

JLSC

ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | <http://jls-publications.org>

Volume 8, General Issue (2020)

Centering Graduate Students' Research Projects in Data Management Education: A Pilot Program

Tina M. Griffin

Griffin, T.M. (2020). Centering Graduate Students' Research Projects in Data Management Education: A Pilot Program. *Journal of Librarianship and Scholarly Communication*, 8(General Issue), eP2365. <https://doi.org/10.7710/2162-3309.2365>

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



© 2020 Griffin. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

Centering Graduate Students' Research Projects in Data Management Education: A Pilot Program

Tina M. Griffin

Assistant Professor, Information Services Librarian/Liaison, University of Illinois Chicago

INTRODUCTION Data management education has been part of library service models for almost 2 decades. This paper describes a pilot graduate student education program whose framework shows interdependencies between data management practices, uses a flipped classroom model to allow maximum time for implementation, and whose primary activities are entirely student research based. **LITERATURE REVIEW** Education in data management encompasses many different formats (in-person, online, synchronous, asynchronous). Within this instruction, Data Information Literacy competencies help define student-learning objectives for data management tasks. Currently data management education is a combination of theory and active learning, with students asking for more hands-on practice. **PROGRAM DESCRIPTION** This program is an 8-week, in-person, flipped classroom series that addresses all data life cycle stages and aligns with many Data Information Literacy competencies. It is entirely student research data focused in that activities require that they use their projects, with significant time allocated to implement these practices while in the classroom. **NEXT STEPS** With a 69% retention rate and student improvement in seven foundational data management concepts, this program is considered a success. Future work involves converting this program to a credit-bearing course.

Received: 02/06/2020 Accepted: 09/22/2020

Correspondence: Tina M. Griffin, 1750 W Polk St, Chicago, IL 60612, tmcg@uic.edu



© 2020 Griffin. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

It is generally agreed that most students learn data management practices ad hoc, influenced by their research environment (Carlson, Fosmire, Miller, & Nelson, 2011; Frugoli, Etgen, Kuhar, 2010). Few students have formal training in data management, and it is rarely mandatory (Carlson, Johnston, Westra, & Nichols, 2013; Federer, Lu, & Joubert, 2016; Johnston & Jeffryes, 2014b). Their graduate and professional programs frequently do not include data management instruction in their curriculum. This means that opportunities for formal instruction are typically through library efforts or campus requirements such as participation in a “responsible conduct of research” seminar series.

Like most universities with large research efforts, the University of Illinois Chicago has similar challenges in delivering data management education. The University of Illinois Chicago is a Carnegie Classification Doctoral University: Highest Research Activity institution. This institution is public and has a current enrollment of 21,000 thousand undergraduates and over 7,000 graduate students. It contains 14 colleges or schools within it, eight of which are STEM or health science focused. The institution is a federally designated Minority Serving Institution and Asian American Native American Pacific Islander-Serving Institution since 2010. In 2016, it received the Hispanic-Serving Institution designation (*Minority-serving Institution Status*, 2020). The University of Illinois Chicago currently receives over \$300 million dollars in research funds, with over \$200 million dollars coming from federal sources (*About OVCR*, 2020).

This paper describes a non-credit program to teach graduate students data management, offered by the University Library, that is novel in several respects, relative to current literature. First, it takes a holistic approach towards learning data management by framing content in discovery, infrastructure, and sustainability themes. Instruction efforts reported in the literature have focused on infrastructure practices almost exclusively, such as file naming conventions and metadata description. By adding discovery-themed work, students will identify the stakeholders or factors that affect their data management choices they must make later; and the sustainability theme helps students identify their own habits and behaviors regarding research work so that they can continue best practices beyond the program. This framework also allows students to realize the interdependencies between data management practices. Second, it uses a flipped classroom model. While flipped classrooms are established in information literacy education generally, they have not been used extensively in data management education. Lastly, this program is entirely focused on the students’ existing research. While the examples and learning content cover a variety of disciplines, all activities use the students’ research projects. Significant classroom time is dedicated for students to implement these practices into their own work.

LITERATURE REVIEW

Data management education efforts have been one of the earliest and most continuous forms of data management support by libraries. Tenopir et al. (2012) reported that 11% of Association of Research Libraries (ARL) survey respondents offered training or education in research data services and 27% planned to offer it (Tenopir, Birch, & Allard, 2012). A follow-up study 3 years later showed no significant change in training or intent to offer training (Tenopir et al., 2015). Similarly, respondents in the ARL SPEC Kit 334 indicated that they saw a role for libraries in data management education for faculty and staff (Fearon, Gunia, Pralle, Lake, & Sallans, 2013). Recent work looking at needs assessments performed by various libraries between 2008 and 2017 indicated that data management education is a continual gap that needs to be addressed (Goben & Griffin, 2019).

Early library involvement in data management education offered non-discipline-specific sessions in order to maximize their audience (Adamick, Reznik-Zellen, & Sheridan, 2012; Carlson et al., 2013; Clement, Blau, Abbaspour, & Gandour-Rood, 2017; Johnston & Jeffryes, 2014b; Kvale & Stangeland, 2017; Muilenburg, Lebow, & Rich, 2014; Whitmire, 2015; Wiljes & Cimiano, 2019). Other libraries designed their education to focus on a particular project such as a water quality field station (Carlson & Stowell-Bracke, 2013), or to address a specific discipline such as engineering (Johnston & Jeffryes, 2014a), natural resources and ecology (Wright & Andrews, 2015), business and public health (Macy & Coates, 2016), meteorology (Frank & Pharo, 2016), or social sciences (Thielen & Hess, 2017).

Further work within the data librarianship education field included development of full curricula and frameworks designed to maximize exposure to data management principles and to accommodate different delivery modes (synchronous to asynchronous, online to in-person) and experience levels (undergraduate to post-doctoral) (Henkel et al., n.d.; Piorun et al., 2012; University of Edinburgh, 2019). Subsequently, there have been so many educational materials developed that a clearinghouse has been created to assist in centrally locating them and making them more accessible (Federation of Earth Science Information Partners, 2016).

Around this same time, Data Information Literacy (DIL) competencies were developed. Carlson et al. interviewed STEM faculty and collected data from students through instruction efforts to describe twelve knowledge topics and skill sets deemed important for research (Carlson et al., 2011). This work continued in the Data Information Literacy Project, an application of these competencies in various higher education settings (Carlson et al., 2013), and culminated in the Data Information Literacy Handbook, which further

analyzed the competencies, provided disciplinary case studies, and outlines of how other institutions could develop their own DIL programs (Carlson & Johnston, 2015). The competencies developed span the potential life cycle of research and include Cultures of Practice that establish standards or norms within the community; Data Conversion/Interoperability for migration and preservation over time; Data Curation and Reuse recognizing the future value of data; Data Management/Organization for creating and tracking data as it is processed; Data Preservation costs and benefits; Data Processing/Analysis tools and techniques; Data Quality and Documentation for preserving context after creation or capture; Data Visualization/Representation for appropriate communication of results; Databases and Data formats for storage and context; Discovery and Acquisition to find and use existing data; Metadata/Description for future understanding and interpretation; and Ethics/Attribution for appropriate understanding of intellectual property, confidentiality and data sharing.

In addition to developing data management specific content within a library service profile, approaches to delivering content also made advances. Within general information literacy instruction, a variety of active learning methods have been established (Grassian & Kaplowitz, 2009) and flipped classroom instruction has been employed. Flipped learning is defined as “a pedagogical approach in which direct instruction moves from the group learning space to the individual learning space, and the resulting group space is transformed into a dynamic, interactive learning environment where the educator guides students and they apply concepts and engage creatively in the subject matter” (Flipped Learning Network, n.d.). Roehling states in a summary of flipped classroom evaluation studies that 71% of them report students perceive flipped classroom instruction as effective and 80% of studies report a student preference for flipped learning over traditional lecture formats (Roehling, 2018b, p. 19). Flipped learning relies on active learning techniques to allow students to engage with the material presented outside of the formal classroom. These techniques can include problem-based learning, discussions, one-minute papers, practicing skills, and peer-teaching (Roehling, 2018a, p. 59; Talbert & Bergmann, 2017, p. 57). In addition, Haak and Theobald separately reported that active learning approaches can reduce achievement gaps of disadvantages students in a stem undergraduate discipline (Haak, HilleRisLambers, Pitre, & Freeman, 2011; Theobald et al., 2020).

Many data management education programs incorporate hands-on work or otherwise encourage students to practice with their own data. The format for these offerings has been a mix of library workshops, either standalone or in a series, or full courses. Within the literature, only Johnson used a flipped classroom approach to teaching data management (Johnston & Jeffryes, 2014b). Despite the format and approach, many learners still reported that they would have liked the content to incorporate more discipline-specific activities

and inclusion of “real world” examples (Adamick et al., 2012; Johnston & Jeffryes, 2014b; Whitmire, 2015; Wiljes & Cimiano, 2019).

Throughout these developments in curriculum and education efforts, data management topics do not appear to be taught with regard to dependencies between practices, such as the impact that file naming decisions may have on folder hierarchies, or how decisions made during data collection about storage will affect data sharing later in the research workflow. This pilot program attempts to address these and other concerns by taking a holistic view of data management needs, focusing entirely on student research products, and using a flipped classroom approach to give learners the most amount of time possible to practice and implement their new knowledge.

PROGRAM DESCRIPTION

Format

The format of this program consisted of 8 continuous weeks of programming delivered in a flipped classroom format. Content was delivered through a LibGuide and released weekly. This program site outlined learning objectives, pre-session readings/videos, preparatory work, and tasks expected to be completed during the synchronous, in-person class. The in-person sessions were 90 minutes, and students were required to bring laptops/tablets or borrow laptops from the library.

The program was delivered in-person, at the Chicago campus only, from June through August 2019. Originally three sections were planned, but the number of students interested in a Friday timeslot justified opening another section. Weekly sessions were scheduled Monday – Wednesday from 4–5:30 pm and Fridays from 2:30–4 pm. Students were assigned to a section according to preferences indicated in their application.

The approach to weekly instruction began with administrative announcements and content overview for the day. The rest of the session was roughly divided by thirds. The first part discussed pre-work content and gave time for questions about the week's (or any week's) content. The second part students engaged in a brief active learning activity as an opportunity to practice and clarify what they may not have understood. The last part students applied what they learned to their own research project. Frequently, the second and third parts were combined to give the students more time to implement and create. Each week the students uploaded documentation of their interaction with the content. This could be a planning sheet, reflection, or other similar object. Submission was through a Box widget embedded into the program LibGuide.

Curriculum

Three themes formed the framework for the program: discovery, infrastructure, and sustainability. While the individual weeks had specific learning objectives, framing the program by theme created an opportunity for students to understand that these data management activities do not exist in isolation and can be interdependent. The themes and topics are outlined in Table 1. Existing literature on data management education, existing curricula (Henkel et al., n.d.; Lamar Soutter Library, n.d.; University of Edinburgh, 2019), personal experience performing consultation-based instruction on data management and 18 years managing data in prior employment influenced framework development. Readings, slide presentations, and video presented to the students were taken from existing materials. No new materials were developed. Weekly learning objectives mapped to Data Information Literacy competencies (Carlson & Johnston, 2015). A full outline describing the weekly learning objectives, learning activity, application activity, and DIL competencies addressed are provided in Appendix A. All program materials (readings, videos, program content links, worksheets, reflection/activity prompts, and instructor outlines) can be found at the instructor site (<https://researchguides.uic.edu/DMIPinstructor>).

Week	Theme	Topic
1	Discovery	Introduction to Data Management
2	Discovery	Discipline Standards and Expectations
3	Discovery	Project Workflow
4	Infrastructure	Building Folder Hierarchies
5	Infrastructure	Creating File Names
6	Infrastructure	Using Tables of Contents, Indexes, Readmes, Dictionaries, and Codebooks
7	Sustainability	Checklists, Protocols, and Standard Operating Procedures for Data Management
8	Sustainability	Productivity and Habits

Table 1. Topic and theme descriptions

The goals of the first theme, discovery, are for the students to identify different potential influencers that may affect their data management practices. The discussion points and activities focus on the students understanding what their research lifecycle looks like and who is involved in it besides themselves. By the end of this theme, students should have a comprehensive picture of their data management needs. DIL competencies covered during this theme include data curation and reuse, ethics and attribution, culture of practice, data management and organization, databases and data formats.

Week one covers the importance and rationale for good data management, an introduction to data management principles, with an emphasis on storage and backup, and some of the historical drivers such as the NSF data management plan requirement (National Science Foundation, 2010). Students reflected and discussed what values are behind data management in the research community and for themselves and then drew their research lifecycle. Discussion points included what is the same and different in their value system versus the research community, and what their data lifecycle contains and why.

Week two content introduced the concept of metadata and standards and began exploration of these and other data management requirements. This week their activity used a modified Deep Dive into Data Management worksheet (Akers, Martin, & Oehrli, 2014) as a tool for students to discover who their stakeholders are and at what stage in their lifecycle they may have obligations to them. This worksheet functioned like a “scavenger hunt” asking students to review their funding agencies, associations, communities of practice, and other sources for data management expectations. This week also explored storage and backup constraints discovered in the previous week, as many students keep some if not all data on their laptops, and they usually do not have complete control over their storage options beyond that. Also, many researchers in STEM and health sciences, including students, work with protected health information or data that is too large to be conveniently stored on a laptop. All students began discussing options for storage and backup for their projects with their mentor/PI or other stakeholders, as this would influence decisions made during the infrastructure building stage.

The goal for the last week of the discovery theme is to bring together all the pieces of their research workflow and expectations or obligations they have been collecting from previous weeks in a visualization so that they have a “map” for creating a data management plan (DMP) for their project. This map also forms the foundation for decisions students will make in the next infrastructure phase. The content for this week outlines and describes data management plans. In addition to the map, students began a formal DMP using the DMPTool.

This week's activity expanded a previously described mapping process (Mattern, Jeng, He, Lyon, & Brenner, 2015) by incorporating what information and data students collected (or expected to collect) in their research, what supplementary information needed to be collected alongside it (i.e. protocols, literature, software for analysis, etc.), who needed access to which data and for how long, and storage needs. These elements were captured on simple office labels and various size sticky notes in a guided, timed process. The students rearranged these elements on a poster-sized sticky note according to different parameters (access, storage, data types, etc.) until they arrived at a workflow that reflected

their current or expected lifecycle. These maps were posted to the wall for the duration of the program as their reference. For the activity, students were paired and asked to review each other's maps guided by these discussion points: Could someone else understand your workflow? Where did they get lost? What may be missing? What needs to be clarified or explained better? Examples of student maps are provided in Appendix B.

In the second theme, infrastructure, students did most of their implementation work. The goals for this theme are for students to develop a personal infrastructure that incorporates requirements by stakeholders and functions with their research project's lifecycle. Each week started with a discussion of best practices, followed by a brief work through of an example, and the rest of the time (greater than 50%) was devoted to implementation with individual attention. This theme avoided using proprietary solutions like electronic lab notebooks because there is not yet institutional support, financial or otherwise, for these tools. Additionally, having students work with existing options did not add to their learning curve or force their research labs to accommodate software or technology into more collaborative workflows. For this reason, students focused on building structured folder hierarchies, file naming, and documentation. Students were encouraged to make intentional decisions regarding their infrastructure choices, with an awareness that there is no "perfect" system that accounts for all the conflicts they may encounter. DIL competencies addressed during this theme were data management and organization, utilizing best practices in preparing for preservation, data quality and documentation, metadata and data description, data curation and reuse.

Weeks four and five were parallel in structure and content. These weeks focused on folder hierarchies and file naming conventions. Content reviewed before class included presentations, chapters, and websites on best practices for folder organization and file naming. Class discussions focused on the tradeoffs between choosing different approaches and how further infrastructure building may have to compensate for deficiencies of a particular approach. For example, a student may organize their data/information in folders according to assays performed, by project scope, or by expected manuscript to be written. If students elected to organize by assay type, the context between pieces of data will be missing should they have to share their project with others. Therefore, this organization approach may necessitate more documentation and readme files. Similarly, if the relationship between folders influences their file naming convention, they may have to compensate for this when sharing files with someone else or depositing them in a collaborative location. These weeks had the greatest amount of time devoted to student activities. They discussions took less than one-third of the class time. While students worked, I held mini consultations with each to answer questions as they decide on a course of action. Students submitted reflections on what practices they thought they do well and where they could improve.

Week six brought back the topic of metadata and standards and included materials on documentation (data dictionaries and readme files) and wayfinding objects (tables of contents, indexes). Taking into consideration their choices over the past 2 weeks, they now needed to consider where there are gaps in context among their organization and workflow or between data objects that have a relationship. The discussion again addressed tradeoffs that may be needed. This week's activity identified their gaps and determined what type of object they needed to create to bridge it. They were required to turn this in as a list of future documentation they need to implement.

The last theme, sustainability, directly addressed habits and behaviors, which may be an obstacle for continuing their data management practices outside of this program. Similar to the infrastructure theme, this theme addressed adjustments they may need to make to compensate for less than ideal circumstances. DIL competencies addressed were data curation and reuse, and data management and organization.

Week seven focused on identifying opportunities for the students to streamline their personal productivity and creating tools that support continued use of data management best practices. Content included functions and uses for templates, standard operating procedures, checklists, and protocols. Discussion started with the differences between these types of objects and how they can be used in a data management context. Students identified a habit they may have trouble sustaining and created an object to support themselves. For example, if they are not yet used to an increased level of documentation for a data type, they will create a checklist that lists all the elements they need to include in order to document their work completely. This checklist will then be nested in their folder hierarchy or in their analog paperwork in a place they can refer to when working with that data. Students submitted the object that they created.

Week eight looks at personal habits and behaviors regarding data management. Content refers to different productivity strategies and long and short-term goal setting. The session began with a discussion on personal productivity preferences and struggles. This conversation addressed the difficulties with collaborative work and differences between work styles, and personal goals as well. The activity guided students to reflect on their current habits and time allocated for data management activities. This process asked them to review their work over the entirety of the program and map out goals and a timeline for completing data management tasks, projecting up to a year out. While this program was designed to give a significant amount of time for implementation, it was not expected that students would be able to complete implementation exhaustively. Depending on their project volume, scope, and degree progression, they would have work to do after the program concluded. These goals and timeline were turned in.

Administration

Advertising for the program involved posting flyers through the STEM and health sciences buildings and sending emails to known departmental listservs. Applications were required for admission to the program. Students filled out a form with contact information, address, general demographics, a statement of why they were interested in the program and what their data management goals were, and a separate statement about their research project. Applications were open to graduate students in the STEM and health sciences that were actively pursuing a research project not associated with coursework. Visiting students, undergraduates, students from other disciplines, and students not actively participating in research were excluded. Students could find the weekly topics, requirements, and minimum expectations posted on the application site before applying.

The initial launch of the program in the spring of 2019 was unsuccessful (0 applicants). It was surmised that the original requirements of statements of support from mentors, strict pursuit of a degree-based research project (e.g. PhD/MS vs capstone), additional application questions, and rolling admission were deterrents. The application was revised to simplify the process and relax the research parameters. Some application questions were moved to program activities and the rolling admissions were dropped in favor of an application deadline and fixed schedule. The program relaunched in late Spring 2019. Thirty-one students applied and 26 were accepted after screening. The majority were enrolled doctoral programs (24) with 2 from masters programs. Most students were from the College of Nursing (9), followed by Medicine (6) and School of Public Health (6), Applied Health Sciences (2), and Engineering (1), Liberal Arts and Sciences (1), and Pharmacy (1). Students were in various stages of completion of their educational goals, from the first year to within 6 months of graduation.

The administrator sent standing calendar invites to the all participants for the duration of the program. Students could attend other sections (described below) during the week if they had conflicts but must give advance notice to the instructor for each instance. Students were allowed two absences, also with advance notice if possible.

Students were compensated for participation and completion of assessments. Compensation occurred in three distributions as VISA gift cards. The first distribution (\$25) occurred after pre-test completion and before the access to the course content. The second distribution (\$75) occurred after completion of the program and post-test. The last distribution (\$50) was given after completion of the 6-month follow up survey, given in February 2020.

Of the 26 students enrolled in the program, six students withdrew prior to program completion, and one was withdrawn by the instructor due to more than two absences. Eighteen students completed the program and post-test, resulting in a 69% retention rate. Sixteen students completed the 6-month follow-up survey.

Assessment

Instruments and IRB

Pre-test, post-test, and follow up surveys were developed with the support of the Survey Research Lab at University of Illinois Chicago. These 3 instruments were administered by Qualtrics and each contained approximately 115 questions about student knowledge, confidence, and behaviors regarding data management. These instruments were tested by a senior graduate student in a health sciences discipline and then revised for clarity and biases. The Institutional Review Board reviewed the instruments and collected artifacts produced during this program and determined them to be exempt (#2019-0048). Selected data about student knowledge from the pre-test, post-test, and artifacts collected will be presented here. The remaining data with statistical analysis will be presented in future papers. The pre- and post-test questions can be seen in Appendix C.

Results

Having Data Information Literacy competencies gives us a guide in assessing student performance in this program. In the discovery, overarching goals were for the students to have a picture of their data management needs, in part, relative to external stakeholders. To this end, students were about their familiarity with documentation standards and sharing obligations required by funders. These aligned with the DIL cultures of practice and data curation and reuse, particularly in that students need to recognize data standards of their field, understand the practices and norms as it relates to their data life cycle, and therefore must be able to “plan activities needed to enable data curation” (Carlson & Johnston, 2015, p. 44). Before the program, students have very little or no knowledge of funder requirements for either of these questions, while afterward there is some familiarity by about half of them (Figures 1 and 2).

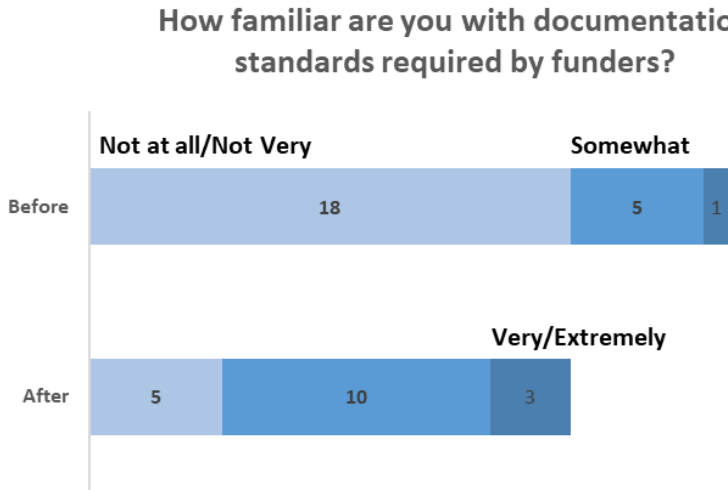


Figure 1. Student familiarity with funder documentation standards.

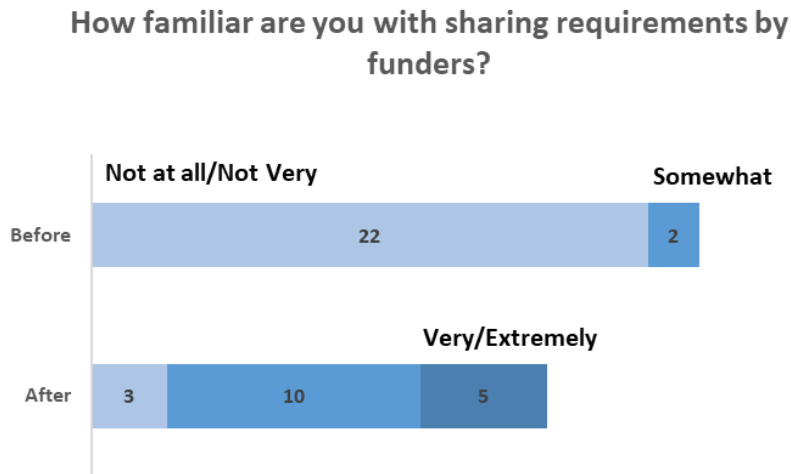


Figure 2. Student familiarity with funder sharing requirements.

Startling increases are not expected here because not all students know if their work is grant funded and if there are concomitant requirements. For those that do know, funders may not have established requirements (likely from smaller federal grants or grants from associations). In addition, certain disciplines may not have cultures of practice established for data management. On several of the Deep Dive worksheets, students indicated that they could not find requirements or standards for their journals, funders, or research community.

In the second theme, students were expected to develop a personal infrastructure that incorporates the data management needs identified in the first theme. Those tasks aligned with data organization and management and “data quality and documentation competencies, with an expectation the students can “keep track of the relation of subset or processed data to the original dataset” and “tracks data provenance and clearly delineates/denotes versions of a dataset” (Carlson & Johnston, 2015, pp. 44–45). When asked how familiar students were with creating folder hierarchies (figure 3) and file naming conventions (figure 4), most responded as somewhat familiar with the concepts, which is not surprising. Students at the graduate level will have had to organize information in various aspects of their lives, from course work to applying for graduate school itself. Afterward, most students reported being very or extremely familiar with these two data management practices.

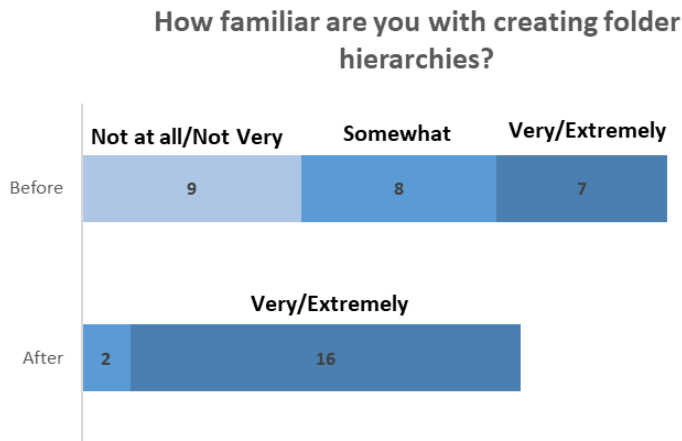


Figure 3. Student familiarity with constructing folder hierarchies.

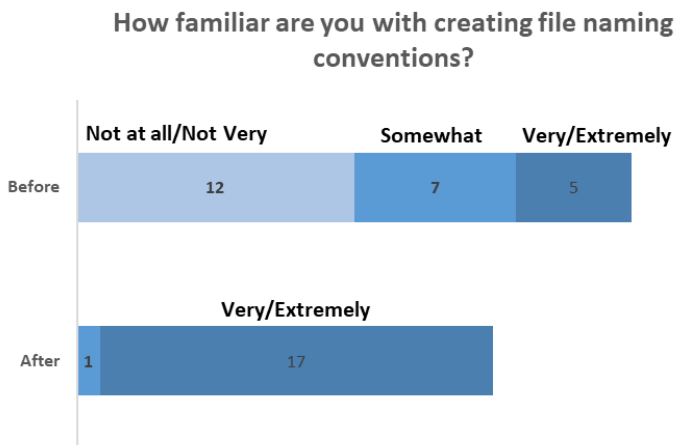


Figure 4. Student familiarity with constructing file naming conventions.

When asked to reflect on their current practices for folder structures and file naming students often described broad categories for folder organization and general file names. Through these program weeks, students realized that both practices needed more refinement to be effective.

“The naming strategies for my files were not very good and some could easily be mistaken for each other if they weren’t in the right place. I also didn’t have my smaller folders properly organized into categories and could definitely stand to make further sub-categorizations to make things easier to find.” Student 4.

“[My] old strategy → trying to give enough detail in names. [My] new strategy is to use project-specific naming strategy. This new naming strategy will help clump things together visually nicely and hopefully make it easier to figure out what a file contains before opening it.” Student 12.

Storage is a constant issue for researchers of all types. Storage was touched upon in all themes of the program as decisions about data management can be deeply affected by storage options or limitations. For this program storage discussions leaned on data curation and reuse and ethics and attribution competencies. The students need to be able to “distinguish which elements are likely to have future value” and demonstrate that they understand issues with privacy or security and can choose appropriate options. Before the programs, most students indicated that they were, at best, “somewhat” familiar with storage practices, while after that shifted to “very or extremely familiar” (Figure 5). In parallel, students were mixed in terms of what their university-specific storage options were spanning all response options. Afterward, almost all were very or extremely familiar (Figure 6).

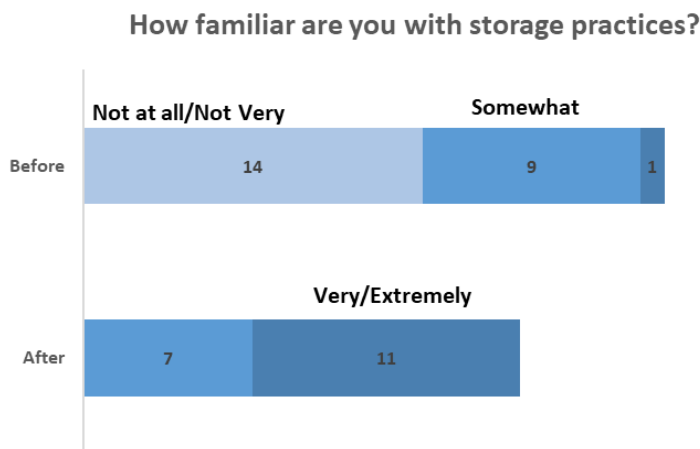


Figure 5. Student familiarity with storage practices.

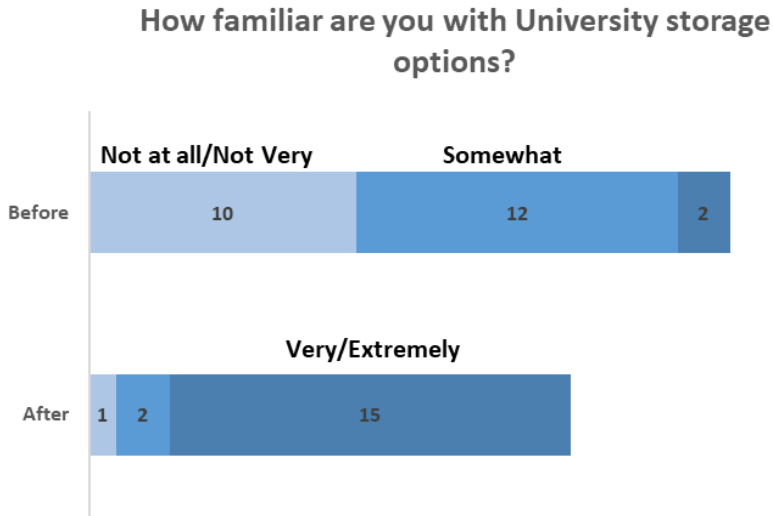


Figure 6. Student familiarity with university storage options.

Lastly, students must be able to put the sum of all data management skills together into a working data management plan (DMP). Students drafted individual plans during the entire 8 weeks. Sections of these plans were addressed weekly as the topic was broached in the curriculum. Irrespective of whether the students were aware of funding requirements, when asked about familiarity with writing DMPs, most students were not. By the end of the program there was improvement, where most students were either somewhat to extremely familiar (Figure 7).

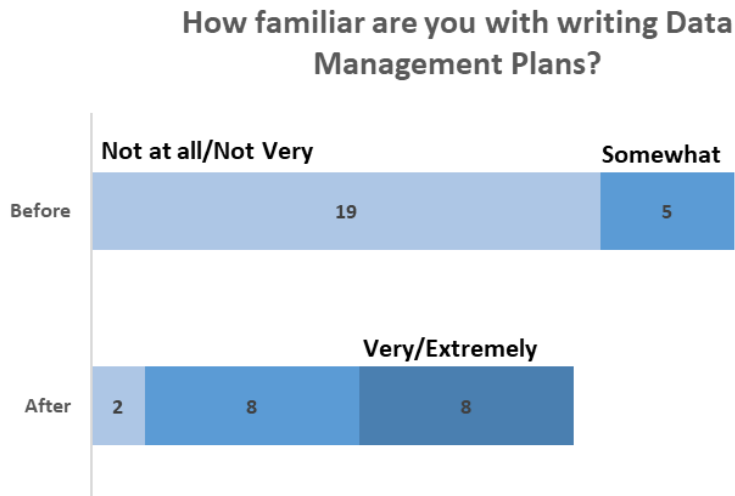


Figure 7. Student familiarity with writing Data Management Plans

Advantages and limitations

While asking students to commit to an in-person, 8-week data management program is time consuming, there are several advantages that may have contributed to the success of this pilot program. One is the flexibility regarding attendance. Summer is typically a time for students to focus on their research work without the competing obligations of didactic work. However, this issue is complicated by conference and vacation schedules. Students frequently took advantage of the ability to attend another session rather than skipping a week of content. Additionally, research work in STEM and health sciences is predictable only to a certain degree. Students who had conflicts with lab work also took advantage of flexible session attendance. Students were aware that weekly program work was interdependent and that missing classes would put them behind in implementation and/or activities. Compensation was also linked to a minimum expected attendance (6 of 8 weeks).

Another advantage is the student-focused outcomes. Previous literature frequently reported that students desire more activities and, specifically, ones that focus on tasks directly applicable to their projects (Adamick et al., 2012; Johnston & Jeffryes, 2014b; Whitmire, 2015; Wiljes & Cimiano, 2019). This is the focus of this program in its entirety. Every class session was designed with a majority of time dedicated to implementing data management for their projects. Specifically, during the infrastructure weeks, students were given greater than 50% of class time to develop the required organization and documentation needed. These weeks were also minimally guided, relying on the flipped classroom format to optimize the in-class time for implementation. Students worked while the instructor held “mini consultations” with each student to answer questions specific to their circumstance. As these were not private, opportunities existed for all students to hear strategies and questions that may apply to their situation.

Since this was a funded research project, it is also likely that compensation influenced retention. A future goal is to turn this program into a graduate seminar course. It will be interesting to see the retention and engagement after the conversion.

NEXT STEPS

At a 69% retention rate and increased familiarity with seven foundational concepts within data management practices, the pilot program is being considered a success. Future directions for this work include a statistical analysis of the full pre-, post-, and follow-up survey data to determine students’ baseline, as well as changes, in their data management knowledge, confidence, and behaviors. Clement’s follow-up survey 8 months after their team-

based education showed sustained awareness of data management practices after education (Clement et al., 2017). A similar result is expected here.

There is student interest in more program cohorts; however, grant funding only supported this effort as a pilot project. Since the initial results are positive, the author will be converting the program into an elective “special topics” course. It may also be adapted to an online format to accommodate regional campuses and to generally be more accessible.

The author welcomes collaborators to see if this program implementation is applicable to other institutions and if student data are consistent across institutions. Please contact the corresponding author with inquiries.

CONCLUSION

This approach to data management education is innovative in that it addresses critical interdependencies between data management practices and concepts through the framework, uses a flipped classroom model to deliver content, and is entirely student research focused with significant time for implementation.

Libraries are still working out what instruction is best for their respective institutions regarding data management. There may be no one approach that fits all disciplines and projects; however, this may not be our burden if we allow the students to use their own projects to determine their data management needs. In this paper, I demonstrate and advocate for an entirely student research data focused approach to data management. This pilot demonstrates that student attendance and engagement remain high, presumably because the program is designed to be directly and immediately applicable to their research. In combination with letting students lead in their data management education, taking advantage of new teaching modes, like flipped classrooms and active learning, offer the opportunity for them to devote larger amounts of time to the tasks needed, which they have difficulty prioritizing against already tight curriculum and research expectations.

This program also attempted to engage students in a more holistic manner. By demonstrating how unintentional decisions regarding data management can have cascading effects for future work, and by including time management and personal behaviors as program topics, students could develop a data management habit that extends beyond the classroom. This approach may also tap into personal values that support data management practices outside of the carrot/stick model that is commonly employed by funders and journals.

Opportunities to teach data management in a regular, structured manner, while necessary, can still be considered a luxury. Approaching this education from a new direction may give us leverage to establish data management education at our institutions more firmly.

ACKNOWLEDGEMENTS

Student willingness and participation drove the success of this program. The author appreciates the work and dedication to self-learning that was demonstrated by all the participants. The author also thanks the first reader of this manuscript Scott Martin, and colleagues Rebecca Raszewski and Abigail Goben for their continued support.

DISCLOSURES

The author declares that they do not have any conflict of interest. Assessment of student knowledge, confidence, and behaviors regarding data management, and artifacts produced during this program were reviewed by the Institutional Review Board at redact and were determined to be exempt (#2019-0048). Informed consent was obtained from all students included in this program.

FUNDING STATEMENT

Program development, execution, student compensation for participation, and assessment instrument development were supported by a grant from the National Networks of Libraries of Medicine-Greater Midwest Region's (NNLM-GMR) Research Data Award and the National Institutes of Health in a cooperative agreement (1UG4LM012346). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

About OVCR. (2020). Office of the Vice Chancellor for Research at the University of Illinois Chicago. <https://research.uic.edu/tools/ovcr-facts-figures/>

Adamick, J., Reznik-Zellen, R., & Sheridan, M. (2012). Data Management Training for Graduate Students at a Large Research University. *Journal of ESience Librarianship*, 1(3), 180–188. <https://doi.org/10.7191/jeslib.2012.1022>

Akers, K., Martin, S., & Oehrli, J. A. (2014). *Deep Dive into Data Workflow*. University of Michigan. <https://deepblue.lib.umich.edu/handle/2027.42/117636>

- 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., 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., & 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. <https://doi.org/10.1353/pla.2013.0034>
- Carlson, J., & Johnston, L. (2015). *Data information literacy: Librarians, data, and the education of a new generation of researchers*. Purdue University Press. https://doi.org/10.26530/OAPEN_626975
- 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>
- Fearon, D., Gunia, B. L., Pralle, B., Lake, S., & Sallans, A. (2013). *SPEC Kit 334: Research Data Management Services*. Association of Research Libraries. <https://doi.org/10.29242/spec.334>
- Federation of Earth Science Information Partners. (2016). *DMT Clearinghouse*. ESIP Data Management Training. <https://dmtclearinghouse.esipfed.org/>
- Federer, L. M., Lu, Y.-L., & Joubert, D. J. (2016). Data literacy training needs of biomedical researchers. *Journal of the Medical Library Association*, 104(1), 52–57. <https://doi.org/10.3163/1536-5050.104.1.008>
- Flipped Learning Network. (n.d.). *Definition of Flipped Learning*. Flipped Learning Network Hub. Retrieved August 6, 2020, from <https://flippedlearning.org/definition-of-flipped-learning/>
- 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. <https://doi.org/10.5860/crl.77.4.536>
- Frugoli, J., Etgen, A. M., & Kuhar, M. (2010). Developing and Communicating Responsible Data Management Policies to Trainees and Colleagues. *Science and Engineering Ethics*, 16(4), 753–762. <https://doi.org/10.1007/s11948-010-9219-1>
- Goben, A., & Griffin, T. (2019). In Aggregate: Trends, Needs, and Opportunities from Research Data Management Surveys. *College & Research Libraries*, 80(7), 903–924. <https://doi.org/10.5860/crl.80.7.903>
- Grassian, E., & Kaplowitz, J. (2009). Critical Thinking and Active Learning. In *Information Literacy Instruction: Theory and Practice* (second, p. 412). Neal-Schuman Publishers.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. *Science*, 332(6034), 1213–1216. <https://doi.org/10.1126/science.1204820>

Henkel, H., Hutchison, V., Strasser, C., Hespanha, S., Vanderbilt, K., Wayne, L., Hampton, S., Budden, A., Wang, Y., Bloom, D., Hodge, A., Steinhart, G., Wright, S., & Mayernik, M. (n.d.). *Education Modules*. DataONE Education Modules. Retrieved from <https://dataoneorg.github.io/Education/>

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

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

Kvale, L., & Stangeland, E. (2017). Skills for research data management—Developing RDM courses at the university of Oslo. *Proceedings of the Association for Information Science and Technology*, 54(1), 728–730. <https://doi.org/10.1002/pr2.2017.14505401134>

Lamar Soutter Library, U. of M. M. S. (n.d.). *New England Collaborative Data Management Curriculum*. New England Collaborative Data Management Curriculum. Retrieved from <https://library.umassmed.edu/resources/necdmc/index>

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>

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. IIf. <https://doi.org/10.1108/PROG-01-2015-0012>

Minority-Serving Institution Status. (2020). University of Illinois Chicago Office of the Chancellor. <https://chancellor.uic.edu/minority-serving-designations/>

Muilenburg, J., Lebow, M., & Rich, J. (2014). Lessons Learned From a Research Data Management Pilot Course at an Academic Library. *Journal of ESience Librarianship*, 3(1), 67–73. <https://doi.org/10.7191/jeslib.2014.1058>

National Science Foundation. (2010, November 30). *Dissemination and Sharing of Research Results*. US NSF - About. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>

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 ESience Librarianship*, 46–50. <https://doi.org/10.7191/jeslib.2012.1003>

Roehling, P. V. (2018a). Creating and Implementing Effective Active Learning Experiences. In P. V. Roehling (Ed.), *Flipping the College Classroom: An Evidence-Based Guide* (pp. 45–78). Springer International Publishing. https://doi.org/10.1007/978-3-319-69392-7_3

Roehling, P. V. (2018b). Student Reactions to and the Effectiveness of the Flipped Classroom Across Learning Domains. In P. V. Roehling (Ed.), *Flipping the College Classroom: An Evidence-Based Guide* (pp. 15–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-69392-7_2

- Talbert, R., & Bergmann, J. (2017). *Flipped Learning*. Stylus Publishing, LLC. <http://ebookcentral.proquest.com/lib/uic/detail.action?docID=4848101>
- Tenopir, C., Allard, S., Frame, M., Birch, B., Baird, L., Sandusky, R., Langseth, M., United States Geological Survey, Hughes, D., University of Tennessee, Knoxville, Lundeen, A., & Michigan State University. (2015). Research Data Services in Academic Libraries: Data Intensive Roles for the Future? *Journal of ESience Librarianship*, 4(2), e1085. <https://doi.org/10.7191/jeslib.2015.1085>
- Tenopir, C., Birch, B., & Allard, S. (2012). Academic Libraries and Research Data Services: Current Practices and Plans for the Future. *Association of College and Research Libraries*. http://www.ala.org/acrl/sites/ala.org/acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf
- Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., ... Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, 117(12), 6476–6483. <https://doi.org/10.1073/pnas.1916903117>
- Thielen, J., & Hess, A. N. (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. lxx. <https://doi.org/10.1080/01639269.2017.1387739>
- University of Edinburgh, I. S. (2019, September). *MANTRA Learning Units*. Mantra Research Data Management Training. <https://mantra.edina.ac.uk/>
- Whitmire, A. L. (2015). Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned. *Journal of Librarianship & Scholarly Communication*, 3(2), 1–22. lxx. <https://doi.org/10.7710/2162-3309.1246>
- Wiljes, C., & Cimiano, P. (2019). Teaching Research Data Management for Students. *Data Science Journal*, 18, 38. <https://doi.org/10.5334/dsj-2019-038>
- Wright, S. J., & Andrews, C. (2015). Natural Resources/Graduate Students/Wright & Andrews/Cornell University/2013. *Data Information Literacy Case Study Directory*, 2(1), 1. <https://doi.org/10.5703/1288284315476>

APPENDIX A Program Outline

Week 1: Introduction

- Learning objectives
 - Identify elements of a research data lifecycle.
 - Identify personal values regarding research data management.
 - Identify other entities' (funder/lab/etc.) values regarding data management.
 - Reflect on the gaps that may exist between value systems as they relate to the data lifecycle.
- Discussion points – themes/points on whiteboard
 - How would you define data management?
 - What are different aspects/elements/activities?
 - What are you responsible for? What are others responsible for?
- In class activity
 - Draw a data lifecycle that represents your project.
 - Talk through your lifecycle to the group.
- Implementation activity (uploaded)
 - 5-minute reflection answering: Why is data management important to you? What do you want to learn in the program? What do you think will be challenging for you regarding data management? What do you think will be easy for you regarding data management?
- DIL Competency and skills addressed
 - DIL 3: Data curation and reuse
 - Recognizes that data may have value beyond the original purpose, to validate research, or for use by others
 - Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community driven e-research
 - DIL 12: Ethics and attribution
 - Develops an understanding of intellectual property, privacy and confidentiality issues, and the ethos of the discipline when it comes to sharing and administering data.

Week 2: Discipline standards and expectations

- Learning objective
 - Students will learn the stakeholders, influencers, and standards that govern data management in their discipline and their research project.
- Discussion points

- What is metadata? How do you define it?
- What are the parameters for good metadata?
- What are examples of bad metadata?
- In-class activity/Implementation activity (uploaded)
 - Work through the modified Deep Dive into Data Worksheet (adapted from (Akers et al., 2014))
- DIL Competency and skills addressed
 - DIL 1: Culture of practice
 - Recognizes the practices, values, and norms of the field, discipline, or subdiscipline as they relate to managing, sharing, curation, and preserving data
 - Recognizes relevant data standards for the field and how these standards are applied.
 - DIL 3: Data curation and reuse
 - Is able to distinguish which elements of a data set are likely to have future value for self and others.

Week 3: Project mapping

- Learning objectives
 - Students will learn the components of a data management plan.
 - Students will map the infrastructure for their own projects.
- Discussion points
 - No discussion.
- In class activity
 - Create a DMPTool account.
 - Overview of the DMPTool and expectations.
 - Peer feedback on project map.
- Implementation activity
 - Facilitated activity building their project map.
 - Submit photo of final map. (uploaded)
- DIL competency and skills addressed
 - DIL 4: Data management and organization
 - Understands the life cycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets.
 - DIL 7: Data curation and reuse
 - Articulates the planning and activities needed to enable data curation, both generally and within their local practice
 - DIL 9: Databases and data formats

- Becomes familiar with standard data formats and types for the discipline.
- Understands which formats and data types are appropriate for different research questions.

Week 4: Folder organization

- Learning objectives
 - Students will learn best practices for folder organization.
 - Students will begin developing an organization structure based on their project maps.
- Discussion points
 - What approaches are there for organization strategies?
 - What are advantages/disadvantages from each strategy?
- In class activity/Implementation activity (combined)
 - Using Wk 3 project map, create folder structure that aligns with workflow.
 - Mini consultation by instructor with each student during class to ask/answer questions about their particular folder strategy.
 - Ten-minute reflection answering the following: What is your current strategy for folder organization? What are you doing well? What could be better and how are you going to get there? (drawing on feedback from the mini consultation) (uploaded)
 - Free time: work on DMP.
- DIL competency and skills addressed
 - DIL 4: Data management and organization
 - Understands the life cycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets.

Week 5: File naming

- Learning objectives
 - Students will learn best practices for file naming.
 - Students will create file naming conventions for their files
- Discussion points
 - What approaches are there for file naming strategies?
 - What are the advantages/disadvantages of each strategy?
- In class activity/Implementation activity (combined)
 - Using Wk 3 project map and Wk 4 folder structure, create file naming conventions that align with workflow.
 - Mini consultation by instructor with each student during class to ask/an-

- Ten-minute reflection answering the following: What is your current strategy for file naming? What are you doing well? What could be better and how are you going to get there? (drawing on feedback from the mini consultation) (uploaded).
 - Free time: work on DMP.
- DIL competency and skills addressed
 - DIL 4: Data management and organization
 - Understands the life cycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets.
 - DIL 5: Data preservation
 - Utilizes best practices in preparing data for its eventual preservation during its active life cycle
 - DIL 7: Data quality and documentation
 - Tracks data provenance and clearly delineates and denotes versions of a dataset
 - DIL 11: Metadata and data description
 - Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by self and others.

Week 6: Tables of contents, indexes, readmes, dictionaries, and codebooks

- Learning objectives
 - Students will learn the principles and best practices for creating and using tables of contents, indexes, readme files, codebooks, and/or dictionaries.
 - Students will begin to create these objects for their research projects.
- Discussion points
 - Review examples of tables of contents, indexes, readme files, data dictionaries, and codebooks.
- In-class activity
 - Small group evaluation of a “messy spreadsheet.” Report out what documentation could be created to make sense of or be able to use the data.
- Implementation activity (uploaded)
 - Using template provided, students list documentation needed to create context between their research components outlined in Wk 1-5 (uploaded).
 - Free time: work on DMP.
- DIL competency and skills addressed
 - DIL 3: Data curation and reuse

- Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community driven e-research
 - DIL 4: Data management and organization
 - Understands the life cycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets.
 - DIL 7: Data quality and documentation
 - Recognizes, documents, and resolves any apparent artifacts, in-completion, or corruption of data
 - Utilizes metadata to sufficiently enable reproduction of research results and data by others
 - Tracks data provenance and clearly delineates and denotes versions of a dataset.

Week 7: Templates: protocols, checklists, Standard Operating Procedures

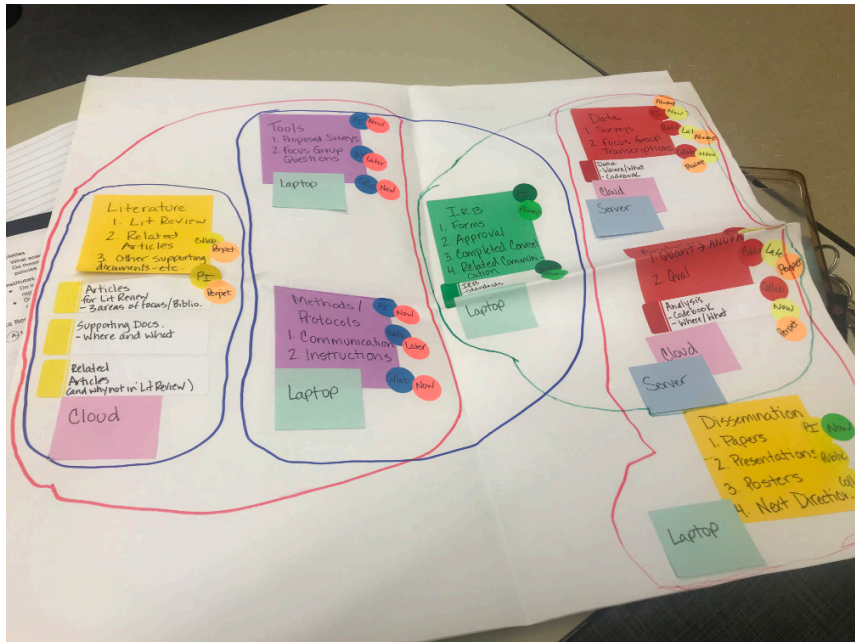
- Learning objectives
 - Students will learn the principles and best practices for creating and using templates, standard operating procedures, and checklists.
 - Students will explore examples of these objects and decide what they need to support their research projects.
 - Students will create one of these objects.
- Discussion points
 - What is the different between protocols, templates, standard operating procedures and checklists?
 - What are the advantages/disadvantages of each?
 - Where can we use each in data management?
- In-class activity/Implementation activity
 - Student selects an area where they are performing a new data management task and creates either a protocol, template, standard operating procedures, or checklist to support it. (uploaded)
 - Mini consultation by instructor with each student during class to ask/answer questions about their support document.
 - Free time: work on DMP.
- DIL competencies and skills addressed
 - DIL 3: Data curation and reuse
 - Articulates the planning and activities needed to enable data curation, both generally and within their local practice.
 - DIL 4: Data management and organization
 - Creates standard operating procedures for data management and

documentation.

Week 8: Productivity and Habits

- Learning objectives
 - Students will identify data management task they would like/need to accomplish going forward in their research.
 - Student will create a plan for accomplishing those tasks.
- Discussion points
 - Difference between productivity strategies.
- In class activity/Implementation activity
 - Facilitated prompts through worksheet answering questions regarding future data management activities and developing a regular habit regarding data management. (uploaded)
 - Submit final DMP. (uploaded)
- DIL competencies and skills addressed
 - DIL 3: Data curation and reuse
 - Recognizes that data must be prepared for its eventual curation at its creation and throughout its life cycle

APPENDIX B Student Map Examples



APPENDIX C

Assessment Questions and Scale

The following questions were selected among the 115 given because they represented foundational concepts in data management relative to the curriculum presented here. Questions were asked before exposure to program content and again within two weeks of program close. Both surveys were distributed by Qualtrics Survey Software.

- How familiar are you with documentation standards required by funders in your discipline?
- How familiar are you with data sharing requirements for funders in your discipline?
- How familiar are you with creating structured organization schemes using folders?
- How familiar are you with creating file naming conventions?
- How familiar are you with practices for storing data?
- How familiar are you with University storage options?
- How familiar are you with writing data management plans?

The same scale was used for these questions.

Extremely (1)

Very (2)

Somewhat (3)

Not Very (4)

Not at All (5)