

Is a Distributed Graduate Seminar a Successful Way to Share Information Management Expertise with Students?

Kristin Vanderbilt^{1,1}, Corinna Gries^{2,2}, Paul Hanson^{2,2}, Andrea Nocentini^{3,3}, and Jonathan Wheeler^{4,4}

¹University of New Mexico, Florida International University

²University of Wisconsin Madison

³Florida International University

⁴University of New Mexico

January 20, 2023

Abstract

In this era of open data and reproducible science, graduate students need to learn where and how to publish their data and to be conversant with the challenges inherent when re-using someone else's data. The Environmental Data Initiative partnered with UNM Libraries and Florida Coastal Everglades LTER to organize a 1-credit, semester-long distributed graduate seminar to learn if this approach could be an effective mechanism for transmitting such information. Each week during the Spring 2021 semester, an informatics specialist spoke remotely to students at University of New Mexico, Florida International University, and University of Wisconsin-Madison on topics ranging from FAIR principles to data security, team science to data provenance. Students prepared for the lecture with one or more readings, and in-class exercises reinforced the material covered. Student assignments included writing quality metadata for their own data and archiving their data in the EDI Repository. The capstone writing assignment, a data management plan for their own research project, allowed the students to integrate much of what they had learned. Student response to this class was positive, and students indicated that they learned a lot of immediately useful information without the course being a significant time-sink. The low registration numbers at UNM and FIU (6 and 7 students, respectively), however, where the seminar was not required, suggest a need to better inform both students and their advisors of the opportunity and the value provided by the training. Instructors also learned that it would be easier to create a cohesive flow to the course, without repetition, if the group of instructors took turns lecturing, rather than bringing in specialists on each subject. It was also apparent from student comments that many felt this information should be integrated, at an introductory level, into undergraduate classes or classes for new graduate students.



Is a Distributed Graduate Seminar a Successful Way to Share Information Management Expertise with Students?

Jonathan Wheeler

University of New Mexico



Motivation

Planning for this course began in the spring of 2020, and was initially motivated by a need for graduate level data management education. Graduate students across disciplines share a recognized need for domain specific instruction in the collection, management, and ethical sharing of research data. Points from recent research include:

- 80% of working with data is pre-processing, wrangling, cleaning, etc., yet few graduate courses teach these skills (Howe et al. 2017),
- In a survey of 131 science grad students about their knowledge of research data management in science, students were least knowledgeable and skilled in data curation and data re-use, metadata,.... and data preservation (Pasek and Mayer 2019).

In our experience, many grad students are also unaware of funder and publisher policies requiring data publication to support the FAIR principles. Further, as some faculty don't have data management skills themselves, a distributed grad seminar is a mechanism to leverage expertise at other universities in those situations.

The figure below (Benedict, 2021) illustrates the diversity of concepts and skills at the intersection of the research and data lifecycles. We designed the course as an integration of common components of data lifecycle models with general data management skills and hands-on practice using ecological data, metadata standards, and repositories.



Administering a Distributed Course

The distributed nature of the course provided an opportunity to engage a group of students with varied and broad interests across the domain of ecology, many of whom work at Long Term Ecological Research (LTER) sites in very different ecosystems. The three universities that piloted the course are:

- University of New Mexico
- Florida International University
- University of Wisconsin, Madison

Regional LTER sites were the Sevilleta in New Mexico, the Florida Coastal Everglades, and the North Temperate Lakes in Wisconsin.

The course was an opportunity to illustrate and apply generalizable data management concepts and skills that could be applied to data of different types and which describe different systems. The distributed nature of the course presented some administrative challenges in terms of aligning the coursework with different requirements around in-class hours, grading requirements, and enrollments at the three institutions:

University of New Mexico (6 enrolled, 26 in-class hours)

Florida International University (7 enrolled, 26 in-class hours)

University of Wisconsin-Madison (35 enrolled, 13 in-class hours)

These issues were addressed through collaborative development and refinement of the syllabus and assignments through bi-monthly meetings. Coursework was evaluated by each instructor according to the standards in place at their respective institutions, with procedures in place to protect student privacy in terms of feedback, grades, and course evaluation.

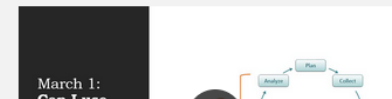
Topics & Assignments

Lecture topics and assignments were organized to align with models of the data lifecycle, allowing students to iteratively build upon foundational concepts and skills in research data management. Students came to the course with varying levels of exposure to and practice with data management concepts and skills, so discussions and assignments were scaffolded to provide more advanced students with opportunities for deeper engagement.

Lectures were delivered by experts in specific areas of ecological data management. Whenever possible, invited speakers participated in the instructors' planning meetings to coordinate readings and assignments.

Hands-on components were platform and application agnostic, and began with the incorporation of Data Carpentry modules focused on tabular data management and an overview of quantitative analytic processes. Other lectures were more conceptual. A complete list of topics and presenters is shown below. Students left the course having had the experience of writing metadata, QA/QCing data, and uploading data to the EDI Data Repository. They also learned to write a data management plan to accompany a proposal or for personal reference.

The video below was shared with students ahead of the course to provide an overview of the seminar.



Assessment & Next Steps

In addition to assessments provided through assignments and course evaluations, students were invited to provide feedback on the course at various points. Student feedback was generally positive:

- Students appreciated learning from experts in the field, which helped make the content more immediately relevant.
- Exercises and hands-on activities during the second hour were engaging and provided structure for further exploring and engaging with lecture topics.
- Having a new guest lecturer most weeks resulted in both some content overlap as well as a need for establishing connections between concepts.

Based on feedback from students, invited speakers, and discussions during the bimonthly debrief among instructors, future iterations of the course will

- Increase continuity and reduce overlap or repetition of concepts by consolidating lectures among instructors, with fewer invited guest lectures
- Integrate hands-on activities into lectures to align data management concepts with concrete skills.