Planting the Seeds for Data Literacy: Lessons Learned from a Student-Centered Education Program

Jake Carlson
University of Michigan

Marianne Stowell Bracke
Purdue University

Abstract

There is an increasing need for graduate students to acquire competencies in managing and curating their data sets as a part of their education. Librarians and other information professionals are beginning to respond to this need by developing programming, but as of yet there are few models to follow and the impact on the practices of students is under-explored. This case study presents a student-centered pilot program on data literacy offered at Purdue University. The program was offered through the College of Agriculture and was structured to be flexible enough to incorporate each student's particular field of study. Exercises and assignments were designed to incorporate the student's own research data to create meaningful, authentic learning experiences. Formative and summative assessment was a critical component of the program, which included interviews with students six months after completion of the program to determine the extent to which the data competencies covered had taken root in students' research practices. The structure of the pilot program, its strengths and weakness, its impact on students, and lessons learned by the instructors are discussed.

Introduction

With researchers facing new requirements and expectations for managing, sharing and curating their data, it is critical that they have the knowledge and skills needed to respond effectively. However, competencies in working with data are often not included as a part of a student's formal education. Students that do acquire proficiencies with data generally gain their skills in an ad hoc manner on the job and at the point of need (Jahnke and Asher, 2012). The disconnect between the changing expectations surrounding research data and the capabilities of researchers to respond is a challenge to realizing the benefits of having ready access to curated and protected research data sets. Addressing this disconnect will require a workforce that is not only able to take advantage of the tools and resources available for analysing data but is capable of producing data sets that are well structured, described and documented so than others can identify, understand and make use of them.

What data competencies do graduate students need to be successful in their eventual careers? How can these competencies be taught to students effectively? This paper presents a case study about the experiences of ten graduate students from the College of Agriculture at Purdue University who enrolled in our semester-long data literacy program. The program was loosely structured to enable students to shape its direction. Instead of providing a static syllabus, we formed weekly lesson plans that were developed as the program progressed. This approach allowed us to create presentations, discussions and exercises that were responsive to student interests and needs.

An under-explored aspect of teaching data literacy is the lasting impact on students. If we seek to foster meaningful change in the cultures and practices surrounding data through data literacy programs then we need to understand how students perceive, process and apply what they are being taught. We employed several formative and summative assessment techniques into our program, including follow-up interviews with students six months after they had completed the program to gauge its long term impact. Our experience in crafting, teaching and assessing this program gave us insight into how graduate students navigate their roles as producers of data, the challenges they face, and how to make data literacy relevant to them.

Background

The lack of data education has been identified as a problem that information professionals are potentially well-suited to address (Haendel et al., 2012). Research initiatives such as the Data Management Skills Support Initiative (Molloy and Snow, 2012) and the Data Information Literacy Project (Carlson and Johnston, 2015) were launched to better understand and respond to the need for data education. Tools and resources are being developed to teach data competencies, such as the New England Collaborative Data Management Curriculum.[1] Additionally, many librarians are offering workshops (Eaker, 2014), online training programs,[2] or even credit bearing courses (Whitmire, 2013). These and other efforts have helped to identify what competencies should be taught and what approaches could be used in teaching them.

---

[1] New England Collaborative Data Management Curriculum: http://library.umassmed.edu/necdmc
[2] MANTRA: http://datalib.edina.ac.uk/mantra/

A challenge to teaching data literacy to graduate students is the need to make the program directly relevant to their environment and specific needs. Graduate students face potentially challenging circumstances as they transition from student to a professional in their field. As a student they assume a great deal of responsibility through taking on multiple roles: researcher, teacher, supervisor, author, data manager and so on, in which they may have little to no previous experience. Furthermore, graduate students often find themselves in stressful situations when carrying out these roles as they lack authority and social standing in the academy (Grady et al., 2014). Graduate students rely heavily on their faculty advisor, as well as other faculty in their department, to guide them through their roles and responsibilities (Austin, 2002). As interviews from the Data Information Literacy project[3] revealed, graduate students are expected to devote themselves to developing a considerable depth of knowledge in their field through intense and focused research, and to produce scholarly publications to serve as a demonstration of their expertise. Any deviation from this intense focus runs the risk of being perceived by the student and their advisor as irrelevant and an unnecessary distraction (Carlson et al., 2013).

In response, we undertook an authentic learning approach in developing our data literacy program. Authentic learning, as defined by Herrington and Oliver (2000), is comprised of nine essential elements that aim to:

1. Provide authentic contexts that reflect the way the knowledge will be used in real life,

2. Provide authentic activities,

3. Provide access to expert performances and the modelling of processes,

4. Provide multiple roles and perspectives,

5. Support collaborative construction of knowledge,

6. Promote reflection to enable abstractions to be formed,

7. Promote articulation to enable tacit knowledge to be made explicit,

8. Provide coaching and scaffolding by the teacher at critical times,

9. Provide for authentic assessment of learning within the tasks.

We believed that an effective way to teach students would be to have them apply the lessons to their own data directly. Class presentations, exercises, and discussions centered on their data and addressed their roles as data producers and managers. Using an authentic learning approach also enabled us to introduce higher level issues and perspectives on data, and to connect them with the students' current practices in their labs.

Program Structure

The Purdue Libraries and faculty in the College of Agriculture have worked together previously on data management and curation issues (Bracke, 2011; Carlson and Bracke, 2013) and as a result developed a solid relationship. The need for an educational

---

3 Data Information Literacy project: http://datainfolit.org

program on data was raised in a conversation with the Associate Dean of Research in the College of Agriculture who was very receptive to the idea. We agreed this initial offering would be run as a pilot with the intent of leveraging what was learned in teaching data literacy towards creating sustainable programming. Although librarians have faculty status at Purdue, we could not teach a for-credit course within the College of Agriculture. Instead, the College agreed to offer a $1,500 stipend to students to enroll in the data literacy pilot program as an incentive.

We sought to take full advantage of our pilot status by creating as many avenues to learn as much about our students, their environment, their data, and their needs as we could. To this end, we required interested students to fill out an application form to get into the program. We asked the applicants to provide us with some information about themselves, express their interest in the program and tell us about their work with research data. Applicants were also required to obtain their advisor's signature on their application indicating the advisor's approval of the student's enrollment. Having students complete an application form provided us with a head start on developing a relevant curriculum.

We received 18 applications for ten available slots. In addition to the quality of their responses, we sought to create a diverse classroom through selecting students from the different fields of the College of Agriculture, who were at different stages in seeking their degree (from first year MA to final year of PhD), some international students and a near equal number of male and female students. Our recruitment and application forms are included as Appendices.

Students in the program met weekly in two hour sessions over 15 weeks. Most of the sessions were comprised of presentations, discussions and exercises in roughly equal measure. Our initial framework for the program was loosely based on the Data Information Literacy competencies (Carlson et al., 2011). However, instead of creating a syllabus at the onset of the program we developed weekly lesson plans as the program progressed and distributed them a week in advance of the session. This flexible structure allowed us to incorporate students' areas of interest into the program as they learned more about data literacy topics and more readily respond to concepts that they found challenging. Our lesson plans and teaching materials are available online[4].

The final session topics were:

1. Introduction to the Data Literacy Pilot Program

2. Data Management Planning

3. Data Lifecycle Models

4. Discovery and Acquisition

5. Description and Metadata

6. Data Security and Storage Issues/Description & Metadata (Part 2)

7. Issues in Copyright and Licensing Data

8. Description and Metadata (Part 3)/Mid-semester Check-in

9. Data Sharing/File Naming Conventions

---

4   See: http://dx.doi.org/10.5703/1288284315518

10. Data Management and Documentation

11. Data Visualizations

12. Data Repositories

13. Data Preservation

14. Data Publication and Curation

15. Data Literacy Course Wrap-Up

We sought to create a community of peers comprised of students in related fields of study to facilitate learning. Although the College of Agriculture is comprised of a wide range of disciplines, such as Agronomy, Biochemistry and Agricultural Economics, we believed that having the shared base of Agriculture would help our students make connections on common data issues and to help each other address them. We also structured the program to encourage our students to reach out to their existing community of peers, their lab groups in particular. Students' work with data is heavily influenced by their environment, but they in turn can act as agents of influence with their advisors and peers.

Assessment

Ongoing assessment was a critical aspect of the course. The "just in time" approach we took to developing the weekly sessions depended on developing a robust framework for communicating with students and assessing their work. Furthermore, as this was a pilot we needed to gather a strong understanding of our students' needs, how these needs were met, or not, and what lasting impact our program had on student practices and behaviors. To this end we developed several formative and summative assessment techniques and incorporated them into the program.

Our formative assessments included "minute papers", brief statements from students describing what they had learned and what was still unclear, and assignments where students would demonstrate what they had learned. For example, for the final assignment we had students create a presentation about something that they had learned from our program and to design it to be delivered to their lab. Although we could not require students to actually give their presentation, several of them took the initiative and did so on their own.

Our summative assessment consisted of a focus group style discussion at the last class session and follow up interviews with students six months after the end of the program. In both the focus group and the follow up interviews we asked them to describe the strengths and the weaknesses of the program, their recommendations for improving the course, and how they applied what they learned into practice.

Strengths of the Program

Overall the program was well received by our students, as evidenced by their feedback in the minute papers and in the summative discussion at the final session. Despite not receiving course credit, all of our students participated fully throughout the semester.

Discussions were robust, with students often providing helpful feedback and encouragement to their peers. Homework assignments were generally completed on time and often exceeded instructors' expectations. Although students were financially compensated, it did not appear as though any of them were participating only to receive the stipend. Several of our students told us that this course had made an impact not just on how they worked with or approached their data, but in how they understood research more generally.

> 'This course really changed the way I think about research. It made me into a better scientist because now I have taken responsibility for my data management. And I've also come to appreciate metadata and organizing your data so that it's in line with the scientific method, so somebody could actually come and repeat your work and understand it. I consider that now to be an essential part of how I see science. So it's had a huge impact on me.'

Students were not confident with their knowledge and skills in data coming into the program, but having a peer group who expressed similar sentiments helped them feel more at ease. Students appreciated seeing and discussing the work of their peers as they found these examples useful in informing their own understanding of the material. Furthermore, the program's focus on students from agriculture disciplines enabled students to make connections to the work and data sets of others in the program. Though engaged in different research projects and working with different types of data, the commonalities between their disciplines helped students to relate to the experiences of their peers and learn from one another.

> 'I felt that today's session was an effective intro class. I liked seeing the average numbers from the survey [of students in the program] because now I don't feel so alone in my deficiency of data management.'

> 'It was good to see the types of information everyone included in their metadata records to get a better idea of what should be included.'

> 'I liked how integrated you were across fields… I saw similarities in other departments that I did not think about [previously].'

Students' primary motivation in enrolling in the program was in acquiring the expertise to address the day-to-day issues and challenges they face in working with their data and learning about the resources that could assist them in their work. Our focus on applying what was taught directly to the students' own data sets in ways that would help address their needs as data producers and managers was definitely seen as an asset. The emphasis on peer discussion was also identified as an important component of the program.

> 'You managed to [teach] it in a way that you were not making it a chore for us. I felt like I was doing it for my own benefit for a lot of the time, and that's the way a course should be. So I ended up getting a lot done because of that attitude.'

> 'Most of this material is abstract, but being able to directly apply the course's material to my own dataset right in class, made it much easier for me to grasp the topics.'

> 'I liked the discussion format. It made me engaged in what otherwise could have been a huge pile of information just shoved down my throat.'

Students reported an increased awareness of the frameworks that are in place or being developed to help them address data management and curation issues. Although students are tasked with working on data they are not generally exposed to concepts, tools or resources that are available to assist them in their responsibilities. Furthermore, even when they are aware of these things they may not have the time to explore them fully or guidance to help them understand how they could impact their work.

> 'I was surprised by the degree to which there is a framework in place for handling these kinds of issues. I guess I had thought it was more of a make it up as you go along sort of thing, but there actually is quite a bit of information out there if you know where to look for it.'

> 'In addition to the tools you introduced to us, it also inspired us to find out what is out there. Are there some other tools that are available that can be used to facilitate our life and study? [The program] was a kind of a seed.'

We recognized that the scope of issues pertaining to data management and curation would exceed our knowledge base. Therefore we recruited multiple guest speakers with expertise on relevant topics. Guest speakers included librarians, IT personnel, faculty, and data curators from subject based repositories. These guest lectures were most successful when the presenter was able to make abstract concepts tangible and relevant to the students, and when were able to key in on student's own needs.

> 'I liked the presenter from Dryad. I thought she was phenomenal. Probably because that would be the way I would like to store my data myself.'

Lessons Learned in Teaching the Program

Our assessments also revealed several areas where the pilot program could be improved for our students. Although students stated that being able to discuss their data sets with peers from related disciplines was helpful, the two students from the Agricultural Economics department did not receive a similar benefit from the program. Their data sets were more centered on Economics than on Agriculture, and the data they used were primarily derived from external sources rather than created in a lab environment. These differences made it more challenging for us to connect the lessons to these students, which ultimately limited the utility of the program for them.

> 'As more of a consumer than an author of data, I didn't feel all the topics were useful to me… Economists are pretty good at data discovery and acquisition, and perhaps need less guidance on data quality/documentation.'

Our students did not have much exposure to the concepts of data management and curation, let alone a sense of how they could be applied into practice. Given their unfamiliarity, students naturally wanted examples to help them understand and apply what they were learning. Early on we were reluctant to provide examples for the DMP session, fearing that students would see an example as a model of how it should be done rather than to focus on developing a plan for their own data. However, students reported not having a sufficient enough understanding of what a completed DMP would look like making it difficult for them to draft their own. In response, students suggested providing a large number of examples to not only provide grounding for their own work, but to demonstrate that there are many different ways to approach data management and curation.

> 'I think your best balance is a lot of examples. Because then we know that there are so many options that there is not a "wrong" way to do it. If you look at one [DMP] and say "that does not fit my data", you can look at others and take the best pieces of them.'

However, delivering the right balance of useful examples proved to be elusive, as some students were overwhelmed by the number of existing data lifecycle models that we shared with them in the next session. We also found that students wanted examples of both "good" and "bad" data practices, which were not always easy to find or evaluate without knowledge of the complete context in which the practices took place.

In a similar vein, finding readings that were suitable for the program was also more difficult than expected. Students want to understand the concepts, but more importantly, they want to know how to apply these concepts to their own work and discipline. Though more practical minded pieces on managing data are appearing, many of them take a fairly general and broad approach to the subject. "Top ten list" style articles written for all researchers provide diminishing returns to students after a while. Materials that are focused on data management from the perspective of Agriculture were scarce.

Finally, in planning the program we knew that we would cover both higher-level conceptualizations, such as data lifecycle modelling, and more practical applied aspects of working with data, such as file naming conventions. We believed that introducing the higher-level conceptualizations earlier in the program would give students useful frameworks to contextualize the more practical components as we introduced them. Although this was the case for some students, others felt overwhelmed by the concepts and were not able to see their relevance immediately without a sense of how they would be beneficial to their work directly. These students suggested introducing higher-level concepts after teaching students some of the more practical tools and resources for context.

> 'I thought it was valuable to see the big picture, but I agree [start with] the hands on. I was really uncomfortable trying to start with the big picture. I still felt lost. I felt like I was drowning.'

Long-Term Impact

We interviewed nine of the ten students six months after the last session to ascertain the long-term impact of the program. Although individual comments varied, two primary themes emerged. The first theme was the value of learning about the data lifecycle as a way to see the big picture of data. Most said that they never gave much thought to the fact that their data was just a small piece of a much bigger process. Seeing their work with data as a part of a lifecycle helped them understand the value of good data management practices.

> 'And I guess just thinking about the data life cycle… I'd never really thought about it past publication and what happens with that but I have to remember that you probably make so much more data than you actually need and it can have value to other people too. So I think that was an eye opener in that regard.'

Secondly, our students commented on the value of learning about file naming conventions and almost all had applied this in either their personal or lab work. The unanimous retention and application of file naming conventions leads us to believe that developing stand-alone lessons or working the subject into the curriculum of other courses could be a starting point to extend the reach of data literacy programming.

> 'Yes, I think the file naming conventions was something that I definitely changed and I definitely developed kind of a similar format which [has] really made my work a lot more efficient because it's, you know, I don't have to reinterpret what every sub-folder means. I just, I automatically know and it's consistent so that's been very helpful.'

> '[The file naming convention is] easy to remember and works great and I'm going to be using that for the rest of my life.'

Metadata was also mentioned as being immediately applicable. The program helped our students understand how a description can help or hinder the discoverability and usability of a data set, which in turn made them reconsider the importance of describing their own data. Teaching metadata was a challenge. Some of our students struggled with the concept and application of metadata during the program. Although we spent more time on metadata than we had initially thought we would, our students came to recognize its importance. Many of them are now actively incorporating metadata into their data practices.

> 'I think I developed an appreciation for metadata documenting the research I do so that it could have value beyond the original intention of the data. And also I feel like it makes me a better scientist because it makes my work more repeatable and I think that in a lot of cases people might do work but it may not be repeatable because they don't document it well enough.'

Finally, students recognized that a lot of the concepts covered in the program were in many ways basic aspects of doing research, even though they were never explicitly covered as a part of their education or as an introduction to doing lab work.

> 'So these are all things that should have been pretty obvious to me as a student, years ago but it's just things that nobody ever tells you about or really teaches you so you don't really think about it until you've got a mess of files.'

> 'I mean I think I had a lot of those ideas floating around in my brain already. Like needing to think about publishing my data and I need to think about how my lab and how I specifically was collecting and storing it but didn't really have the tools necessary to put that into formal practice, I guess. And so the class was really helpful in helping me to do that through the process of creating the management plan or just discussing different options that I had and so forth.'

Students all agreed that this pilot should become a regularly offered class. Suggestions for structuring the class included even more in-class exercises and encouraging graduate students to take this early on so they have some grasp of the data they will be collecting. However, it was noted that students should probably not take this the very first semester, as that semester can already be overloaded with adjusting to graduate life and they may not have enough sense of their own research or data to be able to apply this immediately.

Broader Challenges in Teaching Data Literacy

> 'Do I really have rights to my work as a graduate student?'

We observed a wide variation in the receptiveness of student's faculty advisors to the topics and issues that were being addressed in the program. Many of our student's advisors were open to students bringing back what they had learned to their colleagues and some of our students even led efforts to reassess current practices within their labs. However, some of our students expressed frustration over how difficult it was to incorporate what they were learning without support. A few struggled to even get their advisor to engage in discussions on data issues. Although they are expected to take ownership of their work, graduate students are still under the authority of their advisor and dependent upon their support. This tension between advisor and graduate student about the treatment and disposition of data is an issue that affects the impact of data literacy programming and must be navigated with care.

Beyond issues with their advisor, students are also acutely aware that working with and managing data is typically a group effort. Students themselves often have limited control over the process of generating and administering data in their lab. We recognized this reality and attempted to address it through creating assignments where students would interact with their advisor, peers and others in the lab. However, as we did not have direct access to their colleagues, we had to focus solely the students themselves and open ourselves up to potentially creating undue expectations on the student to change their environment. This balance between empowering individual

students and the desire to foster change in lab culture and practices is another area that needs to be carefully navigated for both the student and the data literacy instructor.

> '…managing the whole lifecycle of data can be really challenging. Maybe it could be like framed more in a way so that it's like a group effort rather than like the responsibility of each individual grad student. I think managing the whole lifecycle, it's too much for one person to do and you aren't even going to be here for more than a few years but, I don't know… it's just kind of overwhelming.'

Conclusion

The demand for data literacy programs will only increase as data management, publishing and curation become a more normative part of scholarship. Although our program is but one instance of an effort to teach data literacy competencies to students, we learned multiple lessons in teaching our course that are likely to be relevant to similar programs.

We found that designing the program around the students' own contexts and experiences in working with and managing data served as a solid foundation to introduce and teach both high-level concepts and practical applications. Although students benefited from exposure to high-level concepts, they had to be connected to practical applications in order for student to understand and appreciate their significance. This balance between concept and application was difficult to achieve and required constant monitoring and correction over the course of the data literacy program. Creating plenty of space and providing time for students to discuss these concepts and to work through their possible applications with each other as peers was a critical component to the success of our program. However, peer networks alone are insufficient for data literacy to take root completely. The perceptions of the faculty advisor of the value of data management, sharing and curation will have an impact on the ability of the student to consider and incorporate data literacy competencies into their own work. Data literacy programs will need to take the student-advisor relationship into account and look for ways to overcome potential barriers and extend education into the work environment.

Finally, we in the data curation community need to pay greater attention to the education of the next generation of researchers. Developing resources and tools to facilitate curation addresses only a part of the challenge and raising awareness of the need for curation is insufficient. In order to realize data curation as a normative function of the research process it needs to be integrated into disciplinary culture and incorporated into lab practices. In other words it needs to be a part of the educational curriculum of graduate students as they assume their identity as research professionals. Though we may have varying levels of responsibility for developing curricula we can and should use our relationships with administration, faculty and students to ensure that the need for education on data management and curation competencies is given due consideration.

References

Austin, A. (2002). Preparing the next generation of faculty: Graduate school as socialization to the academic career. *The Journal of Higher Education, 73*(1), 94–122. Retrieved from http://www.jstor.org/stable/1558449?origin=JSTOR-pdf

Bracke, M.S., (2011). Emerging data curation roles for librarians: A case study of agricultural data. *Journal of Agricultural and Food Information, 12*(1), 65–74. doi:10.1080/10496505.2011.539158

Carlson, J., Fosmire, M., Miller, C., & Sapp Nelson, M. (2011). Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy, 11*(2). 629–657. doi:10.1353/pla.2011.0022

Carlson, J., & Johnston, L. (Eds.) (2015). *Data information literacy: Librarians, data and the education of a new generation of researchers*. West Lafayette, IN: Purdue University Press.

Carlson, J., & Stowell Bracke, M. (2013). Data management and sharing from the perspective of graduate students: An examination of culture and practice at the Water Quality Field Station. *portal: Libraries and the Academy, 13*(4), 343–361. doi:10.1353/pla.2013.0034

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. doi:10.2218/ijdc.v8i1.254

Eaker, C. (2014). Planning data management education initiatives: Process, feedback, and future directions. *Journal of eScience Librarianship, 3*(1), Article 1. doi:10.7191/jeslib.2014.1054

Grady, R.K., LaTouche, R., Oslawski-Lopex, J., Powers, A., & Simacek, K. (2014). Betwixt and between: The social position and stress experiences of graduate students. *Teaching Sociology, 42*(1). 5–16. doi:10.1177/0092055X13502182

Haendel, M.A., Vasilevsky, N.A., & Wirz, J.A. (2012). Dealing with data: A case study on information and data management literacy. *PLoS Biology, 10*(5), e1001339. doi:10.1371/journal.pbio.1001339

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development, 48*(3), 23–48. doi:10.1007/BF02319856

Janke, L.M., & Asher, A. (2012). *The problem of data* (CLIR Publication No. 154). Washington D.C. Retrieved from Council on Library and Information Resources website: http://www.clir.org/pubs/reports/pub154

Molloy, L., & Snow, K. (2012). The data management skills support initiative: Synthesising postgraduate training in research data management. *International Journal of Digital Curation, 7*(2), pp. 101–109. doi:10.2218/ijdc.v7i2.233

Whitmire, A. (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, Purdue University. Retrieved from http://docs.lib.purdue.edu/dilsymposium/2013/posters/6/

Appendix A
Recruitment Form for the Data Literacy Pilot Program

Data Literacy Pilot Program Cohort Recruitment

Accepted students will receive a stipend of $1,500 for participating in this Pilot Program.

Overview

Increasingly we are living in a data-driven world. Smart phones and social media depend on access to well-formed and properly managed data sets. Agriculture is no different. Research in agriculture and related fields is growing more and more dependent on the application of data. You, as an emerging research professional and data producer, will need to understand how to manage, organize, describe, disseminate, and archive your data sets for yourself and for others within and outside of your field of research. These skills will be essential to develop in order to be competitive in today's job market.

The College of Agriculture, with support from the Purdue University Libraries, is seeking to develop a program in data literacy to teach graduate students the knowledge and skills they will need to be successful in as professionals in their discipline. We are seeking to recruit students to participate in a semester long *pilot data literacy program* to be offered in Spring of 2014, with some limited follow up in the Fall. In addition to addressing current trends and thinking in managing, organizing, sharing and curating research data, students will actively participate in shaping the program to address real world challenges and issues they face in developing and working with research data sets. Participants will be applying what they learn to their own data sets as a part of the program.

Through offering this pilot, the Libraries and the College of Agriculture will develop a better understanding of the issues with data that are universally relevant across departments and experience in teaching the knowledge and skills needed for students to address these issues going forward. This cohort of students will benefit from receiving intense, hands-on training in managing their own data. One of the goals for the program is to create a community of students knowledgeable with data management and curation issues who will then be able to spread awareness and best practices of data management with fellow students and others in their own labs.

Students will be able to improve their current data processes through taking this program by working with research data they are already using.

The Application Process

Space in this program is limited. We are only accepting 5-10 graduate students from the College of Agriculture. Students will be selected through an application process. Although PhD students are preferred, the program is open to both PhD and Masters level students provided that they meet the criteria listed below. ***Applications are due November 18th, 2013 by 5pm.***

- Students must be actively working with or developing a research data set as a part of their lab work and/or their scholarship.

- The student's advisor will be asked to participate as a means to gauge student learning from the pilot program. Students will need to get the consent of their advisor to participate and secure an agreement from their advisor to be interviewed at the beginning of the semester and at the end of the semester.

Structure

Meetings will be held once a week as a two hour session. Initially sessions will focus on discussing the nature of the data the students are generating, student responsibilities and actions in developing data sets, and the challenges they face in their work. Based in part on the themes that arise from these discussions, we will then introduce a series of discussion topics and activities pertaining to data management or curation in ways that relate to the student's own work. Activities may include developing a data management plan to guide the development of a data set, documenting and describing data well enough so that others could understand and make use of the data, or taking steps to prepare the data for eventual publication and long term preservation.

Previous research conducted by the Purdue Libraries has identified twelve competencies in working with research data. These twelve competencies will serve as the initial foundation for exploration in the course. The twelve competencies are:

- Discovery and Acquisition

- Ethics and Attribution

- Metadata and Data Description

- Cultures of Practice

- Data Management and Organization

- Data Curation and Reuse

- Data Quality and Documentation

- Data Processing and Analysis

- Data Visualization and Reuse

- Databases and Data Formats

- Data Conversion and Interoperability

- Data Preservation

Each student will work through these issues using the actual data that they are generating in their labs with active consultation from the instructors. Participating students will also have the advantage of having a cohort to discuss particular challenges and to seek out solutions as a group. This will not just be an academic discussion – students will be required to begin implementing these data management techniques into actual practice.

Additional Benefits

Accepted students will receive a stipend of $1,500 for participating in this pilot program.

Students must be able to participate fully in the course sessions and complete the weekly action items satisfactorily. Regular attendance and active engagement in the program is required for the student to receive the stipend.

Questions? Please Contact Us

- Jake Carlson, Associate Professor of Library Science/Data Services Specialist, can be reached at: [Jake's email address]

- Marianne Stowell Bracke, Associate Professor of Library Science/Agricultural Information Specialist, can be reached at: [Marianne's email address]

Appendix B
Application Form for the Data Literacy Pilot Program

Data Literacy Pilot Program Application Form

Name:

Address:

Phone Number:

Email Address:

Department:

Name of your faculty advisor:

Year in your program:

Please describe the research you are conducting:

Please describe the data that you are generating in your research (including the formats, kinds of data, size and other attributes of the data):

Please describe how you are currently managing and documenting this data:

Please provide a brief statement of interest:

Signature of advisor: