Data Management Education from the Perspective of Science Educators

Carol Tenopir
School of Information Sciences
University of Tennessee

Suzie Allard
School of Information Sciences
University of Tennessee

Priyanki Sinha
School of Information Sciences
University of Tennessee

Danielle Pollock
School of Information Sciences
University of Tennessee

Jess Newman
School of Information Sciences
University of Tennessee

Elizabeth Dalton
Department of Communication Studies
and Organizational Communication
Middle Tennessee State University

Mike Frame
U.S. Geological Survey

Lynn Baird
University Library
University of Idaho

Abstract

In order to better understand the current state of data management education in multiple fields of science, this study surveyed scientists, including information scientists, about their data management education practices, including at what levels they are teaching data management, which topics they covering, and what barriers they experience in teaching these topics. We found that a handful of scientists are teaching data management in undergraduate, graduate, and other types of courses, as well as outside of classroom settings. Commonly taught data management topics included quality control, protecting data, and management planning. However, few instructors felt they were covering data management topics thoroughly, and respondents cited barriers such as lack of time, lack of necessary expertise, and lack of information for teaching data management. We offer some potential explanations for the existing state of data management education and suggest areas for further research.

Introduction

Sound management of research data is of increasingly critical importance for science. Researchers have better tools to gather data, the magnitude of data collected has increased, and research reproducibility and accountability are of growing concern. Many funding agencies now require research data management plans (see Burwell, VanRoekel, Park and Mancini, 2013; European Commission, 2015; National Science Foundation, 2011; Wellcome Trust, 2010). Managing research data, including properly describing, archiving, preserving, and enabling access to data, furthers scientific discovery by facilitating data sharing and reuse by scientists (European Commission, 2013; National Science Foundation, 2011; Strasser and Hampton, 2012).

At the same time, scientists are finding it difficult to manage data for a number of reasons. Studies cite several complex factors impeding data management, including the abundance of digital data (Porter, Hanson and Lin, 2012); limited access to datasets, poor data quality, and lack of metadata (Specht et al., 2015); lack of standardization in data description and formats (Volk, Lucero and Barnas, 2014); differences in researchers' willingness to share data (Tenopir et al., 2015a); and a lack of proper data management education for scientists and researchers in the early phases of their careers (Jahnke and Asher, 2012).

This study provides a better understanding of the current state of data management education, with an international survey of scientists from multiple disciplines who teach data management. We asked about their data management education practices and addressed the following research questions:

- **RQ1:** Are scientists teaching data management topics to undergraduate, graduate, or other students?

- **RQ2:** What data management topics are being covered and do instructors feel the coverage is adequate?

- **RQ3:** Are information scientists more likely to teach certain data management topics than other teachers of science?

- **RQ4:** What are the barriers to teaching data management topics?

The Case for Data Management Education

Proper data management yields important benefits for scientists. It enables continued access to data for future scholarly research and communication, and saves time and resources that might otherwise be spent in duplicate data gathering efforts (Doucette and Fyfe, 2013; Shearer, 2010). Several funding agencies around the world, including in the United States, Australia, and Europe have recently stressed the importance of sound data management (Australian Government, 2007; Burwell et al., 2013; European Commission, 2015; Green et al., 2015; NSF, 2011; Wellcome Trust, 2010).

Managing large amounts of data that are available in different formats, however, is a complex and elaborate process and easier to mandate than to accomplish. Data management involves implementing standard scientific practices for accurate data collection, documentation, processing, analysis, and storage throughout the entire data

lifecycle (Strasser, Cook, Michener, and Budden, 2012), as shown in Figure 1. The rules for data collection, processing, and analysis vary across individual disciplines due to differences in research methods, and there is often a lack of understanding among researchers about acceptable practices of data management in disciplines that are outside a researcher's field (Coates, 2014).
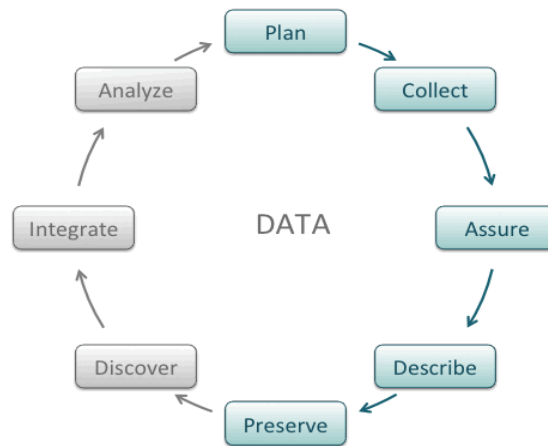
Figure 1. Data lifecycle by DataONE[1].

Data management challenges stem from the disparity of knowledge regarding best practices in data management among scientists. Research indicates a lack of understanding of core data management skills, such as citing datasets, creating metadata, archiving and preserving of data, and data sharing mechanisms among young scientists at the undergraduate and graduate level (Carlson et al., 2011). Even faculty members admit to having a gap in their knowledge of topics related to data management education at their institutions (Carlson et al., 2011). One study of science faculty at a teaching-centred university found that half (50%) of the survey respondents lacked confidence in their data management skills, and many needed guidance on topics such as creating metadata and writing data management plans (Scaramozzino, Ramírez and McGaughey, 2012).

Information scientists and librarians sometimes bridge this knowledge gap. In a recent study of academic libraries in North America 29.7% offer their faculty or students some sort of assistance with data management, which can include reference support for finding and citing data, consulting on data management plans, creating or transforming metadata, and other types of data management support (Tenopir et al., 2015b).

Undergraduate education is a period where students work directly with their instructors and learn core skills of the research process that they will be expected to possess as graduate students and as working researchers, including data collection, writing literature reviews, and analysis, interpretation, and presentation of data. According to Mooney et al. (2014), it is therefore extremely important at this stage for

---

[1] See: https://www.dataone.org/best-practices

them to learn the best practices of data management as a part of their undergraduate research experience.

Strasser and Hampton (2012), in a recent study of instructors across ecology departments at 48 academic institutions, found that many data management topics are not being covered at the undergraduate level. Some of the reasons cited for the absence of such instruction include lack of time, resources, and instructor knowledge. Ecology is an archetype of a highly interdisciplinary field that is data intensive and for which data repositories exist. Therefore, education in ecology could be seen as an exemplar of the state of data management education in the sciences.

The Strasser and Hampton (2012) study focused on understanding instructor and course characteristics, data management education in courses, and perceptions of instructors on the importance of data management topics. Survey results indicated that 100% of the instructors who were also active researchers had been encouraged to share data, 84% had engaged in data sharing, and 71% had reused data from others at some point in their careers. Instructors who placed more importance on data management in their own research also valued data management for undergraduate students. Quality assurance was the most commonly taught data management topic, addressed in 42% of the courses.

The current study expands on Strasser and Hampton (2012) by expanding the scope to multiple fields of science, including information science. This study is part of the National Science Foundation (NSF) sponsored DataONE project and focuses on scientists who also are educators teaching data management topics. The study asks about the variety of topics being taught, the barriers to teaching data management, and educators' satisfaction in teaching data management. The results of the study contribute to identifying best practices in data management education, as well as barriers that must be overcome.

Previous Studies of Data Management Education

Scientists rarely receive training in data management and preservation issues (Heidorn, 2011). This may be because there is a shortage of digital curation professionals with the skills required to train young scientists in the appropriate methods and procedures of digital curation (Poole, 2014).

Previous studies indicate that data management education is lacking at the graduate level. A study of civil engineering graduate students at the University of Minnesota found that most received little data management education and many learned the topics through informal modes of communication (Johnston and Jeffreys, 2014a). A survey of graduate students in the field of environmental sciences found that most had not taken courses in information sciences or advanced data analysis, and lacked both the computational skills necessary for analysing large data sets and experience in creating metadata (Hernandez, Mayernik, Murphy-Mariscal and Allen, 2012).

Although there is little evidence of data management being taught in science courses at the undergraduate level, cases exist of data management skills being taught as part of an undergraduate research laboratory course (Miller et al., 2013), and as a one-hour course for chemistry majors (Reisner, Vaughn and Shorish, 2014). The University of Sydney introduced electronic notebook keeping and data curation skills in undergraduate biochemistry and molecular biology courses (Johnston, Kant, Gysbers, Hancock and Denyer, 2014).

Within library and information science departments, a number of programs in data management and data curation have been developed with a variety of approaches for

educating students (Mayernik et al., 2014). A survey of 52 library and information schools in the United States and Canada found that 16 offered courses of data curation, while seven offered a specialization or concentration in data curation (Harris-Pierce and Liu, 2012). Another survey of 63 information science schools found programs for data professionals – including master's degree programs, certificate programs, and concentrations with an emphasis on data – at 17 institutions (Varvel, Bammerlin and Palmer, 2012). A University of Illinois survey of Graduate School of Library and Information Science alumni who graduated with a master's level certificate in Data Curation found that almost half were currently working in a position related to data curation, while most applied skills from the program in their current position (Thompson, Senseney, Baker, Varvel and Palmer, 2013).

This does not mean that all of those currently working as data professionals have taken courses in data curation during their graduate education, or are satisfied with their current skillset. Many current data professionals have come to their positions by 'accident' and bring a variety of backgrounds and levels of experience with data to their current roles (Mayernik et al., 2014; Poole, 2014). Two studies found most academic libraries in the U.S. and Canada that are offering or planning to offer research data services (RDS) were reassigning existing staff to support these services (Tenopir, Birch and Allard, 2012; Tenopir et al., 2015b). A survey of librarians in the United Kingdom found that skills gaps were a major challenge for Research Data Management (RDM) services, with over half of library staff stating that they did not have the correct skillset (Cox and Pinfield, 2014). Lack of knowledge and levels of anxiety in one study of subject librarians were highest for data-related topics such as data lifecycles, data management plans, and data sharing plans (Bresnahan and Johnson, 2013). As libraries increasingly plan to offer data management and curation services and education to researchers, these findings suggest that continuing education and training for those staff members with data-related responsibilities may be essential (Tenopir et al., 2012; Tenopir et al., 2015b).

Some continuing education programs have been developed for information science professionals who suddenly find themselves with data curation responsibilities. Training opportunities, including consulting services and distance education courses, were developed for information professionals by a national research data archive in the Netherlands (Dillo et al., 2014). The University of Edinburgh EDINA project, in conjunction with the University of North Carolina – Chapel Hill, has developed a MOOC on data management skills for librarians[2]. Kafel (2012) details the creation of data management education tools for librarians, including professional development programs, an e-Science portal, and an active community of interest. Organizations may offer also offer internal training in research data curation issues, though Tenopir et al. (2012) found that only a quarter of academic libraries did offer RDS training for existing staff.

Data Management in the Curriculum

As with any emerging field, educators must explore and define the topics that will comprise a data management curriculum. Data management topics identified in previous research as potentially important for undergraduate and graduate education in the sciences include quality control and assurance; data types and formats; data storage, backup, and security; legal and ethical considerations; metadata creation; data sharing

---

2 See: http://www.ed.ac.uk/news/2016/dataskills-010316

and reuse policies; programming skills and proper use of sensor technology; data archiving and digital preservation; and completion of data management plans (Hernandez et al., 2012; Johnston and Jeffreys, 2014b; Kafel, 2012; Strasser and Hampton, 2012).

Data management practices differ between disciplines and data management education will need to account for these differences. A survey of faculty in 50 graduate research programs in four disciplines related to biomedical sciences designed to assess levels of agreement for possible topics in a responsible conduct of research (RCR) curriculum found less than 50% agreement among respondents on topics such as data sharing and data retention practices (Kalichman et al., 2014). Faculty members in Kafel (2012) recommended the addition of real life research case studies from a range of science, health science, and engineering disciplines to a data management curriculum.

For data curation professionals, curriculum development should take into account real-world data position requirements. Here, technical skills, including those related to repository creation and maintenance, were emphasized frequently (Cox and Pinfield, 2014) as were skills related to the organization and management of data, such as appropriate metadata creation (Mayernik et al., 2014). Among other desired skills were those related to data use, including knowledge of copyright, open access, and proper citation of data (Cox and Pinfield, 2014). Knowledge of current data curation trends was also mentioned (Thompson et al., 2013), and related to this, knowledge of existing repositories (MacMillan, 2014), as well as funders' current data management policies and the creation of appropriate data management plans (Antell et al., 2014). Specific skills needed by data professionals may vary by role or position. Lyon, Mattern, Acker, and Langmead (2015), in a recent curriculum mapping study, found that some skills – including understanding of researcher perspective, knowledge of metadata standards and schema, competence with statistical/analysis software, and knowledge of disciplinary data – were required in for all data science roles under investigation, while other skills varied by position. For example, librarian roles typically required knowledge of funding agency data requirements, while other roles did not.

Methodology

The purpose of this study is to investigate the existing state of data management education from the perspective of scientists globally who actively engage in teaching data management.

Data for this study come from questions that are a subset of a larger worldwide survey of data management practices and opinions among scientists (Tenopir et al., 2015a). Participants in this survey included research scientists and science faculty working in academic institutions, research organizations, federal agencies (e.g., Oak Ridge National Laboratory, National Aeronautics and Space Administration [NASA], National Center for Ecological Analysis and Synthesis), and non-profit organizations in the hard sciences (atmospheric science, biology, geology, hydrology, etc.), social sciences, humanities, and law. The survey was administered using Qualtrics software and distributed by DataONE team members via email to deans, department chairpersons, and research directors at academic institutions and research organizations worldwide. The email contained a link to the survey questionnaire, which these contacts were asked to forward to faculty, lecturers, post-doctoral research associates, graduate

students, undergraduate students, and researchers within their organizations. The survey was also made available on several environmental science blogs and listservs.

Of the 1,015 respondents to the larger survey, 134 indicated that they teach data management. These 134 were given an additional set of questions specifically related to teaching data management. This paper is based on the analysis of responses to these teaching-related questions. Appendix A contains the subset of questions analysed in this paper. The core survey is available at Tenopir et al. (2015a). Data for this study were collected from October 17, 2013 through March 19, 2014. The survey was approved by the (authors' institution's) Institutional Review Board for Human Subjects. Respondents were allowed to skip any question, so not all 134 answered every question. Analysis of each question is based on the number of responses for that question.

Data Analysis and Results

Of the 130 respondents who provided data about their location, most were located in North America (53.8%), followed by Europe (16.2%), Africa (10.8%), Asia (7.7%), South America (7.7%), and Australia/New Zealand (3.8%), as shown in Figure 2.
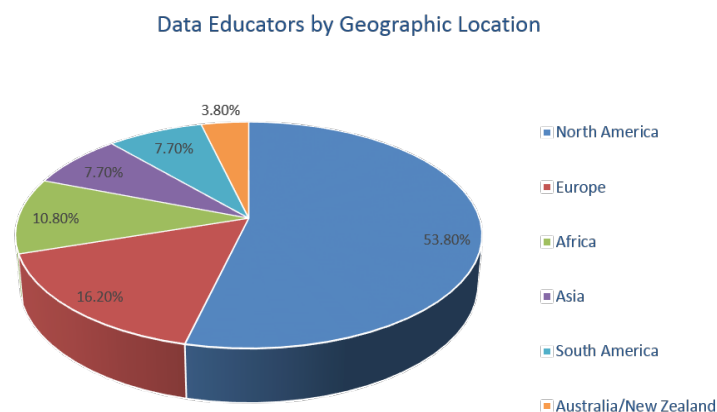

Figure 2. Geographic location of data management educators.

Most of the data management educator respondents reported their primary work sector was academic (78.8 %), followed by government (13.6 %), commercial (3.8 %), non-profit (2.3 %), and other (1.5%). This distribution is similar to the full survey, in which the majority of respondents reported their primary work sector as academic (80.5%) followed by government (12.7 %), commercial (2.6 %), non-profit (2.7 %), and other (1.6%) (Tenopir et al., 2015a).
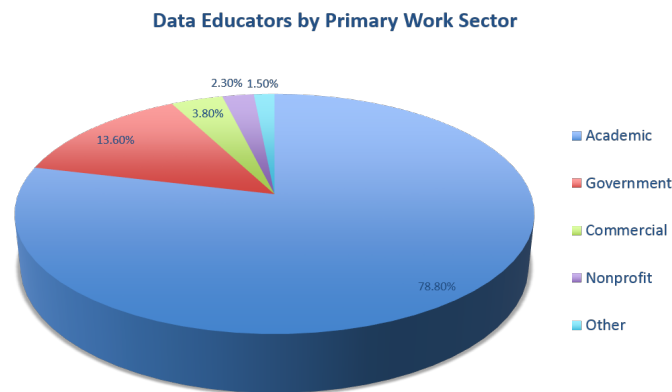
Figure 3. Primary work sector of data management educators.

The primary subject disciplines of a majority of the educator respondents differed from that of the full survey (Tenopir et al., 2015a). The largest primary subject discipline reported for data management educators was information sciences (17.6 %). This was followed by ecology (16.8 %) and environmental science (14.5 %), primary disciplines targeted by DataONE, which made up a large percentage of responses to the full survey (Tenopir et al., 2015a). Other primary subject disciplines reported by data management educators include agriculture and natural resources (9.9 %), social sciences (8.4 %), atmospheric science (5.3%), biology (5.3%), and medicine and health sciences (4.6%). No other primary subject discipline was reported by more than 4% of the respondents, as shown in Figure 4.
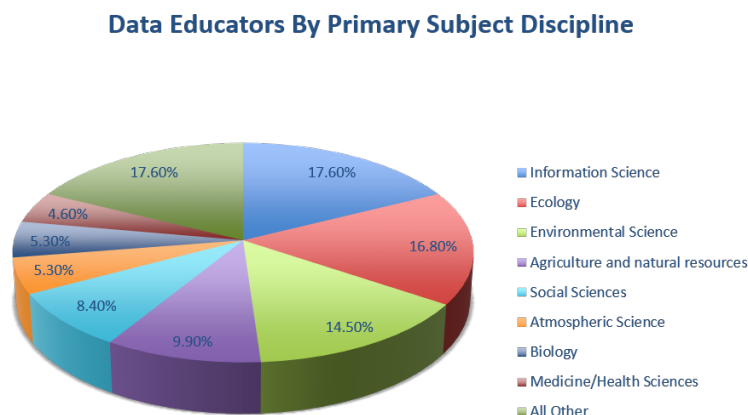
Figure 4. Primary subject discipline of data management educators.

Respondents were first asked whether their data management teaching occurs in courses, outside the classroom, or both. Next, respondents were given a list of data management topics, adapted from the earlier Strasser and Hampton study (2012) and asked to indicate which they include in their teaching in undergraduate, graduate, and other course settings, or outside the classroom. Nearly one quarter (24.8%) of

respondents reported that they teach data management exclusively within graduate, undergraduate or other types of courses, while more (31.6%) reported teaching data management exclusively outside the classroom. Over 40% teach data management both within and outside of courses (see Table 1).

Table 1. Teaching data management.

Data Management Teaching	Frequency Distribution
In Courses	24.8% (n=33)
Outside the Classroom	31.6% (n=42)
Both	43.6% (n=58)

Data management instruction done by these respondents takes place least frequently at the undergraduate level (see Table 2). The most commonly covered data management topics at the undergraduate level are Quality Control (21.6%), File Management (20.1%), and Citing Data (19.4%), yet even these topics are covered by only about one-fifth of the instructors.

Table 2. Data management topics and distribution.

Topics	Undergraduate Courses	Graduate Courses	Other Types of Courses	Outside of Courses
Data Lifecycle	17.2% (n=23)	31.3% (n=42)	10.4% (n=14)	29.9% (n=40)
Data Management Planning	14.2% (n= 19)	32.8% (n=44)	13.4% (n=18)	36.6% (n=49)
Quality Control	21.6% (n=29)	41% (n=55)	12.7% (n=17)	45.5% (n=61)
File Management	20.1% (n=27)	29.9% (n=40)	11.9% (n=16)	39.6% (n=53)
Metadata Generation	16.4% (n=22)	23.1% (n=31)	9% (n=12)	40.3% (n=54)
Workflow	11.2% (n=15)	18.7% (n=25)	9% (n=12)	29.9% (n=40)
Protecting Data	14.9% (n=20)	27.6% (n=37)	8.2% (n=11)	43.3% (n=58)
Data Archiving and Preservation	9% (n=12)	26.1% (n=35)	9.7% (n=13)	35.8% (n=48)
Data Reuse	13.4% (n=18)	26.1% (n= 35)	9.7% (n=13)	29.9% (n=40)
Meta-Analysis	11.2% (n=15)	19.4% (n=26)	5.2% (n=7)	21.6% (n=29)
Citing Data	19.4% (n=26)	34.3% (n=46)	10.4% (n=14)	38.8% (n=52)
Other	2.2% (n=3)	4.5% (n=6)	3% (n=4)	5.2% (n=7)

More teaching of data management topics takes place in graduate courses (Table 2). The coverage of specific topics in graduate courses is similar to those taught at the undergraduate level. The most frequently taught topic at the graduate level is Quality Control (41%), followed by Citing Data (34.3%), Data Management Planning (32.8%), and Data Lifecycle (31.3%). No other topic is taught at the graduate level by more than 30% of the respondents.

Data management topics are also taught in other formal education courses that perhaps do not lead to a degree. A few respondents indicated that they teach data management in courses other than undergraduate or graduate courses. The most

common topics in these 'other' courses are Data Management Planning (13.4%) and Quality Control (12.7%).

A number of respondents indicated they taught data management topics outside formal courses. Topics including Quality Control, Protecting Data, Metadata Generation, and File Management are taught by at least 40% of the total respondents outside of courses, while many other topics are taught outside of courses by one-third or more of the respondents (see Table 2).

The questions in the survey also focused on understanding opinions of educators about data management education. There is a divergence of opinion about whether instructors feel they are covering data management topics sufficiently. Only 9% feel they are covering the subject thoroughly and don't plan to increase coverage, yet a vast majority imply that they could or should be covering topics more thoroughly (see Table 3).

Table 3. Do you feel that you are covering these topics sufficiently?

Opinions	Frequency Distribution
Yes, thoroughly (I wouldn't add anymore)	9.1% (n=12)
Yes, but there is more that I could add	30.3% (n=40)
Yes, minimally	23.5% (n=31)
No, I should add more	34.8% (n=46)
No, and I don't plan to add more	2.3% (n=3)

There is a range of reasons why data management is not taught sufficiently. Respondents were given a list of barriers to teaching data management and asked to indicate any they had experienced. The top three barriers to teaching data management topics reported by these respondents are: lack of time (51.5%); that data management is not the respondent's area of expertise (39.6%); and lack of information to teach data management (30.6%) (see Table 4).

Table 4. Barriers to teaching data management.

Barriers to Teaching Data Management	Frequency Distribution
There is no time to teach data management	51.5% (n=69)
I don't have enough information	30.6% (n=41)
It is not my area of expertise	39.6% (n=53)
It is not appropriate for the level I teach	9.7% (n=13)
It isn't relevant to the courses I teach	10.4% (n=14)
Students are getting this information in other ways	10.4% (n=14)
Other	14.2% (n=19)

Conclusion and Implications

Limitations arise from the fact that the sample size of this study is small (n=134) and represents a subsection of a volunteer sample. Due to the distribution method of the original survey it is impossible to represent a response rate, or to claim that this sample is representative of the entire population of data management educators. Further investigation is required in order to more closely assess the state of data management

education worldwide in all scientific disciplines. However, these results still provide useful insights and implications for science educators and those who work with science educators on data management topics. For instance, major barriers to effective data management education have been identified, as over half (51.5%) of the data management educators cite lack of time as one of the major barriers to teaching data management topics, and over 39% feel that, despite having responsibilities for data management education, they lack the necessary expertise to teach the topics at hand. Over 30% feel they don't have enough information to teach data management topics.

Data management instruction assistance, such as that which can be provided by trained data managers or data librarians, can help science educators with both the lack of time and expertise barriers. The fact that over 17% of the data management educators in this survey reported their primary subject discipline as information science indicates that professionals with information expertise are already involved in teaching science data management topics. Collaborative teaching between trained data management experts and those with expertise in domain sciences can introduce data management topics into a variety of science classes. This expertise and collaboration may not be widely available in all settings. However, shared data education materials are being developed to assist science educators with the lack of expertise and lack of information barriers. For example the data management education modules available on the DataONE website[3] were developed to be used by science educators in a variety of settings. Similarly, a guide has been created to assist those developing data management plans in which several potentially useful tools are identified for various stages of data management planning (Michener, 2015).

Our results are similar to those reported by Strasser and Hampton, in that this survey also found that lack of time is a major barrier for educators in teaching data management (Strasser and Hampton, 2012). However, in Strasser and Hampton, which focused exclusively on undergraduate ecology courses, the beliefs that data management topics were not appropriate for the level of the course being taught and these topics had or should have been covered by a lab section were also cited as barriers, while lack of instructor knowledge was cited by a smaller percentage of the respondents (Strasser and Hampton, 2012). While specific barriers to teaching data management may vary by environment and level of education, our findings indicate that the time devoted to data management in science education overall may be insufficient, and that the content of data management related courses and programs may be lacking. There is a need for more training and resources for data management educators themselves, as well as for the practicing scientists, faculty, and students who are being served by these instructors. The DataONE education modules are a good first step. Another approach being practiced in the United Kingdom is the introduction of research data management education for postgraduates in multi-partner doctoral training centres, such as the Doctoral Training Centre in Sustainable Chemical Technologies at the University of Bath (Pink, 2013).

The most common data management topics being taught at the graduate level and outside of formal courses are: quality control, citing data, and protection of data. This differs slightly from Strasser and Hampton, who also found that quality assurance was the most commonly taught topic at the undergraduate level, but that other common topics were data reuse, data sharing, and reproducibility of data (Strasser and Hampton, 2012). Our review of the literature indicates that there may be differences in the priority assigned to specific data management topics across disciplines and across levels of education, and our results may be indicative of these differences. Additional research

---

3   DataONE Education Modules: https://www.dataone.org/education-modules

with data management educators and students could help determine whether data management topics being addressed in and outside of formal courses are appropriate to the needs of a specific population.

Data management education is an emerging practice, with a majority of those who do teach it offering only limited topics or coverage. Instructors indicate they do not feel comfortable teaching data management topics in which they lack expertise, and they lack the time to add data management topics to their existing courses or workloads. Much of data management education currently is occurring outside of a formal classroom setting. Quality control/quality assurance is one of the most important data management topics taught, mirroring good science practices to ensure the highest quality of data. However, topics such as creating metadata, archiving, and preservation still need more focus.

Awareness regarding best practices in data management is still in its infancy. The continued improvement of the state of data management education depends on the widespread implementation of policies related to creation of metadata, open access to data, data sharing, preservation, and archiving. Many mandates from government and private funding agencies have only recently been enacted, and many institutional libraries and research offices have just started focusing on good data management practices as they are now forced to consider long term data curation. It is perhaps not surprising that there is little data instruction, given that the culture of data sharing and reuse is still in its formative stages for many disciplines.

In some fields, subject data repositories have been available for decades; for others, they are a fairly new phenomenon. Some data repositories are institutional or subject silos with restricted access to data. Some steps are being taken to bridge disciplinary divides created by these silos. For example, specialized training workshops on best practices in data management to scientists and researchers are sometimes offered through libraries. Sharing data across institutions and across the boundaries of subject discipline is another important consideration impacting many data management topics. While this study focused on data management education in the natural sciences, it is important to note that data management is also becoming an increasingly important in the social sciences, arts, and humanities, and the data management education needs of researchers and information professionals working in these disciplines represents an area ripe for future research.

As data management and data sharing are growing concerns across all disciplines of science, the need for appropriate data management education at all levels of scientific education and training is increasing. Information science has taken a lead in this important area of teaching, but it must be a collaborative effort across the sciences. Increasing investment in data management education is needed to benefit scientists, educators, and ultimately, scholarship. With the increase in data management requirements by federal and other funding agencies, sound data management education is imperative.

Acknowledgements

Pjesviac, Grady College of Journalism and Mass Communication, University of Georgia, Athens, Georgia.

References

Antell, K., Foote, J., Turner, J., & Shults, B. (2014). Dealing with data: Science librarians' participation in data management at Association of Research Libraries institutions. *College & Research Libraries*, *75*, 557-574. doi:10.5860/crl.75.4.557

Australian Government, National Health and Medical Research Council, Australian Research Council, and Universities Australia. (2007). *Australian code for the responsible conduct of research*. Canberra: Australian Government. Retrieved from http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf

Bresnahan, M.M., & Johnson, A.M. (2013). Assessing scholarly communication and research data training needs. *Reference Services Review*, *41*, 413-433. doi:10.1108/RSR-01-2013-0003

Burwell, S.M., VanRoekel, S., Park, T., & Mancini, D.J. (2013). Open data policy – Managing information as an asset (M-13-13). Washington, DC: Office of Management and Budget. Retrieved from https://project-open-data.cio.gov/policy-memo/

Carlson, J., Fosmire, M., Miller, C.C., & Sapp Nelson, M.R. (2011). Determining data information literacy needs: A study of students and research faculty. *Libraries Faculty and Staff Scholarship and Research.* Paper 23. Retrieved from http://docs.lib.purdue.edu/lib_fsdocs/23/

Coates, H.L. (2014). Ensuring research integrity: The role of data management in current crises. *College & Research Libraries News*, *75*, 598-601. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-08425-1_23

Cox, A.M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*, 299-316. doi:10.1177/0961000613492542

Dillo, I., Van Horik, R., & Scharnhorst, A. (2014). Training in data curation as service in a federated data infrastructure: The FrontOffice-BackOffice model. In L. Bolikowski, V. Casarosa, N. Houssos, P. Manghi, & J. Schirrwagen (Eds). *Theory and Practice of Digital Libraries - TPDL 2013 Selected Workshops* (pp. 205-215). doi:10.1007/978-3-319-08425-1_23

Doucette, L. & Fyfe, B. (2013). Drowning in research data: Addressing data management literacy of graduate students. Presented at the Association of College and Research Libraries 2013 Conference, Indianapolis, IN. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/DoucetteFyfe_Drowning.pdf

European Commission. (2013). Commission launches pilot to open up publicly funded research data [Press release]. Retrieved from http://europa.eu/rapid/press-release_IP-13-1257_en.htm

European Commission. (2015). *Guidelines on data management in Horizon 2020*. Retrieved from https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Green, E., Krumholz, H., Altman, R., Bauchner, H., Brooks, D., Fridsma, D., … Jorgenson, L. (2015). *National Library of Medicine Working Group Final Report*. Retrieved from http://acd.od.nih.gov/reports/Report-NLM-06112015-ACD.pdf

Harris-Pierce, R.L., & Liu, Y.Q. (2012). Is data curation education at library and information science schools in North America adequate? *New Library World*, *113*, 598-613. doi:10.1108/03074801211282957

Heidorn, P.B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, *51*, 662-672. doi:10.1080/01930826.2011.601269

Hernandez, R.R., Mayernik, M.S., Murphy-Mariscal, M.L., & Allen, M.F. (2012). Advanced technologies and data management practices in environmental science: Lessons from academia. *BioScience*, *62*, 1067-1076. doi:10.1525/bio.2012.62.12.8

Jahnke, L.M., & Asher, A. (2012). *The problem of data: Data management and curation practices among university researchers*. Washington, DC: Council on Library and Information Resources. Retrieved from http://www.clir.org/pubs/reports/pub154/pub154.pdf

Johnston, J., Kant, S., Gysbers, V., Hancock, D., & Denyer, G. (2014). Using an e-portfolio system as an electronic laboratory notebook in undergraduate biochemistry and molecular biology practical classes. Biochemistry and Molecular Biology Education, 42, 50-57. doi:10.1002/bmb.20754

Johnston, L., & Jeffryes, J. (2014a). Data management skills needed by structural engineering students: Case study at the University of Minnesota. *Journal of Professional Issues in Engineering Education and Practice*, *140*(2). doi:10.1061/(asce)ei.1943-5541.0000154

Johnston, L., & Jeffryes, J. (2014b). Steal this idea: A library instructors' guide to educating students in data management skills. *College and Research Libraries News*, 75, 431-434. Retrieved from http://crln.acrl.org/content/75/8/431.full

Kafel, D. (2012). Activities of regional consortia in planning e-science continuing education programs for librarians in New England. In N. Xiao & L. R. McEwan (Eds.), *Special Issues in Data Management, 1110* (pp. 69-96). doi:10.1021/bk-2012-1110.ch005

Kalichman, M., Sweet, M. & Plemmons, D. (2014). Standards of scientific conduct: Disciplinary differences. *Science and Engineering Ethics*. Advance online publication. doi:10.1007/s11948-014-9594-0

Lyon, L., Mattern, E., Acker, A., & Langmead, A. (2015). Applying translational principles to data science curriculum development. In Lee, C., Crabtree, J., Konstantelos, L., McGovern, N., Maeda, Y., Pennock, M., … Zierau, E. (Eds.) *iPres Conference Proceedings* (pp. 109-117). Retrieved from https://phaidra.univie.ac.at/view/o:429524

MacMillan, D. (2014). Data sharing and discovery: What librarians need to know. *Journal of Academic Librarianship*, *40*, 541-549. doi:10.1016/j.acalib.2014.06.011

Mayernik, M. S., Davis, L., Kelly, K., Dattore, B., Strand, G., Worley, S.J., & Marlino, M. (2014). Research center insights into data curation education and curriculum. In L. Bolikowski, V. Casarosa, N. Houssos, P. Manghi, & J. Schirrwagen (Eds). *Theory and Practice of Digital Libraries - TPDL 2013 Selected Workshops* (pp. 239-248). Retrieved from http://link.springer.com/chapter/10.1007%2F978-3-319-08425-1_26

Michener, W.K. (2015). Ten simple rules for creating a good data management plan. *PLoS Computational Biology*, *11*(10). doi:10.1371/journal.pcbi.1004525

Miller, C.W., Hamel, J., Holmes, K. D., Helmey-Hartman, W. L., & Lopatto, D. (2013). Extending your research team: Learning benefits when a laboratory partners with a classroom. *BioScience, 63*, 754-762. doi:10.1525/bio.2013.63.9.11

Mooney, H., Collie, W.A., Nicholson, S., & Sosulski, M.R. (2014). Collaborative approaches to undergraduate research training: Information literacy and data management. *Advances in Social Work*, *15*, 368-389. Retrieved from https://journals.iupui.edu/index.php/advancesinsocialwork/article/view/15089

National Science Foundation. (2011). *Grant Proposal Guide*. Arlington, VA: National Science Foundation. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp

Pink, C.J. (2013). *Research360: Jisc Final Report*. Bath, United Kingdom: University of Bath. Retrieved from http://opus.bath.ac.uk/36362/2/R360_JiscFinalReport_v1.0.pdf

Poole, A.H. (2014). How has your science data grown? Digital curation and the human factor: A critical literature review. *Archival Science*, *15*, 101-139. doi:10.1007/s10502-014-9236-y

Porter, J.H., Hanson, P.C., & Lin, C. (2012). Staying afloat in the sensor data deluge. *Ecological and evolutionary informatics*, *27*, 121-129. doi:10.1016/j.tree.2011.11.009

Reisner, B.A., Vaughan, K.T.L., & Shorish, Y.L. (2014). Making data management accessible in the undergraduate chemistry curriculum. *Journal of Chemical Education,* 91, 1943-1946. doi:10.1021/ed500099h

Scaramozzino, J.M., Ramírez, M.L., & McGaughey, K.J. (2012). A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries*, *73*, 349-365. doi:10.5860/crl-255

Shearer, K. (2010). Research data: Unseen opportunities [Brochure]. Ottawa, ON: Canadian Association of Research Libraries. Retrieved from https://portagenetwork.ca/wp-content/uploads/2016/04/Data_mgt_toolkit_EN.pdf

Specht, A., Guru, S., Houghton, L., Keniger, L., Driver, P., Ritchie, E.G., ... Treloar, A. (2015). Data management challenges in analysis and synthesis in the ecosystem sciences. *Science of the Total Environment*, *534*, 144-158. doi:10.1016/j.scitotenv.2015.03.092

Strasser, C., Cook, R., Michener, W., & Budden, A. (2012). Primer on data management: What you always wanted to know. DataONE. Retrieved from https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

Strasser, C.A., & Hampton, S.E. (2012). The fractured lab notebook: Undergraduates and ecological data management training in the United States. *Ecosphere, 3*(12). doi:10.1890/ES12-00139.1

Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., … Dalton, K. (2015a). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE, 10*(8). doi:10.1371/journal.pone.0134826

Tenopir, C. Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L. … Lundeen, A. (2015b). Research data services in academic libraries: Data intensive roles for the future. *Journal of eScience Librarianship*. *4*(2). doi:10.7191/jeslib.2015.1085

Tenopir, C., Birch, B., & Allard, S. (2012). *Academic libraries and research data services: Current practices and plans for the future* [White paper]. Association of College and Research Libraries.

Thompson, C.A., Senseney, M., Baker, K.S., Varvel, V.E., & Palmer, C.L. (2013). Specialization in data curation: Preliminary results from an alumni survey, 2008-2012. *Proceedings of the ASIST Annual Meeting*, *50*, 1-4. http://dx.doi.org/10.1002/meet.14505001151

Varvel Jr, V.E., Bammerlin, E.J., & Palmer, C.L. (2012). Education for data professionals: A study of current courses and programs. *Proceedings of the 2012 iConference,* 527-529. doi:10.1145/2132176.2132275

Volk, C.J., Lucero, Y., & Barnas, K. (2014). Why is data sharing in collaborative natural resource efforts so hard and what can we do to improve it? *Environmental Management, 53*, 883–893. doi:10.1007/s00267-014-0258-2

Wellcome Trust. (2010). *Policy on data management and sharing*. Retrieved from
http://www.wellcome.ac.uk/about-us/policy/policy-and-position-
statements/wtx035043.htm

Appendix A:
Survey Instrument

Scientists and Research Data: Continuing to Build an Understanding of your Data Needs

You are invited to participate in an NSF-sponsored research study, in which the DataONE (Data Observation Network for Earth) organization is investigating how scientists work. Your responses will help us better understand how scientists manage their data, which will then allow DataONE to better serve their data management needs.

The questionnaire should take about 20 minutes to complete. In addition to demographic information, other questions relate to the data management practices of scientists, the data education practices of scientists who are also educators, and finally how your organization and how designated data managers are involved with your research data. As such, no sensitive items are included in our survey, and therefore we do not anticipate that your participation poses any more than minimal risk. Also, your responses will be recorded anonymously so that no one can link your responses to you personally.

Your participation in this research is voluntary, and you may decline to participate without risk. While it is useful to be complete in your responses to the survey, you may skip any questions, and you are free to withdraw from the study at any time.

If you have any questions about the study or procedures, please contact Dr. Carol Tenopir or Dr. Suzie Allard of the University of Tennessee. If you have questions about your rights as a participant, contact the Office of the Research Compliance Officer.

If you would like to keep a copy of this consent statement, you can save or print this page.

By proceeding to the survey I acknowledge that I have read the above statements, I am 18 years old or older, and I agree to participate.

<Core Survey>

First, we would like to ask you a few questions about yourself.

1) Which one of the following best describes your primary work sector?

- Academic

- Government

- Commercial

- Non-profit

- Other (please specify)

2) Which one of the following best describes your primary subject discipline?

- Agriculture and Natural Resources

- Atmospheric science

- Biology

- Business

- Computer science

- Ecology

- Education

- Engineering

- Environmental science

- Geology

- Hydrology

- Information science

- Law

- Medicine

- Physical sciences

- Psychology

- Social sciences

- Other (please specify)

3) Do you ever teach data management in courses or outside the classroom?

- Yes

- No

3A) You have indicated that you teach data management. Where does this teaching occur?

- In courses

- Outside the classroom

- Both

4) Which of the following data management topics do you teach? (For each topic, choose <u>all</u> that apply.)

	in an undergraduate course	in a graduate level course	in other types of courses	outside of courses	I don't teach this topic
Data life cycle: the stages through which data passes from the inception of a research project to its conclusion					
Planning: Creating a data management plan to control how data are handled throughout the research project					
Quality control: making sure that data are accurate and there are no missing values or errors					
File management: types of files, file naming (such as assigning descriptive file names that indicate spatial and/or temporal information about the data)					
Metadata generation: descriptive information describing data characteristics and software used					
Workflows: detailed description, flow chart, or computer script of how raw data were transformed into final results					
Protecting data: backing up data, creating multiple copies in multiple locations					
Data archiving & preservation: strategies for long-term accessibility of digital information					
Data re-use: using data that was collected for one purpose, for a new or different purpose					
Meta-analysis: statistical synthesis of results of separate studies					
Citing data: how to give attribution and credit for data					
Other data management topics					

If you selected other, please specify

5) Do you feel that you are covering these topics sufficiently? (Choose only the one best answer.)

- Yes, thoroughly (I wouldn't add any more.)

- Yes, but there is more that I could add.

- Yes, minimally

- No, I should add more.

- No, and I don't plan to add more.

- No, I don't cover them.

6) What barriers do you experience in teaching data management? (Choose all that apply.)

- There is no time to teach data management.

- I don't have enough information.

- It is not my area of expertise.

- It is not appropriate at the level I teach.

- It isn't relevant to the courses I teach.

- Students are getting this information in other ways.

- Other (please specify)

End of Core Survey