unable to determine how far along students were in processing data due to lack of a clear files organization system. Although students understood the need to back-up their data, some were uncertain about how to handle backing up different versions. A few wanted a space where they could back-up their files. Faculty recognized this need and were actively seeking a secure solution. One student acknowledged the need for better data documentation to help understand the data. Another discussed data sharing issues; specifically, where to place data so it would be accessible to external collaborators, as well. The students were eager for data management guidelines and standardization, recognizing that it would help make their research data easier to find not just for themselves, but for their faculty advisor, other internal collaborators, and external collaborators.

To address student concerns, the authors were invited to speak at the annual DEES graduate student retreat about files organization, documentation, and metadata. To answer questions about backing-up, data storage, and secure external data sharing, the authors facilitated a meeting between DEES and the Computing Services department on campus to discuss possible solutions.

Graduate Student Annual Retreat

In May 2013, the authors presented a three-hour workshop during the DEES graduate student annual retreat. It was the largest component of the retreat that year, reflecting just how important data management was to the department. In fact, the authors were the only speakers from outside the department. Attendance at the annual retreat is required of all

students receiving departmental support, although all graduate students are encouraged to participate. Faculty are also invited to participate and the authors were pleased to hear from one that he did so mainly because of the data discussion. The workshop addressed specific challenges students indicated on their Data Profile forms — data organization and documentation. The learning objectives of the workshop were to be able to:

1. Connect the importance of proper data management with the continuation of the research lifecycle.
2. Develop a consistent file naming system that is descriptive, but also succinct.
3. Construct a file directory structure to logically organize research-related files.
4. Identify elements of metadata and documentation to include with research data files.

Since the retreat was meant to be interactive, the approach taken by the authors encouraged participation from the students (and faculty). The workshop began with short student presentations about their research workflow and data management strategies. This was one of their “homework” assignments in preparation for the retreat (see Appendix 2 for details). Students were then asked how easy or difficult they thought it would be for someone else to locate their research data or understand how they analyzed their data seven years later. Faculty were asked about their experience regarding these same situations. Unsurprisingly, students were more confident about others’ ability to find and understand their data and analysis than faculty experience showed to be true. One of the graduate students who worked with an old dataset discussed why she did so, what challenges she faced, and how it led her to reconsider her own methods of research data management so that it would be easier for someone else to re-use her data years later. Small group discussions allowed time to reflect on this. Lectures about best practices for file naming, file directories, metadata, and documentation followed (see http://libguides.rutgers.edu/data_EES, Fong and Wang 2013). To wrap up the workshop, attendees once again convened in small groups to apply what they just learned. Students who remembered to bring their laptops and research data files received input from those in their group about how they could improve their file organization and documentation.

Based on verbal feedback from students and faculty present, the workshop was well-received. Plans were made for additional data management training in the future, with suggested topics being to identify appropriate metadata standards to follow and subject repositories for data deposits.

Data Management Plan Assistance

In all instances of data skills training, the authors encouraged faculty and students to contact them for assistance with creating a data management plan — whether for a grant proposal, a specific project, an entire research group, or a department. Several faculty and one graduate student from DEES took advantage of this offer when working on their (primarily NSF) grant proposals.

The graduate student required assistance with her NSF Graduate Research Fellowship application. The authors found that her research was closely related to an ongoing NSF-supported research project and that project website included information about a research sample registry system, research data display file specifications, and shared vocabularies, so these were shared with the student with the recommendation that she consider following them. The student did indeed integrate this information and project-specific data management tools and standards into the data management plan portion of her application.

In the case of a multi-institution research consortium grant proposal a couple of the faculty in DEES were applying for, one of the authors (Physical Sciences Librarian) was invited to be the data manager and the other (Data Services Librarian) was invited to serve on the data team. This would have been at 15% time for each librarian. Responsibilities would include providing training to all researchers — especially graduate students — involved in the grant. It is expected that faculty and students will continue to seek advice from the authors in this capacity.

Future Directions

Building on their experience, the authors will persist in assisting DEES faculty and students with DMPs for their grant proposals and plan to do the same for other departments on campus. Perhaps this will lead to opportunities for required data management training for students in other programs. It is the authors' hope that researchers in other departments will hear about the training for DEES graduate students and approach them for assistance to improve data management practices within their own research teams. Given the positive experience training the DEES students, the authors would use it to serve as a model for expanding similar services to other interested departments.

There has also been some discussion with the DEES chair about the possibility of establishing data management guidelines with standardized procedures for all DEES members to follow for new and/or existing projects. Of course, with the variety of research performed in the department, it might be more helpful to address the specific data management needs of each research group. This is an approach the authors are considering; however, sustainability is a concern due to the limited time librarians have to be deeply involved with research projects.

Summary

Data information literacy skills have become increasingly important for researchers — especially those in the sciences — due to changing grant, publisher, and/or institutional requirements. Oftentimes the primary data collectors, graduate students must be properly trained on data management and ideally this would occur early on in their research careers.

When there is no room in the graduate curriculum to add a data management component and there is not enough demand to sustain a separate course on the topic, alternative means for teaching graduate students data information literacy skills must be explored. Two possible venues are weekly seminar series and departmental retreats where graduate student attendance is required. This ensures all students are proficient in important techniques. Seminars are more suitable for lecture-style teaching, whereas retreats can allow for hands-on practice and reflective small group discussions.

Buy-in from students and their faculty advisors for such training is vital, and requires the belief that the training will truly benefit both parties. The first step towards this goal is to gain a better understanding of what exactly their data needs are — which can be accomplished through conversations with the students and faculty and/or completion of Data Profile forms by the same groups. This must then be followed by a clear addressing of specific concerns. The authors believe this is the primary reason why their work with DEES has been so well-received. Once there is buy-in as to the importance of data skills training and the expertise of librarians in this field, librarians can expect continual contact.

Of course, while subject specialist librarians may offer expertise in disciplinary knowledge, data librarians have a broader vision and a general sense of data trends. In addition to their background in multi-disciplinary research data handling, data librarians also have the ability to do in-depth research into data management in specific subject areas. Together, the data and subject specialist librarians can identify the most important data management best practices, standards, and systems for their researchers, resulting in a data support team that is able to properly address the data management training needs of graduate students and their faculty.

Supplemental Content

Appendices 1 and 2

An online supplement to this article can be found at <http://dx.doi.org/10.7191/jeslib.2015.1067> under “Additional Files”.

Acknowledgement

The authors would like to acknowledge the faculty and graduate students in the Rutgers University – Newark Department of Earth & Environmental Sciences who participated in this initiative and thank them for their enthusiasm.

Disclosure

The authors report no conflict of interest.

References

- Adamick, Jessica, Rebecca C. Reznik-Zellen, and Matt Sheridan. 2012. “Data Management Training for Graduate Students at a Large Research University.” *Journal of eScience Librarianship* 1:180–188. <http://dx.doi.org/10.7191/jeslib.2012.1022>
- American Association for the Advancement of Science. 2014. “General Information for Authors: Data Materials and Availability.” *Science*. Accessed December 9. http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail
- Antell, Karen, Jody Bales Foote, Jaymie Turner, and Brian Shults. 2014. “Dealing with Data: Science Librarians’ Participation in Data Management at Association of Research Libraries Institutions.” *College & Research Libraries* 75:557–574. <http://dx.doi.org/10.5860/crl.75.4.557>
- Carlson, Jacob, Michael Fosmire, C.C. Miller, and Megan Sapp Nelson. 2011. “Determining Data Information Literacy Needs: A Study of Students and Research Faculty.” *portal: Libraries and the Academy* 11:629–657. <http://dx.doi.org/10.1353/pla.2011.0022>

- Carlson, Jake, and Marianne Stowell-Bracke. 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* 13:343–361. <http://dx.doi.org/10.1353/pla.2013.0034>
- Dryad Digital Repository. 2014. "Joint Data Archiving Policy (JDAP)." Last modified November 21. Accessed December 9. <http://datadryad.org/pages/jdap>
- Fong, Bonnie, and Minglu Wang. 2013. "Earth & Environmental Sciences – Data Management." *Rutgers University Libraries*. Last modified May 14. http://libguides.rutgers.edu/data_EES
- Holdren, John P. 2013. "Increasing Access to the Results of Federally Funded Scientific Research." *White House Office of Science and Technology Policy*. Released February 22. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- Hswe, Patricia, and Linda Musser. 2014. "Data Management Plan Tutorial." *Penn State University Libraries*. Accessed December 9. <https://www.e-education.psu.edu/dmpt/node/508>
- Macmillan Publishers Limited. 2015. "Availability of Data, Material, and Methods: Availability of Data." *Nature*. Last modified April 15. <http://www.nature.com/authors/policies/availability.html#data>
- Massachusetts Institute of Technology. 2014. "Data Management: Workshops." Accessed December 9. <http://libraries.mit.edu/data-management/services/workshops>
- National Institutes of Health. 2003. "Final NIH Statement on Sharing Research Data." Last modified February 26. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
- National Science Foundation. 2010. "Dissemination and Sharing of Research Results." Last modified November 30. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- Piorun, Mary E., Donna Kafel, Tracey Leger-Hornby, Siamak Najafi, Elaine Martin, Paul Colombo, and Nancy R. LaPelle. 2012. "Teaching Research Data Management: An Undergraduate/Graduate Curriculum." *Journal of eScience Librarianship* 1:46-50. <http://dx.doi.org/10.7191/jeslib.2012.1003>
- PLOS ONE. 2014. "PLOS ONE Manuscript Guidelines: Data Reporting Guidelines." Accessed December 9. <http://www.journals.plos.org/plosone/s/submission-guidelines#loc-data-reporting>
- Purdue University. 2012. "Data Curation Profiles Toolkit." <http://datacurationprofiles.org>
- Qin, Jian, and John D'Ignazio. 2010. "Lessons Learned from a Two-Year Experience in Science Data Literacy Education." *International Association of Scientific and Technological University Libraries, 31st Annual Conference* Paper 5. <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>
- Starr, Joan, Perry Willett, Lisa Federer, Claudia Horning, and Mary Linn Bergstrom. 2012. "A Collaborative Framework for Data Management Services: The Experience of the University of California." *Journal of eScience Librarianship* 1:109-114. <http://dx.doi.org/10.7191/jeslib.2012.1014>
- Tenopir, Carol, Ben Birch, and Suzie Allard. 2012. "Academic Libraries and Research Data Services: Current Practices and Plans for the Future, An ACRL White Paper." *Association of College and Research Libraries (ACRL)*. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf
- University of Minnesota. 2014. "Data Management Course: Winter 2014 Workshop Series." <https://sites.google.com/a/umn.edu/data-management-workshop-series>