Capturing their "first" dataset: A graduate course to walk PhD students through the curation of their dissertation data

Megan Sapp Nelson[1]; Ningning Nicole Kong[2]

Abstract

The data set accompanying theses is a valuable intellectual property asset, both from the viewpoint of the PhD student, who can procure employment and build publications and research grants from the work for years to come, and the university, which owns the data and has invested in the work. However, the data set has generally not been captured as a finished product in a similar manner to the published thesis. A course has been developed which walks PhD students through the process of identifying an archival data set, selecting a repository or long term storage location, creating metadata and documentation for the data package, and the deposit process. A pre- and post assessment has been designed to ascertain the level of data literacy the students gain through curating their own dataset. PIs for the projects have input into the repositories and metadata standards selected. The university thesis office was consulted as the course was developed, so that accurate procedures and practices are reflected throughout the course. This first of a kind class is open to students of any discipline at a Research-1 university. The resulting mixture of data types creates a unique course every time it is offered.

Keywords

Data curation, instruction, curriculum, data literacy, dissertation, thesis

Introduction

Identifying the appropriate timing within the capture, management, and preservation of a research project's data for the just-in-time education in data curation has been discussed briefly in the literature but remains unresolved. This lack of resolution is in part due to the necessity of practicing skills in situ with existing data sets as conceptual skills are taught or transferred.

PhD students, by virtue of the point in their career that they have attained, have produced viable data sets, and have reached a conceptually open space where they understand the necessity of practicing data curation skills. The alignment of timing, existing data sets of intrinsic importance to the individual learner, a self concept within data management (either in a current role as a data manager in a lab or a perception of themselves as a future data steward responsible for the production of data), and extrinsic motivation to capture as many metrics as possible to demonstrate impact of nascent research careers, all align to make metriculating PhD candidates prime targets for data curation education.

A need for this data curation education was also determined by requests from PhD candidates' academic advisors, particularly when the dataset was not captured but was a valuable output of the research process. This often reflected the fact that the traditional university dissertation submission requirement is not enough to capture the value of the research data.

 Sapp Nelson, Megan & Kong, Ningning Nicole (2020) Capturing their "first" dataset: A graduate course to walk PhD students through the curation of their dissertation data, IASSIST Quarterly 44(3), pp. 1-13. DOI: https://doi.org/10.29173/iq971

Dedicated time and appropriate guidance is needed to guide the students through the process of organizing their final datasets, creating data documentation, and sharing their data either within their research lab or with the public. Coupled with the recent increasing demands emphasizing the data management skills from the U.S. job market, the curriculum designers developed a data curation course for graduate students to fulfill this need.

Given this likely audience is also among the busiest and most burdened, providing instruction in data curation of a highly valued data set in an efficient and targeted manner is necessary for knowledge transfer. In the design phase of this course, the curriculum developers have collected information from disciplinary faculty and PhD students about their expectations. Then their needs were mapped with the data curation process with inputs from many data librarians, to develop the course content and lab activities. While this course is being offered for the first time, the curriculum developers also are closely observing the classroom dynamics and collecting feedback from students in order to improve it for future iterations.

Brief Synopsis of Curricular Innovation

This article reports on an innovative implementation of a graduate level 16 week three credit hour course, designed to assist PhD students in the final semester before they deposit their thesis to prepare their data for deposit. The initial pilot enrolled six students from the Colleges of Agriculture, Education, Engineering, Liberal Arts, and Polytechnic. The types of data deposited during the pilot included flat files, audio files, 3D image files, GIS data, image files, and text files.

Review of Literature

Research data management has been identified as a key educational need for graduate and post-graduate students. (Carlson et al, 2011, Doucette and Fyfe, 2013)There have been credit courses developed targeting graduate students that teach data management and data literacy, but frequently with specific audiences in mind. Basic introductions to data management and data literacy have been described in the literature, tailored to graduate students in Agriculture (Carlson and Bracke, 2015), engineering (Jeffryes and Johnston, 2013), science (Qin and D'Ignazio 2010; Frank and Pharo, 2016), social sciences (Thielen and Hess 2017) and health (Macy and Coates, 2016). Many courses focus on data storage and reuse competencies. However, a few courses are now moving into data wrangling using data science tools (Pascuzzi and Sapp Nelson 2018). Generally, if data publication and sharing is addressed in these courses, it is the topic of one course session.

Theses are a special category of institutional and scholarly communication that have required libraries to take special steps to catalog and preserve knowledge for decades, due to limited publication runs. In the past two decades the electronic manifestation of theses and dissertations (ETDs) have required specific research and technological support to ensure the long term preservation and access of content that was previously available only in print on local institutional shelves.(Fineman, 2003) When those dissertations were sitting on the shelves, the data frequently were packaged in the form of tables in text or supplemental data on CD-ROMs or floppy disks tipped

into the binding of the dissertation itself, and shelved with the text in the library. (Schöpfel et al, 2015a) This made the data available for the purpose of reproducibility, but was not sufficiently flexible to facilitate reuse. With the explosion of digital data, the dissertation is widely available on electronic platforms, but the institution has lost control of the co-linking of the data set with the published dissertation.(Collie and Witt, 2011) The question of where the responsibility lies for the curation of the dissertation data sets remains unresolved, but generally institutions do not have the staffing to provide hands-on support for full service curation for dissertation data sets. Instead, institutions are training PhD candidates to curate their own data sets in advance of the deposit of their dissertation (Schöpfel et al, 2015b).

Disciplinary Needs for Dissertation Data Curation

The initial idea for the development of this course was inspired by a civil engineering faculty member who felt that data was being lost due to PhD students and Masters level students graduating after publishing theses without the attendant data sets curated in appropriate repositories. Library consultations were provided at that time to help his graduating student publish datasets specific to the faculty members' research lab to a repository, as well as create a web page to guide users to access the data. As part of the library's data curation support, the data usage has been monitored after the publication. Below are visit statistics for the dataset one year after the data was published.

Figure 1: Visit statistics for published dataset used to help make use case for data publication course

This data curation and publication process (along with the data set altMetrics) helped the faculty to demonstrate his research findings to colleagues and funding agencies, as well as to secure the longevity of the dataset for further research. This process is a good use case showing that libraries can provide data curation education for graduate students at the final stage of their research, providing positive outcomes for themselves and their research advisors.

Initial Proposal

Following the suggestion from this faculty member, the initial conversation with him identified learning outcomes and a general outline for content that should be included in a course, should it be developed. This initial proposal represents the first tentative thoughts as to what might be appropriate for students who are approaching the problem of archiving their data for long term access and sharing, as prioritized in the Civil Engineering professor's original vision.

The proposal as written is included here:

This class is a practicum for final semester graduate students who have data sets to deposit prior to graduation. In this course, the students will receive instruction and mentoring on preparing their thesis data set for deposit and sharing. Topics include protecting the intellectual property of the data set, preparing the data set for deposit, and maximizing scholarly impact of the data set.

Learning Objectives

Students will

- Critically evaluate their thesis data set for quality in order to enable reproducibility and reuse
- Compile readme files and documentation in order to enable others understand and use their data sets
- Create metadata in order to facilitate publication of their data set in a data repository
- Prepare files with human readable file names in order to share data with thesis supervisors and others.
- Select data publication venues based upon future impact in order to maximize their scholarly output from their data.
- Select an appropriate data license in order to preserve their intellectual property.

General Outline

1. Critical evaluation of thesis data sets
2. Human readable file names/file structure and file conversion to non-proprietary file formats
3. Introduction to documentation as context (Readme as Index)
4. Identification of publication files (subset of thesis data set)
5. File level documentation (All files in thesis data set)
6. File level documentation peer review
7. Selecting a data repository
8. Metadata (subset of thesis data set)
9. Metadata peer review
10. Data Licensing
11. Submission Process overview
12. Preparation of submission package for repository
13. Preparation of submission package for repository
14. Peer review of submission package for repository

15. Critical evaluation of thesis data set part two/ Course evaluation

The proposal did not proceed at that point in time, in part due to a changing landscape in graduate education around data science at the university. A year long process was developed to create an Integrated Data Science Initiative, which seeks to instill data science principles within each curriculum at the university at some level, but does not articulate fixed curricular outcomes for those data science principles.

This landscape within the university means that individual departments are looking closely at how their existing curricula currently purvey data science, but that there are not many locations that are synthesizing data management and curation from a holistic perspective. Given the direction that the university has taken in the development of this program, the Libraries has initiated programs to develop curricula to support these underpinning management and curation skills from the undergraduate to graduate levels.

In this newly developing landscape, the previously proposed course represented an innovative solution to serve the entire Graduate College in educating future data stewards, as well as preserving and protecting the intellectual property produced by the students of the university.

Given these incentives, the course development process began in Summer 2018, and the course was offered for the first time in Spring Semester 2019.

Curricular Development

For the pilot, the proposed course was limited to final semester PhD students. This narrowed the scope for the course design: the course designers could focus on the very practical needs of taking a pre-existing data set coming from any discipline on campus that may or may not have had metadata or documentation prepared for an external audience, and bring that data set to publishable state in 16 weeks (1 full semester). Additionally, the course developers wished to keep nearly all course activities within the confines of in-class time (one 50 minute lecture and one 120 minute laboratory). Therefore, ensuring that students were not simultaneously collecting data, analyzing data, etc. was a priority so that the tasks for the class could be completed within the time scope designed for the class.

Additionally, since all course components were intended to be concluded within the confines of the class meetings, the format of the course was by definition active learning. Any lecture or theoretical content had to be relayed to the students efficiently, clearly, and in the context of practical, hands on, applied tasks that logically come next in the chain of events that are linked together to create a curated data set.

The hands on activities also need to dovetail with the thesis deposit process as dictated by the University's Graduate College. At the time of the course development, a new dissertation repository was being developed and procedures were being developed simultaneously with the course. Therefore, a series of meetings were held with the thesis repository manager to ensure that the curriculum was accurate and supported the current best practices of the Graduate School.

These constraints meant that the course development of the lecture materials and lab activities started from the initial outline, and then solidified into a "data set designed for re-use" focus that was comprehensive of both reuse within the local (but future) research laboratory, or reuse by an unknown third party. The selling point for students taking the class focused on building their professional portfolios through sharing their data for reuse.

Ultimately, the course objectives and learning objectives were articulated as:

This course walks students through the process of preparing a data set for sharing with both internal and external audiences. Students will select authoritative data sets from the data sets that they have prepared in the process of doing their thesis work and/or research projects for sharing and publication, apply metadata to those data sets, create documentation for end users of the data sets, and publish the data sets to internal or external data repositories or storage as appropriate.

The learning objectives include:

- Recognize and evaluate the value of research datasets and the needs for preservation.
- Prepare dataset packages for sharing and reuse that describes the documentation, workflows, and data enclosed in ways that allow users to determine currency, relevance, authority, accuracy, and purpose of the data set.
- Understand basic metadata fields, metadata standards, and be able to apply standard metadata to make the data set available to others and, if applicable to comply with disciplinary norms.
- Recognize disciplinary practices, values and norms related to organizing, sharing in disciplinary data repositories, curating and preserving data.
- Post data sets with recommended citations, including Digital Object Identifier.
- Share data in a repository or appropriate storage as agreed upon project primary investigator.

In some occasional cases, pre-existing lecture materials created for other courses at our institution were appropriate for use. In one case, the DataOne lecture on Metadata was considered to be at the correct level of specificity and comprehensiveness that there was no reason to develop a lecture from scratch. However, in most cases, lectures were developed to meet the specific needs in terms of practicality and theory that are unique to this course.

The activities are custom created for the course according to the content of each week. Overall, each course activity can be considered as one component toward a larger project of curating and sharing the student's valuable research data during their PhD study. The activities are not graded on a weekly basis. In some cases, the assignments accumulate over the course of several weeks prior to submission, due to the work required to compile a completed component of the data submission package. Of note, the documentation and metadata each require multiple weeks to create in the class, and therefore those sections of the semester are covered over about one and a half months.

Assignments Graded Towards the Final Grade Total

The course begins with a review of the known information about the students' research projects and data sets to familiarize the entire class with everyone else's projects and datasets. Students were then introduced to methods of determining the value of data and the research data life cycle so that they can identify the most valuable data records for the purposes of preservation and reuse. After that point, they learn the process of creating meaningful documentation, machine-readable metadata, identify relevant data repositories or data sharing spaces, and prepare the data package for preservation. The concepts of data sharing policies, licensing, embargo, and selecting a data repository are introduced during the semester "just in time" so that students can make appropriate decisions according to the nature and stage of their projects.

The course performance is evaluated by seven assignments, two peer review experiences, and attendance. Attendance is key to the completion of the final deliverables due to the reliance on in-class time to complete tasks. Without actively attending the course, especially the first time when it is offered, it is hard to ensure each student will get enough guidance for their discipline specific dataset.

The list of evaluations include:

- Data Information Sheet - An introduction to the dataset as well as an agreement with the student's academic supervisor regarding the intent to share the data either internally or externally
- Data of Record - An inventory of datasets used or generated for the research
- Final Data of Record/ Authoritative Data - An evaluation of the data records using the value of data rubrics
- Documentation - Documentation at the project, folder, and data levels, including ReadMe file, data dictionary, and code book

The curriculum developers adapted the ReadMe file template in use in the class from https://cornell.app.box.com/v/ReadmeTemplate. Additional fields were added to integrate a data dictionary to the template.

- Metadata - Standards are selected as appropriate to each project/discipline/repository/dataset and applied as appropriate.
- Data Submission Package - A completed package of documentation, metadata, data and a preferred citation.
- Citation - A deliberately designed, specified citation for the data set.
- Peer Review of Documentation
- Peer Review of Submission Package
- Attendance

Peer review processes were built into this course during the data documentation stage and data submission package stage. These are two critical stages for data curation. Through a peer review process, students can get helpful feedback from another set of eyes and make sure their data documentation and organization can be understood by others who are not inculcated within their research project.

It was not necessary that all the lab activities were evaluated for the course, since there is no right or wrong decision that could be made at multiple points throughout the semester. The curriculum designers were present throughout the semester to provide feedback as decisions were made and to point to best practices and pros and cons of decisions. We only chose to grade the activities where our feedback or suggestions could be helpful in the students' data curation process and to provide momentum toward the creation of the data submission package.

Students' academic advisors' signatures are required at the beginning and end of the course to make sure the data curation and sharing efforts made through this course align with their research labs' data management needs and ethics, so that the students can maximize the benefits of the course by handing down best practices to other personnel in their home labs during this process.

Class Schedule (Subject to modification)

Week	Class	Topic	Activities
Week 1 Jan 7	Lecture	Introduction	Gather individual data information and get PI to fill out data profile section before Lab.
	Lab	Data Profile	
Week 2 Jan 14	Lecture	Value of Data	Identifying the Data of Record for preservation
	Lab		
Week 3 Jan 21	Lecture	No class – Martin Luther King, Jr. Holiday	
	Lab		
Week 4 Jan 28	Lecture	Evaluation of Data of Record	Initial evaluation of data set
	Lab		
Week 5 Feb 4	Lecture	Introduction to Documentation	1 st draft of documentation
	Lab		
Week 6 Feb 11	Lecture	Peer Review of Documentation 1	Peer review of Documentation
	Lab		
Week 7 Feb 18	Lecture	Documentation, cont.	2nd draft of documentation
	Lab		
Week 8 Feb 25	Lecture	Identifying Repository (Guest Lecture)	Identifying repository, discussion with data steward
	Lab		
Week 9 Mar 4	Lecture	Metadata	Draft of metadata
	Lab		
Spring Break			
Week 10 Mar 18	Lecture	Sharing Policies and Licensing/ Embargo (Guest Lecture)	License selection in collaboration with PI/ Identification of embargo if any.
	Lab		
Week 11 Mar 25	Lecture	Pulling Together a Data Package	Inventorying data objects to include in data package
	Lab	Creating a Data Package	
Week 12 Apr 1	Lecture	Creating a Data Package, part 2	Peer review of data package, including PI

	Lab		
Week 13 Apr 8	Lecture	Submission to PURR	Identify required information to submit; don't submit yet.
	Lab	Submission to Storage/Disciplinary Repository	
Week 14 Apr 15	Lecture	Attribution/Citation/ PURL	Develop a suggested citation; make certain a DOI will be assigned for a published data set; Get final sign off from Data Steward/PI
	Lab		
Week 15 Apr 22	Lecture	Submission Week/ Final Proofread/ Hit submit	Record your DOI in your thesis.
	Lab		

Table 1: Course Schedule with Topics and Deliverables

Data curation professionals who work with research data management within the Libraries were enlisted to present the lectures and activities in areas of their specialization that relate to the specific activities the learners are undertaking at a specific point within their curation process.

Additionally, professional data curators were enlisted to provide feedback on the curriculum, and then to provide specific technical support for data types that require additional layers of access or curation, such as MatLab files or GIS layers as indexes to multiple data files.

Assessment

The class was developed with a pre-assessment and post-assessment as part of the curricular model. The assessments were based on two pre-existing documents. The first was the Data Curation profiles, a long-form structured interview modality designed to elicit data curation practices from researchers.(Witt, Carlson, Brandt, and Cragin, 2009) The second pre-existing tool that was used as a starting point for the pre and post- assessments was a self assessment tool designed for post-docs and early career faculty members to identify research data management skills that they may need to develop in order to be successful data stewards. (Carlson, Nelson, Johnston, and Koshoffer, 2015) The resulting assessment contains a brief demographics section that links school, department, years pursuing their PhD, and types of funding and support received. A section records number and types of research outputs that exist from the research project independent of the thesis, including conference proceedings and journal articles, and the role that the student participating in the class played in the authorship of these works (solo author, first author, etc.).

Finally, an extensive section details the research data management process and parameters for the data set that will be curated in the course of the class. This section includes information on grant funding agencies and data sharing requirements, data management plans, licensing agreements for any data that was reused in the process of carrying out the PhD research project, preferences for data preservation and sharing, and any requirements that are mandated by local research groups or funding agencies.

The pre-assessment and post-assessment questions are similar in topic. The post-assessment measures both the skills the students have put in practice and attitudinal indicators regarding the importance of the practice for the long term preservation and curation of their data set. The intent

is that the assessments will measure changes in each of the cognitive, affective, and psychomotor domain, when combined with the graded submissions for the course.

With the size of the pilot course, nothing will be able to be said about statistical significance or correlation. It is the hope of the curriculum developers that (over the course of multiple semesters) we will be able to determine the impact of the curation of the data sets on altMetrics and and traditional citation metrics (which have not traditionally represented the impact of dissertations well).

Lessons Learned from Pilot Course

Each data set brought to the pilot course is unique and represents a different discipline. Not only are the data fundamentally different (audio data, microscopy images, GIS data, MATLAB data, open source software outputs, text) but the goals of the students for the data sets are different as well. The articulated goals include creating a reference data set of images, setting up a curation protocol for an ongoing project that will be completed in three years, and sharing for the purpose of developing policy.

The students also bring a variety of baseline knowledge of research data management skills and curation principles. Some students have been working with high performance computing, while others are using desktop computing software for all aspects of their dissertation project. One has been the designated data manager for a large research lab, others have been working independently on their research throughout their career.

All of this is to say that the curriculum designers focus on fundamental principles of data publication was correct because the principles are the only thing the course participants have in common. Tools are not held in common. Processes are not held in common. If examples are needed for the course, they can be selected from any discipline because connections will have to be drawn for many other people in the class from that example to their work.

Course preparation for specific aspects of the class such as metadata standards, packaging for software, identifying repositories, and other disciplinary specific topics will have to be customized to each semester's roster of students and their specific data sets. In that way, the course will never truly be a completed curricula, and will always require a higher workload for the instructor on an ongoing basis.

The initial instinct of the course designers were to ensure that the primary investigators signed a document indicating their consent for the PhD students to share data. It turned out that this was too limited for large research groups. Many primary investigators consider themselves to not be the data stewards/sole proprietors of the data, but to hold it in common with all members of the research laboratory. This in turn means that the PI wants to have conversations with all members of the laboratory about what portions of the data can be shared along with the dissertation. This is a very positive ripple effect from the class, but the one week turnaround time for this deliverable is too short, given the amount of coordination that has to happen for the large labs.

There was some question whether three credit hours were too many. However, in the final semester prior to deposit, the participants are very busy with final edits on their dissertation, defending their thesis, and a number of other details needed to get the document done. Were the class time not provided to carry out the curation activities, it is unlikely that the steps of the process would be completed. Therefore, though reducing the number of credit hours to two was an option, it has been rejected.

Issues Remaining to be Addressed

An unanticipated issue that was identified early in the course was that of the inflexible nature of the software profile of the computers in our computer lab. An image is loaded at the beginning of the semester, before the students are enrolled in the course. No software is permanently installed after that point. However, the instructors don't actually know what the curation requirements for the course will be, including metadata editors, software packages, etc, until the first week of class. This timeline mismatch is problematic, and leads to the computer lab being significantly less useful than it would otherwise have been.

There was not enough time, even with 16 weeks in the class, to do comprehensive image level or audio file level metadata for the largest of the student projects. Good quality metadata really requires the time investment of a full time data curator, which is just not available under the current design of the course. Unless the students select a data repository that employs a full time data curator that can backfill that level of metadata, the data sets will have collection level metadata, but at some level will still not be machine readable at the individual datum level. Relying on the individual researcher to create this level of metadata may well be wholly unrealistic, however.

Conclusion

This small pilot indicates that the format of a three credit hour course is an appropriate venue for data curation and publication education. However, educational research is ongoing regarding the efficacy of the intervention itself. Whether the data sets that are produced will be of sufficient quality for reuse; whether repositories will be happy with the level of documentation and metadata produced by the students; whether students will be ready to serve in a directive role as data stewards in their future endeavors; and whether primary investigators will feel that the data has been captured sufficiently are all areas of ongoing research. Once these areas have been established, a primary problem that will have to be addressed is that of the scale of the educational intervention. How can institutions of higher education teach all graduating PhD and Master's students to capture their own research data in conjunction with the writing of their thesis? It is the question that the curriculum developers started this project with, and it remains an outsized problem that this class does not resolve. If anything, this class points to the complicated nature of providing the customized data curation education that individual students with their own data sets need. Further research and publication will be released in the future as these questions are investigated.

References

Carlson, J. and Bracke, M. (2015). Planting the Seeds for Data Literacy: Lessons Learned from a Student-Centered Education Program. *International Journal of Digital Curation*, 10(1), https://doi.org/10.2218/ijdc.v10i1.348

Carlson, J, Fosmire, M, Miller, C, and Sapp Nelson, M. (2011). Determining Data Literacy Needs: A Study of Students and Research Faculty. Portal: Libraries & the Academy, 11(2) p. 629-657. DOI: https://doi.org/10.1353/pla.2011.0022

Carlson, J., Sapp Nelson, M., Johnston, L. and Koshoffer, A. (2015). Developing Data Literacy Programs: Working with Faculty, Graduate Students and Undergraduates. Bulletin of the Association for Information Science and Technology. http://doi.org/10.1002/bult.2015.1720410608

Collie, W. and Witt, M. (2011). A Practice and Value Proposal for Doctoral Dissertation Data Curation. *International Journal of Digital Curation*, 6(2), pp.165-175.

Doucette, L and Fyfe, B. (2013). Drowning in Research Data: Addressing Data Management Literacy of Graduate Students. In Imagine, Innovate, Inspire: The Proceedings of the ACRL 2013 Conference (pp. 165-171).

Fineman, Y. (2003). Electronic Theses and Dissertations. *portal: Libraries and the Academy*, 3(2), pp.219-227. https://doi.org/10.1353/pla.2003.0032

Frank, E. and Pharo, N. (2016). Academic Librarians in Data Information Literacy Instruction: A Case Study in Meteorology. http://hdl.handle.net/10642/3470

Jeffreys, J. and Johnston, L. (2013). An e-Learning Approach to Data Information Literacy Education. In Proceedings of the ASEE Annual Conference Proceedings, 2013. http://hdl.handle.net/11299/156951

Macy, K and Coates, H. (2016). Data Information Literacy in Business and Public Health: Comparative Case Studies. *IFLA Journal* 42(4), 313-327. https://doi.org/10.1177/0340035216673382

Qin, J. and D'Ignazio, J. (2010). Lessons Learned From a Two-Year Experience in Science Data Literacy Education. In the Proceedings of the 31st Annual IATUL Conference. Retrieved from https://docs.lib.purdue.edu/iatul2010/conf/day2/5/

Research Data Management Service Group. (2019). "Author_Dataset_ReadmeTemplate.txt" Retrieved from https://data.research.cornell.edu/content/readme

Schöpfel, J., Primož, J., Prost, H., Malleret, C., Češarek, A., & Koler-Povh, T. (2015a). Dissertations and Data: keynote address. In GL17 International Conference on Grey Literature (hal-01285304). Amsterdam, Netherlands. Retrieved from https://hal.univ-lille3.fr/hal-01285304/document

Schöpfel, J., Prost, H., & Malleret, C. (2015b). Making Data In PhD Dissertations Reusable for Research. In 8th Conference on Grey Literature and Repositories (p. hal-01248979). Prague, Czech Republic. Retrieved from https://hal.univ-lille3.fr/hal-01248979/document

Thielen, J and Hess, A. (2017). Advancing Research Data Management in the Social Sciences: Implementing Instruction for Education Graduate Students into a Doctoral Curriculum. *Behavioral & Social Sciences Librarian* 36(1), pp 16-30. http://doi.org/10.1080/01639269.2017.1387739

Witt, M., Carlson, J. , Brandt, D. , & Cragin, M. (2009). Constructing Data Curation Profiles. International Journal of Data Curation, 4(3), pp. 93–103. http://doi.org/10.2218/ijdc.v4i3.117

Endnotes

[1] Megan Sapp Nelson is a Professor of Library Science and Science and Engineering Data Librarian at Purdue University Libraries. She can be reached by email: msn@purdue.edu.
[2] Ningning Nicole Kong is an Associate Professor of Library Science and Geographic Information Specialist at Purdue University Libraries.