

Incorporating Research Data Management into an Existing Graduate Course on Theory and Methods of Research

Joseph H. Holles

Department of Chemical Engineering, University of Wyoming
Laramie, WY 82071

Abstract: Training and practical knowledge in Research Data Management (RDM) is a mandatory need for today's researchers. Thus, graduate school, when students are focused on research is the time to obtain this education. To address the RDM need of graduate students, several approaches have been taken from workshops to stand-alone courses. The content of these approaches varies from the high-level RDM coverage of a short workshop to the practical experience of developing a full Data Management Plan (DMP) during a course. However, the literature indicates that a one-size fits all approach isn't required as students balance their RDM needs with the time required to obtain the training and knowledge.

As a middle approach between workshops and full courses, RDM was incorporated into an existing required graduate course on the Theory and Methods of Research. Thus, all students in the program would be exposed to some preliminary concepts and practices in RDM. Depending on the student's needs, they can then obtain additional RDM exposure through subsequent stand-alone workshops or full courses. Material from a recently developed semester-long RDM course was used to deliver three lectures on: 1) RDM and Sharing Mandates, 2) What is Data? and the Data Lifecycle, and 3) Planning for Your Research Project and DMPtool. This provides the students with the big picture requirements along with practical education on data, the data lifecycle, data in your research project, and using DMPtool to develop a DMP. Motivation for and details on these three lectures will be provided. Lessons learned and proposed modifications to these lectures will be discussed.

INTRODUCTION

Practical knowledge and training in Research Data Management (RDM) has become a requirement for today's researchers. Best practices in RDM are the foundation of good research for two reasons: (1) RDM is a required part of federally funded research, and (2) efficient RDM leads to more thorough use of the data for publication and sharing. As a result, the National Academy of Sciences (NAS) has advocated for data management training for all researchers.[1] Carlson et al. follow this up by stating that "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 road." [2] Many federal agencies (e.g., NSF,[3] NIH,[4] USGS[5]) also require the submission of Data Management Plans (DMP) when submitting proposals for research funding. Assuming that RDM is complete when the DMP is submitted with the proposal will only result in subsequent issues. Clearly, since the graduate students are the researchers in the lab dealing with the data on a day-to-day basis, they are also the ones to be trained in and execute the RDM.[6] Finally, graduate students are the future leaders in their institutions of employment. They will be the ones tasked with developing and executing RDM while on the job. Thus, a thorough graduate education will start the process of providing graduate students with the RDM tools that they need.

Two general approaches have developed for educating graduate students on RDM: seminars/workshops and standalone courses. The workshops/seminars are typically offered outside of student's departments (by libraries)[7-11] while the standalone courses are offered by departments to address specific needs.[12-16] The workshops usually require a smaller time commitment while the standalone courses provide in-depth material coverage. The drawback of workshops is that they are usually not for-credit and suffer retention issues.[7] For standalone courses, the time commitment is greater but the students then receive course credit towards graduation requirements. Standalone courses on RDM are currently being offered through the graduate school and taught by librarians[14] or offered by specific departments and taught in teams by faculty and librarians, each bringing their area of expertise to the classroom.[13, 17] The drawback with both the seminar/workshop and the stand alone course option is that as long as the seminar, workshop, or standalone course is optional, not all students will receive an exposure to RDM.

The goal of this paper is to describe the incorporation of a subset of RDM into an existing graduate course on the Theory and Methods of Research. Since this course is required, all graduate students will have an initial exposure to foundational reasons behind RDM along with best practices and available tools for RDM.

METHODS

To provide initial exposure to reasons for, best practices, and available tools in RDM, three lectures were added to an existing required graduate course on Theory and Methods of Research. This course is 3 credits and meets 42 times for 50 minutes during the semester. Details on the existing Theory and Methods of Research course were previously published.[18] RDM material for these three lectures was drawn from a previously developed and offered

semester long 3 credit graduate elective course in RDM,[15] and all course material is available at <https://doi.org/10.15786/M28D50>. The three lectures were on: (1) RDM and Sharing Mandates, (2) What is Data? and the Data Lifecycle, and (3) Planning for Your Research Project and DMPtool.

RESULTS

The first RDM lecture was broadly focused on the needs and requirements for data management. This lecture was based on Day 3 lecture notes from the full semester course. The OSTP document mandating data management and sharing was presented. The goals and objectives of this document were presented. Next, RDM requirements for both NIH[4] and NSF[3] were presented, compared, and contrasted. Depending on your field, examination of RDM mandates from other agencies would be appropriate.

The second lecture covered two topics: (1) What is Data?, and (2) Data Lifecycle. These were based on Day 4 lecture notes from the full RDM course. What is Data defined what data is and what data isn't considered data. Data types (observational, experimental, simulation, etc.) and data sets were also discussed. Finally, the data lifecycle (data management planning, data acquisition, data analysis etc. through data re-use) and management of data through the lifecycle were discussed. Students were then asked to provide examples from their lab that address management of data through the lifecycle.

For the final lecture, the topics were (1) Planning for Your Research Project a (2) DMPtool. These lectures were developed from Day 5 and Day 8 lecture notes of the semester long RDM course. The first topic included a group discussion on the student research project. Students were asked what kinds of data would be generated or re-used and how they planned on managing it (notebooks, spreadsheets, relational databases, etc.). This topic also included file naming, file structures, and other organizational approaches. The students responded very appreciatively to this seemingly low level discussion. The importance of this seemingly low-level topic has also been noted by Phillips et al. in their work to determine specific RDM needs of graduate students.[19] Finally use of DMPtool[20] was covered. This tool is available at many universities to assist faculty in complying with the DMP requirement associated with grant proposals. Access is typically available through the university's research office. In this tool, the user specifies the targeted funding source and the program then asks questions to address the agency specific DMP requirements. The tool also provides background information on what should go into each section as well as user supplied examples of prepared and submitted DMPs. This database is searchable on funding agency. While the students do not typically prepare this type of funding DMP, exposure to its requirements encourages their reflection on existing RDM and potential improvements to their project's RDM.

In order to include three lectures of new material in an existing course, other material must be removed or coverage must be reduced. For the existing Theory and Methods course, two student presentations are included as assignments. To create room for the three new lectures, class periods previously devoted to student presentations were replaced with RDM lectures. In order to still accommodate all of the student presentations, the time for each

presentation was shortened. Instead of two 20-minute presentations per class period, three 15-minute presentations were scheduled. This was acceptable based on previous instructor experience teaching this class as students often did not use the full amount of time allotted for their presentation. In addition, the shorter time constraint also more closely mimicked a conference presentation with a limited time length and forced the students to tighten their presentation to convey the necessary material.

DISCUSSION

The changes to this course were made during the first offering of this course following the development and initial offering of the semester long RDM course. The existing Theory and Methods of Research course already includes a number of topics that help the students become more productive in the laboratory while also helping them to proceed towards their thesis or dissertation. This includes topics such as literature search skill, reference managing software, and research notebooks. Thus, RDM education and skills fit well within this existing course.

Since the Theory and Methods course is required for all graduate students, no student will complete the program without being exposed to at least an initial discussion of RDM. Depending on the interests of the student and the RDM requirements associated with their research, the student may elect to take the subsequent semester long RDM course or other seminars/workshops available on campus. This approach is similar to many other topics in graduate school where all students take the required and broader thermodynamics or graduate math course and those students whose research requires more advanced analysis will take additional specialized thermodynamics or mathematics courses.

Topics for the three RDM lectures were developed by the instructor from the previous semester long RDM course and through discussions and feedback from students who had taken the RDM course and students who had not taken the RDM course. Students who had taken the RDM course suggested more practical, hands-on concepts and tools and techniques to help them manage their data while student who had not taken the RDM course suggested broader topics such as what is data and what are the requirements. Thus, students already exposed to RDM suggested material that resulted in the “Planning for Your Research Project and DMPtool” lecture while students not previously exposed to RDM suggested topics resulting in the other two lectures on “RDM and Sharing Mandates” and “What is Data? and the Data Lifecycle.”

As instructor, my biggest frustration was, as expected, simply not enough time to provide more coverage and depth of RDM topics. Unfortunately, there is little that can be done to address this issue. It would be hard to determine that other topics in the Theory and Methods course should receive reduced coverage in order to expand the coverage of RDM. Second, the third lecture on Planning for your Research and DMPtool should be slightly revised to not only talk about DMPtool as a tool for developing a DMP for submission with a funding proposal, but to also discuss a separate DMP that specifies and is used for the day-to-day RDM of a laboratory or project. Since the product from DMPtool is a broad high level plan of future work (funding DMP), it is very different from the details and depth needed for day-to-day RDM in the laboratory (a project DMP).[21]

The three lectures on RDM should involve the students more actively through an assignment. While asking the students to prepare a complete DMP for their project is not appropriate based on the limited coverage of the three lectures, an assignment to encourage the students to start thinking about their RDM and the existing RDM requirements of their laboratory/research project may provide an appropriate assignment. For future offerings of the course, a RDM assignment will be incorporated with the following requirements:

- Determine if your laboratory has any RDM protocols. If so, provide some examples.
- Ask your advisor if there are any DMP commitments in your funding. If so, provide examples,
- Ask you lab mates for examples of RDM (file naming, data back-ups, on-line data repositories, data sharing, etc.) that they are doing.
- How can you apply any aspect of RDM from these lectures to your research project?
- What aspects of RDM do you feel you need to gain more knowledge about in order to apply it to your research project?

While this approach is more qualitative than practical, hands-on examples, it should prod the students to think about their RDM, take charge of the RDM for their project, and to determine what aspects of RDM they will need to self-educate themselves on. Finally, the third item to ask lab mates for examples of their RDM may partially address the student feedback to discuss more aspects and examples of day-to-day laboratory RDM needs. Student feedback indicated that this assignment would be useful and appropriate for the course.

Based on student feedback and instructor self-reflection, the incorporation of the three lectures on RDM into the existing Theory and Methods course met its goals of exposing the students to RDM and prompting them to start examining their own RDM protocols.

Informal student feedback was obtained from several students completing the course. Students indicated that this course was an appropriate place for the RDM material, that they liked the broad exposure to RDM, and that they were fine with only three lectures because of the need to cover other topics in the course. All students asked indicated that this small exposure to RDM has prompted them to ask further questions and examine their own RDM such use of a laboratory notebook, examining data backup, and file naming conventions. One noted that despite the broadness of the lectures, when students asked questions in class the instructor was able to provide them with additional information and resources. Several followed this up by stating a desire for additional data management education through more frequent offerings of the semester long RDM course or other seminars, workshops, and on-line offerings. Many students also wanted additional details, instruction, and examples of day-to-day RDM needs on naming conventions (for example). Several students indicated that based on these lectures they had implemented data back-up (both electronic and research notebooks) for their research data.

CONCLUSIONS

Three lectures on Research Data Management (RDM) were incorporated into an existing required graduate course on Theory and Methods of Research. Since this existing course includes a number of other topics designed to improve the research skills of the graduate

students, this course is a strong fit for the RDM topics. The three lectures were on: (1) RDM and Sharing Mandates, (2) What is Data? and the Data Lifecycle, and (3) Planning for Your Research Project and DMPTool. In this way the students were exposed to broad, high-level concepts such as sharing mandates and the data lifecycle while also getting practical, low-level instruction on file naming, meta-data, and preparing a funding DMP. Future offerings of the Theory and Methods course with RDM will incorporate a student assignment to investigate their labs current RDM practices and start considering how to apply RDM to their research.

ACKNOWLEDGEMENTS

The author would like to acknowledge Larry Schmidt from the University of Wyoming Libraries for his efforts to develop and co-teach the semester long RDM course. Lecture material from that course was subsequently modified for this course.

REFERENCES

1. National Academy of Sciences, Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age* 2009, Washington, D.C.: The National Academy Press.
2. Carlson, J., M. Frosmire, C.C. Miller, and M. Sapp Nelson, Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Portal: Libraries and the Academy*, 2011. 11(2): p. 629-657.
3. National Science Foundation, Proposal and Award Policies And Procedures Guide, 2019.
4. National Institutes of Health, NIH Data Sharing Policy. 2017 April 17, 2007 [cited 2017 March 16, 2017]; Available from: https://grants.nih.gov/grants/policy/data_sharing/.
5. United States Geologic Survey, USGS Data Management: Data Management Plans. 2017 March 7, 2017 [cited 2017 March 16, 2017]; Available from: <https://www2.usgs.gov/datamanagement/plan/dmplans.php>.
6. Carlson, J., L. Johnston, B. Westra, and M. Nichols, Developing an Approach for Data Management Education: A Report from the Data Information Literacy Project. *The International Journal of Digital Curation*, 2013. 8(1): p. 204-207.
7. Muilenburg, J., M. Lebow, and J. Rich, Lessons Learned from a Research Data Management Pilot Course at an Academic Library. *Journal of eScience Librarianship*, 2014. 3(1): p. e1058.
8. Emory University, DMPTool can make applying for grants easier. 2018 [cited 2018 3/14/2018]; Available from: http://news.emory.edu/stories/2012/10/er_take_note_dmptool_class/campus.html.
9. San Jose State University, Writing Data Management Plans. 2018 [cited 2018 3/14/2018]; Available from: https://events.sjsu.edu/EventList.aspx?fromdate=2/28/2018&todate=2/28/2018&display=Day&type=public&eventidn=24591&view=EventDetails&information_id=40849.
10. The University of Oklahoma Libraries, Introduction to Data Management and Data Management Planning Fall 2015. 2015 [cited 2018 3/14/2018]; Available from:

- <https://libraries.ou.edu/content/introduction-data-management-and-data-management-planning-fall-2015>.
11. University of South Carolina Libraries, Library Workshops: Data Management Planning and DMPTool. 2018 [cited 2018 3/14/2018]; Available from: <http://guides.library.sc.edu/c.php?g=410310&p=5036248>.
 12. Thielen, J., S.M. Samuel, J. Carlson, and M. Moldwin, Developing and Teaching a Two-Credit Data Management Course for Graduate Students in Climate and Space Science. *Issues in Science and Technology Librarianship*, 2017. 86(Spring).
 13. Wright, S.J. and C. Andrews, Natural Resources / Graduate Students / Wright & Andrews / Cornell University / 2013. *Data Information Literacy Case Study Directory*, 2015. 2(1): p. Article 1.
 14. Whitmire, A.L., Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned. *Journal of Librarianship and Scholarly Communication*, 2015. 3(2): p. eP1246.
 15. Schmidt, L.O. and J.H. Holles, A Graduate Class in Research Data Management. *Chemical Engineering Education*, 2018. 52(1): p. 52-59.
 16. Holles, J.H. and L.O. Schmidt, Implementing a Graduate Class in Research Data Management for Science/Engineering Students. in 2018 ASEE Annual Conference & Exposition. 2018. Salt Lake City, UT.
 17. Schmidt, L.O. and J.H. Holles. Teaching Research Data Management; It takes a team to do it right! in 2018 ASEE Annual Conference & Exposition. 2018. Salt Lake City, UT.
 18. Holles, J.H., A Graduate Course in Theory and Methods of Research. *Chemical Engineering Education*, 2007. 41(4): p. 226-232.
 19. Phillips, G.A., R.S. Miller, and C.F. Miller. Understading Gradaute Students' Knowledge About Research Data Management: Workflows, Challenges, and the Role of the Library. in *Proceedings of the Association of College and Research Librarians Conference*. 2019.
 20. University of California Curation Center, Data Management Plan Tool. 2014; Available from: <https://dmptool.org/>.
 21. Holles, J.H. and L.O. Schmidt. Graduate Research Data Management Course Content: Teaching the Data Management Plan (DMP). in 2018 ASEE Annual Conference & exposition. 2018. Salt Lake City, UT.