

Authentic Learning in the Research Data Curation Classroom

Karen S. Baker, University of Illinois

Ruth E. Duerr, University of Colorado at Boulder

Abstract

We explore the provision of authentic learning through curation of scientific research data collections as preparation for information professionals. Hands on experience with curating a research data collection is provided in a graduate level classroom. Students gain insight into work with research data through online exploration of a data repository as well as via contact with a repository information professional. Four major elements of a student data collection curation project are described: selecting a data collection, developing a draft data curation plan, keeping a data collection activity log, and summarizing via formative and summative reports. The data curation project provided an experience mix of the curation culture and its services with data generating research cultures and their emergent practices.

Keywords: data curation, data collection, authentic learning, education

Citation: Baker, Karen S.; Duerr, Ruth E. (2015). Authentic Learning in the Research Data Curation Classroom. In *iConference 2015 Proceedings*.

Copyright: Copyright is held by the author(s).

Acknowledgements: Thanks to the participating repositories: the Inter-University Consortium for Political and Social Research (ICPSR), the Advanced Cooperative Arctic Data and Information Service (ACADIS), the National Oceanic and Atmospheric Administration (NOAA), the UK Earth Observation Data Centre as part of the Centre for Environmental Data Archival (NEODC, NERC) and the British Atmospheric Data Centre (BADC). This project was supported by the Institute of Museum and Library Services grant #IMLS RE-02-10-004-10.

Contact: karenbaker@gmail.com, RDuerr@colorado.edu

1. Introduction

Fulfilling the vision of open data and a web of data repositories requires the education of information professionals. As guides and curricula are being written to support teaching data curation (e.g. Varvel et al, 2012; Piorun et al, 2012; Lee, 2009), we explore the provision of authentic learning experiences in the classroom as preparation for a workforce able to meet the growing flood of research data. This example of experiential learning is made possible by partnership with a number of data repositories. Students are paired with data collections in need of curation.

2. The Program, the Class, and the Repositories

Curation of a data collection as a semester-long classroom project was introduced as an assignment in the Foundations of Data Curation course at the Graduate School of Library and Information Science (GSLIS) at the University of Illinois Urbana-Champaign. The data curation project was piloted in Fall 2013 (Duerr and Chao, 2014) and expanded in Fall 2014 to include a formal Data Set Request Form for repository participation, multiple assignments focusing on curation throughout the entire period of the class including a weekly data collection activity log, a data curation plan and additional reporting assignments.

The fifteen-week Foundations of Data Curation course content covers topics such as the research environment, data practices, metadata, scientific research data, data in the humanities, collection management, metadata, data organizations, data repositories, models, and infrastructure. The course is a required class for the two-year Specialization in Data Curation certificate program. A recent survey summarizes the program content and subsequent job experience of graduates (Thompson et al, 2013).

Partnering research repositories provided data for the classroom project resulting in a diversity of collections offered to students. The repositories included social science data from Interuniversity Consortium for Political and Social Research (ICSPR), local data/knowledge data from Exchange for Local Observations and Knowledge of the Arctic at the National Snow and Ice Data Center (ELOKA, NSIDC), arctic data from the Advanced Cooperative Arctic Data and Information Service (ACADIS), earth science data from both NOAA's National Geophysical Data Center and the UK Earth Observation Data Centre as part of the Centre for Environmental Data Archival (NEODC, NERC), and climate model data from the British Atmospheric Data Centre (BADC). A data set request form was developed in order to formalize the request, provide details about the class, and provide a summary of expected activities. The form also points out: "in addition to contributing to data availability, this approach represents an opportunity for a) students to provide a link on their vitae to an example of data curation work and b) repositories to enhance visibility of some data as well as to highlight their contributions to education and training of a much needed workforce in data curation." The collections offered are typically those

queued up awaiting attention at data repositories due to a deluge of data creating large backlogs of data requiring curation.

3. Authentic Learning

The data collection activity meets the four criteria identified as supporting authentic learning (Rule, 2006): an activity that 1) involves real-world problems and that mimics the work of professionals; 2) uses open-ended inquiry, thinking skills and metacognition; 3) prompts students to engage in discourse and social learning in a community of learners; 4) ensures students are self-directed in their learning from project work. Knowledge construction reflects elements of the scientific process including discovery and inquiry as well as design thinking and multi-situational awareness.

The kinds and forms of data in the collections are quite diverse ranging from a single excel spreadsheet to collections of derived products, satellite images, and local knowledge. The data may have metadata and codebooks missing and/or may need reformatting for ease of use or sustainability. As a result, a variety of different assessment and investigative activities are needed. Further, each collection has an associated repository contact. Students must demonstrate through posts in their weekly logs that they have prepared adequately and formulated questions prior to approach their contact. The curation project puts students in direct communication with information professionals about matters such as needs assessment, data description, data provenance and data presentation as well as missing or inconsistent information. An interpersonal interaction for a student apprentice with repository contact as mentor provides a window into the larger curation community (Lombardi, 2007). A contact outside the classroom meets the authentic learning recommendation for discussion of findings in venues outside the classroom so that the task is no longer seen as merely an exercise but understood as engagement with real-world issues and situations.

Activities and discussions are planned to support students becoming a community of learners. They are encouraged to post questions in open forums and to read each other's posts so they create a cross-comparative dialogue that elicits similarities and differences in their data circumstances. Such a comparative approach broadens their understanding of the data landscape and deepens their understanding of their own dataset.

4. Curating the Collection: Project Elements

There are four major elements in the data collection curation project.

1) Selecting the Data Collection

Browsing the collections affords students the opportunity to experience the diversity of the data landscape with data from field observations and experiments, from computational efforts and modeling, as well as from health sciences and humanities.

2) Developing a Draft Data Curation Plan

Drafting a curation plan helps students adopt the role of data curator who must describe the collection selected, assess its state, and write a plan for its curation. The plan is written early in the semester as a draft (1500 words with references) that will be revised throughout the term. The goal of the assignment is to move students into the curator role by having them get their hands on the data as well as the context of the collection and the repository.

3) Keeping a Data Collection Activity Log

Establishing an informal, weekly log of data collection activities ensures immediate and ongoing work with the data. The log is in the form of a post that they reply to each week so becomes a diary of their work with the data. Posts are to be a few sentences or a short paragraph that summarizes progress, questions, and barriers. Learning the data curation vocabulary and articulating collection activities are critical to professionals working with data.

4) Summarizing in Formative and Summative Reports

In order to assimilate, apply and reflect upon their understanding of data curation, additional assignments were developed to ensure continuing learning and provide opportunities for feedback. A curation assignment forum is available for questions and discussion. In addition, a 10-minute presentation to the class by each student is made about their collection midway through the semester. Finally, a written report is given at the end of class summarizing the data, the curation activities undertaken and any future next steps.

5. Learning by Doing Curation

This piloting of an authentic learning project in the classroom would benefit from a more formal measure of success. Surveys and interviews are frequently used to evaluate research projects involving mentoring. Until such an evaluation can be undertaken in conjunction with classroom work, we report on the case and project findings.

Students in this class had a wide range of skills in working with digital resources in the library realm and also came from diverse backgrounds (chemistry, law, sociology, – from novice to experienced but had little familiarity with research data needs and management). The need for significant help for some students was identified with the data curation plan assignment. These students reported using statements such as “There should” and “I would want to” while students who were taking steps to manage the data declared “I plan to” and “I am inquiring about”. An active versus passive framing of work with the data is indicative of actual engagement in data curation.

Given the often disorderly and complex set of activities associated with data curation, the need for patience became evident as students shared the many ill-defined and unanticipated tasks they encountered. An initial bewilderment was evident as students came to terms with the open-ended learning that comes from doing curation rather than learning about curation. Further, students were faced with addressing a data curation project before having had a full class on the subject. Yet, curation became in time an iterative activity involving numerous revisits to the repository website and related materials. Within the digital realm, the data curation project for students provided authentic learning about the service-oriented curation culture as well as the data-generating research cultures.

References

- SDCS, (n.d.). UIUC Specialization in Data Curation Certificate. Retrieved September 23, 2014 from http://www.lis.illinois.edu/academics/degrees/specializations/data_curation)
- Duerr, R., & Chao, T. (2014). Curating data collections in the classroom: lessons learned. Poster presented at the 9th International Digital Curation Conference, San Francisco, CA. Retrieved September 23, 2014 from http://www.dcc.ac.uk/sites/default/files/documents/IDCC14/198Duerr_Chao_idcc_14.pdf
- Lee, Christopher A. (2009). Curation Junction, What's your Function? Defining What it Means to Do Digital Curation in Order to Teach It. In Proceedings of the 2009 Annual Conference of the Association for Library and Information Science Education (ALISE).
- Lombardi, M. M. (2007). Authentic learning for the 21st century: An overview. *Educause learning initiative*, ELI Paper 1, 1-12.
- Piorun, M. E., Kafel, D., Leger-Hornby, T., Najafi, S., Martin, E. R., Colombo, P., and LaPelle, N. R.. (2012). Teaching Research Data Management: An Undergraduate/Graduate Curriculum. *Journal of eScience Librarianship* 1(1), Article 8. <http://dx.doi.org/10.7191/jeslib.2012.1003>
- Rule, A. (2006). The Components of Authentic Learning. *Journal of Authentic Learning* 3(1), 1-10.
- Thompson, C. A., Senseney, M., Baker, K. S., Varvel, V. E., & Palmer, C. L. (2013). Specialization in data curation: Preliminary results from an alumni survey, 2008–2012. *Proceedings of the American Society for Information Science and Technology*, 50(1), 1-4.
- Varvel Jr, V. E., Bammerlin, E. J., & Palmer, C. L. (2012). Education for data professionals: a study of current courses and programs. *Proceedings of the 2012 iConference*. doi: 10.1145/2132176.2132275