Knowledge and Practice Changes Following a Student Data-Focused Data Management Education Program

Tina M. Griffin

This article underwent fully anonymous peer review in accordance with JLSC's peer review policy.

Knowledge and Practice Changes Following a Student Data-Focused Data Management Education Program

Tina M. Griffin

Information Services and Liaison Librarian, University of Illinois Chicago

ABSTRACT

Introduction: It is known that graduate students work with research data more intimately than their faculty mentors. Because of this, much data management education is geared toward this population. However, student learning has predominantly been assessed through measures of satisfaction and attendance rather than through evaluating knowledge and skills acquired. This study attempts to advance assessment efforts by asking students to report their knowledge and practice changes before, immediately after, and 6 months following education.
Method: Graduate students in science, technology, engineering, and mathematics (STEM) and health science disciplines self-enrolled in an 8-week pilot data management program that used their research projects as the focus for learning. Three surveys were administered (pre-, post-, and 6 months following) to determine changes in students' knowledge and practices regarding data management skills through self-assessment. The survey consisted of approximately 115 Likert-style questions and covered major aspects of the data life cycle.
Results and Discussion: Overall, students increased their data management knowledge and improved their skills in all areas of the data life cycle. Students readily adopted practices for straightforward tasks such as determining storage and improving file naming. Students improved in but struggled with tasks that were more involved, such as sharing data and documenting code. For most of these practices, students consistently implemented them through the 6-month follow-up period.
Conclusion: Impact of data management education lasts significantly beyond immediate instruction. In-depth assessment of student knowledge and practices indicates in which areas this education is effective and in which areas it needs further support. It is likely that this effect is due to the program length and focus on implementation.
Implications for practice

- An increase in data management knowledge and practices continues to at least 6 months with student-research-focused data management education.

- Students adopt and keep up simpler practices easily, including file-naming conventions, but need more time and support to adopt complicated practices such as data sharing.

- Data management education is effective when presented in a longer format, with compensation/incentives and time for practice implementation.

INTRODUCTION

In previous work, the author described an approach to data management curriculum that was novel, relative to the literature, in three respects: by framing instruction content in three themes (discovery, infrastructure, and sustainability) that highlight interdependencies between data management practices, by centering all active learning activities on students' research projects, and by using a flipped classroom to give students time for implementing what they learned into their own research projects (Griffin, 2020). This paper follows that work and analyzes the changes in students' knowledge and practices up to 6 months after completing the data management program.

LITERATURE REVIEW

It is recognized that graduate students interact with research data more than faculty (Fong and Wang, 2015) and can have a shared or total responsibility for managing research data (Lage et al., 2011; Peters and Dryden, 2011). In their supervision of students, faculty acknowledge that students cannot yet manage data well (Carlson et al., 2011, 2015). But this is not solved during the course of graduate education, as industrial partners who employ these students do not consider them data fluent as they enter the workforce (Van Tuyl and Whitmire, 2018). Students report that they learn data management skills from peer lab members or through their mentor/ supervisor (Pouchard and Bracke, 2016). However, researchers admit that they often do not know how to properly manage data (Tenopir et al., 2011) or do not have good habits (Whitmire et al., 2015). Although researchers are willing to get assistance, they are not always willing to pay for these services (Steinhart et al., 2012). Students are then left on their own to develop data management skills (Carlson et al., 2013; Johnston and Jeffryes, 2014a). Some may learn these skills, as they are associated with research ethics courses or seminars or in laboratory methods coursework, but they are not associated with direct research practice (Johnston and Jeffryes, 2014a). Ultimately, this leaves students without comprehensive formal training and with patchwork informal training (Bishop and Borden, 2020; Carlson and Johnston, 2015; Johnston and Jeffryes, 2014a).

Identifying specific graduate student needs

Graduate students' needs have been indicated in various environmental scans and needs assessments. However, this population has rarely been the focus of these investigations. At least several studies have shown a need for data management training for students, sometimes teased out from a larger population. The majority occurred from 2011 to 2015 and address various aspects of a data life cycle.

Similar to what has been seen in faculty needs assessments, the most common need identified for or by students involves the issues of storage and backup. These issues include what storage options are available within the institution, lab, or work group (Fong and Wang, 2015; Mattern et al., 2015); how long to store active or backup data (Bishop and Borden, 2020; Johnston and Jeffryes, 2014a); dealing with storage constraints and file types (Lage et al., 2011; Marcus et al., 2007); and what practices should be used for proper storage and backup (Adamick et al., 2012; Fong and Wang, 2015; Weller and Monroe-Gulick, 2015). This is not surprising, as Peters noted that students and lab personnel are, in many cases, the ones in control of storage and backup (Peters and Dryden, 2011).

Organization and general data management skills have also been identified as a high priority. In some studies, this was a general, blanket need (Adamick et al., 2012; Bishop and Borden, 2020; Marcus et al., 2007; Peters and Dryden, 2011; Pouchard and Bracke, 2016). In others, specific skills were mentioned, such as file versioning (Carlson et al., 2011; Fong and Wang, 2015; Johnston and Jeffryes, 2014a), file-naming practices (Carlson et al., 2011; Fong and Wang, 2015; Johnston and Jeffryes, 2014a), and collocating data or documentation and/or metadata (Johnston and Jeffryes, 2014a; Valentino and Boock, 2015).

Interestingly, finding and working with data was next identified as a need, and there were a variety of skills associated with it. Data analysis and visualization was mentioned frequently (Adamick et al., 2012; Bishop and Borden, 2020; Carlson et al., 2011; Johnston and Jeffryes, 2014a; Marcus et al., 2007; Pouchard and Bracke, 2016; Valentino and Boock, 2015) as well as digitization (Weller and Monroe-Gulick, 2015), data collection practices (Adamick et al., 2012), data citation (Carlson et al., 2011; Valentino and Boock, 2015), data ownership (Valentino and Boock, 2015), and general finding data and/or working with repositories (Adamick et al., 2012; Johnston and Jeffryes, 2014a; Mattern et al., 2015).

Standards, documentation, and policy were sometimes mentioned as a student need. Many times, this was in relation to understanding how and when to apply data management skills and in what context. Understanding federal and university policies were mentioned (Adamick et al., 2012; Johnston and Jeffryes, 2014a), as were understanding lab policies and understanding values and norms for a given discipline (Pouchard and Bracke, 2016; Valentino and Boock, 2015). Standards were also mentioned in relation to applying metadata and creating codebooks (Carlson et al., 2011; Johnston and Jeffryes, 2014a; Pouchard and Bracke, 2016; Valentino and Boock, 2015).

Last, and mentioned least, were the issues of curating, preserving, and sharing data. Needs assessments mentioned that students share data, but it is mostly within the research community or within laboratory or collaborative groups (Marcus et al., 2007; Peters and Dryden, 2011).

When mentioned, developing sharing skills is an anticipated need, presumably for when students are done with their research work (Bishop and Borden, 2020; Carlson et al., 2011; Fong and Wang, 2015; Peters and Dryden, 2011; Valentino and Boock, 2015; Weller and Monroe-Gulick, 2015). Similarly, curation and preservation are mentioned as anticipated needs that are not well defined (Bishop and Borden, 2020; Carlson et al., 2011; Johnston and Jeffryes, 2014a; Pouchard and Bracke, 2016; Valentino and Boock, 2015; Weller and Monroe-Gulick, 2015).

Workshop-based data management education and assessment

Libraries have been continually offering data management education in the form of workshops, embedded lectures, seminar guest lectures, and similar events for over 10 years to address these needs (Fearon et al., 2013; Tenopir et al., 2012, 2015). The reported education efforts have taken a variety of formats. Most often, these are workshops (Adamick et al., 2012; Clement et al., 2017; Eaker, 2014; Johnston et al., 2012; Macy and Coates, 2016; Otto, 2012; Surkis et al., 2017; Thielen and Nichols Hess, 2017; Valentino and Boock, 2015), but they have also been integrated into seminar series (Fong and Wang, 2015), embedded in courses (Thielen and Nichols Hess, 2017), or taken place during other events such as a department annual retreat (Fong and Wang, 2015). Also, some were targeted to a specific discipline (Adamick et al., 2012; Fong and Wang, 2015; Macy and Coates, 2016; Thielen and Nichols Hess, 2017), whereas most were discipline agnostic or open to all disciplines. Standalone workshops tended to focus on a single topic such as writing data management plans (Johnston et al., 2012), finding and using data (Otto, 2012), or an overview of data management principles (Adamick et al., 2012; Fong and Wang, 2015; Valentino and Boock, 2015), whereas extended workshop series or day-long workshops tended to dive into a selection of data management practices (Clement et al., 2017; Eaker, 2014; Fong and Wang, 2015; Surkis et al., 2017; Thielen and Nichols Hess, 2017).

As many of these efforts were first attempts at offering data management education in some form, many of the assessments were limited to determination of participant satisfaction. Questions included whether the stated learning objectives were met, how useful the content was, whether the sessions would be recommended to others, what participants learned or still had questions about (or what they liked/did not like), and suggestions for improvement (Adamick et al., 2012; Eaker, 2014; Johnston et al., 2012; Macy and Coates, 2016; Surkis et al., 2017; Valentino and Boock, 2015). A few studies attempted to quantify learning or determine change in understanding. Thielen and Nichols Hess (2017) administered a student self-assessment that rated prior knowledge and perceived change. Clement et al. (2017) surveyed participants 8 months following the workshop, and participants reported they were still able to apply data management principles. Macy and Coates (2016) were able to review the course deliverables for data integration and quality and reported improvement. Finally, Otto (2012)

was able to report an increase in finding and using data through both a pre-post test and examination of course assignments.

Attendance was often used as a measure of success, but attendance was not consistently documented or was sometimes recorded as incomplete. An initial number compiled from these studies counts 700 attendees. A total of 86 of those were graduate students, but, because of incompleteness, that number is unreliable. For workshop series, a common complaint was an initial high attendance with attrition over time.

Course-based data management education and assessment

Beginning in 2010, documentation of credit-bearing courses focused on data management began to emerge in the literature. Offering credit-bearing courses solves some problems seen with workshops and other similar formats. Courses encourage students to remain engaged with the materials because of the grading incentive, and they allow a breadth of data management principles to be taught with some depth. These courses also allow a more substantial assessment to be attempted through evaluating course assignments and other student products.

Qin and D'Ignazio (2010) offered a 14-week course that covered topics such as databases and programs, data formats, data description/metadata, managing data, cleaning, storage, access, organization/workflows, curation, archiving, funder requirements, visualization, and ethics. This effort was National Science Foundation funded and was offered for 2 years. Assessment included student feedback on whether assignments were useful or interesting. The authors did administer a pre-post survey asking students to evaluate their attitudes about science and being a data manager. From those results, Qin and D'Ignazio reported that perceived self-competence did not change.

Whitmire (2015) offered a similar course that ran 11 weeks for two credits. The course covered a comparable range of topics and was partially assessed based on student satisfaction (how well sessions are preparing them to meet learning outcomes, what they liked most/least, suggestions for improvement). The summative assessment was an evaluation of a data management plan as their final project. Whitmore reports that "students rated the course highly" and "quality of data management plans indicated that they met learning objectives" (p. 10).

Thielen et al. (2017) offered a two-credit, 14-week course designed for engineering graduate students. The content aligned with Accreditation Board for Engineering and Technology accreditation standards. An environmental scan of students and faculty showed that related data information literacy competencies were not currently covered in their education. Course content included writing data management plans, storage, documentation, metadata,

visualization, sharing, curation, preservation, Internet Protocol/licensing, and ethics, all with a climate and space science focus. Assessments were based on student reflections (what they learned, whether they had more questions, and instruction effectiveness). Similar to Whitmire (2015), a full data management plan, this time with an accompanying presentation, was the final project. A pre-post survey of student perceptions of importance and current knowledge in nine data information literacy competencies, as outlined in the course content, showed improvement in all competencies assessed. Although the course was meaningful, unfortunately enrollment in the course was low ($n = 5$).

Schmidt and Holles (2018) offered a three-credit course to science and engineering graduate students for a full semester. Similar to previous efforts, topics covered a similar range of data management skills. For assessment, students rated their self-knowledge in eight areas (data management planning, archiving/preservation, sharing/reuse, legal/ethical concerns, documentation/metadata, storage/backup/security, organization, and types/formats) on a five-point Likert scale (poor = 1 to excellent = 5). In the pre-course assessment, most areas averaged between 2 and 2.5, and only organization was rated 3. Afterward, every area averaged a minimum of 3, and organization knowledge averaged nearly 4. Increases were found for each area of education.

Last, Wiljes and Cimiano (2019) taught 15 data management courses over 6 years in Germany. The course was initially offered as an elective, but, in 2016, two master's programs made it mandatory. Course content remained comparable to what has already been discussed. Also, similar to many workshops and several courses, an assessment reported student-perceived value of content, their willingness to recommend the course to others, and their judgement on the importance of the topic. Attendance was mostly master's students with some bachelor's and doctoral students. All disciplines were welcome, but enrollees were predominantly from science, technology, engineering, and mathematics (STEM) disciplines.

Regardless of the format, most of these education efforts, particularly workshops, assess their success by participant satisfaction and attendance numbers. Coates et al. (2018) recognize that data management services by libraries are maturing, and efforts at more substantial assessment could be attempted. This is seen in the work by Schmidt and Holles (2018) and by Thielen et al. (2017), but only Clement's (2018) study evaluated whether the education had any lasting effect.

This paper advances assessment efforts by asking students to self-report their knowledge and practice changes before, immediately after, and 6 months following data management education. Because of the design of the program, this study is also able to obtain more granular information about student learning data management skills beyond improvement in broad foundational concepts.

METHODS

As previously described, students participated in an 8-week pilot data management program that used a flipped classroom approach, centered the activities entirely on their current research projects, and dedicated time for participants to implement data management infrastructure into their projects (Griffin, 2020). A pretest was distributed after acceptance to the program and was required to be completed before the sessions started. The posttest was distributed and completed within 2 weeks after the program ended. The 6-month follow-up was distributed February 2020 and was completed within 2 weeks. The full survey consisted of about 115 questions regarding knowledge, confidence, and practices regarding data management and took 30 to 45 minutes to complete. Data regarding student knowledge and practices are reported here.

Surveys were developed with the support of our Survey Research Lab. A senior graduate student not participating in the program also reviewed the questions, and the instruments were subsequently edited for bias and clarity. The Institutional Review Board reviewed the instruments and determined them to be exempt (#2019-0048). The survey was built in Qualtrics (Qualtrics, Seattle, Washington and Provo, Utah) and distributed to participants by email. The survey instruments can be found in Appendix A. Participants were compensated for participation in the program and completion of the surveys for a total of $150 in three distributions based on completion.

This pilot program aligned with the data information literacy competencies (Carlson and Johnston, 2015), and assessment was guided by those standards. Directly measuring skills is difficult at this time because of the differences in subject discipline needs regarding data management, the lack of standards in any discipline, and lack of infrastructure within institutions. Developing a single, measurable assessment instrument that directly assessed skills across these gaps and with granularity beyond what has been published to date was impossible. Therefore, participants were asked to assess their knowledge and practice changes and, at the end of the program, were asked to evaluate whether they over-, under-, or correctly estimated their knowledge in these areas. Likert scales were used for most questions. Participants were asked about their familiarity with data management concepts or information ("Not At All," "Not Very," "Somewhat," "Very," "Extremely") and how often they practiced a given skill ("Never," "Rarely," "Sometimes," "Often," "Always"). Questions were organized by data management domain and roughly follow different stages of a data life cycle. These domains include standards and documentation, organization and file management, storage and backup, and sharing.

Responses were downloaded and coded to anonymize individuals. Incomplete surveys were not included (participants who did not "submit" at the end), but individuals could skip questions. A total of 26 participants submitted the pretest, 18 submitted the posttest, and 16

submitted the 6-month follow-up. Analysis within this paper includes basic counts, percentages, and descriptive statistics. Responses regarding knowledge and practice changes before, after, and 6 months following data management education are reported.

RESULTS

Program participants were restricted to students in STEM or health science disciplines who were in a graduate-level program working on original research. A total of 26 participants enrolled in the program. Most were obtaining doctoral degrees (24), and two were pursuing a master's. Seven disciplines were represented: nursing, medicine, public health, applied health sciences, engineering, liberal arts and sciences, and pharmacy. Students' progress toward their degrees were from the first year of study to within 6 months of graduation. Before beginning, students reported that only three of them had previous data management education, and this was through a credit-bearing course (3/26 [11%]); the rest had none.

Standards and documentation

When asked how familiar students are with documentation standards required by journals, funders, or their research community, on average, most replied that they were "Not Very" familiar. After the program, they reported being at least "Somewhat" familiar (Figure 1). They also reported an increased familiarity with established metadata standards, ontologies, data elements, and similar items averaging from "Not Very" to at least "Somewhat" familiar. This was retained through the 6-month follow-up. Relatedly, when asked to give a definition for metadata, about half could give a correct or partially correct one, but, after education, almost all could give a solid, correct definition (Figure 2).
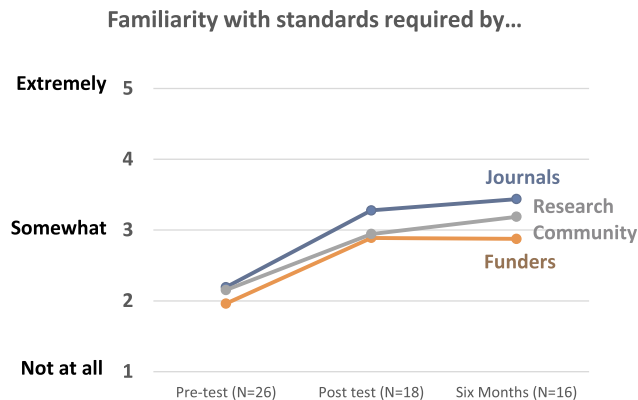

Figure 1. Students' familiarity with existing standards.

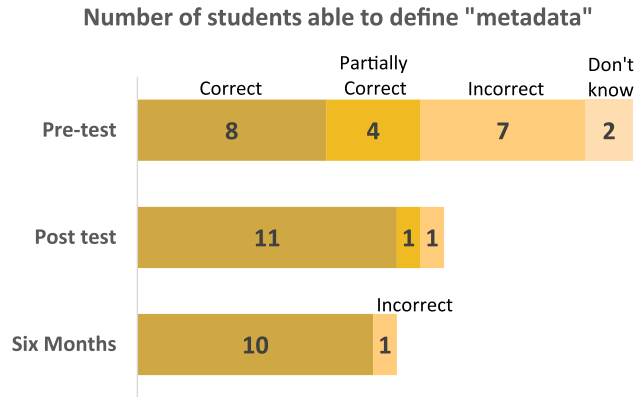**Number of students able to define "metadata"**

Figure 2. Number of students able to define the term "metadata."

When asked more specific questions regarding their documentation knowledge, they increased familiarity ("Somewhat" to "Very") with wayfinding tools such as tables of contents and creating identification systems for data and increased from "Not at All" to "Somewhat" with creating inventories.

When asked about practices, over 80% of participants were already including written project descriptions in their work (20/24 [83%]) and continued to do so, but, after education, they increased use of data dictionaries and codebooks from 29% to 61% (7/24 to 11/18) and increased documentation of analysis methods from 68% to 83% (16/24 to 15/18). Dictionary/codebook use were sustained 6 months after (11/16 [69%]), but documentation of analysis methods decreased (9/16 [56.3%]). Some students were coding as part of their projects and keeping documentation of their code (10/24 [42%]), but that activity neither significantly increased nor decreased immediately after (7/18 [38.9%]) or 6 months after (6/16 [37.5%]).

For the infrastructure work that they did create and implement, students reported that they were creating more continuity and context throughout their projects (Table 1).

Organization and file management

Students were, on average, "Somewhat" familiar with folder organization practices before the program began. Afterward, they reported being "Very" familiar with them both immediately after education and 6 months following. Half of these students (13/25 [52%]) already had an intentional folder organization scheme, and, after the program, 89% had one, and this continued through the 6 months. Of those with an intentional folder organization at each time point, most said that they are "Often" consistent with following this scheme (pretest = 61.5% [8/13], posttest = 69% [11/16], and 6 months = 77% [10/13]).

		Pretest ($n = 26$)	Posttest ($n = 18$)	6 months ($n = 16$)
Do you create tables of content, indexes, or similar tools to help find your data?	Y	5	12	9
	N	21	6	7
Do you create standardized vocabulary, definitions, acronyms, and/or values to use across experiments/assays, etc.?	Y	7	12	12
	N	19	6	4
Do you use a standardized identification scheme/code to create context between files?	Y	3	10	10
	N	23	8	6
Is there continuity between your digital files for your research project and analog materials?	Y	11	15	13
	N	15	3	3

Table 1. Students' creation and use of documentation.

Students reported being "Not Very" familiar with file management practices, which include file-naming conventions, versioning files, and working with file management software (Figure 3). After the program, they reported being "Very" familiar with naming and versioning but still "Not Very" familiar with software that manages files such as electronic lab notebooks (ELNs). Most students "Rarely" or "Never" use software to manage files even after the program and through the 6-month follow-up.
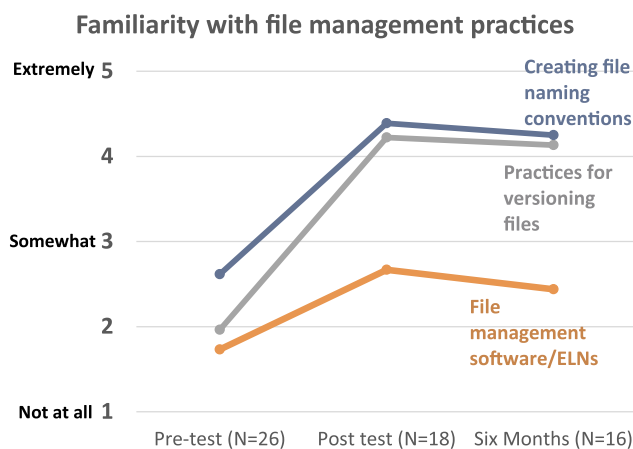

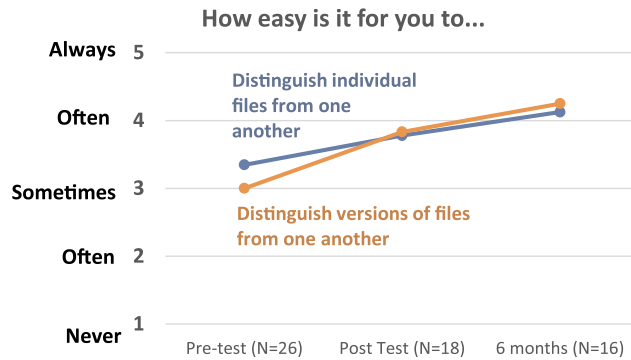**Figure 3.** Students' familiarity with file management practices.

Figure 4. Students' ability to disambiguate files and file versions.

When asked, only 24% (6/25) of students reported having a standardized file-naming convention initially. After the program, 83% did (15/18), and, 6 months following, 94% did (15/16). If the students did have a file-naming convention, most of them reported that the conventions were consistent across projects, across all time points asked. For those that had one, most consistently followed that convention "Often" but not "Always." Immediately after and during the follow-up, they still consistently followed it "Often," but one-third were now "Always" following it.

Because including dates in file names is common practice, students were also asked about their date conventions. Half of the students already used a standard date format in their file names, and, immediately after the program, 77% (13/18) and later 81% (13/16) did. Similar to what was seen with following their organization scheme, if students included or had started including a standard date format, students were split between following this format "Often" or following it "Always," at all times asked; no students followed this date format "Sometimes" or less frequently.

Students, on average, reported that they "Sometimes" can distinguish files from one another or distinguish different versions of the same file data. After the program and 6 months following, they were "Often" able to distinguish files and file versions from one another (Figure 4). Interestingly, at any time point, only about 25% of students reported needing to open files in order to know what is in them.

Storage and backup

Within the domain of storage and backup, students were "Somewhat" to "Not Very" familiar with their options and best practices. After the program, they were "Very" familiar with all of these (Figure 5). Most students also reported backing up their data (20/26, 77%), but after
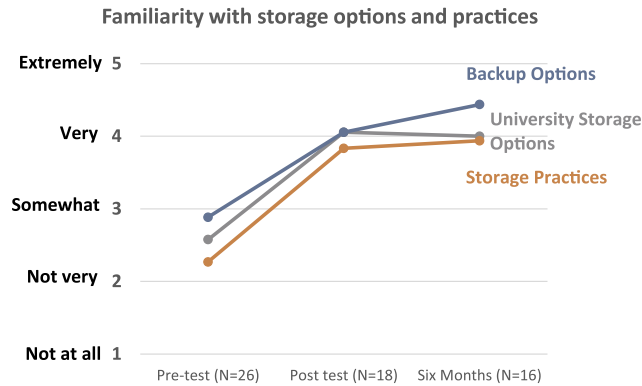
Familiarity with storage options and practices

Figure 5. Students' familiarity with storage options and practices.
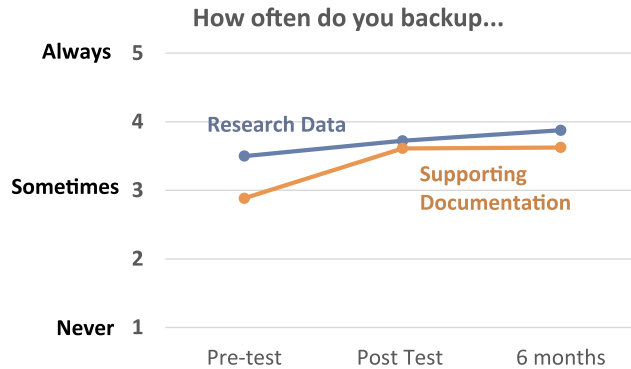
How often do you backup...

Figure 6. Students' backup practices regarding data and documentation.

the program and 6 months following, all students who completed the assessment were backing up their data.

They were also consistent in frequency of backing up their data from "Sometimes" to "Often," but after the program, they improved in also backing up supporting documentation that explains their research data (Figure 6).

Students were also asked what the relationship was between storage locations and the data that they contain in order to see whether those practices had changed as well. Figure 7 demonstrates that, before the program, most students who "backed up data" stored the data in many locations, but there was little relationship between storage location and individual or redundant data files, and after the program, more students had their data compiled at one location and copied to other locations, with no students needing to know which location their data was in in order to find it.

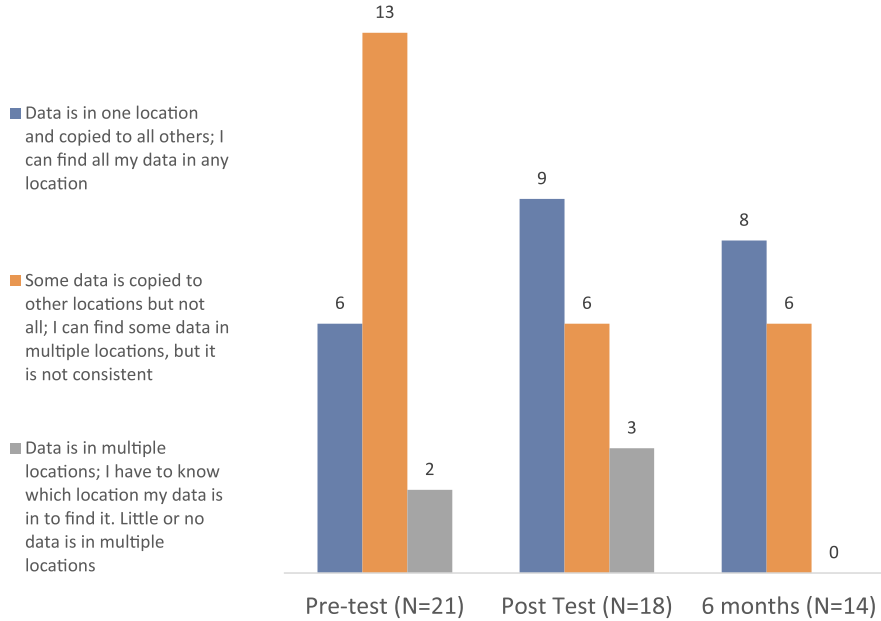**What is the relationship between locations where your data is stored?**

■ Data is in one location and copied to all others; I can find all my data in any location

■ Some data is copied to other locations but not all; I can find some data in multiple locations, but it is not consistent

■ Data is in multiple locations; I have to know which location my data is in to find it. Little or no data is in multiple locations

Figure 7. Students' practices regarding storage locations and backup.

Sharing

Similar to what was seen in the "Standards and documentation" section, students were "Not Very" familiar with data sharing requirements by various stakeholders but became more so over time, particularly with journals (Figure 8).
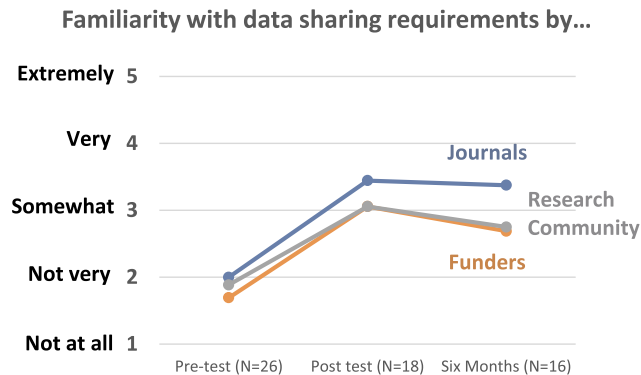
Familiarity with data sharing requirements by…

Figure 8. Students' familiarity with data sharing requirements.
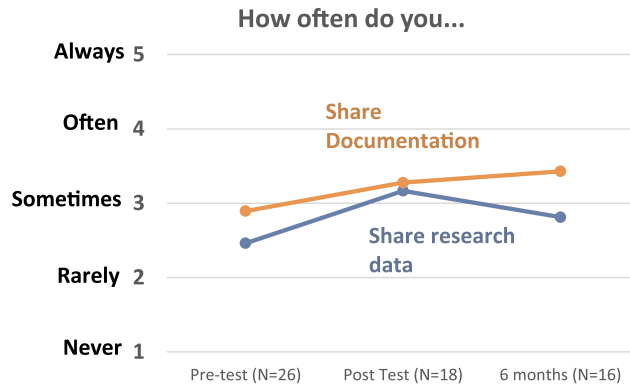
How often do you...

Figure 9. Students' practices regarding data sharing.

Regarding sharing practices, students were more often sharing their documentation than their research data. As they progressed, they increased sharing documentation through the 6 months, whereas sharing research data initially increased and dropped off, but not to the original levels (Figure 9).

Self-estimate

At the end of the program, but not at 6 months, students were asked to provide a self-estimate of their knowledge of data management practices across the different domains. Most students reported that they overestimated their knowledge of organization and file management. Documentation, sharing, and storage knowledge were split between students who correctly estimated and those who overestimated. Whereas most students correctly estimated their knowledge of standards and policies, very few students underestimated their abilities in any domain (Figure 10).
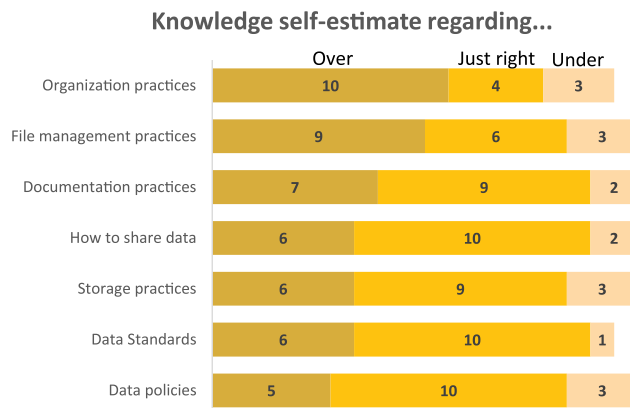
Knowledge self-estimate regarding...

Figure 10. Students' self-estimate of data management practices

DISCUSSION

Because data management education is not yet mandatory in higher education, students are exposed to these concepts and practices in an inconsistent manner or through education that is disassociated from their day-to-day research work. Because of this, it has been difficult to determine instruction effectiveness. This study aimed to measure whether differences could be seen in data management knowledge and practices over time after data management education that was specifically aimed at learning and practicing within their research projects.

This pilot program shared similarities with both workshop series and data management courses. It allows data management principles to be taught in-depth because it takes place over 8 weeks, similar to many workshop series and academic quarters. The topics in this course are comparable in that they cover most of the data life cycle and the skills associated in each step. It also incentivizes student engagement—in this case, with compensation rather than with grades. This program differs in that the curriculum was designed to teach data management skills as interrelated and dependent skills, which has not been done before. It also uses a flipped classroom model, which has only been documented once (Johnston and Jeffryes, 2014b). These approaches may be contributing factors to increased knowledge and skills over the time surveyed.

Assessment for this program is also more specific than what has been typically administered and moves beyond measures of attendance and satisfaction. Similar to other authors' work (Schmidt and Holles, 2018; Thielen et al., 2017; Thielen and Nichols Hess, 2017), this study relies on students' self-assessment, which is in line with the student research focus of the curriculum. The diversity in disciplines and research topic made it impractical to develop tests that objectively measure student knowledge. Considering that the students are usually accurate in estimating their knowledge (Figure 10), it is reasonable to trust their self-perception of knowledge and skills, which has been done in-depth here. This assessment also indicates that these skills and practices are retained beyond the immediate instruction period. Only Clement et al. (2017) have reported whether skills were sustained but that evaluation was in broader terms. The data presented here can be helpful to demonstrate library education's impact on a longer time scale regarding more specific skills.

Overall, this study shows improvement of students' knowledge and practices in almost all data management domains. For all categories investigated, students reported being initially "Not Very" or "Somewhat" familiar with foundational data management concepts. After education, the improvement was typically one rank higher to "Somewhat" or "Very" familiar, and this was maintained over time.

Looking granularly at acquisition and retention of individual practices within these domains, beginning with standards and documentation, this study shows that students increased familiarity with a variety of standards and expectations across stakeholder types. They also increased the creation of tools and various documentation types such as codebooks but struggled with maintaining documentation of analysis or coding. Not all students use code in their projects, but all are doing some sort of analysis. Future education will need to understand that, as students learn these skills, they need additional support to implement and maintain them. They are having difficulty on their own, even with initial time for implementation.

Similarly, within organization and file management, students pick up these skills readily and typically have some sort of scheme already in place. The question is whether those skills can be refined to be more effective for their projects. It appears that students acquire file-naming skills easily and use them consistently. On the other hand, students did not improve knowledge of software options or use. This topic was discussed in the program but not encouraged as a solution, as there is no institutional support for these products at any level at this institution, and the program did not want to impose a technical solution to a potentially collaborative workflow within the students' labs. Because of this, this is not a surprising result. However, this may be an area that needs to be developed. Research into marketplace needs regarding software use (academic, industrial, or other) would help align this education area.

Storage and backup also showed significant improvement. Most students were able to discern appropriate store options and backup practices. This is likely because these activities are a "set it and forget it" type, until circumstances change. This can certainly happen within a student project, but with options known and understood, or even anticipated, the decision becomes simpler to execute. Students are often constrained regarding storage and backup. It is known that most students store data on their personal devices, whether appropriate or not. With guidance, it seems that students will readily change their habits in this area.

Last, students' familiarity with data sharing does improve with education. Unsurprisingly, this aligns with stakeholders that are closest to them, such as journals rather than funders. They do not sustain sharing research data as much as sharing supporting materials. This study did not delve into who they share with and how. Further work might explain why sharing research data is not as strong. One factor may be where they are within the timeline of their projects. Within this program, students range between just starting their research to being within 6 months of graduation. It is possible that this data reflects those who have not progressed through their work enough to share. However, they are aware of the need to do so.

Based on the student self-estimate, students are accurate in their ability to determine their needs and estimate their knowledge and practices. They overestimated knowledge in domains

that were already familiar to them but correctly estimated within domains that were less familiar. Because students elect to participate in a program like this one, they have already identified that there could be improvements, and so this is an unsurprising result. If programs like this become mandatory, it would be interesting to see the changes in estimates, as students who believed that they already have adequate knowledge and practices (and would therefore not take a data management course) would now be included in the data.

This work supports the hypothesis that centering students' research projects in data management education has lasting effects, but further studies directly comparing programs that center students work versus those that do not will be needed.

Limitations

Because of the small number of participants, stronger statistical assessment of the data was not appropriate. The small number and the focus on STEM and health science graduate students from one institution prevent the data from being generalized to all graduate students. Students were being compensated for participation in the program and knew that there would be a follow-up survey at 6 months. It is possible that the compensation and the expectation of follow-up affected their practices. Because of the diverse discipline representation among the student participants and the lack of data management standards across disciplines, a direct assessment of students' skills was not possible at this time.

CONCLUSION

Thielen et al. (2017) summarize well some of the challenges associated with data management education.

> "Time is needed to explore data management topics in-depth…Shorter education opportunities such as workshops or seminars cannot cover the same breadth of topics, and do not give students the opportunity to apply or practice…Through the progression of the course, students see how topics are connected and related" (p. 9).

Because data management education is elective, is offered as workshops, and is not linked to other incentives, it can be difficult to evaluate learning. Whereas workshops serve their primary purpose to teach data management principles and practices, they also serve a secondary purpose to raise awareness of the library as a data management service provider. This is seen in assessments that measure attendance and general satisfaction, as well as ask for suggestions for improvement. These assessment types rely on satisfaction in order to justify spending personnel time and/or money but often do not directly assess student skills.

Data management education has progressed to the point at which competencies have been developed (Calzada Prado and Marzal, 2013; Carlson and Johnston, 2015; Piorun et al., 2012; Qin and D'Ignazio, 2010; Schneider, 2013). Also, a pilot competency matrix for scaffolding data management skill development has been proposed (Sapp Nelson, 2017). Most workshops and formal courses on data management align with one of these competency frameworks, with the Association of College and Research Libraries framework for information literacy and/or with discipline accreditation.

Aligning with competencies and accreditation standards, embedding data management in courses, or offering them as standalone courses can help to determine whether data management skills are being acquired. However, this is hindered by the lack of data management infrastructure at any level and the lack of established standards to measure against. It is likely that we will have the opportunity to further measure data management skills in the coming years, as expectations for data management increase and the developed competencies are applied. This program and granular assessment of student knowledge and practice change contributes to this effort.

Requirements for data sharing and open science are increasing through all levels of the academic enterprise. Funders are requiring data sharing to create transparency and trust with the public or stakeholders and to reduce financial waste, journals are requiring it to bolster the peer review process and create a pipeline for "value added" services, and researchers themselves are engaging in it, as the complexity of data prohibits "single author" science to exist, among other reasons. However, despite all these drivers, data management education has not yet been prioritized. Libraries have stepped up to fill this gap as best they can but are hindered by the need to justify their place in this arena and to entice students to participate. Therefore, measures of satisfaction can only go so far. Steps that move us to more accurately measure students' data management skills may lead us to data that can be leveraged within our institutions to argue for data management education to be considered a core skill and a contributor to student success overall.

DISCLOSURES

The author declares that they do not have any conflict of interest, financial or otherwise.

HUMAN SUBJECTS RESEARCH

Assessment of student knowledge, confidence, and behaviors regarding data management and artifacts produced during this program were reviewed by the Institutional Review Board at the University of Illinois Chicago and were determined to be exempt (#2019-0048). Informed consent was obtained from all students included in this program.

FUNDING STATEMENT

REFERENCES

Adamick, J., Reznik-Zellen, R., & Sheridan, M. (2012). Data management training for graduate students at a large research university. *Journal of EScience Librarianship*, *1*(3), 180–88. https://doi.org/10.7191/jeslib.2012.1022

Bishop, B. W., & Borden, R. M. (2020). Scientists' research data management questions: Lessons learned at a data help desk. *Portal: Libraries & the Academy*, *20*(4), 677–692. https://doi.org/10.1353/pla.2020.0032

Calzada Prado, J., & Marzal, M. A. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, *63*(2), 123–134. https://doi.org/10.1515/libri-2013-0010

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. https://doi.org/10.1353/pla.2011.0022

Carlson, J., & Johnston, L. (2015). *Data information literacy: Librarians, data, and the education of a new generation of researchers*. Purdue University Press.

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 Digital Curation*, *8*(1), 204–217. https://doi.org/10.2218/ijdc.v8i1.254

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

Clement, R., Blau, A., Abbaspour, P., & Gandour-Rood, E. (2017). Team-based data management instruction at small liberal arts colleges. *IFLA Journal*, *43*(1), 105–118. https://doi.org/10.1177/0340035216678239

Coates, H. L., Carlson, J., Clement, R., Henderson, M., Johnston, L. R., and Shorish, Y. (2018). How are we measuring up? Evaluating research data services in academic libraries. *Journal of Librarianship and Scholarly Communication*, *6*(1). https://doi.org/10.7710/2162-3309.2226

Eaker, C. (2014). Planning data management education initiatives: Process, feedback, and future directions. *Journal of EScience Librarianship*, *3*(1). https://doi.org/10.7191/jeslib.2014.1054

Fearon, D., Gunia, B. L., Pralle, B., Lake, S., & Sallans, A. (2013). SPEC kit 334: Research data management services (July 2013). *SPEC Kit*. https://doi.org/10.29242/spec.334

Fong, B., & Wang, M. (2015). Required data management training for graduate students in an earth and environmental sciences department. *Journal of EScience Librarianship*, *4*(1), e1067. https://doi.org/10.7191/jeslib.2015.1067

Griffin, T. (2020). Centering graduate students' research projects in data management education: A pilot program. *Journal of Librarianship and Scholarly Communication*, *8*(1), 2365. https://doi.org/10.7710/2162-3309.2365

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), 05013002. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000154

Johnston, L., and Jeffryes, J. (2014b). Steal this idea. *College & Research Libraries News*, *75*(8), 431–434. https://doi.org/10.5860/crln.75.8.9175

Johnston, L., Lafferty, M., and Petsan, B. (2012). Training researchers on data management: A scalable, cross-disciplinary approach. *Journal of EScience Librarianship*, *1*(2), e1012. https://doi.org/10.7191/jeslib.2012.1012

Lage, K., Losoff, B., and Maness, J. (2011). Receptivity to library involvement in scientific data curation: A case study at the University of Colorado Boulder. *Portal: Libraries and the Academy*, *11*(4), 915–937. https://doi.org/10.1353/pla.2011.0049

Macy, K. V., & Coates, H. L. (2016). Data information literacy instruction in business and public health: Comparative case studies. *IFLA Journal*, *42*(4), 313–327. https://doi.org/10.1177/0340035216673382

Marcus, M., Ball, S., Delserone, L., Hribar, A., & Loftus, W. (2007). *Understanding research behaviors, information resources, and service needs of scientists and graduate students: A study by the University of Minnesota Libraries.* University of Minnesota Libraries.

Mattern, E., Jeng, W., He, D., Lyon, L., & Brenner, A. (2015). Using participatory design and visual narrative inquiry to investigate researchers' data challenges and recommendations for library research data services. *Program*, *49*(4), 408–423. https://doi.org/10.1108/PROG-01-2015-0012

Otto, J. (2012). Assessing and improving data literacy: A study with urban and regional planning students. *PNLA Quarterly, 76*(4), 5–23.

Peters, C., & Dryden, A. R. (2011). Assessing the academic library's role in campus-wide research data management: A first step at the University of Houston. *Science & Technology Libraries*, *30*(4), 387–403. https://doi.org/10.1080/0194262X.2011.626340

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), e1003. https://doi.org/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*. https://doi.org/10.5062/F4057CX4

Qin, J., & D'Ignazio, J. (2010, June 22). Lessons learned from a two-year experience in science data literacy education [Conference session]. *31st Annual International Association of Scientific and Technological University Libraries Conference*, Purdue University, West Lafayette, IN, United States. https://docs.lib.purdue.edu/iatul2010/conf/day2/5/

Sapp Nelson, M. R. (2017). A pilot competency matrix for data management skills: A step toward the development of systematic data information literacy programs. *Journal of EScience Librarianship*, *6*(1), e1096. https://doi.org/10.7191/jeslib.2017.1096

Schmidt, L., & Holles, J. (2018). A graduate class in research data management. *Chemical Engineering Education*, *52*(1), 52–59.

Schneider, R. (2013). Research data literacy. In S. Kurbanoğlu, E. Grassian, D. Mizrachi, R. Catts, & S. Špiranec (Eds.), *Worldwide commonalities and challenges in information literacy research and practice* (pp. 134–140). Springer International Publishing.

Steinhart, G., Chen, E., Arguillas, F., Dietrich, D., & Kramer, S. (2012). Prepared to plan? A snapshot of researcher readiness to address data management planning requirements. *Journal of EScience Librarianship*, *1*(2), e1008. https://doi.org/10.7191/jeslib.2012.1008

Surkis, A., Zametkin LaPolla, F. W., Contaxis, N., & Read, K. (2017). Data day to day: Building a community of expertise to address data skills gaps in an academic medical center. *Journal of the Medical Library Association*, *105*(2), 185–191. https://doi.org/10.5195/jmla.2017.35

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS One*, *6*(6), 1–21. https://doi.org/10.1371/journal.pone.0021101

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.

Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Sandusky, R., Langseth, M. L., & Lundeen, A. (2015). Research data services in academic libraries: Data intensive roles for the future? *Journal of EScience Librarianship*, *4*(2), e1085. https://doi.org/10.7191/jeslib.2015.1085

Thielen, J., & Nichols Hess, A. (2017). Advancing research data management in the social sciences: Implementing instruction for education graduate students into a doctoral curriculum. *Behavioral & Social Sciences Librarian*, *36*(1), 16–30. https://doi.org/10.1080/01639269.2017.1387739

Thielen, J., Samuel, S., Carlson, J., and 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*. https://doi.org/10.5062/F42Z13HQ

Valentino, M., & Boock, M. (2015). Data management services in academic libraries: A case study at Oregon State University. *Practical Academic Librarianship: The International Journal of the SLA Academic Division*, *5*(2), 77–91.

Van Tuyl, S., & Whitmire, A. (2018). Investigation of non-academic data management practices to inform academic research data management. *Research Ideas and Outcomes*, *4*, e30829. https://doi.org/10.3897/rio.4.e30829

Weller, T., & Monroe-Gulick, A. (2015). Differences in the data practices, challenges, and future needs of graduate students and faculty members. *Journal of EScience Librarianship*, *4*(1), e1070. https://doi.org/10.7191/jeslib.2015.1070

Whitmire, A., Boock, M., and Sutton, S. C. (2015). Variability in academic research data management practices. *Program*, *49*(4), 382–407.

Whitmire, A. (2015). Implementing a graduate-level research data management course: Approach, outcomes, and lessons learned. *Journal of Librarianship & Scholarly Communication*, *3*(2), 1–22. https://doi.org/10.7710/2162-3309.1246

Wiljes, C., & Cimiano, P. (2019). Teaching research data management for students. *Data Science Journal*, *18*(1), 38. https://doi.org/10.5334/dsj-2019-038

APPENDIX A

The three surveys were identical except that students were asked to estimate their knowledge and practices at the immediate program end. The full survey is approximately 120 questions. Not all questions were analyzed for this manuscript. Questions regarding confidence and associated behaviors will be reported in future work and are also excluded from this appendix. The questions used for this analysis are included below. (Please note, the questions numbers were not visible to the participants and are not sequential. Questions are displayed here in the order they were presented to the participants.)

Q1 This survey will compare the change in confidence, knowledge, and practices you currently have regarding data management relative to when you began the program.

Q12 Have you had previous education or training in data management, or have you not had data management education/training?

- ○ Previous data management education/training (1)
- ○ No data management education/training (2)

Q13 Which of the following types of data management education/training have you received? *(Select all that apply.)*

- ☐ . Credit course (1)
- ☐ . Seminar or workshop (2)
- ☐ . Self-education (online, books, or other materials) (3)
- ☐ . Other – please specify
    (4) _____

Q121

Before the data management implementation project, did you overestimate or underestimate your …

Knowledge of data standards?

- ○ Overestimate (3)
- ○ Neither overestimate or underestimate (2)
- ○ Underestimate (1)

Q122

Knowledge of documentation practices?

- Overestimate (3)
- Neither overestimate or underestimate (2)
- Underestimate (1)

Q123

Knowledge of data policies?

- Overestimate (3)
- Neither overestimate or underestimate (2)
- Underestimate (1)

Q124

Knowledge of organization practices?

- Overestimate (3)
- Neither overestimate or underestimate (2)
- Underestimate (1)

Q125

Did you overestimate or underestimate your …

Knowledge of file management practices?

- Overestimate (3)
- Neither overestimate or underestimate (2)
- Underestimate (1)

Q126

Knowledge of <u>storing</u> data well?

- Overestimate (3)
- Neither overestimate or underestimate (2)
- Underestimate (1)

Q127

Knowledge of <u>sharing</u> data appropriately?

- Overestimate (3)
- Neither overestimate or underestimate (2)
- Underestimate (1)

Q128

Your practice in implementing data standards?

- Overestimate (3)
- Neither overestimate or underestimate (2)
- Underestimate (1)

Q129

Did you overestimate or underestimate your …

Your practice in <u>implementing</u> data standards?

- Overestimate (3)
- Neither overestimate or underestimate (2)
- Underestimate (1)

Q130

Your practice in <u>following</u> data standards?

- ○ Overestimate (3)
- ○ Neither overestimate or underestimate (2)
- ○ Underestimate (1)

Q131

Your practice in creating organization structure?

- ○ Overestimate (3)
- ○ Neither overestimate or underestimate (2)
- ○ Underestimate (1)

Q132

Your practice in documentation?

- ○ Overestimate (3)
- ○ Neither overestimate or underestimate (2)
- ○ Underestimate (1)

Q133

Did you overestimate or underestimate your …

Your practice in managing individual files?

- ○ Overestimate (3)
- ○ Neither overestimate or underestimate (2)
- ○ Underestimate (1)

Q134

Your practice in <u>storing</u> data well?

- ○ Overestimate (3)
- ○ Neither overestimate or underestimate (2)
- ○ Underestimate (1)

Q135

Your practice in <u>sharing</u> data appropriately?

- ○ Overestimate (3)
- ○ Neither overestimate or underestimate (2)
- ○ Underestimate (1)

Q136

Your consistency in sustainability practices?

- ○ Overestimate (3)
- ○ Neither overestimate or underestimate (2)
- ○ Underestimate (1)

Q45

How familiar are you with…

documentation standards required by <u>journals</u> in your discipline?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q46

documentation standards required by <u>funders</u> in your discipline?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q47

How familiar are you with…

documentation standards developed by your <u>research community at large</u>?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q40

established metadata, ontologies, definitions, common data elements, or other similar schemes in your discipline?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q43 What does the term "metadata" mean to you?

Q44

<u>Organization</u>

How familiar are you with…

creating structured organization schemes using folders?

- Extremely familiar (5)
- Very familiar (4)
- Somewhat familiar (3)
- Not very familiar (2)
- Not familiar at all (1)

Q45

creating structured identification schemes/codes for files?

- Extremely familiar (5)
- Very familiar (4)
- Somewhat familiar (3)
- Not very familiar (2)
- Not familiar at all (1)

Q46

How familiar are you with…

creating indexes, tables of contents or other tools?

- Extremely familiar (5)
- Very familiar (4)
- Somewhat familiar (3)
- Not very familiar (2)
- Not familiar at all (1)

Q47

working with file management software such as Git/GitHub or Electronic lab notebooks?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q48

How familiar are you with…

creating inventories to track reagents/materials?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q49

practices for versioning files?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q50

How familiar are you with …

creating file naming conventions?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q51

<u>Storing</u>

How familiar are you with …

university storage options?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q52

personal storage options?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q53

How familiar are you with…

backup options?

- Extremely familiar (5)
- Very familiar (4)
- Somewhat familiar (3)
- Not very familiar (2)
- Not familiar at all (1)

Q54

data repositories?

- Extremely familiar (5)
- Very familiar (4)
- Somewhat familiar (3)
- Not very familiar (2)
- Not familiar at all (1)

Q55

How familiar are you with…

code repositories?

- Extremely familiar (5)
- Very familiar (4)
- Somewhat familiar (3)
- Not very familiar (2)
- Not familiar at all (1)

Q56

practices for storing data?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q58

How familiar are you with…

data sharing requirements for <u>journals</u> in your discipline?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q59

data sharing requirements for <u>funders</u> in your discipline?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q60

How familiar are you with…

your discipline's <u>research community at large</u> data sharing requirements?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Somewhat familiar (3)
- ○ Not very familiar (2)
- ○ Not familiar at all (1)

Q64

<u>Other Practices</u>

How familiar are you with….

creating checklists?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Moderately familiar (3)
- ○ Slightly familiar (2)
- ○ Not familiar at all (1)

Q65

creating workflow protocols or standard operating procedures?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Moderately familiar (3)
- ○ Slightly familiar (2)
- ○ Not familiar at all (1)

Q66

How familiar are you with…

writing data management plans?

- ○ Extremely familiar (5)
- ○ Very familiar (4)
- ○ Moderately familiar (3)
- ○ Slightly familiar (2)
- ○ Not familiar at all (1)

Q70

Do you have an intentional scheme for organizing your files, or do you not have an intentional scheme?

- ○ Have intentional scheme (1)
- ○ Do not have intentional scheme (2)

Q71

How often do you consistently follow your organizational scheme?

- ○ Always (5)
- ○ Often (4)
- ○ Sometimes (3)
- ○ Rarely (2)
- ○ Never (1)

Q72

Do you create tables of content, indexes, or similar tools to help find your data, or do you not create these tools?

- ○ Create tools (1)

- ○ Do not create tools (2)

Q73

How often do you update these tools?

- ○ Always (5)
- ○ Often (4)
- ○ Sometimes (3)
- ○ Rarely (2)
- ○ Never (1)

Q74

Is there continuity between your digital files for your research project and analog materials, or is there not continuity between the two?

- ○ Continuity between digital files for project and analog materials (1)
- ○ No continuity between digital files for project and analog materials (2)

Q75

Do you create standardized vocabulary, definitions, acronyms, and/or values that you can use across different assays, experiments, tests, or stimulations, or do you not created such vocabulary or values?

- ○ Create standardized vocabulary (1)
- ○ Do you not create standardized vocabulary (2)

Q76

Which of the following are included in the documentation for your project? *(Select all that apply.)*

- ☐. Data dictionaries/codebooks that explain abbreviations and acronyms (1)

☐ . A project description or outline (2)

☐ . Documentation of analysis methods (3)

☐ . Documentation of code/scripts (4)

Q83

Do you have a standardized file name convention that you use, or do you not have a standardized file name convention?

○ Have a standardized file name convention (1)

○ Do not have a standardized file name convention (2)

Q84

Is that file name convention consistent across projects, or is it not consistent across projects?

○ Consistent across projects (1)

○ Not consistent across projects (2)

Q85

How often do you consistently follow your naming convention?

○ Always (5)

○ Often (4)

○ Sometimes (3)

○ Rarely (2)

○ Never (1)

Q86

Do you use a standardized identification scheme/code to create context between files, or do you not use a standardized scheme identification scheme?

○ Use standardized identification scheme (1)

○ Do not use standardized identification scheme (2)

Q87

How often do you follow this identification scheme?

- ○ Always (5)
- ○ Often (4)
- ○ Sometimes (3)
- ○ Rarely (2)
- ○ Never (1)

Q88

Do you use a standard date format with your files, or do you not use a standard date format?

- ○ Use standard date format (1)
- ○ Do not use standard date format (2)

Q89

What date format do you most often use?

_____

Q91

How often do you follow your date format?

- ○ Always (5)
- ○ Often (4)
- ○ Sometimes (3)
- ○ Rarely (2)
- ○ Never (1)

Q90

How easy is it for you to distinguish your files from one another?

- ○ Extremely easy (5)
- ○ Very easy (4)
- ○ Somewhat easy (3)
- ○ Not very easy (2)
- ○ Not at all easy (1)

Q93

How easy is it for you to distinguish <u>various versions of your files</u> from one another?

- ○ Extremely easy (5)
- ○ Very easy (4)
- ○ Somewhat easy (3)
- ○ Not very easy (2)
- ○ Not at all easy (1)

Q92

Do you usually know what information your files contain only by opening them, or do you know without opening them?

- ○ Know information only by opening files (1)
- ○ Know information without opening files (2)

Q94

How often do you use software to manage files?

- ○ Always (5)
- ○ Often (4)
- ○ Sometimes (3)

- ○ Rarely (2)
- ○ Never (1)

Q95 What software do you use to manage files? *(Select all that apply.)*

- ☐. R (1)
- ☐. Git/GitHub (2)
- ☐. SAS (3)
- ☐. Access/SQL (4)
- ☐. Other (5) _____

Q96

Is all your data stored in one place, or is it stored in more than one place?

- ○ Data stored in one place (1)
- ○ Data stored in more than one place (2)

Q98

Which of the following <u>best</u> describes the arrangement of data between these storage places?

- ○ Data is in one location and copied to all others; I can find all my data in any location (1)
- ○ Some data is copied to other locations but not all; I can find some data in multiple locations, but it is not consistent (2)
- ○ Data is in multiple locations; I have to know which location my data is in to find it. Little or no data is in multiple locations. (3)

Q100

Do you back up your data, or do you not back up your data? *(Please include backups you do personally as well as backups by your department/unit.)*

- ○ Back up data (1)

- Do not back up data (2)

Q103

How often do you backup your data?

- Always (5)
- Often (4)
- Sometimes (3)
- Rarely (2)
- Never (1)

Q106

How often do you backup your <u>documentation and other supporting information</u>?

- Always (5)
- Often (4)
- Sometimes (3)
- Rarely (2)
- Never (1)

Q109

How often do you share your research data?

- Always (5)
- Often (4)
- Sometimes (3)
- Rarely (2)
- Never (1)

Q110 How often do you share documentation (protocols or other supporting information) when you share data with someone?

- Always (5)
- Often (4)
- Sometimes (3)
- Rarely (2)
- Never (1)

Q112

How often do you …

use templates or checklists?

- Always (5)
- Often (4)
- Sometimes (3)
- Rarely (2)
- Never (1)

Q117 Finally, we'd like to close with some demographic information.

- Name (1) _____
- College (2) _____
- Department (3) _____
- Degree Program (4) _____