

**EScience in Action****Dinner and Data Management: Engaging undergraduates in research data management topics outside of the curriculum**

Cameron Cook, Tobin Magle, Heather Shimon, and Trisha Adamus

University of Wisconsin-Madison, Madison, WI, USA

Abstract

Researchers are faced with unprecedented challenges due to the size and complexity of data, and libraries are stepping in to help by providing guidance on research data management primarily to graduate students and faculty. Currently, many universities are encouraging an undergraduate research experience where students engage in research projects in the classroom and in research labs, yet research data management is often not included as part of these opportunities. At UW-Madison, we piloted researchERS (Emerging Research Scholars), a program for undergraduates from all disciplines to learn data management skills. Focusing on core concepts as well as data ethics, reproducibility, and research workflows, the format of the program included seven evening workshops, two networking events, and one field trip. Each workshop invited campus and community speakers relevant to the workshop's theme as a way to introduce the students to the network of available resources and data expertise and provided food for attendees. The workshops also built in customized activities to show students how to incorporate best practices into their work. Local businesses provided a tour of their facilities as well as a talk on how they leverage data. This paper will describe this program as well as the benefits and drawbacks of tailoring a research data management program toward undergraduates.

Correspondence: Cameron Cook: cameron.cook@wisc.edu**Received:** May 8, 2020 **Accepted:** July 22, 2020 **Published:** September 10, 2020**Copyright:** © 2020 Cook, Magle, Shimon, and Adamus. This is an open access article licensed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/).**Disclosures:** The authors report no conflict of interest.

Introduction

Researchers today are faced with unprecedented data challenges that will only grow in scope and scale in the coming years. Managing born digital data and storing and transferring large files, combined with changing funding and publisher requirements, necessitate an entirely new research skill set than ever before. Graduate students and faculty have been well supported at University of Wisconsin-Madison in the research data management realm through the work of Research Data Services (RDS). RDS is a collaborative enterprise among employees at the libraries, information technology, and the graduate school who respond to research data management questions, offer consultative services, and teach workshops on data management topics. Historically, undergraduate researchers have not been a sustained focus of RDS efforts; however, we have observed an increase in the number of undergraduates pursuing research even as early as their sophomore and junior year. We have also seen that this increase has not been accompanied by a proportional increase in research data management education in the undergraduate curriculum. This paper will describe the UW-Madison Libraries researchERS (Emerging Research Scholars) program as well as the benefits and drawbacks of tailoring a research data management program for undergraduates.

Developing researchERS

Our first effort at targeting undergraduates was piloting an embedded introductory undergraduate data management curriculum in 2015-2017 in Biocore, the undergraduate honors biology program at UW-Madison. The instructors of Biocore were interested in planting the seed of research data management as a lifelong research skill and felt the format of the program, through which students conduct unique research in small groups, lent itself to incorporating the skills. During the pilot, each semester RDS would give an introductory data management talk to the students early in the semester while students were designing their research projects. Students were then asked to include a short data management plan as part of their research proposal, with RDS providing in-class feedback. At the end of their projects, a short survey was included as part of the course surveys, to assess whether the assignments were effective and students felt that the skills learned were relevant to their work. This pilot continued over the course of a year and a half, but ultimately found that it was hard to pinpoint an appropriate point in the student's progress through Biocore in which to fit the curriculum. Second semester students found that they did not have enough unique data yet for the content to feel deeply relevant, while third semester students really enjoyed the content and asked that it be introduced earlier in the program. This pilot was also deeply dependent on the willingness of the instructor to give large amounts of class time, which was hard to sustain over time.

In exploring other avenues for providing data management instruction to undergraduates, we found that a credit course was not possible. Instead, RDS focused on planning a series of one-shot workshops on specific data management topics. Based on this decision, we applied for an innovation grant through the UW-Madison Libraries to fund a research data management program called researchERS (Emerging Research Scholars) targeted toward undergraduates across the campus community. Without the structure of a credit course, the workshop series was created with the understanding that each session would need

to stand alone. We could not require students to attend, so we were not able to scaffold learning across workshops. However, we also had students who attended multiple workshops, and communicated an interest in continued learning, so we created a different series of workshops in the spring from the previous fall. As a way to encourage attendance, we scheduled the workshops in the evening and provided dinner.

Our primary goals in developing the researchERS program were to determine whether undergraduate students were interested in research data management and, as in the Biocore course, to plant the seed of data management as a research skill.

We developed three general learning objectives that we wanted all of our sessions to address: understanding that data has value and requires diligent organization and management; recognizing that data collection, use, and sharing include ethical, social, and legal considerations; and developing confidence in finding resources to help manage data. In planning the program, we were mindful of choosing topics that would resonate with undergraduates and fill gaps in their research experiences in courses or as assistants in research labs. The program drew upon existing work and resources from both the data librarian and teaching and learning communities that had been used as models for prior RDS workshops. To narrow down topics, we reviewed the DataOne modules, which we had often incorporated into previous RDS instruction. To set the learning objectives for each individual workshop, we consulted what we considered to be the seminal work in this area: Sapp-Nelson's "Data Management Skills Competency Matrix" (2017). For our Research Reusability and Reproducibility workshop, we adapted the LEGO reproducibility activity from Pullman and Zilinski's "A Multi-Framework Approach to Teaching Data: A Case Study in Modern Languages" (2016), which we had also already used in a previous RDS presentation. We also identified experts on campus and in the community who would be able to present on the workshop topics we selected. These included campus librarians, professors, and staff, as well as data professionals from local businesses, public libraries, and government agencies. We also reached out to local businesses to see if they would host a field trip for students to observe data management and use in a workplace setting.

After securing funding, we ran the researchERS program during Fall 2018 and Spring 2019. Participants in the program were asked to enroll through a course management system, which we used to communicate with participants and store workshop materials. In total, the researchERS program included seven evening workshops, two networking events, and one field trip. Most of these sessions were held in the BioCommons, a space housed in Steenbock Library focused on bringing together bioscience students from across different disciplines. The space is a partnership between the library and the Educational Innovation initiative of WISCIENCE, and hosts both social and academic events throughout the year. The program funding allowed us to provide dinner at all of the events held in the BioCommons and for the field trip transportation.

*Workshop sessions (in chronological order)***Fall 2018**Data in the World: Impact & Ethics | *Attendees: 55*

In the first workshop of the program, we introduced students to the real-world impact of using data and a primer on the ethics involved. The City of Madison Data Projects Coordinator presented on community's usage of the city's data portal. We followed her presentation with the video "The Coded Gaze" by the Algorithmic Justice League and a discussion on data bias.

Data Discovery, Acquisition & Citation | *Attendees: 18*

This workshop provided an introduction to finding datasets and what to do with them once you've found them, including processing the data and properly giving credit. We invited campus librarians to speak about data citation and GeoData@Wisconsin, the state geoportal. Students then participated in a hands-on data management activity, cycling through five stations that covered data storage, data organization, dataset searching, discipline specific repository searching, and data citation.

Data Visualization | *Attendees: 11*

UW-Madison DesignLab (designlab.wisc.edu) consultants showed students how to choose the right chart or visualization for their data. The consultants also covered general best practices for conveying information visually and then ran an activity where the students had to interpret and critique existing visualizations. This workshop was especially well received by the students.

Data Story Slam | *Attendees: 39 including non-student campus staff (10 students)*

Presenters were invited to share their experiences with data in the form of five minute, free-form short stories in this story slam style session. Presenters talked about obsolete data formats, challenges they face working with their data, and more. This event was open to the campus community.

Epic Field Trip | *Attendees: 25*

Epic, an electronic medical record software company in Verona, WI, is one of the largest employers in Madison. On this field trip, participants toured the Epic campus' many themed buildings, networked with Epic staff about internships or job opportunities, and learned about how data and data science is being leveraged to improve health outcomes at Epic. This event was also open to students who were not enrolled in the researchERS program.

Spring 2019

Collecting Data & Quality Control | *Attendees: 38*

In the first workshop of the spring semester, we covered the best practices of collecting data in lab notebooks, organizing spreadsheets, and better ways to wrangle data. The students participated in a hands on activity of data collection and cleaning.

Research Reusability and Reproducibility | *Attendees: 15*

Karl Broman, data scientist and Professor in the Department of Biostatistics & Medical Informatics spoke about the importance of research reproducibility and his experience advocating for it. We also covered steps to incorporate reproducibility into the research workflow through the adapted LEGO activity from Pullman and Zilinski's "A Multi-Framework Approach to Teaching Data: A Case Study in Modern Languages" (2016).

Data and Your Research Question | *Attendees: 16*

In this workshop, we examined the anatomy of a good research question and identified ways to ensure that students have the right data to answer their research questions. The Data Science Facilitator from the UW-Madison Data Science Hub ran an activity in which the students had to answer research questions using an example dataset.

Data Workflows: How to Manage Big Data | *Attendees: 10*

Staff from the Center for High Throughput Computing (CHTC), ran an activity with students to demonstrate how high throughput computing works and then spoke to students about how to leverage the computing infrastructure at UW-Madison. After the talk, we provided a brief introduction to considerations for managing big data.

Local Business Data Day | *Attendees: 34*

We invited five guests from local organizations to speak about how they work with data in their day-to-day activities, and trends they are seeing in their industries. Speakers were from the following organizations: American Family Insurance, IBM, Illumina, Mirus Bio LLC, and South Central Library System. This event was also open to students who were not enrolled in the researchERS program.

In total, 179 students enrolled in the program using the course management system. Of those enrolled, 95 unique students attended at least 1 workshop, and 35 students attended 2 or more workshops. Before the first workshop, we surveyed students enrolled in the program to benchmark their experiences with data, their majors, and current year in school. Students were also encouraged to introduce themselves in the course management system discussion section where they talked about their research experiences and interest in the program. The

majority of students who attended were from STEM disciplines and wanted to learn about data collection and computational tools. We used this feedback to choose topics for the spring semester workshops, since we had already planned the fall workshops before we had heard from the students. At the end of each session, participants filled out evaluations that gauged overall satisfaction with the workshop, whether the stated learning objectives were met, and whether the material presented would help them in their research. The evaluations indicated that respondents were satisfied with the quality of the workshops in general and that the stated learning objectives were met. End-of-semester surveys were distributed to determine barriers to attendance, what students valued most in the program, and if this was connected to their current experiences with data. An end-of-the-year survey was sent to students who attended two or more workshops to see how the program influenced their data practices. Students indicated that the largest barrier to attendance was scheduling conflicts. End-of-the-year surveys indicated that the students found that the workshop had made them interrogate their current data practices, and that they wanted deeper learning that was discipline specific.

Discussion and Lessons Learned

One of the biggest lessons learned from the researchERS program was that undergraduates are actively interested in research data management, based on workshop attendance and satisfaction with the workshop as a whole. In feedback received on the program, one comment that we received from an attendee encapsulated the impact that we had:

"I used to think of data as something that would just belong to me or a small group of people I knew personally. The researchERS events have helped me see that data is something that is commonly shared among different groups of people around the world, and it's important to think about the way I handle data when sharing it with people. I'll try to be more organized and follow a set format when collecting my data from now on!"

Another lesson is that a program of this nature is hard to sustain and faces many logistical challenges. Attendance waxed and waned throughout the semester, despite the large number of undergraduates who expressed interest by enrolling in the program. This indicates that while undergrads are invested in learning about data, they are busy and have many competing priorities, which made scheduling the sessions at times and locations that were convenient challenging. It also indicates that students need more structure and more tangible incentives to prioritize participation, such as extra credit, a certificate that would appear on their transcripts, or content tied to a class assignment.

Guest speakers from useful campus resources were often a draw for students to attend workshops. However, scheduling those guest speakers proved to create an issue for the flow of the series as a whole. We chose to provide a year-long series with different topics instead of repeating the same content in the fall and spring semesters to encourage repeat attendance from students. This proved challenging for us to create clear transitions between each session as the order of session content was largely determined by when guest speakers could attend. Thus, content was sometimes forced to jump between data management concepts rather than building subsequently on one another.

This program also had challenges that are common to research data management initiatives in general. First, the phrase “research data management” is interpreted differently depending on the user community, and early on some students indicated that they had enrolled expecting the topics to be focused on data analysis and specific analysis tools. Furthermore, we received some feedback that the content was “too general,” which is also common feedback heard with workshops that are geared toward graduate students and faculty, due to the diverse nature of research data and the inability to provide attendees with one-size-fits-all solutions outside a classroom setting.

Future Directions

Though the researchERS program was valuable for both the attendees and the organizers, we will not be continuing the program in its current form due to the one-time funding source and lack of success obtaining new funding for the recurring costs of these workshops for items such as food and materials. The costs to run the program and maintain interest from students are considerable, making it difficult to sustain without commitment from the Libraries or other partner organizations long-term.

The program also had unexpected future benefits for RDS consultants. Because of the outreach needed to recruit guest speakers and the buzz that was generated by the program, we were able to establish relationships with many organizations in the campus community that will benefit us over time. For example, this program strengthened our relationship with CHTC and the Data Science Hub by connecting them with undergraduates, a population that they typically don’t interact with but are continually interested in engaging. We have also developed industry partnerships with companies like Epic Systems that can be leveraged for future data-related events.

We will be carrying the lessons learned from this program to the future in several ways. The content as developed is modular and allows for easy reuse across contexts. We plan on repurposing the content and activities into individual one-off workshops for researchers and students and also see opportunity to use them to engage library staff for staff development workshops. We also see this content being useful in a potential new collaboration with the data science major at UW-Madison. Some components of the workshops have also already been incorporated into newly funded projects for a data management board game and a research data management escape room for the UW-Madison Libraries.

Conclusion

Undergraduates are an engaged community and a fantastically ripe opportunity for data management education; however, without options for embedded work with undergraduates, sustaining engagement is difficult. This suggests a couple of opportunities - first, leveraging partnership for more embedded work would allow for students to deepen their understanding and immediately implement what they are learning either in the classroom or in their assignments and second, students responded positively to the interactive and unique ways we provided to engage with these concepts and we can continue to build those types of activities in more modular ways to allow them to be run in a classroom setting or for students to run in their own time. This would help avoid conflicting with students’ many competing

priorities while still providing opportunities to learn about concepts that can provide immediate impact to their work. This second identified opportunity has helped shape our forthcoming projects, the data management escape room and data management board game, as previously mentioned. These projects provide physical, unique, and interactive ways to explore data management concepts and may be perceived as more 'fun' by students than typical workshop structures. We are hoping these projects can also be checked out of the library by a student group or instructor to be run in the classroom, with the option to have an RDS member provide wrap-up and context for the concepts demonstrated in the games.

Acknowledgments

The authors would like to acknowledge the University of Wisconsin-Madison Libraries for the funding to pilot the researchERS program.

References

"Education Modules | DataONE." n.d. Accessed July 29, 2019.
<https://www.dataone.org/education-modules>

Pullman, Ethan, and Lisa Zilinski. 2016. "A Multi-Framework Approach to Teaching Data: A Case Study in Modern Languages." *LOEX Quarterly* 44: 8-10.
<http://commons.emich.edu/loexquarterly/vol44/iss2/5>

Sapp Nelson, Megan R. 2017. "A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs." *Journal of eScience Librarianship* 6(1): e1096. <https://doi.org/10.7191/jeslib.2017.1096>