Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned

Oregon State
UNIVERSITY

Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned

Amanda L. Whitmire

Implementing a Graduate-Level Research Data Management Course:
Approach, Outcomes, and Lessons Learned

Amanda L. Whitmire
Data Management Specialist, Oregon State University

INTRODUCTION As data-driven research becomes the norm, practical knowledge in data stewardship is critical for researchers. Despite its growing importance, formal education in research data management (RDM) is rare at the university level. Academic librarians are now playing a leadership role in developing and providing RDM training and support to faculty and graduate students. This case study describes the development and implementation of a new, credit-bearing course in RDM for graduate students from all disciplines. **DESCRIPTION OF PROGRAM** The purpose of the course was to enable students to acquire foundational knowledge and skills in RDM that would support long-term habits in the planning, management, preservation, and sharing of research data. The pedagogical approach for the course combined outcomes-centered course design with active learning techniques. Periodic course assessment was performed through anonymous student surveys, with the objective of gauging course efficacy and quality, and to obtain suggested modifications or improvements. These assessment results indicated that the course content and scope were appropriate and that the active learning approach was effective. Assessments of student learning demonstrated that all major learning objectives were achieved. **NEXT STEPS** Information derived from the student surveys was used to determine how the course could be modified to improve student experience and the overall quality of the course and the instruction.

NOTE: Some material included herein was drawn from poster presentations given at the *Data Information Literacy Symposium* in September 2013 (A. L. Whitmire, 2013) and the *6th Annual University of Massachusetts & New England Area Librarian e-Science Symposium* in April 2014 (A. L. Whitmire, 2014).

Correspondence: Amanda L. Whitmire, 121 The Valley Library, Oregon State University, Corvallis, OR 97331-4501, amanda.whitmire@oregonstate.edu

INTRODUCTION

This paper describes the development and implementation of a new course for graduate students in selected aspects of data information literacy. The idea of creating a credit-bearing, graduate-level course was conceived during a meeting between two members of the library's Center for Digital Scholarship and Services and two high-level administrators of the Graduate School. Our library-based data services were just becoming established, and we were beginning to open lines of communication with university stakeholders. A major point of discussion during this meeting was the idea of preserving the datasets that underpin graduate student theses and dissertations in our institutional repository (IR). Our Graduate School mandates the deposit of an electronic copy of the thesis or dissertation (ETD) into our IR in order to graduate, and those of us at the meeting were discussing adding a mandate for the deposit of student datasets as well. We all had to concede, however, that many graduate students would probably not be prepared to deposit a dataset into the IR because they would not have received any formal training in data management during the course of their programs, and as such, their data would not be in sufficient condition to be shared (i.e., be well organized and sufficiently documented). If preserving and sharing the datasets produced by graduate students was to be a shared goal of the Graduate School and the library, then we had to develop and provide a mechanism that would empower students to be able to realize that goal.

Our library partnered with the Graduate School to facilitate the process of establishing a new credit-bearing course in the OSU catalog titled "Research Data Management." The course is now open to graduate students from all disciplines without any prerequisites and is taught by the Libraries' data management specialist. Curricular materials were drawn from existing resources where possible, but much of the lecture and computer laboratory content were developed from scratch. The pedagogical approach for the course (A. L. Whitmire, 2013) combined outcomes-centered course design (Nilson, 2010) with active learning techniques. The approach was predicated on the idea that getting the students actively engaged with the content and techniques of data management would be more effective than lecture alone. This approach was also driven by the practical reality that while the course content was necessarily discipline-agnostic, successful learning outcomes depended on the students' ability to apply the material to discipline-specific standards of practice (e.g. metadata and data documentation, data sharing formats, and methods, etc.). The goal was that after taking the course, the students would successfully incorporate data management best practices into their daily workflow. Such behavioral change takes self-reflection (i.e., "*How does this material relate to me?*") and practice, and opportunities for both were integrated into class meetings. This paper describes the pedagogical approach, content, and assessment of the first offering of this course during the winter term of 2014 and describes how the course could be modified to improve future offerings.

LITERATURE REVIEW

As data-driven research becomes the norm, practical knowledge in data stewardship is critical for researchers (Jahnke, Asher, & Keralis, 2012; Ogburn, 2010). Carlson, Fosmire, Miller, and Nelson (2011) capture the sentiment nicely by saying, " …it is not simply enough to teach students about handling incoming data, they must know, and practice, how to develop and manage their own data with an eye toward the next scientist down the line." In a series of semi-structured Data Curation Profile interviews with faculty from a broad range of disciplines, Witt, Carlson, Brandt, and Cragin (2009) found that faculty often felt that graduate students were lacking skills in data management and curation. Just as importantly, they also observed that the faculty interviewees often admitted that they were unprepared to provide adequate training for their students in RDM because they themselves lacked knowledge and skills in that area. Despite its growing importance and a clear need for training, however, credit-bearing coursework for graduate students in research data management (RDM) is still rare at the university level.

Academic librarians are now playing a leadership role in developing and providing RDM training and support to faculty and graduate students alike (ACRL Research Planning and Review Committee, 2012; Cox, Verbaan, & Sen, 2012; Tenopir, Birch, & Allard, 2012). These instructional programs range from standalone workshops (Coates, 2013; Konkiel, Marshall, & Polley, 2013) and workshop series (Muilenburg, Lebow, & Rich, 2015) to online guides and tutorials (Earth Science Information Partners, n.d.; EDINA & University of Edinburgh, 2011), non-credit flipped classroom instruction (Johnston & Jeffryes, 2015), and discipline-specific credit-bearing courses (Wright & Andrews, 2015). Many of these resources are discipline-specific or are targeted to a specific audience. Discipline-agnostic materials tend to either be in the form of online guides created by librarians for a university-wide audience or workshop or course content created by librarians for librarians (Creamer, Morales, Kafel, Crespo, & Martin, 2012; Martin, Creamer, & Kafel, 2013), although at least one other RDM course for a broad audience is available (Borgman, 2015). A more complete, curated list of data information literacy instructional content and materials is available for those who wish to find and reuse these resources ("DIL Course Materials," 2015).

DESCRIPTION OF PROGRAM

1. Course Structure, Content and Activities

The primary goal of the course, titled Research Data Management, was to enable graduate students to acquire foundational knowledge and skills in selected data information literacy

(DIL) core competencies (Carlson et al., 2011; Carlson, Johnston, Westra, & Nichols, 2013) that would support long-term habits in planning, management, preservation, and sharing of research data. During the winter term of 2014 the inaugural course had an official enrollment of 11 students, including one faculty member enrolled for credit and two as non-credit auditors. There was also an additional faculty member sitting in on lectures who did not complete any of the outside assignments or assessments. The disciplinary range of the students was broad: six students from the College of Public Health and Human Sciences, two from the College of Forestry, and one each from the Colleges of Veterinary Medicine, Science, and Agriculture. Aside from the faculty members, student degree paths ranged from non-thesis masters to Ph.D., with some of the students having a very well defined research project already planned and others much less so. With all of the variability in student disciplinary background and experience, the in-class learning activities and the homework, midterm, and final exam assignments were relied upon to facilitate application of the generalized course content to their individual, discipline-specific circumstances.

The 2-credit, eleven-week class met twice per week for 50 minutes. For each meeting a lesson plan was created that included anticipated timing, learning outcomes, lecture content, teaching strategies, the DIL core competency addressed, and assessment approach, if any (A. Whitmire, 2014b). Lesson plans also included the associated readings and homework assignment, if any. In order to develop learning outcomes for the course, the learning outcomes from Piorun et al. (2012) were sorted into the categories of DIL core competencies and then combined with the DIL competencies themselves. There was some duplication between the DIL competencies and the Piorun et al. learning outcomes, but surprisingly little (see Appendix 1). The next steps were to merge areas of overlap and then whittle down and refine the learning outcomes to fit the scope of the course. The DIL competencies are broader in scope than the topic of research data management (they include data visualization, for example), so not all DIL competencies were addressed in the course curriculum. The final set of merged, whittled down learning outcomes formed the foundation of the ten-week series of lectures, with computer laboratory exercises and in-class activities blended in. The course content was then determined based upon the learning outcomes. That is, I first decided what I wanted the students to learn, and then developed content to support those learning outcomes. This approach is called outcomes-centered course design (Nilson, 2010) and was a major aspect of the pedagogical approach for the development of this course.

Lecture materials were mostly created from scratch, but were also drawn from existing resources. I drew mostly from the New England Collaborative Data Management Curriculum (NECDMC; Kafel, Creamer, & Martin, 2015; Piorun et al., 2012), the DataONE Education Modules (Henkel et al., 2012), and the MANTRA online course materials (EDINA & University of Edinburgh, 2011) (see Appendix 2 for a list of lectures

with source materials attributed). The final set of lecture slide decks is available (A. Whitmire, 2014a), but has since been refined for the second course offering (updated slides are posted at: A. Whitmire (n.d.)).

The midterm and final exams for the course were intended to help transform and apply the discipline-agnostic course content to widely varying discipline-specific student experiences. The midterm assignment for the course was a scaled back Data Curation Profile (DCP), and the final exam assignment was to create a data management plan (DMP) for their research project. The rationale behind using a DCP as an assignment was to give the students an opportunity to hear from a researcher in their discipline about how data is managed in the "real world." The lectures, class activities, and homework can only go so far in linking discipline-agnostic course content with the realities of discipline-specific practices. In the ideal scenario, interviewing a practitioner in their field would provide the student with insights into discipline-specific methods for creating metadata, for example, or reveal a community-accepted data sharing platform or archive. What actually transpired was a reality check for most of the students, who discovered that the data management habits of their mentor were ad hoc at best and nonexistent at worst. Students reported that having an opportunity to discuss data management with a practitioner in their field was a valuable exercise because it revealed, in very specific ways, how a course like Research Data Management could be useful.

Assigning a DMP as the final exam project served many purposes. First, like the DCP, it was intended to engage the students in adapting and applying course material to their discipline and their individual workflow. Second, it was intended to provide the students with a tangible, practical product that they could each take away from the course and refer back to for the duration of their graduate studies. This course was intended to be practical and applied; I wanted students to finish the course knowing the things that they needed to do in order to manage their data well and how to do them. Having them create a DMP was the most straightforward mechanism for facilitating student self-reflection on how theoretical data management best practices were directly related to their research processes. Lastly, the topics covered in a data management plan almost fully address the major learning outcomes for the course. I had planned to use an assessment of student performance on the DMP assignment as a means to evaluate how well the students ultimately achieved the course-level learning outcomes. A discussion of this assessment follows in a subsequent section.

2. Instructional Approach

The average student cannot remember the factual content of a lecture fifteen minutes after it ends (Nilson, 2010). As such, lecturing students who passively listen is not a

very effective means to impart knowledge and affect behavioral change. An instructional approach that engages the student with lecture content through reading, writing, talking, and reflecting, has been shown to result in better retention of material. This instructional approach is called active learning, and it was the approach that was taken in developing Research Data Management. As mentioned in the previous section, DIL core competencies were matched with learning outcomes and teaching strategies that used active learning techniques. For example, under the DIL core competency "Databases and data formats" is the following learning outcome (from Piorun et al. (2012)): "Explain what a research data set is, and the range of data types." After lecturing on the topic, I had the students execute one minute of reflective writing wherein they wrote down which types of data they would be generating and, if known, which formats. I then asked a few of the students to report out to the class for a short discussion on the topic. More examples of active learning teaching strategies for DIL competencies and learning outcomes are shown in Table 1 (following page), and the complete set are laid out in the course lesson plans (A. Whitmire, 2014b). The learning outcomes drove the development of the content, which then informed the creation of the active learning exercises.

3. Assessment

In an effort to gain information about the quality of the course and how well the students achieved the learning outcomes, I conducted both formative and summative assessments. Formative assessment is generally conducted as a means to track student learning and provide instructors with information on how to improve their teaching. The formative assessment approach for Research Data Management was to anonymously survey the students twice during the course, once at halfway through, and again during the final week of classes. I asked targeted questions about how well sessions prepared them to meet specific learning outcomes and requested written feedback on what they liked the most and the least about the course. I also asked what they thought would be the single most significant improvement to the course thus far. I also used the results from some of the active learning exercises as "check-ins" to see how the students were grasping individual topics. For example, in the case of the reflective writing exercise discussed above regarding data types, I collected the notecards that the students wrote on and reviewed them after class. It was clear from reviewing the cards the extent to which the students had understood, for example, the difference between data type and data format (a key point during the lecture). Summative assessment is conducted in order to evaluate student learning and is generally captured at the major milestones of a course (e.g. via the midterm or final exams). My approach to summative assessment was to use student performance on the DMP assignment as an indicator of how well they had achieved the major course learning outcomes. The content and quality of the students' DMPs were direct indicators of how well they had successfully grasped concepts, and, as such, the DMP was a valuable resource in performing summative assessments.

Core Competency	Learning Outcomes	Teaching Strategy	Strategy Description
Data Types & Formats	Becomes familiar with standard data types and formats for their discipline	Write-Pair-Share	Group students by discipline, if possible; have students individually write a response to a question about common data types & formats for their discipline or for their research; have students share and compare responses within groups
Data Management & Organization	Understands the life-cycle of data	Buzz Groups	Instructor provides a blank flowchart of the research lifecycle; get students into small groups; have students work together to fill in the lifecycle phases; bring class back together and share results
Metadata & Data Description	Understands the concept of, rationale for, metadata	Muddiest Point	Immediately following lecture topic introducing metadata, ask students to write down what they found unclear about the concept of, or rationale for, metadata; results are shared with the class, recorded and tabulated
	Develops the ability to read and interpret metadata from external disciplinary sources	Cooperative Learning, Integrating Technology	Group students randomly; each group visits an online repository or database; group evaluates the scope, quality, and overall level of adequacy of the metadata; results are reported back to the whole class
Data Curation & Re-use	Articulates the planning and actions needed to enable data curation	One Minute Paper	Have students write a brief response to the question, "What obstacles might you encounter in sharing your data?"; students report back to the whole class; responses are recorded to identify common themes
Discovery & Acquisition	Locates and utilizes disciplinary data repositories	Cooperative Learning, Integrating Technology	Group students by discipline, as much as possible; task them with identifying a disciplinary data source (repository or database); ask them to download and open data; summarize process as a group and report back to whole class about the process

Table 1. Examples of how DIL core competencies and learning outcomes can be linked with various teaching strategies. All of the teaching strategies listed are a form of active learning, which are intended to help students better retain the material compared with lecture alone. Adapted from Whitmire et al. 2013.

4. Lessons Learned

The anonymous student surveys were critical in being able to gauge which aspects of the course were successful and where there was room to improve. There were several major takeaways from the surveys. First, the students expressed a desire for me to connect them and the content more to the "real world." Not surprisingly, the students most enjoyed aspects of the course that involved grounding theoretical topics, practices, or ideas in reality. Examples of reality-based aspects that they most enjoyed about the course included hands-on activities in class, opportunities to learn about software tools and resources in the computer lab, examining case studies in data management success and failure, and having guest lecturers visit the class. It was clear from their feedback that I could improve the course by incorporating more real-world cases into lecture content. One student suggested that I use a real research project as a case study that we could follow across topics throughout the course. This is a fantastic concept, but one that may be difficult to implement completely (for reasons that are explored in the Next Steps section below).

In the survey responses the students expressed an interest in gaining more hands-on experience with metadata. They were eager to learn more about the theoretical concept of metadata and the tools and methods for creating it. We had a computer lab period devoted to demonstrating metadata tools (Colectica and DataUP) led by our Metadata Librarian, but the students wanted more time with the material and something more interactive. Metadata format and creation are very discipline-specific, and this is one area where I was less successful in meeting the learning needs of my discipline-diverse students. In future iterations of the course, I need to add another computer laboratory session on metadata in order to give the students more time and experience with the tools and to design an assignment that would help to clarify both the metadata creation process and the desired products (a codebook, for example). Instead of just demonstrating the tools, it makes sense to give the students the opportunity to actually use them.

Another suggestion that came directly from a student was to use the data management plan as framework for the course. The goal of the course was to give students knowledge and skills in data management that apply directly to their research workflow. Their final assignment, a data management plan (DMP), was the culmination of the class and was intended to provide them with a guiding document for the remainder of their graduate research. While I verbally related course content to the DMP throughout the course, they did not actually create one until the end of the quarter. A better approach may be to have them create sections of the DMP as homework assignments throughout the course. This would better facilitate the self-reflective process of applying largely discipline-agnostic course content to their highly individual research practices.

There were a few suggestions from the students to offer at least two versions of this course, one for humanities and social sciences and one for natural and applied sciences. This suggestion makes sense on many levels, as the research methods, data types and formats, and cultural practices related to things like metadata and data sharing are vastly different across that broad disciplinary divide. A more discipline-specific course, even over such broad areas, would enable an instructor to work more subject-specific content into lectures and would permit spending a bit more time and going into greater depth on some topics. It would also make weaving a single case study through the course a more approachable objective. A major constraint to splitting the course however is clearly that such an approach would require twice the teaching workload. As such, this is an idea that would require considerably more effort to implement relative to the other suggested changes to the course, and despite being an excellent idea, is less likely to be carried out in the near term.

One lesson learned during the course that was not derived from the mid- and end-of-course student surveys was related to where the students were in terms of their thesis or dissertation research process. In a course pre-assessment, I asked the students how well defined their graduate research topic was. Since the main purpose of the course was to enable the students to manage their own data effectively, it was important that they had a good idea of what they would be working on. The reason being that they would then have a practical framework upon which they could apply the major course concepts. A good example of applying course content to an individual's workflow is the practice of file naming. After introducing the concept of having a file naming convention and providing some best practices and examples, I asked the students to develop a file naming convention for their data files. Even if they hadn't collected any data yet, if they understood what they would be working on and the kinds of data they would be generating, they could make a best guess at a file naming convention. A student who didn't have at least a topic fleshed out would be at a loss in developing a file name. The course pre-survey results indicated that two students had no idea what they would be working on, and three had only a general idea. A few lectures into the course, it became clear that these students were having a hard time connecting course content to their individual situations. For the fifth lecture, I decided to deviate from my syllabus and added a brief lecture that described how to approach outlining a research project. I introduced the major elements that they needed to consider, including a description of the research question, the process for addressing it, the data types and formats produced, and who would be responsible for the various aspects of the work. I revealed the parallels between these topics and the first section of a data management plan and then assigned section one of a DMP as homework, along with describing their research question and why it was important. Following this lecture and homework assignment it was much clearer to the students (and to me) how they would apply the concepts that they were

learning in class. In subsequent course offerings I need to be more explicit up front about the need for students to have a relatively well-defined research topic and plan to give this homework assignment earlier on.

NEXT STEPS

The students rated the course highly, and many expressed in anonymous comments how useful they thought it had been. The quality of the data management plans that they submitted at the end of the course demonstrated that they had successfully met the course learning objectives and had been able to apply the course concepts to their individual research workflows. Nevertheless, there is always room for improvement in instruction, especially after the first offering of a newly developed course. Feedback from anonymous student surveys was extremely useful in charting a path forward. The active learning exercises were successful and generally enjoyed by students, but more active learning opportunities need to be developed and integrated into the lectures where they were lacking. A few lectures only had one active learning component, and two others had none. It is well understood that after about fifteen minutes of lecture, student attention span and level of engagement drop precipitously (Nilson, 2010, and references therein). The addition of active learning exercises was meant to directly address that problem. Active learning also included the hands on learning that students experienced in the computer laboratory. It's clear that students felt a need for more applied instruction on metadata. During the metadata and other sessions that we conducted involving computers, it was evident that the students were not very well versed in Excel. Adding a computer laboratory session on spreadsheet best practices and how to use Excel most efficiently would directly support the course goals of enabling the students to document, organize, and share their data effectively.

The idea of using a single research case study to examine data management topics throughout the course is a good one, difficult though it may be to implement. In order to use a single research study as an example for all of the major course content, the project would have to involve both the use of human subjects and have intellectual property (IP) ramifications and also be relatable for students across a broad disciplinary range. That is a quite a challenge. Despite the potential difficulty, there are several possible solutions ranging from steadfastly searching until the perfect case is discovered to deciding to use alternative case studies for special topics like sensitive human subjects data or IP issues. Using a variety of cases may well be a better strategy for a course that is designed to meet the needs of students from a variety of disciplines, and there are many existing resources to pull from. For example, the NECDMC currently provides eleven case studies from actual research projects across several disciplines and includes resources for how to incorporate these case studies into instruction as well as a list of which case studies specifically address certain data management topics

(Gaudette & Kafel, 2013; "New England Collaborative Data Management Curriculum," 2012). Beyond the resources available through the NECDMC, other potential sources of case studies are the Data Curation Profiles Directory (Witt et al., 2009), topic-specific data management examples from a growing set of blog entries ("Data Stories," n.d., "Your Data Stories," n.d.), and possibly even data management plans that have been made public (see a growing list at the DMPTool; https://dmptool.org/public_dmps).

All in all, the feedback derived from asking the students to complete anonymous surveys was invaluable and is a practice that will be continued in future course offerings. As more feedback comes in, the course will continue to be adapted and improved upon with each iteration.

CONCLUSION

Research data management continues to be an area of growing involvement for academic libraries, and providing instruction in data information literacy is one of the primary areas of engagement. Librarians are currently offering a wide range of instructional opportunities for diverse audiences. These range from online guides, to workshops and workshop series' for graduate students and faculty, and discipline-specific credit-bearing courses. This case study presented the course structure and instructional approach for developing a discipline-agnostic, credit-bearing course in research data management for graduate students. The course was constructed using a pedagogical approach that combined outcomes centered course design and active learning. Major course topics were: the research lifecycle and data management planning; types, formats, and stages of data; storage, backup and security; metadata; legal and ethical considerations; data sharing and reuse; and archiving and preservation. Student assessment was carried out during the course and was used to solicit anonymous feedback. The lessons learned from those assessments were extremely helpful in determining how the course could be improved. The next steps will be to incorporate student feedback in a few critical areas where changes in the course design or content will improve the student experience and ideally, their success.

While it is too early to have robust evidence of the impact of this course on the research data management habits of the students, their informal feedback has been encouraging. One student left survey feedback that said, "Very helpful, I feel like I am walking away prepared to start my data collection for my thesis." More than a year after the fact, students still tell me that the class was useful. One student shared with me that she has strongly influenced other members of her research group in using the file-naming conventions that she developed during the course. That kind of feedback tells me that I was on the right track in developing and offering this course. As a direct result of student feedback, the second offering of the course in 2015 saw some great improvements. I look forward to continuing to adapt and enrich the course in the future.

ACKNOWLEDGEMENTS

REFERENCES

ACRL Research Planning and Review Committee. (2012). 2012 top ten trends in academic libraries A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, *73*(6), 311–320. Retrieved from http://crln.acrl.org/content/73/6/311

Borgman, C. L. (2015). *Syllabus for Data Management and Practice, Part I, Winter 2015*. Retrieved from http://works.bepress.com/borgman/358

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. http://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. http://doi.org/10.2218/ijdc.v8i1.254

Coates, H. L. (2013). *Developing a data management lab: Teaching effective methods for health and social sciences research*. Retrieved from https://scholarworks.iupui.edu/handle/1805/3569

Cox, A., Verbaan, E., & Sen, B. (2012). Upskilling liaison librarians for research data management. *Ariadne*, (70). Retrieved from http://www.ariadne.ac.uk/issue70/cox-et-al

Creamer, A., Morales, M., Kafel, D., Crespo, J., & Martin, E. (2012). A sample of research data curation and management courses. *Journal of eScience Librarianship*, *1*(2). http://doi.org/10.7191/jeslib.2012.1016

Data Stories. (n.d.). [Blog]. Retrieved March 9, 2015, from https://notebooks.dataone.org/data-stories/

DIL Course Materials. (2015). Retrieved February 16, 2015, from http://esciencelibrary.umassmed.edu/DIL_Course_Materials

Earth Science Information Partners. (n.d.). *Data management short course for scientists*. Retrieved March 8, 2015, from http://commons.esipfed.org/datamanagementshortcourse

EDINA, & University of Edinburgh, D. L. (2011). *Research data MANTRA* [online course]. EDINA. Retrieved from http://datalib.edina.ac.uk/mantra #sthash.kvwJvIEw.dpuf

Gaudette, G., & Kafel, D. (2013). A case study: Data management in biomedical engineering. *Journal of eScience Librarianship*, *1*(3). http://doi.org/10.7191/jeslib.2012.1027

Henkel, H., Hutchison, V., Strasser, C. A., Rebich Hespanha, S., Vanderbilt, K., & Wayne, L. (2012, May). *DataONE education modules*. DataONE. Retrieved from https://www.dataone.org/education-modules

Jahnke, L., Asher, A., & Keralis, S. D. C. (2012). *The problem of data* (No. 154) (p. 43). Washington, D.C.: Council on Library and Information Resources.

Johnston, L., & Jeffryes, J. (2015). Teaching civil engineering data information literacy skills: An e-learning approach. *Data Information Literacy Case Study Directory*, *3*(1). http://doi.org/10.5703/1288284315479

Kafel, D., Creamer, A., & Martin, E. (2015). Building the New England Collaborative Data Management Curriculum. *Journal of eScience Librarianship*, *3*(1). http://doi.org/10.7191/jeslib.2014.1066

Konkiel, S., Marshall, B., & Polley, D. E. (2013). Integrating data management literacies with data visualization instruction: A one-shot workshop. Retrieved from https://scholarworks.iu.edu/dspace/handle/2022/16814

Martin, E., Creamer, A., & Kafel, D. (2013). *Scientific Data Managament LIS 532G-01*. Retrieved March 9, 2015, from https://sites.google.com/a/vt.edu/srdm/home

Muilenburg, J., Lebow, M., & Rich, J. (2015). Lessons learned from a research data management pilot course at an academic library. *Journal of eScience Librarianship*, *3*(1). http://doi.org/10.7191/jeslib.2014.1058

New England Collaborative Data Management Curriculum. (2012). Lamar Soutter Library, University of Massachusettes Medical School. Retrieved from http://library.umassmed.edu/necdmc/index

Nilson, L. B. (2010). *Teaching at its best : A research-based resource for college instructors* (3rd ed.). Hoboken: Wiley.

Ogburn, J. L. (2010). The imperative for data curation. *portal: Libraries and the Academy*, *10*(2), 241–246. http://doi.org/10.1353/pla.0.0100

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

Tenopir, C., Birch, B., & Allard, S. (2012). *Academic libraries and research data services: Current practices and plans for the future* (White paper) (p. 54). Association of College and Research Libraries. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Whitmire, A. (2014a, April). *GRAD 521 Research Data Management Lectures*. PowerPoint. Retrieved from http://dx.doi.org/10.6084/m9.figshare.1003835

Whitmire, A. (2014b, April 18). *GRAD 521 Research Data Management Syllabus and Lesson Plans*. figshare. Retrieved from http://dx.doi.org/10.6084/m9.figshare.1003834

Whitmire, A. (n.d.). *GRAD 521 - Research Data Management*. Lectures. [LibGuide]. Retrieved June 10, 2015, from http://guides.library.oregonstate.edu/grad521lectures

Whitmire, A. L. (2013, September). Pedagogy for data management: curricular design strategies for teaching research data management to graduate students. Poster presented at the *Data Information Literacy Symposium*, West Lafayette, IN. Retrieved from http://ir.library.oregonstate.edu/xmlui/handle/1957/42703

Whitmire, A. L. (2014, April). Implementing a graduate-level data information literacy curriculum at Oregon State University: approach, outcomes and lessons learned. Poster presented at the *6th Annual University of Massachusetts & New England Area Librarian e-Science Symposium*, Worcester, MA. Retrieved from http://escholarship.umassmed.edu/escience_symposium/2014/posters/7

Witt, M., Carlson, J., Brandt, D. S., & Cragin, M. H. (2009). Constructing data curation profiles. *International Journal of Digital Curation*, *4*(3), 93–103. http://doi.org/10.2218/ijdc.v4i3.117

Wright, S., & Andrews, C. (2015). Developing a for-credit course to teach data information literacy skills: A case study in natural resources. *Data Information Literacy Case Study Directory*, *2*(1). http://doi.org/10.5703/1288284315476

Your Data Stories. (n.d.). [Blog]. Retrieved March 9, 2015, from http://datastories.jiscinvolve.org/wp/

APPENDIX 1. Developing learning outcomes for a graduate-level research data management course by merging existing outcomes and creating new ones.

Under the data information literacy (DIL) core competency main categories (Carlson et al. 2012), I sorted the learning outcomes (LO) from the Piorun et al. (2011) syllabus and the more granular DIL competencies. I distinguish between the two sources of LOs by using different fonts, as follows:

Data Information Literacy Main Categories
- Data Information Literacy Core Competencies
- Piorun et al. learning outcomes

Learning outcomes that made it into the final course lesson plans have an asterisk (*). This effort is a "first-pass", and would benefit from community feedback.

Databases and Data Formats
- Understands the concept of relational databases, how to query those databases
- *Explain what research data is/are
- *Explain what a research data set is and the range of data types
- *Understands which data types are appropriate for answering different types of research questions
- Chooses appropriate data format for data management action (e.g. sharing or preservation) or audience
- Becomes familiar with standard data types and formats for their discipline
- *Identify stages of research data

Data Management and Organization
- *Understands the lifecycle of data, and which DM actions are associated with lifecycle stages
- Describe how data should be managed differently in different phases of the life cycle
- Explain the need for managing/sharing research data and identify relevant public policies
- *Explain the lifecycle continuum to manage and preserve research data
- *Describe the value and relative importance of data management to the success of a research project

- *Develops data management plans
- *Formulate an abbreviated data management plan or data curation profile to manage their research project data and define roles/responsibilities of research staff
- *Identify data management plan (DMP) requirements used to characterize and plan for the lifecycle of research data
- Keeps track of the relation of subsets or processed data to the original data sets
- *Creates standard operating procedures for data management and documentation, including proper use of a field or laboratory notebook (if applicable)
- Describe best practices for starting and keeping a laboratory notebook
- *Understand why data storage, backup and security of research data are important
- Understand data storage, backup and security methods for research data
- *Understand best practices for research data storage, access control, migration to newer storage media and security of research data
- *Formulate an approach to creating a data storage, backup and security plan for a project

Quality Assurance and Documentation
- Recognizes and resolves any apparent artifacts, incompletion, or corruption of data sets
- Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community-driven research
- Utilizes metadata to facilitate understanding of potential problems with data sets
- Compare relevant quality control techniques/technical standards

Metadata and Data Description
- *Understands the concept of, and rationale for, metadata
- Proficiently annotates and describes data so it can be understood and used by self and others
- *Develops the ability to read and interpret metadata from external disciplinary sources
- Understands the structure and purpose of ontologies in facilitating better sharing of data

- Understand what metadata is
- Understand why metadata is important
- *Identify applicable standards for documenting and capturing metadata
- Understand disciplinary practices associated with the collection and sharing of metadata
- *Formulate an approach to creating metadata for a project
- *Understand benefits of using a unique researcher ID in metadata (e.g. ORCID or ISNI)

Data Curation and Re-use
- 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 research
- Recognizes that data must be prepared for its eventual curation at its creation and throughout its lifecycle
- Articulates the planning and actions needed to enable data curation
- Determine common potential storage formats for data that will be accessible in the future and non-proprietary where possible (NP)
- Identify who can share/access your data and for what purpose
- Determine requirements for pre/post publication access for project phases of the research
- Determine temporary or permanent access policy
- Define process steps and access levels for gaining access
- *Understand options for maximizing data reuse
- *Understand data identifiers and their use
- Identify issues related to discovery of relevant data sets for reuse

Ethics, including citation of data
- Develops an understanding of intellectual property, privacy and confidentiality issues, and the ethos of the discipline when it comes to sharing data
- Appropriately acknowledges data from external sources
- *Explain ownership considerations related to data sharing
- *Explain and evaluate potential legal issues connected to your data; intellectual

property, copyright claims, licenses needed for use, monetary charges for data

- *Explain ethical considerations related to data sharing
- *Understand privacy levels for research data as required by potential funding agencies
- *Recognize the importance of privacy with some forms of research data (HIPAA)
- *Understand the importance of removing key personal identifiers to facilitate confidentiality
- Understand the need for data attribution and citation
- Identify issues related to discovery, reuse, and sharing
- *Discuss issues/obstacles related to reuse and sharing
- *Understand publisher and licensing restrictions on re-use of data, analysis software and instrumentation
- Understand Open Access requirements
- Understand controversies surrounding open science, open data
- *Address re-use/sharing requirements from granting agencies or sponsors

Data Preservation
- Recognizes the benefits and costs of data preservation
- Understands the technology, resource, and organizational components of preserving data
- Utilizes best practices in preservation appropriate to the value and reproducibility of data
- *Explain options for a long-term sustainable preservation strategy/policy for your data (e.g., discipline specific, institutional, departmental)
- *Identify types of repositories/archives (discipline-based, institutional, etc.)
- *Choose appropriate subject repository for long-term storage of data
- *Understand process issues for depositing data in repository
- Explain data management tools and services available for preservation and discovery
- *Understand costs for data storage, management tools and services

Data Conversion and Interoperability
- Becomes proficient in migrating data from one format to another

- Understands the risks and potential loss or corruption of information caused by changing data formats
- Understands the benefits of making data available in standard formats to facilitate downstream use
- *Address the need for conversion to standard formats needed for re-use
- Understand different types of collaborative workspaces for sharing data

Discovery and Acquisition
- Locates and utilizes disciplinary data repositories
- Not only identifies appropriate data sources, but also imports data and converts it when necessary, so it can be used by downstream processing tools
- Understanding the need for querying and retrieval methods - discovery aids for multiple user communities to find the data they want to reuse

Data Processing and Analysis
- Becomes familiar with the basic analysis tools of the discipline
- Uses appropriate workflow management tools to automate repetitive analysis of data.

Data Visualization
- Proficiently uses basic visualization tools of discipline
- Avoids misleading or ambiguous representations when presenting data
- Understands the advantages of different types of visualization, for example, maps, graphs, animations, or videos, when displaying data

Cultures of Practice
- Recognizes the practices, values, and norms of his/her chosen field, discipline, or sub-discipline as they relate to managing, sharing, curating, and preserving data
- Recognizes relevant data standards of his/her field (metadata, quality, formatting, etc.) and understands how these standards are applied
- Identify methods of recording data that are specific to student's discipline and research interests
- Define data collection recording policies/procedures for student's research

Newly developed learning outcomes

- Understands US federal agency context for data management and sharing

- Develops appropriate file- and folder-naming conventions

- Organizes data and documentation in a logical manner

- Familiar with tools for creating metadata appropriate for your discipline and data type

- Identify and list the types of information typically included in metadata records

- Familiar with tools for creating metadata appropriate for your discipline and data type

- Recognize inconsistencies that can make a dataset difficult to understand and/ or manipulate

- Identify data entry tools

- Identify validation measures that can be performed as data is entered

- Describe steps to make data sharing easier

- Perform basic archival processes: checksum, auditing, format migration, etc.

APPENDIX 2. Research Data Management, Winter 2014 meeting schedule with source materials for lectures and laboratory exercises.

Meeting	Title or Topic	Source(s)
1	Introduction to research data management	Henkel et al., 2012
2	Data management plans and planning	Henkel et al., 2012
3	Types, formats and stages of data	"New England Collaborative Data Management Curriculum," 2012
4	Organizing your data	Purrington, n.d.
5	Data management planning for your research	None
6	Data Curation Profiles: an introduction to the midterm assignment	Carlson, J. and Brandt, D.S. (2010) "Introduction to the Data Curation Profile" [workshop] *International Digital Curation Conference (IDCC)*. Chicago, IL
7	Data storage, backup and security	EDINA & University of Edinburgh, 2011; "New England Collaborative Data Management Curriculum," 2012
8	Version control theory and methods	None – guest lecture
9	Data documentation through metadata	Henkel et al., 2012
10	Computer laboratory: Metadata Exercise	None
11	Ethical considerations for research data: human subjects	None – guest lecture
12	Legal considerations for research data: intellectual property	None – guest lecture
13	Data sharing	None – guest lecture
14	Data sharing and reuse	None
15	No lecture	N/A
16	No lecture	N/A
17	Plan for archiving and preservation of data	UK Data Archive, n.d.; Henkel et al., 2012; Research Data Management Team, UK Data Archive, University Of Essex (2012). Managing and Sharing Data: Training Resources. UK Data Service. Retrieved 29 May, 2012 from: http://data-archive.ac.uk/media/335419/trainingresources.zip
18	Computer laboratory: data repositories	Adapted from *Curating and Managing Research Data for Re-Use*, ICPSR Summer Program 2013
19	Open period	N/A
20	DMP peer review	N/A

WORKS CITED

EDINA, & University of Edinburgh, D. L. (2011). *Research data MANTRA* [online course]. EDINA. Retrieved from http://datalib.edina.ac.uk/mantra #sthash.kvwJvIEw.dpuf

Henkel, H., Hutchison, V., Strasser, C. A., Rebich Hespanha, S., Vanderbilt, K., & Wayne, L. (2012, May). *DataONE education modules*. DataONE. Retrieved from https://www.dataone.org/education-modules

New England Collaborative Data Management Curriculum. (2012). Lamar Soutter Library, University of Massachusettes Medical School. Retrieved from http://library.umassmed.edu/necdmc/index

Purrington, C. B. (n.d.). *Maintaining a laboratory notebook*. Retrieved from http://colinpurrington.com/tips/lab-notebooks

UK Data Archive. (n.d.). *UK Data Archive - TRAINING RESOURCES*. Retrieved September 14, 2012, from http://www.data-archive.ac.uk/create-manage/training-resources

Research Data Management Team, UK Data Archive, University Of Essex (2012). *Managing and sharing data: Training resources*. UK Data Service. Retrieved 29 May, 2012 from http://data-archive.ac.uk/media/335419/trainingresources.zip