



## **Graduate Research Data Management Course Content: Teaching the Data Management Plan (DMP)**

**Dr. Joseph H. Holles, University of Wyoming**

Associate Professor, Department of Chemical Engineering

**Mr. Larry Schmidt, University of Wyoming**

Larry Schmidt is an associate librarian at the University of Wyoming and is the current Head of the Brinkerhoff Geology Library. He holds BS degree's in Chemistry and Biology, MS Degree in environmental engineering from Montana State University and received an MLS from Emporia State University in 2002. His interests lie in providing undergraduate and graduate students with information, data and science literacy skills that will allow them to succeed in a global economy.

Graduate Research Data Management Course Content: Teaching the Data  
Management Plan (DMP)

## **Introduction:**

The need for effective education of graduate students in the area of Research Data Management (RDM) has been demonstrated through the number and types of recently developed courses on this topic.[1-6] These courses tend to take one of two general forms: 1) a full-term, for-credit standalone course and 2) a workshop/seminar approach. However in both forms of the RDM course, one topic that is common to both approaches is the Data Management Plan (DMP). The DMP typically addresses the following major topics: 1) Data that will be created, 2) Data documentation and organization, 3) Data storage and security, 4) Data management and preservation after project completion, and 5) Data accessibility for reuse and sharing. These RDM courses typically include a student assignment or term project to develop a DMP. This DMP assignment is usually focused on the graduate student's research project with a goal of helping the student manage their research data.

While the broad array of RDM education approaches and courses from the literature have focused on the consensus of teaching graduate students how to write a DMP, there is not consensus in the RDM courses discussed in the literature over exactly what the DMP is, how it is used, and what DMP education should include. For example, one common DMP is a short, high level 1-2 page document prepared by the primary investigator (PI) as part of grant request to a federal funding agency. In contrast, a second type of DMP is a larger, more comprehensive, multi-page document prepared as part of a research project and intended as a working document to guide the researcher and research team through all aspects of their research data management. In addition, a Data Curation Profile (DCP) is a document used as a tool to develop the larger DMP, while DMPtool [8] is a different tool used to develop the shorter DMP. These very different types of DMPs and similar sounding acronyms (DMP and DCP) then lead to the question: what is the difference in these DMPs, what type of DMP is being taught in the literature courses on RDM, how does the DCP fit in, and what should be covered in the DMP topic of a RDM course?

The focus of this work will be an examination of the different types of DMPs and how and what should be taught in a graduate RDM course. Examples of how RDM courses from the literature teach DMPs will be compared and contrasted. The differences between the two types of DMPs as well as the reasons behind them will be delineated and discussed. Terminology to differentiate the DMPs ("funding DMP" and "project DMP") will be developed and presented. Tools to prepare both types of DMPs will be discussed. The approaches and outcomes of teaching the two types of DMPs in our course will also be described and compared with the literature. Results and lessons learned from this approach will be discussed.

## **Methods:**

A RDM graduate course was developed and co-taught by a librarian and faculty member with an active research program. The course was designed to provide students with broad concepts and best practices of RDM and also to provide the students a focused application of RDM to active research projects.[2, 7] Full details of the course can be found in the references. The course also included assignments on both types of DMPs: an individual assignment to develop a funding agency type DMP using DMPtool.com[8] and an end-of-semester, team project to develop a project based DMP for an active research project using the Purdue Data Curation Profiles toolkit[9] (a DCP) as a tool.

## Results:

Early in the course, DMPtool was used by the students to develop a short (2-page) DMP for their research project. DMPtool is designed to help the researcher prepare the short DMP required as part of a funding proposal submission. DMPtool allows the user to select the desired funding agency and then requests information for the DMP consistent with that agency's requirements. For this assignment, the students selected the agency that was providing their funding or selected NSF-Engineering if university funded. The goal of this assignment was to provide an initial, high-level exposure of the students to DMPs. Consistent with funding agency space limitations, these DMPs ranged in length from 1 to 2 pages.

The second DMP application involved the development of a Data Curation Profile (DCP) and the application of the DCP to develop a full length DMP for an active research program. The purpose of a DCP is to "provide a foundational base of information about a particular data set that may be curated..." and is intended to "address the needs of an individual researcher or research group with regards to the 'primary' data generated or used for a particular project." [9] In this way, the DCP serves as a tool to gather and collate information about research data which then allows subsequent development of a DMP. For this assignment, the class developed a DCP based on the Data Curation Profiles toolkit from Purdue University. [9-12] Once developed, the DCP was applied to an ongoing research activity as the Final Project for the course. For the Final Project, the class was divided into teams of two or three students. Each team then worked with a research active faculty member. For the project, the team then prepared for the interview, interviewed the faculty member, and developed a DMP with best practices for the faculty member/research project. The research faculty were volunteers and chosen broadly from across campus. Two faculty were selected based on a desire to develop DMPs for their labs while the other two were selected due to known specific considerations about their research data. DMPs resulting from this final project were up to 10 pages in length. The goal of the faculty interview and the DMP development was to become familiar with data management issues, standards, and best practices in the researcher's field. Each team then delivered a class presentation on their final project. This allowed the issues, standards, and best practices for each research project to be further shared by the entire class. Following each presentation, each student also completed a reflection assignment to consider how the issues, standards, and best practices from the presentation could be applied to their own research project.

In order to clarify the confusion associated with DMPs, we propose the following terminology. A "funding DMP" is the short, often no more than 2 pages, document prepared for and submitted with a funding proposal to, for example, a federal agency. This is usually a broad, high level document outlining the data management that will be required and accomplished if the project receives funding. The funding DMP, for example, may specify that research results will be periodically submitted to an on-line archive without providing any additional detail. In contrast, a "project DMP" is a longer document prepared for and used in conjunction with a functioning research project. This is usually an in-depth and specific document associated with the day-to-day data management of a research project. The project DMP may, for example, specify exactly what data is to be archived, the format for archiving, and the schedule for archiving the data.

## Literature Comparison:

While our course included class discussion and assignments on both the funding DMP and the project DMP, other approaches to teaching DMPs in the classroom, as discussed in the literature, have taken a variety of approaches.

Fong and Wang developed a Data Profile Form (in essence a DCP).[3] This form was adapted from the Purdue DCP based on discussions between the graduate program director, a faculty member, and the authors. Graduate students were expected to complete this form based on their current research project. This student assignment is an example of the project DMP. The authors also offered additional assistance in creating DMPs, and several faculty and students received assistance with funding DMPs. However, funding DMPs were not covered in the classroom.

Thielen et al. developed and taught a for credit course on Data Management for graduate students in climate and space sciences.[5] For this course, students had weekly assignments to write a section of a DMP based on that week's classroom activities. The final project then consisted of combining these sections in a comprehensive DMP. This student assignment is an example of the project DMP. The authors note that: 1) the DMP "was not tied to any funding agency requirement", and 2) the assignment may be revised in the future and limit the DMP to two pages to "more accurately reflect federal funding agency requirements".[5] Thus, the authors acknowledge the funding DMP (i.e., a desire to limit the DMP to 2 pages) and that it is worth teaching this concept to students. However, they don't clearly distinguish between the two types of DMPs and the need to educate students on both types of DMPs.

Wright and Andrews delivered a for-credit course on data management for graduate students in natural resources.[4] As part of this course, students evaluated a DMP to observe successful examples. Students also created a DMP to 1) "manage and curate their own data, and 2) to meet funding agency requirements".[4] For this course, there does not appear to be a distinction between the two types of DMPs. While there is a recognition of the need for DMPs to meet funding agency requirements, the DMP discussion in the course appears to focus on the project DMP.[4]

Whitmire taught a for-credit graduate course in Research Data Management.[1] This course included a mid-term project using a scaled back version of the DCP with the student interviewing an active researcher to gain insights into examples and RDM standards in that field. In addition, the "final exam" for this course was for the students to create a DMP for their research project. Thus, this course was focused on the project DMP and there does not appear to be any discussion of the funding DMP.

Qin and D'Ignazio delivered a Science Data Management course.[13] For this course, the students were paired and then developed a DMP based on a subject scientist's research and work practices. Based on the development of the DMP from interviews with the subject scientist, this DMP also appears to be a project DMP.

Using a modularized e-learning format, Johnston and Jeffryes instructed graduate students in a Data Management Course.[6] As a component of this course, each student wrote a DMP focused on the student's data using a template (in essence a DCP) developed by the

instructors. The DMP template was aligned with the modules used for course delivery. Based on the template, this DMP meets the definition of a project DMP.

A focus on using DMPtool in the classroom, and correspondingly teaching the funding DMP seems to be most commonly delivered through the university library. For example, Muilenberg et al. (all librarians) developed and offered a modular 7 week course on RDM.[14] While the course covered the standard high level RDM topics (from Briney[15], for example), they also had hands-on time for the students to explore DMPtool. However, this course did not provide direct instruction on creating either a funding or project DMP. Similar short educational opportunities are commonly offered by university libraries as workshops/talks/presentations and focus on using DMPtool to generate funding DMPs.[16-19]

Additionally, even outside of the research data management education literature, there is some confusion over what a DMP entails. McClure et al. conducted a study of researcher practices and needs in the development of DMPs.[20] The authors investigated the feasibility of adapting the Purdue DCP protocol[9] for use in helping researchers develop DMPs. Responses to this study revealed that researchers had varied perspectives on what a DMP entails and whether it was a formal plan (funding DMP) or may also include procedural workflows embedded in their current research projects (a project DMP).[20]

### **Discussion:**

As described above, there have been numerous efforts in the literature to teach graduate students how to develop DMPs associated with the student's research project. Developing DMPs will also be required in the student's future professional careers. However, there is not a clear consensus in these literature RDM courses on what exactly a DMP is and how it is used. As a result, there is not clarity on what to teach graduate students in a RDM course. Thus, some works focus on the project DMP while other works focus on the funding DMP.

In contrast with the previous efforts, the course described by the authors[2] clearly delineates between the funding DMP and the project DMP. To complement this distinction, student assignments are associated with each type of DMP. The use of DMPtool[8] allowed the students to learn the topic of funding DMPs. Thus, DMPtool is a tool to develop a "funding DMP". In addition, the final project involved an interview of an active researcher using a DCP followed by development of a DMP subsequently delivered to the researcher that covers the topic of project DMP. Thus a DCP is a tool to develop a "project DMP". This approach allows the students to clearly differentiate between the two types of DMPs. It also allows the students practical experience generating both types of DMPs. Since many graduate students will become the next generation of PIs, experience with this tool will help them in their future careers. Examination of the Partners list at DMPtool[8] indicates in excess of 200 institutions that provide their researchers access to this tool for developing funding DMPs. Therefore, this tool has become broadly recognized and student exposure should be considered a standard experience.

### **Conclusion:**

Literature reports on instruction of graduate students in the preparation of data management plans (DMPs) were reviewed and compared. While both the "funding DMP" and "project DMP" documents have been taught in these courses, there is minimal discussion about

the differences between the two documents and the needs that they serve. In fact, most courses reviewed from the literature teach only one or the other type of DMP. In our recently developed research data management course for graduate students, students were educated on both types of DMPs. The funding DMP is a short, high level document associated with a funding proposal. A project DMP is a longer, more in-depth document associated with the day-to-day management of an ongoing research project. DMPtool serves as a tool to help the researcher prepare a funding DMP while the Purdue DCP serves as a tool to help the researcher or graduate student develop a project DMP. The goal of the DCP and DMPtool is to assist the researchers, librarians, and other stakeholders in the preparation of the appropriate project or funding DMP.

## References:

1. 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.
2. Schmidt, L.O. and J.H. Holles, *A Graduate Class in Research Data Management*. Chemical Engineering Education, 2018. **52**(1): p. 52-59.
3. Fong, B.L. and M. Wang, *Required Data Management Training for Graduate Students in an Earth and Environmental Sciences Department*. Journal of eScience Librarianship, 2015. **4**(1): p. e1067.
4. 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.
5. 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).
6. Johnston, L. and J. Jeffryes, *Civil Engineering/ Graduate Students/ Johnston & Jeffryes/ University of Minnesota/ 2012*. Data Information Literacy Case Study Directory, 2015. **3**(1): p. Article 1.
7. Holles, J.H. and Schmidt L.O., *Implementing a Graduate Class in Research Data Management for Science/Engineering Students*. in *2018 ASEE Annual Conference & Exposition*. 2018. Salt Lake City, UT.
8. University of California Curation Center, *Data Management Plan Tool*. 2014; Available from: <https://dmptool.org/>.
9. Carlson, J., *The Data Curation Profiles Toolkit: User Guide*, in *Data Curation Profiles Toolkit*, 2010, available from: <http://dx.doi.org/10.5703/1288284315650>.
10. Carlson, J., *The Data Curation Profiles Toolkit: Interviewer's Manual*, in *Data Curation Profiles Toolkit*, 2010, available from: <http://dx.doi.org/10.5703/1288284315651>.
11. Carlson, J., *The Data Curation Profiles Toolkit: Interview Worksheet*, in *Data Curation Profiles Toolkit*, 2010, available from: <http://dx.doi.org/10.5703/1288284315652>.
12. Carlson, J., *The Data Curation Profiles Toolkit: The Profile Template*, in *Data Curation Profiles Toolkit*, 2010, available from: <http://dx.doi.org/10.5703/1288284315653>.
13. Qin, J. and J. D'Ignazio, *The Central Role of Metadata in a Science Data Literacy Course*. Journal of Library Metadata, 2010. **10**(2-3): p. 188-204.

14. 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.
15. Briney, K., *Data Management for Researchers: Organize, maintain and share your data for research success*. Research Skills Series. 2015, Exeter: Pelagic Publishing, UK.
16. 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](http://news.emory.edu/stories/2012/10/er_take_note_dmptool_class/campus.html).
17. 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](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).
18. The University of Oklahoma University 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>.
19. University of South Carolina University 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>.
20. McLure, M., A.V. Level, C.L. Crabston, B. Oehlerts, and M. Culbertson, *Data Curation: A Study of Researcher Practices and Needs*. *Portal: Libraries and the Academy*, 2014. **14**(2): p. 139-164.