Putting Research Data into Context: Scholarly, Professional, and Educational Approaches to Curating Data for Reuse

Ixchel M. Faniel  
OCLC Research  
6565 Kilgour Place  
Dublin, OH 43017-3395  
fanieli@oclc.org

David Minor  
University of San Diego Library  
University of San Diego  
9500 Gilman Dr.  
La Jolla, CA 92093-0699  
dminor@ucsd.edu

Carole L. Palmer  
Graduate School of Library and Information Science  
University of Illinois at Urbana-Champaign  
501 E. Daniel Street  
Champaign, IL 61820-6212  
clpalmer@illinois.edu

ABSTRACT
In this panel session we will discuss approaches to data curation for data reuse, with particular emphasis on the contextual information needed for reuse. We invite three panelists to present perspectives from communities that are 1) conducting scholarly research in the area of data reuse, 2) performing data curation and preservation work, and 3) educating information professionals. Given the shared interests among the three communities, one of the major objectives of the panel session is to deepen the engagement and connection among these groups. We have planned interaction with the audience to integrate the panelists’ different perspectives and approaches, outline common objectives, and explore ways we might create a common agenda in the area of data curation for data reuse.

Keywords
Data curation, data curation education, data reuse

INTRODUCTION
Data curation is an important, emerging topic in academia, given federal funding agency mandates, federal policies, and recent initiatives within higher education and research communities. We have seen federal funding agency mandates for data sharing and requirements for data management plans (National Institutes of Health, 2003; National Science Foundation, 2009, 2011). In February 2013, the White House Office of Science and Technology Policy issued a memorandum for science agencies to develop public access policies for research outputs, including data (Holdren, 2013). Most recently the Association of Research Libraries (ARL), the Association of American Universities (AAU) and The Association of Public and Land-grant Universities (APLU) have partnered around a SHared Access Research Ecosystem (SHARE) initiative. “The SHARE partnership is seeking ways in which these universities can be in compliance while transforming aspects of the research ecosystem to better meet the information and data-related needs of its stakeholders” (Walters & Ruttenberg, 2014, p. 56).

Data curation is a complex phenomenon as well. According to the Digital Curation Centre, it “involves maintaining, preserving and adding value to digital research data throughout its lifecycle” (The Digital Curation Centre, 2014). Consequently curation should occur at the inception of a research project with data management planning and be carried through to the selection, appraisal, deposit, preservation, and reuse of the data (Jahnke, Asher, Keralis, & Henry, 2012; The Digital Curation Centre, 2014). Given the complexity of the topic, this interactive panel session focuses on one aspect of data curation – curation for reuse. We are particularly interested in the contextual information needed to reuse the data. This is perhaps one of the thorniest issues given there are few studies of data reuse practices, few models of how to address the differences in data within and across disciplines and varied stakeholder needs, and limited resources for developing data management and curation skills. Even in the face of these challenges, we believe that three key communities within academia are making progress – scholarly researchers, information professionals, and educators. We call on representatives from these three communities to respond to the following question posed during this interactive panel session: When putting research data into context, how are scholarly researchers, information professionals, and educators approaching data curation for data reuse?
PUTTING RESEARCH DATA INTO CONTEXT: APPROACHES TO DATA CURATION FOR DATA REUSE

Scholars have been examining the data sharing, management, and reuse needs of faculty and students to understand their data curation needs, including the social, organizational, and technical infrastructure needed to support them (Faniel, Kansa, Whitcher Kansa, Barrera-Gomez, & Yakel, 2013; Tenopir et al., 2011). In practice, academic librarians and to a lesser extent archivists have taken a central role on their college and university campuses by offering research data services to meet the needs of faculty and students (McLure, Level, Cranston, Oehlerts, & Culbertson, 2014; Ray, 2014). In the area of education, several data or digital curation degree concentration and certificate programs have been developed at major universities across the United States, including the University of North Carolina at Chapel Hill, and the University of North Texas, to prepare a future workforce of information professionals (Harris-Pierce & Liu, 2012; Palmer, Weber, Muñoz, & Renear, 2013). Continuing education programs and workshops focused on skill development and community building are also being offered (Jahnke et al., 2012).

Given the overlapping interest of scholarly researchers, information professionals, and educators, one would expect the connections among the communities to be strong and well integrated, and in some cases they are. For instance, the curriculum at the University of Illinois at Urbana-Champaign is grounded in empirical work on data practices conducted by the faculty and other scholarly researchers. In addition, students gain hands on experience working with information professionals in the field during internships. This type of integration is important, but more often than not scholarly researchers, information professionals, and educators tend to observe, rather than deeply engage with each another. We need to develop additional ways in which we can be more coordinated across research, practice, and education.

The purpose of this interactive panel session is to begin the work to deepen the engagement and connections. We do this by inviting three panelists to present their approaches to data curation for data reuse, with an emphasis on putting research data into context. Each panelist will take one of three perspectives: that of the scholarly researcher, the information professional, or the educator. After the presentations, we will begin a conversation with the audience that centers on integrating the panelists’ perspectives and approaches in order to create the synergies needed to outline common objectives and an integrated agenda.

The major focus of this panel is well aligned with the conference theme, “Connecting Collections, Cultures, and Communities,” because it aims to forge better connections among the communities conducting scholarly research, performing data curation and preservation work, and educating information professionals. We believe these connections will result in better, more integrated and accessible data collections.

A Scholarly Researcher’s Perspective

Dr. Ixchel M. Faniel is an Associate Research Scientist at OCLC Research. She will take the perspective of the scholarly researcher. Her presentation will draw from her work as Principal Investigator on the Dissemination Information Packages of Information Reuse (DIPIR) project (http://www.dipir.org). The DIPIR project is investigating data reuse in three disciplines – quantitative social science, archaeology, and zoology – to identify how to best create and preserve contextual information that supports data reuse. She will discuss findings from an analysis of data reuse practices within and across the disciplines and address two research questions:

1. What are the significant properties of quantitative social science, archaeological, and zoological data that facilitate reuse?
2. How can these significant properties be expressed as representation information to ensure the preservation of meaning and enable data reuse?

An Information Professional’s Perspective

David Minor is Director of the Research Data Curation Program at the UC San Diego Library. He will take the perspective of the information professional performing data curation work. His program has been developing a campus-based approach to data curation for a wide range of research disciplines. This development began with a two year pilot program, examining the needs and expectations of selected campus research labs. From the pilot, a set of tools and services has been put into place to address the management of the entire data lifecycle, from data creation to discovery, publishing and preservation. He will discuss the core issues that have surfaced in this process, including:

1. What commonalities can be found among the needs of diverse research communities? What significant differences exist, and how can they be addressed effectively?
2. How can multiple campus organizations, including the Library, central computing and large technology institutions, work together to provide the needed spectrum of services?

An Educator’s Perspective

Carole L. Palmer is Director of the Center for Informatics Research in Science and Scholarship (CIRSS) and Professor in the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign. She leads research teams investigating data curation needs across disciplines and the reuse value of data, and has been integrating that work into educational programs in data curation for nearly a decade. She currently serves on the National Academy of Sciences study...
committee on preparing the digital curation workforce and is an active speaker and organizer of data curation initiatives at the intersection of research, practice, and education. She will present on the critical need for data curation curriculum to be grounded in empirical studies of the context of data production and reuse and the advances made by data service providers in libraries and other organizations supporting and promoting data reuse.

THE CONVERSATION

To ensure that there is enough time to engage with the audience, each presenter will have 15-20 minutes to present their approach to Putting Research Data into Context. The remaining time will be used to engage the audience in a brainstorming conversation to address the following questions.

- How do the three approaches complement each other? In what ways do they conflict?
- Are there major themes that we might use to begin to establish a common agenda?
- How can we begin to build connections and community among research investigators, information professionals, and educators?
- What are the challenges to building connections and community given past experiences and how might we overcome them?

We do not expect to reach the perfect solution for “Connecting Collections, Culture, and Communities” by the conclusion of this interactive panel session. Instead, we anticipate that this session will serve as a kickoff to ongoing conversations among scholarly researchers, information professionals, and educators during ASIS&T and other meetings. We expect that the ongoing conversation will lead to a better understanding about where we can 1) build a shared agenda, 2) develop future collaborative projects and grant proposals, and 3) co-author conference presentations and journal articles that result in a more synergistic approach to data curation.

ACKNOWLEDGMENTS

The DIPIR Project was made possible by a National Leadership Grant from the Institute for Museum and Library Services (IMLS), LG-06-10-0140-10, “Dissemination Information Packages for Information Reuse.” Carole Palmer wishes to acknowledge support from the National Science Foundation for the Data Conservancy initiative (OCI-0830976) and IMLS for the Site-Based Data Curation project (LG-06-12-0706-12) and the Data Curation Education at Research Centers (DCERC) project (RE-02-10-0004-10).

REFERENCES


The columns on the last page should be of approximately equal length.