



EScience in Action

Research Data Management Instruction for Digital Humanities

Willow Dressel

Princeton University, Princeton, NJ, USA

Abstract

eScience related library services at Princeton University started in response to the National Science Foundation's (NSF) data management plan requirements, and grew to encompass a range of services including data management plan consultation, assistance with depositing into a disciplinary or institutional repository, and research data management instruction. These services were initially directed at science and engineering disciplines on campus, but the eScience Librarian soon realized the relevance of research data management instruction for humanities disciplines with digital approaches. Applicability to the digital humanities was initially recognized by discovery of related efforts from the history department's Information Technology (IT) manager in the form of a graduate-student workshop on file and digital-asset management concepts. Seeing the common ground these activities shared with research data management, a collaboration was formed between the history department's IT Manager and the eScience Librarian to provide a research data management overview to the entire campus community. The eScience Librarian was then invited to participate in the history department's graduate student file and digital asset management workshop to provide an overview of other research data management concepts. Based on the success of the collaboration with the history department IT, the eScience Librarian offered to develop a workshop for the newly formed Center for Digital Humanities at Princeton. To develop the workshop, background research on digital humanities curation was performed revealing similarities and differences between digital humanities curation and research data management in the sciences. These similarities and differences, workshop results, and areas of further study are discussed.

Correspondence: Willow Dressel: wdressel@princeton.edu

Keywords: research data management instruction, digital humanities

Rights and Permissions: Copyright Dressel © 2017



All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Introduction

eScience related library services at Princeton University started with data management plan consultations in response to the National Science Foundation's (NSF) data management plan requirements. After several years, in anticipation of needs beyond NSF requirements, one of Princeton University Library's science librarians was appointed as a half-time eScience Librarian and services grew to include data management plan consultation from all funders, assistance with depositing into a disciplinary or institutional repository, and research data management instruction. Initially the primary audience for these services was assumed to be the data-intensive fields of science and engineering.

During the planning and outreach stages of eScience services development, a connection was made with a colleague on campus engaged in related instruction for a different audience. Carla Zimowsk, the Information Technology (IT) Manager for Princeton's History Department had been giving a workshop on file and digital asset management concepts to history graduate students for many years. The complementary nature of these concepts with research data management led to a collaborative presentation on best practices for file and data management as part of the campus wide "Productive Scholar" series. The eScience Librarian was then invited to participate in the history department workshop on file management given to incoming graduate students.

Encouraged by the success of the collaboration with history graduate students, the eScience Librarian offered to develop a workshop on research data management for newly formed Center for Digital Humanities at Princeton (CDH). Opened in 2015, the Center for Digital Humanities is an interdisciplinary research center that provides consultation, teaching, and fosters research in the digital humanities. The Center provides workshops on topics such as data visualization, introduction to Text Encoding Initiative (TEI), and more. The Associate Director for Center for Digital Humanities was enthusiastic to add a research data management workshop and, having some experience with research data management, provided input and resources about specific needs in the digital humanities.

Literature Review

Over the last several decades scholarly research has become more electronic and data driven (Bell, Hey, and Szalay 2009). As the amount of research data has grown, so has attention to its importance as a scholarly product (Borgman 2008). In order to ensure research data are properly maintained, many funding agencies require data management plans with grant applications. Seeing an opportunity to share their expertise in these areas, many librarians have started providing data management plan and/or research data management training. In 2013, 61% of ARL libraries surveyed offered data management plan training and data management best practices (Fearon et al. 2013).

To help librarians apply their existing expertise to this new field of instruction, a variety of educational materials, curricula, and handbooks have been developed. Goben and Raszewski's *Research Data Management Self-Education for Librarians* provides a list of curricula, programs, online courses, and more that librarians can use to learn about research data management (Goben 2015). The *Data Literacy Course Materials* web page from the eScience Portal for Librarians provides a listing of materials aimed at faculty and students

(“Data Literacy Course Materials” 2017). The materials on the eScience Portal for Librarians are listed as either general or disciplinary resources. However, most of the general course materials have more of a science, engineering, and/or social science focus. Only two of the ten disciplinary course materials listed are for humanities disciplines, DataTrain¹ for archeology and Curating Artistic Research Output (CAiRO)².

The traditionally data-driven fields of the sciences and social sciences have also been the main focus for research data management library services. Of the libraries in the ARL Research Data Management Services survey that offer these services, most estimate their services are used by the humanities only occasionally (62%) or never (38%) (Fearon et al. 2013). However, humanities research is also increasingly becoming more electronic and data driven, especially with the rise of the digital humanities. The National Endowment of the Humanities (NEH) Office of Digital Humanities started requiring data management plans for grant applications in 2011. The reason for the requirement is to, among other things, consolidate information about data to a consistent format and “to encourage applicants and reviewers to think in broader terms about humanities data” (National Endowment for the Humanities 2017). Data management plan requirements for grants from the NEH Office of Digital Humanities are based on and contain elements similar to that of the NSF including types of data, how data will be maintained, access and sharing, and information about documentation and metadata (National Endowment for the Humanities 2016).

Although the NEH data management plan requirements are similar to those for the NSF, the emphasis in digital humanities literature has been on digital curation. According to Alex Poole, the American Council of Learned Societies “crystalized the unprecedented urgency of digital data curation in the humanities” with the 2006 report *Our Cultural Commonwealth* (Poole 2013). Lee and Tibbo define digital curation as involving “the management of digital objects over their *entire lifecycle*, ranging from pre-creation activities wherein systems are designed, and file formats and other data creation standards are established, through ongoing capture of evolving contextual information for digital assets housed in archival repositories” (Lee and Tibbo 2007).

By invoking the entire lifecycle of a digital object, this definition illustrates how digital curation encompasses and also extends beyond many of the activities involved in research data management. The Digital Curation Centre (DCC) provides the most robust description of the full curation lifecycle which covers the actions of conceptualize, create or receive, access and use, appraise and select, dispose, ingest, preservation action, reappraise, store, access, use and reuse, and transform (Digital Curation Centre 2017). The NSF and NEH data management plan requirements can serve as a good example of typical research data management activities including types of data, documentation and metadata, how data will be maintained, access and sharing, re-use, and archiving or preservation. Jake Carlson helps illuminate the differences between curation and research data management, describing curation as what takes place not just during the active phase of the research, but afterward when data may be stewarded by a third party. Data management on the other hand “takes place during the “active” stages of the data life cycle, when researchers are generating and making use of the

1 Lindsay Lloyd-Smith, “Archaeology Data Service,” Archaeology Data Service, accessed August 18, 2017, <http://archaeologydataservice.ac.uk/learning/DataTrain.xhtml>

2 Stephen Gray, Paul Clarke, and Sarah Jones, “CAiRO: Curating Artistic Research Output,” accessed August 18, 2017, <http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain/cairo.aspx>

data themselves for their own purposes” (Carlson 2014).

Since *Our Cultural Commonwealth* was published in 2006, several treatments on curation in the digital humanities have been written, but only two cover aspects of research data management. *The Digital Humanities Curation Guide* from Flanders and Muñoz is an online guide with an introduction and articles covering digital curation aspects particular to the digital humanities such as data representation, metadata standards, digital classics, digital collections, and legal aspects. *Digital Curation in the Digital Humanities* by Sabharwal provides a conceptual framework for digital humanities curation. Research data management concepts are mentioned in these but the treatments are high level, looking at the entire curation lifecycle and do not provide details on the basics needs of research data management in the digital humanities.

However, a disciplinary perspective and the ideal of digital curation should not be overlooked. Flanders and Muñoz’s introduction identifies specific curation challenges presented by prominent research objects in the digital humanities, such as scholarly editions, text corpora, text with markup, thematic research collections, data with accompanying analysis or annotation, and finding aids. They further identify unique treatments of humanities data that also require curatorial attention. Interpretive layering, information about how data is captured and prepared, as well as responsibility, editorial voice, and debate are aspects of the research methods that need to be retained for the long-term preservation and reuse of a work. (Flanders and Muñoz 2011) Given that the data management plan requirements for the NEH are similar to those for the NSF, applying this perspective to existing general training tools such as the New England Collaborative Data Management Curriculum (NECDMC)³ can be an effective way of developing materials for research data management instruction in the digital humanities.

Workshop Development

The best practices in data management workshop delivered to the broader campus community was developed based on three existing materials: NECDMC’s Module 1⁴, Kristin Briney’s ACRL DCIG presentation “Practical Data Management,”⁵ and the Create and Manage Data section of best practices on the U.K. Data Archive’s website.⁶ The general presentation covered reasons for managing research data, data management plans, file management, documentation, storage and backup, long term planning and included specific resources available on campus. This presentation served as the backbone for both the workshop to history department graduate students and the Center for Digital Humanities. The workshop for history department graduate students was co-presented with the History Department IT Manager and Princeton’s History Librarian, Elizabeth Bennet. The research data management

3 Lamar Soutter Library, University of Massachusetts Medical School, “New England Collaborative Data Management Curriculum | Lamar Soutter Library - University of Massachusetts Medical School,” <http://library.umassmed.edu/necdmc/index>

4 Elaine R. Martin et al., “Module 1: Overview of Research Data Management,” *Lamar Soutter Library - University of Massachusetts Medical School*, accessed February 10, 2017, <https://library.umassmed.edu/necdmc/modules>

5 Kristin Briney, “Practical Data Management - ACRL DCIG Webinar.” (April 30, 2014) <http://www.slideshare.net/kbriney/practical-data-management-acrl-dcig-webinar>

6 UK Data Archive, “UK Data Archive - CREATE & MANAGE DATA.” <http://www.data-archive.ac.uk/create-manage>

portion for this workshop was limited to NEH data management requirements, documentation, storage, back-up, and long-term preservation. The workshop for the Center for Digital Humanities included the core set of data management concepts used as the backbone for previous presentations but also included NEH data management requirements and information on special considerations for digital humanities drawn from *Introduction to Humanities Data Curation* including the importance of “interpretive layering”, provenance and documentation, and “capturing responsibility, editorial voice, and debate” (Flanders and Muñoz 2011) (See supplementary file, RDM-DigitalHumanities Presentation).

Results

The history department graduate student workshop on file and digital asset management was attended by around 20 incoming graduate students. The information presented on file and digital asset management by the departmental IT manager and history librarian was grounded in real-world examples of creating and managing digital objects in the reading room. This helped students contextualize application of the research data management concepts of documentation, storage, back-up, and long-term preservation to an early foundational stage of research. Student follow-up questions revolved around documentation and copyright issues.

The Center for Digital Humanities research data management workshop was attended by seven people ranging from Center for Digital Humanities staff and patrons, library staff, and one member of the general public. After the session, the eScience Librarian was invited to be paired with one of Center for Digital Humanities’ research fellows to consult on research data management throughout the life of a fellow’s digital humanities project. Unfortunately, this opportunity did not materialize due to a change in the eScience Librarian’s position.

Discussion

Program-level assessment was hoped to be achieved by pairing with a Center for Digital Humanities fellow. Such a pairing would provide an opportunity to directly observe pain points experienced in digital humanities work to develop more targeted instruction. Student questions and interest from both workshops suggests licensing and re-use are an area of need that should be developed further.

Research on assessment of student learning in research data management is still in its early stages and as with educational curriculum is largely focused on science and technology. The assessment built into the lesson plan for Module 1 of NECDMC calls for participants to discuss data management issues in their research setting or examine one of NECDMC’s sciences based case studies. Based on these early workshops, a few similar activities for assessing digital humanities student learning present themselves. Participants potentially involved in reading room research might be asked to discuss data management issues in this setting. For a more general digital humanities audience, assessment could include having students view existing digital humanities projects, discuss what type of data may be involved and special considerations that may apply to the project.

Conclusion

There is much interest in data curation in the digital humanities. Data curation represents a full

range of actions on a digital object over its lifecycle and includes the basics of data management. Despite this, little attention has been paid to providing research data management instruction to the digital humanities. While the data produced in scientific research differ from the digital humanities in many ways, both disciplines have similar themes when it comes to research data management as exhibited by the similarities in the data management plan requirements of the funding agencies NSF and NEH. Thus, with an understanding of what makes digital humanities data unique, existing research data management education curricula can be applied to the digital humanities.

Supplemental Content

RDM-DigitalHumanities Presentation

An online supplement to this article can be found at <http://dx.doi.org/10.7191/jeslib.2017.1115> under “Additional Files”.

Acknowledgement

The author would like to thank Jean Bauer, Associate Director of the Center for Digital Humanities at Princeton, Elizabeth Bennett, Librarian Emerita for History and History of Science at Princeton University, and Carla Zimowsk, Technology Services Manager for the History Department at Princeton University.

Disclosure

The author reports no conflict of interest.

References

- Bell, Gordon, Tony Hey, and Alex Szalay. 2009. “Beyond the Data Deluge.” *Science* 323(5919): 1297-1298. <http://www.dx.doi.org/10.1126/science.1170411>
- Borgman, Christine L. 2008. “Data, Disciplines, and Scholarly Publishing.” *Learned Publishing* 21(1): 29-38. <http://dx.doi.org/10.1087/095315108X254476>
- Carlson, Jake. 2014. “The Use of Life Cycle Models in Developing and Supporting Data Services.” In *Research Data Management: Practical Strategies for Information Professionals*, edited by Joyce M. Ray, 63-86. Purdue University Press. <http://www.thepress.purdue.edu/titles/research-data-management-practical-strategies>
- Digital Curation Centre. 2017. “What Is Digital Curation?” Accessed August 24. <http://www.dcc.ac.uk/digital-curation/what-digital-curation>
- E-Science Portal for Librarians. 2017. “Data Literacy Course Materials.” Accessed August 17. <http://esciencelibrary.umassmed.edu/data-literacy/course-materials>
- Fearon, David Jr, Betsy Gunia, Sherry Lake, Barbara E. Pralle, and Andrew L. Sallans. 2013. “SPEC Kit 334: Research Data Management Services.” *Association of Research Libraries*. <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334>
- Flanders, Julia, and Trevor Muñoz. 2011. “An Introduction to Humanities Data Curation.” *DH Curation Guide: a community resource guide to data curation in the digital humanities*. <http://guide.dhcuration.org/contents/intro/>
- Goben, Abigail and Rebecca Raszewski. 2015. “Research Data Management Self-Education for Librarians: A Webliography.” *Issues in Science & Technology Librarianship* 82. <http://dx.doi.org/10.5062/F4348HCK>

National Endowment for the Humanities. 2017. "Data Management Plans for NEH Office of Digital Humanities Proposals and Awards." <https://www.neh.gov/content/data-management-plans-pdf>

———. 2015. "Data Management Plans From Successful Grant Applications (2011-2014) Now Available." <https://www.neh.gov/divisions/odh/grant-news/data-management-plans-successful-grant-applications-2011-2014-now-available>

Poole, Alex H. 2013. "Now Is the Future Now? The Urgency of Digital Curation in the Digital Humanities" *Digital Humanities Quarterly* 7(2). <http://www.digitalhumanities.org/dhq/vol/7/2/000163/000163.html>