Education Needs in Research Data Management for Science-Based Disciplines:

Self-Assessment Surveys of Graduate Students and Faculty at Two Public Universities

Judith E. Pasek
STEM Liaison Librarian
University of Wyoming
Laramie, Wyoming
jpasek@uwyo.edu

Jennifer Mayer
Head of Library Research Services
University of Northern Colorado
Greeley, Colorado
jennifer.mayer@unco.edu

Abstract

Research data management is a prominent and evolving consideration for the academic community, especially in scientific disciplines. This research study surveyed 131 graduate students and 79 faculty members in the sciences at two public doctoral universities to determine the importance, knowledge, and interest levels around research data management training and education. The authors adapted 12 competencies for measurement in the study. Graduate students and faculty ranked the following areas most important among the 12 competencies: ethics and attribution, data visualization, and quality assurance. Graduate students indicated they were least knowledgeable and skilled in data curation and re-use, metadata and data description, data conversion and interoperability, and data preservation. Their responses generally matched the perceptions of faculty. The study also examined how graduate students learn research data management, and how faculty perceive that their students learn research data management. Results showed that graduate students utilize self-learning most often and that faculty may be less influential in research data management education than they perceive. Responses for graduate students between the two institutions were not statistically different, except in the area of perceived deficiencies in data visualization competency.

Introduction

Quality research entails attention to details of collecting, documenting, and analyzing data in systematic, verifiable, and reproducible manners, following standards of practice within fields of study. Researchers typically learn methodologies and disciplinary practices through graduate

study and involvement in original research projects under the guidance of expert faculty researchers. In recent years, accompanying the widespread availability of networked communication technologies, expectations for research outputs have increased beyond article publication to include dissemination of the underlying data in accessible and usable formats and in open data repositories. Consequently, the need for education in research data management (RDM) concepts and practices has expanded. Graduate students need to develop understanding and proficiency in RDM as they enter scholarly professions.

Our exploratory study was an attempt to gather additional information regarding perceived RDM knowledge levels, with an emphasis on graduate education needs. The scope encompassed disciplines in natural, agricultural and life sciences, including health sciences and science education, as well as engineering and some data-intensive social sciences. The primary objective of this study was to identify faculty and graduate student perceptions of RDM learning and skills in order to inform recommendations about RDM education at two medium-sized, public universities: the University of Northern Colorado (UNC) and the University of Wyoming (UW). At both universities, graduate students and faculty received similar but separate online surveys. This study sought to answer the following research questions:

1. How do graduate students and faculty rate the importance of RDM competencies?
2. How do graduate students and faculty rate RDM knowledge and skill levels of graduate students?
3. How do graduate students learn RDM concepts and practices?
4. Do self-reported assessments by graduate students regarding RDM education differ from faculty perceptions of their graduate students?
5. Are there differences in RDM education needs between the two institutions studied?

Literature Review

Many academic librarians have begun developing RDM services and infrastructure support to help address data sharing requirements of funders (Tenopir et al. 2012). RDM services promoted by librarians (sometimes in collaboration with disciplinary faculty) have increasingly incorporated a variety of instructional approaches, such as workshops and credit courses, to educate faculty and graduate students about RDM concepts and practices (Schmidt & Holles 2018). Some RDM education case studies include preparatory information gathered largely from interviews of small samples of faculty or students. Surveys of larger groups typically have focused more on technical aspects of managing research data, such as data file formats and file sizes (Weller & Monroe-Gulick 2014; Sheehan et al. 2015; Whitmire et al. 2015).

Several preliminary studies, involving small samples, focused on graduate students' RDM needs. Challenges uncovered for engineering graduate students included a lack of consistent terminology regarding RDM, limited awareness of campus RDM services, and disconnect with faculty expectations (Wiley & Kerby 2018). Structural engineering students indicated they had several unmet needs at various stages of the data life cycle, which could hinder them in future research positions (Johnston & Jeffryes 2014). An examination of the data sharing practices of graduate students at a water lab (Carlson & Stowell-Bracke 2013) resulted in creation of data curation profiles that document needed support at various stages of the data management cycle. In general, these studies suggest that education of graduate students in RDM remains insufficient across a wide array of science-based disciplines.

Other studies have focused primarily on RDM needs of science faculty, although some researchers included graduate students within their study populations. Interviews of early career social science researchers (Jahnke et al. 2012) indicated that many of them understood the importance of RDM, but their challenges included developing a data management plan, a variety of RDM technical issues, and the notion that RDM training does not take precedence because it does not directly contribute to publication production. In a land grant university analysis of RDM needs and practices (Fernandez et al. 2016), almost ninety percent of agricultural faculty and graduate student respondents felt they could benefit from assistance working with data or were unsure if they would benefit from assistance. Gaps in faculty knowledge and confidence in concepts of RDM likely contribute to weaknesses in educating graduate students in RDM principles and practices. Interviews of meteorology graduate students indicated they learned RDM mostly informally from other students, faculty, IT personnel, and independently (Frank & Pharos 2016). Atmospheric scientists and engineering faculty indicated that they feel limited on how to provide instruction on best RDM practices (Mischo et al. 2017). Interviews of agriculture students who completed a course in RDM revealed that designing a program in the context of the students' work in order to teach higher level concepts was important, vigilance was required to connect students to practical applications of RDM work, and advisor buy-in to RDM education was crucial (Carlson & Bracke 2015). Insights about data management practices of students at a medical and polytechnic school led to development of seven learning modules for a curriculum in RDM (Piorun et al. 2012). Uncertainty remains as to the best approaches for RDM education and identification of which RDM concepts to emphasize.

A pivotal work for our research project is "Determining Data Information Literacy Needs: A Study of Students and Research Faculty" (Carlson et al. 2011), in which the authors created a proposed list of core Data Information Literacy (DIL) competencies. The authors sought to triangulate the needs related to data information literacy through interviews with research faculty and by analyzing the results of their own data literacy course. They found that some faculty have little knowledge of how to manage data and expect students to learn as they go. Additionally, differences between faculty and graduate students were discovered regarding the DIL 12-point needs assessment (Carlson et al. 2015a). In further study, the authors (Carlson et al. 2015c) concluded that graduate students do not have a lot of preparation or education in data management, educators should not assume graduate students learned even basic data management skills in undergraduate programs, students are motivated to obtain DIL skills for the job market, and flexible educational opportunities like modules may be the best approach for students' busy schedules. Thielen et al. (2017) used nine of the 12 DIL competencies and professional standards to develop a course in RDM for graduate students in climate and space sciences. While a significant time commitment, the authors found the credit course to be a successful way to teach RDM.

Examining the existing literature indicated that there is a heavier focus on faculty RDM needs, and there is a larger gap regarding understanding graduate student RDM needs. Few studies directly compare graduate student and faculty perceptions of RDM education needs. Our literature review set the foundation for our decision to use separate, but similar, online surveys to facilitate comparisons of responses by graduate students and faculty in science-based disciplines. Survey questions were intentionally designed to broadly explore education needs for RDM in the context of existing curricula. Not highlighted were specific library services relating to RDM, although results of this study may inform future development of RDM library services at the universities included in this study.

Methods

The two academic institutions included in this study are both public, doctoral universities with similar enrollments. Spring 2018 enrollment numbers at UNC were 11,954 total students, which includes 3,103 graduate students (UNC 2018). UNC originated as a teacher's college and today maintains strong programs in an array of education subjects and applied health science disciplines, complemented by a core of natural and life science disciplines. Its graduate instructional program is described as research doctoral, comprehensive. Official enrollment for Spring 2018 at UW was 11,833 students, including 2,543 graduate and professional students (UW 2018). UW is a land-grant university and serves as a flagship institution, being the only public higher education institution offering four-year degrees and graduate education within the state of Wyoming. It offers programs of study in agricultural sciences and engineering not available at UNC. However, many programs in the natural and life sciences, applied health sciences, and education are similar to those offered by UNC. The graduate instructional program at UW is described as research doctoral, STEM (Science, Technology, Engineering, and Math) dominant. Collaborating on identical surveys at these two neighboring institutions provided advantages of increasing the participant pools, while offering a chance to examine potential differences in needs that may result from differences in graduate instructional programs.

Surveys were conducted at UW and UNC in an effort to gather input from a greater number of respondents than is practical with interviews or focus groups. Institutional Review Boards at both universities approved this study (UW Protocol 20180122JP01835 and UNC Project 1186741-1) and verified its status as exempt. Separate surveys for graduate students and for faculty and researchers were created using Qualtrics software. The same surveys were distributed at both institutions. An initial screening question limited respondents to current affiliates of UNC or UW. The graduate student survey further limited responses to those enrolled in degree programs requiring original research and management of data. Non-thesis options are common for Master's degrees at both universities, but it was not possible to identify and exclude individuals pursuing non-thesis degrees prior to survey distribution. The faculty and researcher survey was inclusive of all ranks, but excluded classified staff positions. Respondents were allowed to skip any of the questions and cease participation at any time.

Survey questions are presented in Appendices A (graduate students) and B (faculty). Questions about perceived importance and knowledge levels of RDM concepts were based upon the 12 DIL competencies categorized by Carlson et al. (2011), and mirrored questions developed by Thielen et al. (2017) with some modification. The Data Management and Organization competency was retitled as Data Planning and Organization to reduce confusion, given that all 12 competencies deal with data management concepts. Competency labels received short descriptions to facilitate understanding of terminology and reduce ambiguity, without entailing lengthy textual instruction. The competency questions were similar in the two survey instruments except that graduate students were asked to self-assess, whereas faculty were asked to provide ratings in relation to their graduate students. Questions about where students learn RDM skills were adapted from Pouchard and Bracke (2016). Questions about topics and programs of interest were derived from Peters and Vaughn (2014). A question in the graduate student survey about their role in RDM was based on data management life cycle steps as described by McLure et al. (2014). Questions about areas of struggle were modified from Wiley and Mischo (2016). Similarities in questions between the graduate student survey and the faculty survey were intended to facilitate comparisons of responses.

Email distributions for current faculty and researchers and graduate students in STEM, education disciplines, and data-intensive social science departments were used to recruit participants. At UNC, email addresses were obtained from the Office of Assessment for the College of Natural and Health Sciences and the School of Psychological Sciences (within the College of Education and Behavioral Sciences). Mail merge was used to send letters containing a link to the appropriate survey to UNC individuals on March 26, 2018. Reminders were sent from Dean and Director offices about a week later and additional email reminders were sent to faculty. At UW, letters containing a survey link were distributed to various departmental email lists via contacts with office associates beginning April 3, 2018, and continuing over a week due to the decentralized nature of the distribution process. UW departments exclusively offering non-thesis professional degrees, such as Pharmacy, were omitted. Follow-up distributions were sent to Department Heads on April 19, 2018, with a request to encourage greater participation from their faculty members. Both surveys were available online through May 17, 2018. The graduate student survey included a chance to win a Popsocket grip as an incentive for participating. The faculty survey did not feature an incentive.

The broad classifications of the National Science Foundation (NSF 2013) fields of study provided the basis for aggregating survey responses into disciplinary categories. For Education disciplines where specific STEM fields were identified, the responses were counted within STEM categories. For example, Nursing Education was counted in the Health Sciences.

Survey responses were analyzed quantitatively and qualitatively, as appropriate to the question formats. Quantitative data were exported from Qualtrics to Excel for data summary and analysis. Ordinal rating scales were converted to point values for purposes of calculating averages and determining rank order. Responses to the graduate student survey questions were subdivided by institution for additional statistical comparisons. Chi-square tests of homogeneity were conducted to determine whether responses of graduate students differed for UNC versus UW for questions using ordinal scales. Separate chi-square tests were run for the institutional comparisons for each of the 12 DIL competencies included in student survey questions that self-rated importance or knowledge/skill abilities. Faculty responses were not subdivided by institution for comparative analysis due to a low number of returned surveys. However, chi-square tests of homogeneity were conducted for ordinal data to compare responses of all faculty versus all graduate students for parallel questions about RDM knowledge and skill levels of graduate students and where RDM learning occurs. Post-hoc tests were calculated using chi-square tests of independence and Bonferroni's correction for the alpha value for all 66 possible paired comparisons of DIL competencies for graduate student responses to questions about importance and of knowledge and skills. Percentages of responses for combined answers of Needs Improvement and Deficient were calculated for each of the 12 DIL competencies in the questions about knowledge and skills. These figures along with their corresponding sample counts were used to conduct two-sample t-tests comparing UNC versus UW graduate students. Similar t-tests were run to compare responses of all faculty versus all graduate students. Additional two-sample t-tests were conducted for percentages of positive answers for each categorical choice in questions about training interests to compare responses of all graduate students to what all faculty thought would interest their students. A one-sample t-test and one-tailed probability was used to test whether faculty were less interested in suggested training topics for themselves as compared to interest for their graduate students. The Statistics Calculator available from StatPac, Inc. was used to conduct statistical analyses. Qualitative responses to open-ended questions were examined by one of the authors for informative content, hand-coded, and categorized into the twelve DIL concepts used in other survey questions.

Results

Response Demographics

There were 63 valid graduate student surveys received from UNC and 68 from UW, for a total of 131 included respondents. Calculated response rates for graduate students, based on the numbers of email addresses included in survey distributions were consistent at 7% (63/867) for UNC, 7% (68/931) for UW, and 7% (131/1,798) overall. The calculated response rates are likely underestimates relative to the eligible populations of graduate students pursuing degrees with research requirements, given that students pursuing non-thesis graduate degrees (other than professional degrees) could not be precluded from receiving survey solicitations. Doctoral students returned a higher number of surveys than master's degree students (Figure 1), perhaps reflective of greater research requirements for doctoral students.
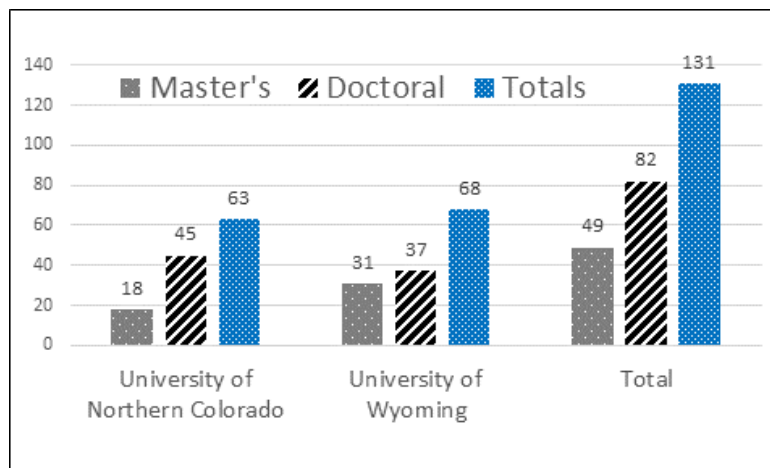

Figure 1. Degree types of graduate student survey respondents.

For the faculty survey, there were 39 returned from UNC and 40 from UW, for a total of 79 valid respondents. The returned surveys represent response rates of 13% (39/294) from UNC, 9% (40/452) from UW, and 11% (79/746) overall. Assistant, Associate, and Full Professors accounted for 81% of faculty and researcher respondents.

Nearly 3/4 of graduate student respondents identified their program discipline. Health Sciences and Psychology dominated at UNC, while the majority of responses from UW were in Life and Agricultural Sciences (exclusive of Health Sciences). Less than 2/3 of faculty respondents identified their primary research discipline; however, the pattern of discipline categories mirrored that of graduate student responses.

DIL Competencies

Based on average scores of response ratings, graduate students and faculty agreed that ethics and attribution was the most important competency for graduate students to be knowledgeable about by the time they graduate, followed by data visualization and quality assurance (Figure 2). Graduate students assigned differing levels of importance to the competencies, with significant differences revealed in 21 of 66 post-hoc pairwise comparisons. Unfortunately, a middle score choice of Important was inadvertently left out of the graduate student question, resulting in a 4-point scale rather than the 5-point scale used in the faculty survey, precluding direct comparison

of responses between these populations. No significant differences were detectable between UNC and UW graduate student responses for importance of any of the 12 DIL competencies.
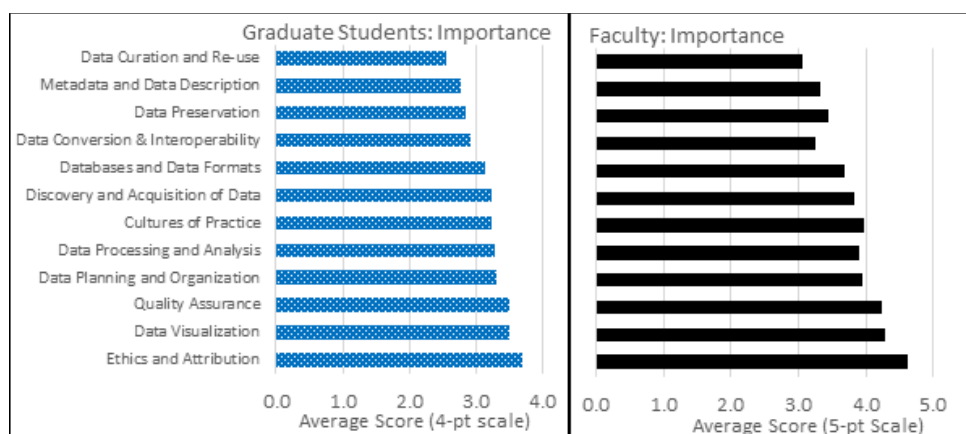

Figure 2. Average ranking of importance of 12 DIL competencies by graduate students versus faculty respondents.

When asked to rate knowledge and skill level for the 12 DIL competencies, graduate students and faculty again ranked ethics and attribution the highest, based upon average scores. The rank order of other competencies were similar for graduate student and faculty responses, with average scores being lowest for competencies of data curation and re-use, metadata and data description, data conversion and interoperability, and data preservation (Figure 3). Graphical representation of average scores of graduate student and faculty responses hint that graduate students may display more confidence in their knowledge and abilities for all competencies relative to perceptions of faculty. However, no statistical differences in the proportions of answers by graduate students versus faculty were detectable in chi-square tests. No significant differences were detectable between UNC and UW graduate student responses for self-reporting of knowledge and skill levels of any of the 12 DIL competencies either.
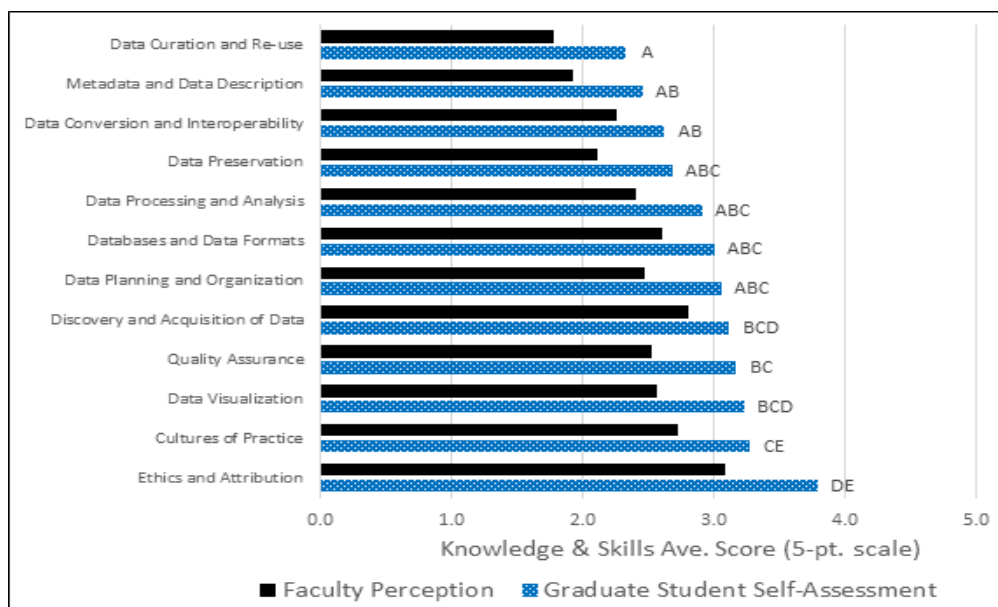

Figure 3. Comparison of faculty and graduate student assessments of knowledge and skills in 12 DIL competencies.

Post-hoc chi-square analyses did reveal significant differences in proportions of answers by graduate students for 14 of 66 possible paired comparisons of competencies. This indicates that graduate students were not equally confident in their knowledge and skill levels for all 12 competencies. Responses for data curation and re-use (the DIL competency with the lowest average score) differed significantly from the five DIL competencies with the highest average scores: discovery and acquisition of data, quality assurance, data visualization, cultures of practice, and ethics and attribution. For the DIL competency with the highest average score, i.e., ethics and attribution, responses differed significantly from the other DIL competencies with the exception of cultures of practice, data visualization, and discovery and acquisition of data. Responses for cultures of practice differed significantly from those of data conversion and interoperability, and metadata and data description, in addition to data curation and re-use.

Another way to look at the responses to the knowledge and skills questions was to determine the percentages of answers that fell into the Needs Improvement (NI) and Deficient (D) categories (Figure 4). Only 10% of graduate students and 18% of faculty chose either of those responses for the category of ethics and attribution. More than 40% of students indicated they were deficient or needed improvement for competencies of data curation and re-use, metadata and data description, data conversion and interoperability, and data preservation. No significant differences were detectable between UNC and UW graduate student responses for self-reporting of NI/D levels for 11 of the 12 DIL competencies. NI/D responses for data visualization were significantly higher ($\chi^2 = 2.159$, df = 104, p = 0.0332) at UNC than at UW. For faculty, more than 40% indicated their graduate students were deficient or needed improvement in metadata and data description, data planning and organization, and data preservation. Differences between graduate student self-assessments and faculty ratings of their graduate students' knowledge gaps were not significant for any of the 12 DIL competencies.
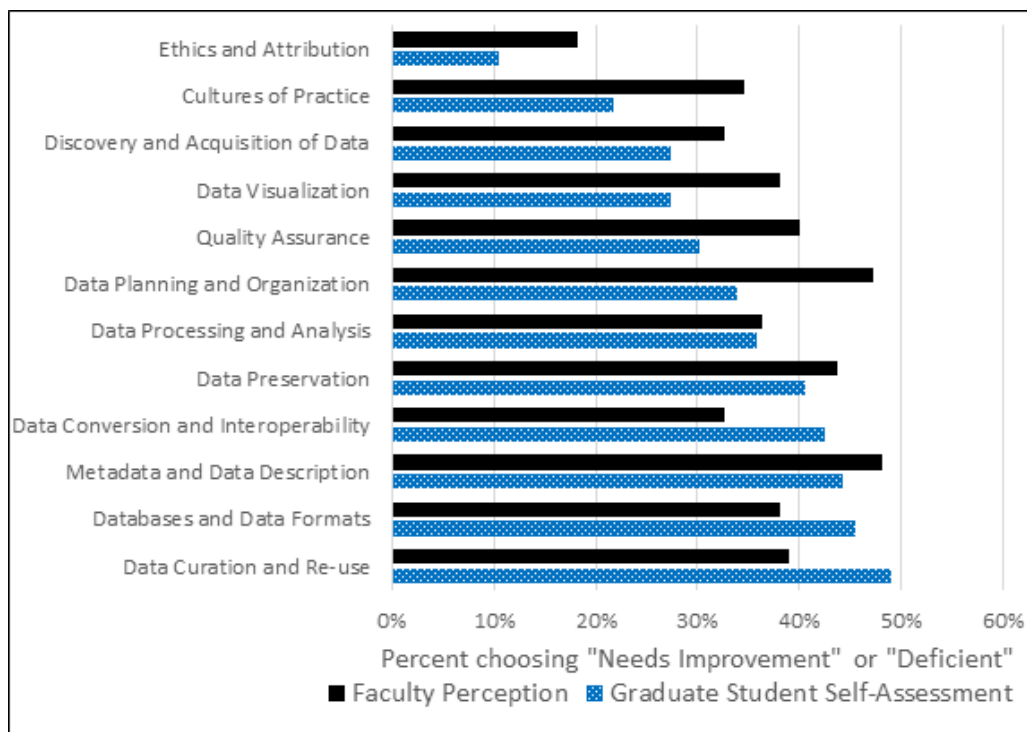

Figure 4. Comparison of faculty perceptions of graduate students' knowledge and skills versus graduate students' self-assessment for combined ratings of Needs Improvement or Deficient in the 12 DIL competencies.

The relative importance graduate students assign to RDM competencies and the knowledge levels they attain may be influenced by the roles they play in research. Graduate students indicated they typically have responsibilities for RDM roles involving the conduct of research: planning (63%), organizing (65%), creating (68%), and analyzing (71%). Fewer indicated they were involved with security (39%), training others (27%), transfer and preservation (21%), and sharing data (20%).

Engagement with RDM Learning

Graduate students indicated that self-instruction was their most frequent means of learning RDM skills (49% Often; 29% Sometimes). Faculty or advisor, peers and fellow students, and courses were also common sources for learning (Figure 5). Services of the research office, libraries, and information technology were infrequent sources indicated by graduate students (2-3% Often; 20-27% Sometimes). Free text entries of other sources of learning were split between work and internship experiences and specific types of self-instruction, such as internet searching and reading.
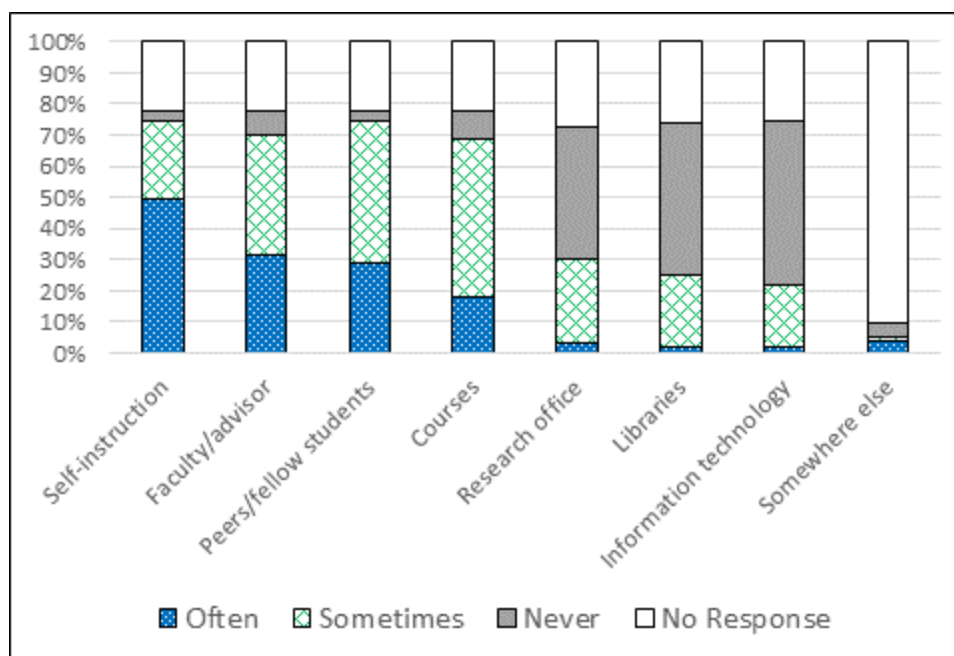

Figure 5. Where graduate students say they learn RDM skills (N = 131).

Faculty responded that they (faculty or advisor) are the primary source of learning about RDM by their graduate students (44% Often; 23% Sometimes), while recognizing that peers and fellow students, self-instruction, and courses also were important learning sources (Figure 6). Faculty generally perceived that services of the research office, information technology, and libraries were less common sources of RDM learning for their graduate students. Some faculty identified professional conferences, meetings, and workshops, or work and project experiences, or grants and Institutional Review Board (IRB) training as additional sources for student learning about RDM.
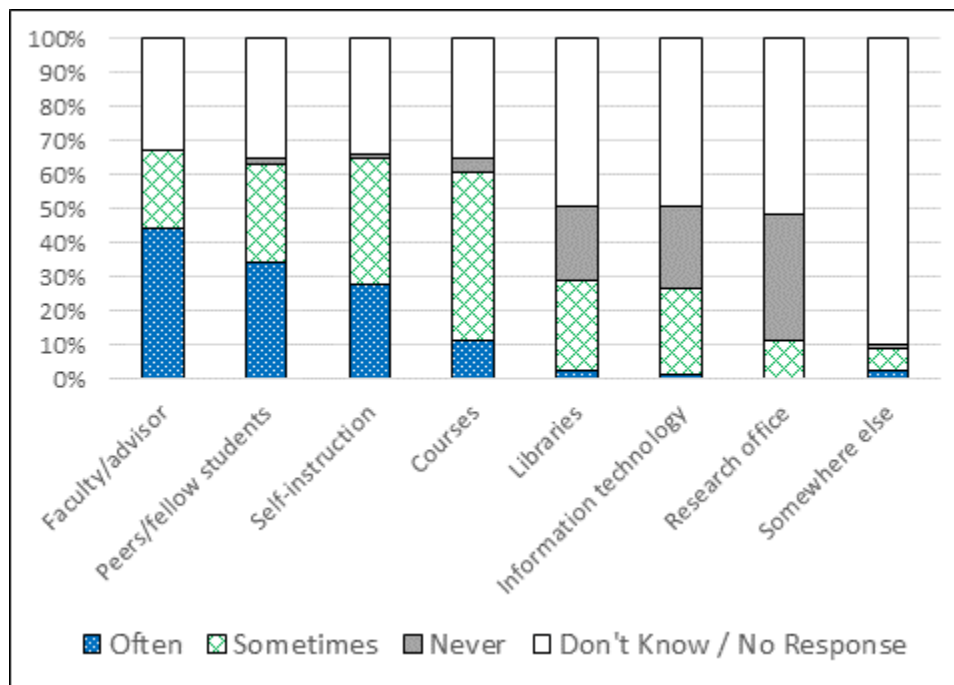
Figure 6. Where faculty say their graduate students learn RDM skills (N = 79).

Faculty perceptions of where their students learn RDM differed from what graduate students self-reported. Responses of graduate students versus faculty for self-instruction differed significantly (χ^2 = 11.192, df = 3, p = 0.0107). A significant difference also was found for responses of graduate students versus faculty for instruction from faculty or advisor (χ^2 = 14.428, df = 3, p = 0.0024). Graduate students chose answers of Often in greater proportions than faculty for self-instruction. Faculty chose answers of Often in greater proportions than graduate students for faculty or advisor.

For teaching methods utilized by faculty, one-on-one mentorship was the most common means of providing RDM instruction as identified by faculty respondents (58% of surveys). Instruction in class (32%) and in lab or field (27%) were also commonly identified approaches. Online tutorials and other teaching methods were less frequently indicated. The prevailing topics of RDM instruction (from choices given) were creating data sets, data storage practices, file naming conventions, and privacy. Less common topics included metadata and open data concepts.

Training Interests

When graduate students were asked to indicate which suggested topics for RDM instruction they would be highly interested in, positive responses ranged from 55% for data organization and best practices to 27% for licensing research data (Figure 7). Faculty indicated similar levels of interest for their graduate students to learn about the RDM topics (Figure 7). No significant differences were identified between responses of graduate students and what faculty thought would interest their graduate students for each of the topics presented in survey questions.
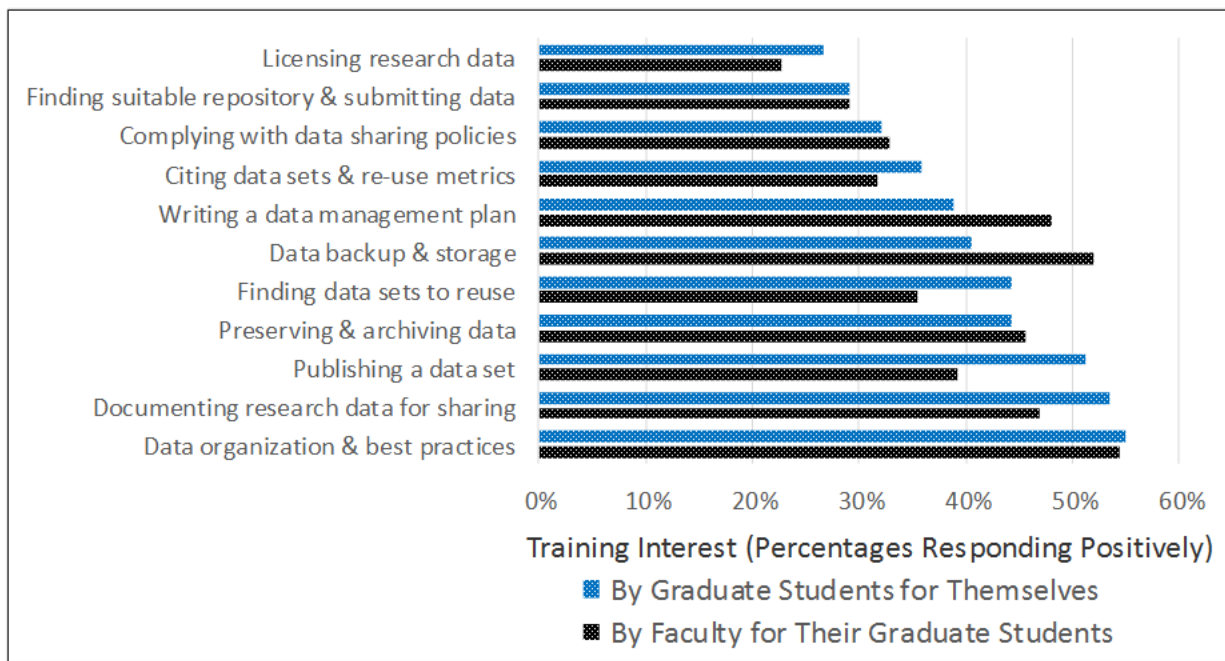
Figure 7. Training interest in RDM topics for graduate students, comparing responses of graduate students (N = 131) and faculty (N = 79).

Faculty and researcher respondents in non-tenure positions indicated little interest in RDM training for themselves. However, some faculty in tenure lines (i.e., Assistant, Associate, and Full Professor ranks) did indicate interest in obtaining RDM training for themselves. Question responses among tenure ranks ranged from 49% for writing a data management plan or data backup and storage to 27% for how to find a suitable repository and submit data (Figure 8).
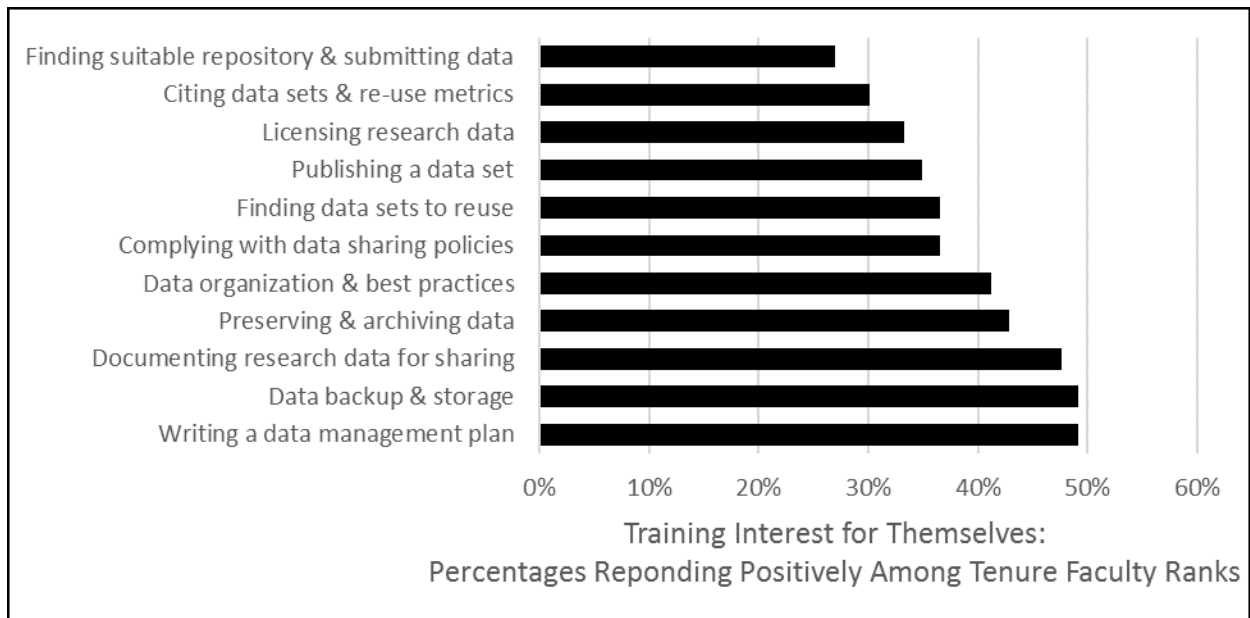

Figure 8. Training interest in RDM topics for faculty in tenure ranks (i.e., Assistant Professor, Associate Professor, and Professor) (N = 63).

Challenges for Graduate Students and Faculty

Graduate students and faculty members noted in their response to the question, "What areas are you struggling with or wish you knew more about, if any?", that they struggle in various areas related to managing their research data. Figure 9 illustrates the frequency of responses relative to the twelve DIL competencies.
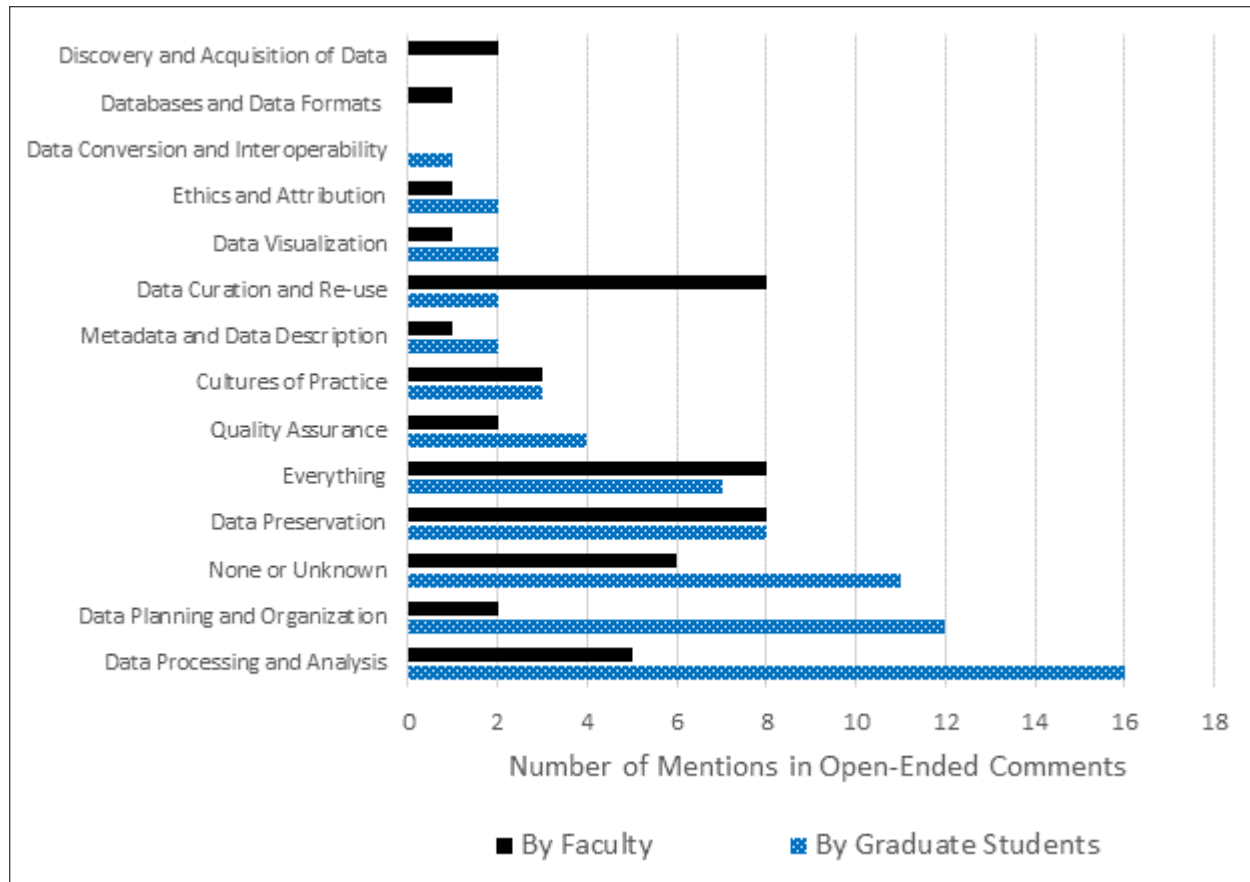

Figure 9. Challenges in Research Data Management (RDM) for faculty (N=35) and graduate students (N = 44).

Some respondents elaborated on particular areas of difficulty. One doctoral student shared: "Despite being finished with PhD coursework--and having conducted and published qualitative and quantitative studies--I feel uncertain of my processes for my dissertation...No faculty member has EVER shown me, in a class setting, his/her actual approach to managing data. So, even though I've read about some of the concepts, I feel like I lack adequate skills to develop my own structures. The valuable notion of sharing data (through repository/publication) is only now (i.e., in the past year) occurring to me as a goal I should WANT to achieve for the data I've collected. Especially given the recent and unsuccessful efforts in some fields to replicate landmark studies…"

A faculty member observed: "Until recently, it's been hard communicating ideas about record-sensitivity (including state-level requirements) to partners. They've recently started seeing the light with regard to sane ways of sharing some data in a safe and responsible way, but it's been

quite a road getting there with many twists, turns and reversals. Fear of legal liability and unclear state laws have been major factors."

The identified challenges echo the quantitative data regarding graduate students' self-assessment of knowledge gaps in planning, organizing, processing, and analyzing data. Conversely, faculty identified that they struggle most with data curation and reuse, and data preservation.

Teaching Challenges

Faculty responses varied to the question: "What research data management concepts or skills are you uncomfortable teaching or find difficult to teach, if any?" The top response with six comments indicated discomfort or difficulty teaching data processing and analysis. Metadata and data description had four comments, as did data planning and organization. Databases and data formats, data curation and re-use, data processing and analysis, and data visualization all had two comments a piece. Faculty shared a single comment for each of the following categories: discovery and acquisition of data, data conversion and interoperability, and cultures of practice. Categories that had no mention regarding teaching concern included quality assurance and ethics and attribution. One faculty respondent is concerned that graduate students receive "inconsistent messages regarding data management." Three faculty respondents indicated no challenges teaching any of the 12 DIL areas, and six shared they are uncomfortable teaching all areas. One faculty member stated, "I feel I am incompetent to teach any data management skills;" others said they need to brush up on all these concepts; and another said they rely completely on older graduate students to teach these topics.

Who Should Teach RDM?

There was no consensus among faculty responding to the question, "Who do you think should be responsible for educating graduate students in regard to data management skills?" Twenty-two responses indicated the faculty advisor, principal investigator, or faculty mentor should be responsible for RDM education of students. Five responses indicated faculty should be responsible, without specifying a research or program relationship. Three individuals suggested statistics and economics faculty provide RDM education. Three responses suggested a required course in RDM. Seven respondents recommended a more collaborative approach to where responsibility for RDM training lies, including the home department, faculty, liaison librarians, university data services, information technology, and the research office. Three respondents stated that they were not sure who should be responsible.

Additional Faculty Comments

Faculty were given the opportunity to share any additional comments regarding training/education in research data management. One individual said that the university needs to make RDM courses a priority, and provide resources in order to offer the courses on a regular basis. Another faculty member stated in their experience, students need assistance with presenting their data in tables and figures that are clear, and this is an area for needed instruction. Basic training in statistics—e.g., understanding of variance—was another identified need. Two responses indicated they know there is a knowledge gap, but they do not know what it is, and could use assistance themselves in many aspects of RDM. One respondent shared, "If your data aren't accessible to the outside world, they might as well not exist," and pointed to the need for faculty to have training on how to put this notion into practice without needing technological

assistance. One faculty member commented that faculty need more assistance in teaching data management when they are already managing large graduate student loads.

Additional Graduate Student Comments

When given the opportunity to provide any additional comments regarding training/education in research data management, graduate students shared more feedback. Eight comments contained observations about the lack of RDM education at their institution. One student shared, "None of my courses, except a brief introduction to qualitative software, has really addressed these issues. I am largely self-trained in handling vast amounts of data I am using for my study, and I worry that 'I don't know what I don't know.'" Another student describes learning advanced data management software such as R (for statistical computing) and ArcGIS (for spatial representation) as a race to learn because there is no formal training, and acknowledges that RDM building blocks or basics are "not well in place."

Two comments indicated a real need for education on the basics of RDM, which neither of them received; two individuals suggested a credit course on RDM would be the best approach to learning these skills, including courses focused on SAS software (a statistical analysis system) and R. One of these respondents added, "A suite of courses or even workshops would be lovely." One student observed credit courses should be electives, since post-graduate plans vary widely among graduate students. Yet another student remarked, "Professionalism in data management seems to have a culture of upperclassmen teaching underclassmen. I've learned more in a lab from the tech than any other source." Alternately, three responses advocated for a non-credit course approach in the form of sources available as needed, picking up information from advisors, and a menu approach to provide various RDM options, so students could pick and choose an area as it relates to their research data.

One respondent made the observation that there are many knowledge gaps in RDM for graduate students, and the health care's future relies on data that is easy to use and report outcomes. Another student also tied the importance of RDM skills to future jobs: "I received in-depth training at a past job over a few weeks and the knowledge I gained was invaluable. Knowing what has been provided at my current university (very little) in comparison, I see a lot of room for improvement, as this skill is paramount to learn in today's research settings." One student shared, "The lack of training and education in research data management in my graduate program is disappointing. There is no formal guidance given to students on the best practices in our field on how data should be stored, processed, and analyzed." The graduate student comments revealed a clear need for increased educational opportunities regarding data management.

Discussion

Implications

In general, responses of graduate students and faculty to survey questions about RDM education needs of graduate students involved in research indicate a dichotomy between competencies that relate to the conduct of research compared to those relating to longer-term preservation, sharing, and use or re-use of data. Graduate students as well as faculty rated both importance and knowledge/skill levels lower relative to other competencies for data curation, metadata and data description, data preservation and data conversion and interoperability. Activities involving these competencies may be left to the end of the research process, if done at all. An emphasis in

graduate education has traditionally focused on training in how to do research, with an end goal of academic publication. The research roles graduate students identify as their responsibility also align with an emphasis on the conduct of research. Consequently, it is understandable that graduate students would be most interested in gaining knowledge and skills in those competencies directly relating to successful completion of their research and reporting of results. Similarly, Whitmire et al. (2015) determined that graduate students rarely are involved in data sharing activities beyond their research group.

At least some graduate students appear to recognize that they may benefit from additional training in RDM topics beyond what is needed to obtain a diploma. Similar to our results regarding interest in RDM topics, Peters and Vaughn (2014) found that students who had attended a basic RDM workshop were most interested in additional workshop topics relating to conduct of research, i.e., "data storage, backup and security," and "types, formats, and stages of data," as opposed to data sharing topics of "archiving & preservation," metadata, "data sharing and re-use policies," and "legal and ethical considerations."

The preferences of graduate students may be reinforced by the RDM topics faculty choose to teach, which also emphasize the conduct of research over data sharing competencies. Faculty appear not to have embraced teaching about evolving trends in sharing research data. This may indicate that sharing data is not highly ingrained in the research process or scientific culture yet. For example, faculty and researchers in focus groups at Colorado State University identified "the plan, create, produce, and transfer stages" of the data management life cycle as most important (McLure et al. 2014). Opportunities to work on national and global projects were given as compelling reasons to pursue data sharing and dissemination. Given that faculty and researchers are primarily responsible as the principal investigators for obtaining and complying with grants, there remains a gap in RDM education for graduate students as they move on to more responsible positions in the work force.

Compared to results in a Carlson et al. study (2013; 2015a) and in pre-course and post-course surveys (Thielen 2017), where almost every category of DIL competencies was ranked high in importance, our study reveals more variance in responses. This may reflect differences in methodological approaches. The anonymous nature of our surveys may have avoided inadvertent researcher influence on responses that can occur during interviews. In a survey of agricultural students and faculty (N = 136), most respondents indicated that the DIL competencies were very important or somewhat important (on a 3-point scale) (Pouchard & Bracke 2016). However, there was greater differentiation for a rating of very important alone, with more choosing "data analysis" and "ethics, including citation of data," and fewer choosing "data curation and re-use," "data conversion and interoperability," "metadata" and "data preservation." These results are strikingly similar to ours; however, their more limited rating scale may have masked some distinctions in importance.

It is unclear whether disciplinary differences in populations sampled influenced results. Carlson et al. (2015b) interviewed a small (N = 25), targeted sample of students and faculty in person, drawn predominantly from engineering and natural resources disciplines. Our larger samples were drawn from a broader array of disciplines, although dominated by life and health sciences, and therefore likely captured a wider array of experience in RDM concepts. The disciplinary focus in agriculture of Pouchard and Bracke (2016) was also narrower than our samples, but somewhat similar in that agriculture was included within our broader category of life and health sciences.

The passage of time between previous studies (e.g., Carlson et al. 2015a) and ours likely influence levels of awareness and acceptance of data sharing policies, as suggested by results of surveys by Tenopir et al. (2015). Interest, by at least some faculty, in learning more about topics related to sharing of data (such as documenting research for sharing or writing a data management plan) could be due to the increased mandates during the past few years specifying that U.S. government funded research be openly available and usable. As these faculty become more confident in their own knowledge, they may be more willing to translate data sharing concepts into instruction of graduate students.

Graduate students and faculty rated RDM ethics and attribution as most important, and also as the DIL competency in which graduate students are most knowledgeable. These results may be due to training requirements of Institutional Review Boards for researchers, including graduate students. Need for additional training in this area may therefore be comparatively low.

Graduate students will likely need to be introduced to advanced research data management concepts, including those pertaining to data sharing, given that they use self-instruction most often to learn RDM. Students focused primarily on learning how to conduct their research may miss exploring the bigger picture when their learning is largely self-directed. Previous interview studies (Carlson et al. 2013) and surveys (Pouchard and Bracke 2016) identified dependence on informal learning and self-instruction in RDM but this phenomenon is even more evident in our study. Significant differences between responses of graduate students versus faculty in ratings for self-instruction and from faculty or advisor suggest that graduate students avail themselves of self-instruction more than faculty may realize, and indicate that faculty may not be the most common source for learning RDM by graduate students. Possible explanations include that formal instruction is generally inadequate to meet specific student needs, faculty may be unaware of student needs for RDM instruction, students may have a preference for independent learning, and point-of-need learning may be more conducive to addressing needs of differing research areas.

Finding more effective means for teaching RDM concepts and skills may require some trial and error and adaptation for individual institutions. Peters and Vaugh (2014) found that students' confidence in their knowledge of data management did not improve sufficiently through attendance at workshops, perhaps because prior knowledge was too low for some students or the content was not specific enough to attendees' disciplines. Promoting use of tutorials on RDM concepts is an alternative approach. Although faculty indicated that they infrequently use tutorials for instruction in RDM, they may be more willing to incorporate tutorials if they do not have to spend time locating quality content or creating materials themselves. Librarians with data management expertise can fill a role in creating tutorials and encouraging their use by faculty. The reluctance of students to contact librarians for help could thereby be side-stepped by greater collaboration with faculty on content delivery. Providing a menu of tutorials as options for point-of-need learning might concurrently satisfy student preferences for self-study. Offering exercises to accompany tutorials that would encourage students to apply principles to their own data could enhance and reinforce learning. Outreach and education for faculty, such as through collaboration of librarian experts with a faculty teaching and learning center, may be a necessary component for success.

Another approach may be to develop a formal train the trainer model, given that graduate students indicated that they also commonly learn RDM from peers. Experienced graduate students could be designated as points of contact to teach their peers about RDM. Pairing these peer mentors with personnel in libraries, research offices, and information technology could

better leverage the expertise available in these units. Peer mentoring has been successfully used with undergraduates to promote learning in information literacy (O'Kelly et al. 2015), and might serve as a template for development of similar programs for graduate learning.

Previous studies describe the potential role in RDM for librarians (McLure et al. 2014; Bracke & Fosmire 2015; Whitmire et al. 2015; Pouchard & Bracke 2016). Recommendations from McLure et al. include providing education related to RDM data curation and sharing, promoting more widespread awareness of the libraries' data management plan templates and repository, enhancing awareness of data sharing considerations, and facilitating the use of metadata standards. In their experience, Bracke and Fosmire found that providing three librarian-led instruction sessions during a semester lab section was effective. Whitmire et al. suggested librarians can fill a gap in RDM instruction by teaching metadata topics. Pouchard and Bracke advise librarians to be explicit in promoting their RDM services, and to have an answer ready for why libraries should have a role in RDM. There is a potential role for librarians as interdisciplinary educators for both graduate students and faculty in the arena of data organization, metadata, and re-use portions of the RDM cycle. For example, an area of emphasis might be learning how to document research data for sharing and data preservation. These areas dovetail well with library professionals' commitment and knowledge surrounding open access and inclination to provide a leadership role in this area.

Limitations

Respondents to our surveys were self-selected samples from among those who received emailed requests to participate. Time pressures and survey fatigue may have discouraged participation. Those responding may have had a greater interest in data management than non-respondents. Therefore, it is unclear as to how representative our results are relative to the populations engaged in graduate education in science-based research within the two universities surveyed. Furthermore, the results could differ from that of graduate students and faculty at other academic institutions.

The pool of graduate student responses (N = 131) was sufficient to conduct some comparative statistical analyses by subdividing the dataset by institution. Similar comparative analyses by institution were not possible for faculty responses due to the smaller sample size (N = 79). Direct comparisons of responses by discipline category were not possible for either survey group due to low numbers of responses in data subsets, especially for particular disciplines. Therefore, we could not examine or discern differences in RDM education between disciplines, although other researchers have claimed that some disciplinary differences in RDM practices do exist (Akers & Doty 2013; Weller & Monroe-Gulick 2014). Similarly, data subsets by major or academic rank were too small to allow statistical analyses of possible differences by these demographic characteristics.

A mismatch of response scales in the two surveys for parallel questions about the perceived importance of 12 DIL competencies prevented us from conducting statistical comparisons between graduate students and faculty. The error precluded us from being able to draw conclusions about differences in perceptions of importance of DIL competencies.

Questions on the graduate student survey required self-assessment. Responses involving perceptions may involve bias for a number of reasons, including over-confidence or lack of confidence in one's own abilities, unfamiliarity or misunderstanding of RDM terminology and concepts, or inaccurate memories about education behaviors. Self-assessments about knowledge

and skill levels for RDM, in particular, may not represent an accurate measure of actual abilities. Further research to compare self-assessment of RDM knowledge with tests of understanding of RDM concepts could provide further insight into knowledge gaps and education needs.

Results of this study raise additional questions about RDM education. Although students identified self-instruction as the primary means by which they learn RDM skills, it remains unclear as to the types and quality of sources students locate and use on their own. Additional research could delve into examining what pedagogical approaches (e.g., online modules, discipline-specific courses, or workshops taught by librarians or information technologists) are most effective for educating graduate students, and even faculty, in various RDM concepts. Disciplinary differences need to be explored further and identified in order to better tailor instruction. Of particular interest to follow up this study is a question of whether significant differences found between UNC and UW graduate students in their self-reported limitations in levels of knowledge and skills in data visualization are a reflection of disciplinary and academic program differences between the two institutions.

Conclusion

The needs of faculty researchers for support services and training in research data management has been widely written about in the library science literature in recent years as recognition of the importance for sharing of research data to scientific advancement in the digital age increased. However, the needs of graduate students for RDM education has been much less studied and defined. Given that graduate students may become the research leaders of tomorrow, more attention should be paid to preparing them for their future roles. Additionally, few studies have compared the RDM needs of graduate students and faculty.

Our study found RDM ethics and attribution was the competency identified as most important by both graduate students and faculty, followed by data visualization and quality assurance. Graduate students and faculty also identified ethics and attribution as the competency that graduate students are most knowledgeable about. Consequently, additional training in this competency may not be necessary. Data sharing competencies of data curation and reuse, metadata and data description, data conversion and interoperability, and data preservation were ranked lowest in both importance and knowledge levels of graduate students by both groups of respondents, graduate students and faculty.

A dichotomy between competencies related to conduct of research versus data sharing was evident for interest in training in specified RDM topics. Graduate students generally expressed a higher level of interest in training topics relating to the conduct of research, corresponding to the roles and responsibilities they have in RDM during their programs of study. By contrast, faculty in professorial ranks who expressed interest in additional RDM training for themselves were most interested in learning about writing a data management plan or in data backup and storage.

The greatest differences in responses by graduate students versus faculty were in frequencies of answers to questions about where graduate students learn about RDM. Our results suggest that graduate students rely more on self-instruction for RDM learning than faculty may realize. Additional study is needed to discern whether these relationships represent preferred approaches to learning or are responses to inadequate coverage of RDM concepts and competencies within curricula.

Graduate student respondents who provided comments recognized they were lacking knowledge in a number of RDM concepts. Ultimately, graduate students surveyed at both institutions indicated that RDM skills are important to them, and with the exception of ethics, RDM education is insufficient in the curriculum at both universities. These findings provide a basis for exploring ways librarians can work with other campus partners to help graduate students across the sciences and science education improve certain RDM skills. Faculty may be best suited to teach the disciplinary side of RDM, such as the cultures of practice competency.

Librarians can offer support and education in the areas of data planning, organization, metadata and preservation and help ensure that datasets are better indexed and accessible in open access repositories. Collaboration between faculty, especially those overseeing graduate student research, and librarians with expertise in RDM, particularly in data sharing competencies, can result in synergies that enhance learning. As such, librarians can play a key role in the knowledge creation process leading to long-term preservation and access to scientific information. Graduate students indicated preferences for a variety of learning approaches for RDM competencies; therefore, developing a variety of instruction approaches through collaboration of faculty and librarian experts may be necessary.

References

[UNC]. 2018 Spring Final Enrollment Profile [Internet]. Greeley (CO): University of Northern Colorado, Institutional Reporting and Analysis Services; 2018 [cited 2018 Jul 26]. Available from: http://www.unco.edu/institutional-reporting-analysis-services/pdf/enrollment-stats/Spring2018Final.pdf.

Akers, K.G. & Doty, J. 2013. Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation* 8(2):5-26. DOI: 10.2218/ijdc.v8i2.263.

Bracke, M.S. & Fosmire, M. 2015. Teaching data information literacy skills in a library workshop setting: A case study in agricultural and biological engineering. In: Carlson, J. & Johnston, L., editors. *Data information literacy: Librarians, data, and the education of a new generation of researchers*. West Lafayette (IN): Purdue University Press. p. 129-148. Available at https://www.jstor.org/stable/j.ctt6wq2vh.11.

Carlson, J. & Bracke, M. 2015. Planting the seeds for data literacy: Lessons learned from a student-centered education program. *International Journal of Digital Curation* 10(1):95-110. DOI: 10.2218/ijdc.v10i1.348.

Carlson, J., Fosmire, M., Miller, C.C. & Nelson, M.S. 2011. Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy* 11(2):629-657. DOI: 10.1353/pla.2011.0022.

Carlson, J., Jeffryes, J., Johnston, L.R., Nichols, M., Westra, B. & Wright, S.J. 2015a. An exploration of the data information literacy competencies: Findings from the project interviews. In: Carlson, J. & Johnston, L.R., editors. *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*. West Lafayette (IN): Purdue University Press. p. 51-70. Available at https://www.jstor.org/stable/j.ctt6wq2vh.8.

Carlson, J., Johnston, L., Westra, B. & Nichols, M. 2013. Developing an approach for data management education: A report from the Data Information Literacy Project. *International Journal of Data Curation* 8(1):204-217. DOI: 10.2218/ijdc.v8i1.254.

Carlson, J., Johnston, L.R. & Westra, B. 2015b. Developing the data information literacy project: Approach and methodology. In: Carlson, J. & Johnston, L.R., editors. *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*. West Lafayette (IN): Purdue University Press. p. 35-50. http://www.jstor.org/stable/j.ctt6wq2vh.7.

Carlson, J., Nelson, M.S., Johnston, L.R. & Koshoffer, A. 2015c. Developing data literacy programs: Working with faculty, graduate students and undergraduates. *Bulletin of the American Society for Information Science and Technology* 41(6):14-17. DOI: 10.1002/bult.2015.1720410608.

Carlson, J. & Stowell-Bracke, M. 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(4):343-361. DOI: 10.1353/pla.2013.0034.

Fernandez, P., Eaker, C., Swauger, S. & Davis, M.L.E.S. 2016. Public progress, data management and the land grant mission: A survey of agriculture researchers' practices and attitudes at two land-grant institutions. *Issues in Science and Technology Librarianship* 83. DOI: 10.5062/F49P2ZNN.

Frank, E.P. & Pharo, N. 2016. Academic librarians in data information literacy instruction: A case study in meteorology. *College & Research Libraries* 77(4):536-552. DOI: 10.5860/crl.77.4.536.

Jahnke, L., Asher, A. & Keralis, S.D.C. 2012. The problem of data [Internet]. Washington (DC): Council on Library and Information Resources [cited 2018 Jul 27]. Report No.: CLIR Publication 154. Available from: https://www.clir.org/pubs/reports/pub154/.

Johnston, L. & Jeffryes, J. 2014. 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):05013002. DOI: 10.1061/(ASCE)EI.1943-5541.0000154.

McLure, M., Level, A.V., Cranston, C.L., Oehlerts, B. & Culbertson, M. 2014. Data curation: A study of researcher practices and needs. *portal: Libraries and the Academy* 14(2):139-164. DOI: 10.1353/pla.2014.0009.

Mischo, W.H., Wiley, C.A., Schlembach, M.C. & Imker, H.J. 2017. An integrated data management plan instructional program [Internet]. [cited 2018 Jul 27]. *2017 ASEE Annual Conference & Exposition*; 2017 Jun 24; Columbus (OH). American Society for Engineering Education. Available from: https://peer.asee.org/27572.

O'Kelly, M., Garrison, J., Merry, B. & Torreano, J. 2015. Building a peer-learning service for students in an academic library. *portal: Libraries and the Academy* 15(1):163-182. DOI: 10.1353/pla.2015.0000.

Peters, C. & Vaughn, P. 2014. Initiating data management instruction to graduate students at the University of Houston using the New England Collaborative Data Management Curriculum. *Journal of eScience Librarianship* 3(1):e1064. DOI: 10.7191/jeslib.2014.1064.

Piorun, M., Kafel, D., Leger-Hornby, T., Najafi, S., Martin, E., Colombo, P. & LaPelle, N. 2012. Teaching research data management: An undergraduate/graduate curriculum. *Journal of eScience Librarianship* 1(1):46-50. DOI: 10.7191/jeslib.2012.1003.

Pouchard, L. & Bracke, M.S. 2016. An analysis of selected data practices: A case study of the Purdue College of Agriculture. *Issues in Science and Technology Librarianship* 85. DOI: 10.5062/F4057CX4.

Schmidt, L. & Holles, J.H. 2018. Teaching research data management: It takes a team to do it right! [Internet]. [cited 2018 Jul 16]. *2018 ASEE Annual Conference & Exposition*; 2018 Jun 23; Salt Lake City (UT). American Society for Engineering Education. Available from: https://peer.asee.org/31061.

[NSF]. Science and engineering degrees: 1966-2010, Appendix B: Classification of fields of study [Internet]. Arlington (VA): National Science Foundation, National Center for Science and Engineering Statistics (US); 2013 [cited 2018 Sep 17]. Available from: https://www.nsf.gov/statistics/nsf13327/content.cfm?pub_id=4266&id=4.

Sheehan, J., Kenning, A., Mannheimer, S., Knobel, C. & Llovet, P. 2015. Data-intensive science and campus IT [Internet]. *EDUCAUSE Review* [2015 Sep 28; cited 2019 Apr 18]. Available from: https://er.educause.edu/articles/2015/9/data-intensive-science-and-campus-it.

Tenopir, C., Birch, B. & Allard, S. 2012. Academic libraries and research data services: Current practices and plans for the future [Internet]. Chicago (IL): Association of College and Research Libraries. p. 1-54 [cited 2018 Jul 27]. Available from: http://www.ala.org/acrl/issues/whitepapers.

Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D. & Dorsett, K. 2015. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE* 10(8):e0134826. DOI: 10.1371/journal.pone.0134826.

Thielen, J., Samuel, S.M., Carlson, J. & Moldwin, M. 2017. Developing and teaching a two-credit data management course for graduate students in climate and space sciences. *Issues in Science and Technology Librarianship* 86. DOI: 10.5062/F42Z13HQ.

[UW]. University of Wyoming Enrollment Summary Spring 2018 [Internet]. Laramie (WY): University of Wyoming Office of Institutional Analysis; 2018 [cited 2018 Jul 27]. Available from: http://www.uwyo.edu/oia/student/eos/enroll-sum/.

Weller, T. & Monroe-Gulick, A. 2014. Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech* 32(3):467-482. DOI: 10.1108/LHT-02-2014-0021.

Whitmire, A.L., Boock, M. & Sutton, S.C. 2015. Variability in academic research data management practices: Implications for data services development from a faculty survey. *Program* 49(4):382-407. DOI: 10.1108/PROG-02-2015-0017.

Wiley, C. & Mischo, W.H. 2016. Data management practices and perspectives of atmospheric scientists and engineering faculty. *Issues in Science and Technology Librarianship* 85. DOI: 10.5062/F43X84NJ.

Wiley, C.A. & Kerby, E.E. 2018. Managing research data: Graduate student and postdoctoral researcher perspectives. *Issues in Science and Technology Librarianship* 89. DOI: 10.5062/F4FN14FJ.

Appendix A. Data Management Training: Graduate Student Survey

I am a student at:

- University of Northern Colorado
- University of Wyoming
- Other [Skip to end of survey]

What degree are you pursuing?

- Master's (M.S., M.A., M.Ed.)
- Doctoral (Ph.D., DNP, Ed.D.)
- Professional (MD, Pharm.D.)
- Other [Skip to end of survey]

- My degree program requires me to conduct original research and manage data (e.g., for a thesis or dissertation project).
- Yes
- No [Skip to end of survey]

Please indicate how important you believe it is for you to be knowledgeable in each competency by the time you graduate. [Matrix: Essential, Very Important, Somewhat Important, Not Important, I don't know or N/A]

- Databases and Data Formats: appropriate data types and formats, concepts of relational databases
- Discovery and Acquisition of Data: locating external sources of data, downloading and using data files
- Data Planning and Organization: developing data management plans, creating standard operating procedures, file naming conventions, file versioning, tracking components of data sets
- Data Conversion and Interoperability: standard data formats, migrating data from one format to another
- Quality Assurance: data consistency and completeness, data corruption/loss, data security and backup
- Metadata and Data Description: annotation for data understanding and re-use, metadata schemas, controlled vocabularies and ontologies/classifications, reproducibility
- Data Curation and Re-use: data lifecycle from raw stage to outputs, value beyond initial purpose, funder data sharing policies, identifiers
- Cultures of Practice: discipline-specific norms, standards and practices for managing research data
- Data Preservation: benefits and costs, technical and resource considerations for long-term storage of data
- Data Processing and Analysis: workflows and analysis tools, repetitive tasks automation, data summary and calculations
- Data Visualization: types of visual data representations, avoiding misrepresentation
- Ethics and Attribution: privacy and confidentiality, intellectual property, citing data

How would you describe your knowledge and skills in…? [Matrix: Outstanding, Proficient, Acceptable, Needs improvement, Deficient, I don't know or N/A]

- Databases and Data Formats: appropriate data types and formats, concepts of relational databases
- Discovery and Acquisition of Data: locating external sources of data, downloading and using data files
- Data Planning and Organization: developing data management plans, creating standard operating procedures, file naming conventions, file versioning, tracking components of data sets
- Data Conversion and Interoperability: standard data formats, migrating data from one format to another
- Quality Assurance: data consistency and completeness, data corruption/loss, data security and backup
- Metadata and Data Description: annotation for data understanding and re-use, metadata schemas, controlled vocabularies and ontologies/classifications, reproducibility
- Data Curation and Re-use: data lifecycle from raw stage to outputs, value beyond initial purpose, funder data sharing policies, identifiers
- Cultures of Practice: discipline-specific norms, standards and practices for managing research data
- Data Preservation: benefits and costs, technical and resource considerations for long-term storage of data
- Data Processing and Analysis: workflows and analysis tools, repetitive tasks automation, data summary and calculations
- Data Visualization: types of visual data representations, avoiding misrepresentation
- Ethics and Attribution: privacy and confidentiality, intellectual property, citing data

Where do you learn data management skills? [Matrix: Often, Sometimes, Never, N/A]

- In courses
- In special instruction from the Libraries
- In special instruction from campus Information Technology services From interactions with the Research Office From faculty/advisor
- From peers/fellow students
- Through self-instruction (e.g., tutorials, reading/literature review)
- Somewhere else (Please specify) [Comment box]

Which of these topics/program offerings would be of high interest for you? (Please click all that apply.)

- How to write a data management plan
- Data organization and best practices
- Data backup and storage
- How to document research data for sharing
- Licensing research data
- Complying with federal, sponsor and publisher data sharing policies
- How to preserve and archive data
- How to find a suitable repository and submit data
- Publishing a data set
- Citing shared data sets and measuring impact of re-use
- Finding data sets to reuse in research
- Other (Please describe) [Comment box]
- None

What is your role in managing research data? (Please click on all that apply.)

- Planning (methods for data collection, standards for data recording and metadata/documentation, grant proposals)
- Training (explaining procedures and practices to team members, sharing protocols)
- Creating (producing/acquiring data, quality control, storage and backup)
- Organizing (file types, file naming conventions, lab notebooks)
- Security (privacy, confidentiality concerns)
- Analysis (data summary, data visualizations, publication)
- Transfer/Preservation (creating metadata/documentation, file type conversions, long-term storage)
- Sharing (depositing data into an institutional or subject repository, controlling access, licensing, publishing datasets, assuring funder policy compliance)
- Other (Please describe) [Comment box]

What areas are you struggling with or wish you knew more about related to managing your research data, if any? [Comment box]

Any additional comments regarding training/education in research data management (e.g., needs, knowledge gaps, preferred approaches)? [Comment box]

What is your major/disciplinary degree program? [Comment box]

For education disciplines, do your subject areas include STEM (science, technology, engineering, math)?

- Yes
- No
- Not applicable

Are you interested in participating in a drawing for a Popsocket grip for a mobile device?

- Yes
- No [Skip to end of survey]

We thank you for your time spent taking this survey. Your response has been recorded. Please click here to register for the drawing. This form is not connected to your survey responses.

Appendix B. Data Management Training: Faculty Survey

My employer is (choose primary):

- University of Northern Colorado
- University of Wyoming
- Other [Skip to end of survey]

My current position is (choose primary):

- Administrator
- Professor
- Associate Professor
- Assistant Professor
- Clinical Professor or Professor of Practice or Professional-in-Residence
- Lecturer or Instructor (Assistant, Associate, Senior)
- Researcher/Research Associate/Research Assistant
- Post-Doctoral Fellow
- Other [Skip to end of survey]

Please indicate how important you believe it is for your graduate students to be knowledgeable in each competency by the time they graduate. [Matrix: Essential, Very important, Important, Somewhat important, Not important, Don't know or N/A]

- Databases and Data Formats: appropriate data types and formats, concepts of relational databases
- Discovery and Acquisition of Data: locating external sources of data, importing and converting data files
- Data Planning and Organization: developing data management plans, creating standard operating procedures, file naming conventions, file versioning, tracking components of data sets
- Data Conversion and Interoperability: standard data formats, migrating data from one format to another
- Quality Assurance: data consistency and completeness, data corruption/loss, data security and backup
- Metadata and Data Description: annotation for data understanding and re-use, metadata schemas, controlled vocabularies and ontologies/classifications, reproducibility
- Data Curation and Re-use: data lifecycle from raw stage to outputs, value beyond initial purpose, funder data sharing policies, identifiers
- Cultures of Practice: discipline-specific norms, standards and practices for managing research data
- Data Preservation: benefits and costs, technical and resource considerations for long-term storage of data
- Data Processing and Analysis: workflows and analysis tools, repetitive tasks automation, data summary and calculations
- Data Visualization: types of visual data representations, avoiding misrepresentation
- Ethics and Attribution: privacy and confidentiality, intellectual property, citing data

On average, how would you describe your graduate students' knowledge and skills in…?
[Matrix: Outstanding, Proficient, Acceptable, Needs Improvement, Deficient, Don't Know or N/A]

- Databases and Data Formats: appropriate data types and formats, concepts of relational databases
- Discovery and Acquisition of Data: locating external sources of data, importing and converting data files
- Data Planning and Organization: developing data management plans, creating standard operating procedures, file naming conventions, file versioning, tracking components of data sets
- Data Conversion and Interoperability: standard data formats, migrating data from one format to another
- Quality Assurance: data consistency and completeness, data corruption/loss, data security and backup
- Metadata and Data Description: annotation for data understanding and re-use, metadata schemas, controlled vocabularies and ontologies/classifications, reproducibility
- Data Curation and Re-use: data lifecycle from raw stage to outputs, value beyond initial purpose, funder data sharing policies, identifiers
- Cultures of Practice: discipline-specific norms, standards and practices for managing research data
- Data Preservation: benefits and costs, technical and resource considerations for long-term storage of data
- Data Processing and Analysis: workflows and analysis tools, repetitive tasks automation, data summary and calculations
- Data Visualization: types of visual data representations, avoiding misrepresentation
- Ethics and Attribution: privacy and confidentiality, intellectual property, citing data

Where do your students learn data management skills? [Matrix: Often, Sometimes, Never, Don't Know]

- In courses
- In special instruction from the Libraries
- In special instruction from campus Information Technology services
- From interactions with the Research Office
- From faculty/advisor
- From peers/fellow students
- Through self-instruction (e.g., tutorials, reading/literature review)
- Somewhere else (Please specify) [Comment box]

How do you teach your students to manage research data? (Click on all that apply.)

- One on one mentorship
- Online tutorial
- In class
- In lab/field
- Other/please elaborate [Comment box]

What do you specifically teach regarding research data management? (Click on all that apply.)

- File naming conventions
- Metadata
- Privacy
- Open data
- Storage data practices
- Creating data sets
- Other/please elaborate [Comment box]

What research data management concepts or skills are you uncomfortable teaching or find difficult to teach, if any? [Comment box]

Which of these topics/program offerings would be of high interest? (Please check all that apply.) [Matrix: For your students? For yourself/faculty?]

- How to write a data management plan
- Data organization and best practices
- Data backup and storage
- How to document research data for sharing
- Licensing research data
- Complying with federal, sponsor and publisher data sharing policies
- How to preserve and archive data
- How to find a suitable repository and submit data
- Publishing a data set
- Citing shared data sets and measuring impact of re-use
- Finding data sets to reuse in research
- Other (describe)
- None

Who do you think should be responsible for educating graduate students in regard to data management skills? [Comment box]

What areas are you struggling with related to managing your research data, if any? [Comment box]

Any additional comments regarding training/education in research data management (e.g., needs, knowledge gaps, approaches)? [Comment box]

What is your primary discipline for research? [Comment box]

For education disciplines, do your subject areas include STEM (science, technology, engineering, math)?

- Yes
- No
- Not applicable