

## EDITOR'S SUMMARY

Three postdoctoral fellows in a program sponsored by the Council on Library and Information Resources/Digital Library Foundation are exploring and contributing to the field of digital curation through very different perspectives. With a neuroscience background, Katherine Akers is encouraging scientists to preserve and share research datasets and analyzing the use of library resources. For Inna Kouper, building cyberstructure and facilitating and promoting user engagement are primary. Matthew Lavin is working to make digital tools and approaches serve the needs of humanists, focusing on digitally conveying the physical features and histories of books. With different definitions of data and a variety of research goals, the scholars apply hybrid professional approaches to digital curation, stimulating expanded information, intellectual cross fertilization and a broader view of data, research and knowledge.

## KEYWORDS

data curation  
 research and development  
 interdisciplinarity  
 cross disciplinary fertilization  
 research data sets  
 information infrastructure  
 usage studies  
 engagement  
 books  
 materials preservation

# Data Curators at Work: Focus on Projects and Experiences

by Inna Kouper, Katherine Akers and Matthew Lavin

The Council on Library and Information Resources/Digital Library Federation (CLIR/DLF) postdoctoral fellowships in academic libraries and data curation encourage the development of digital curation as a hybrid profession by hosting fellows with a variety of disciplinary backgrounds, roles and responsibilities. As the authors of the grant proposal that partially supported the new CLIR/DLF fellowships explained:

“Facing ... a shortage of experienced data scientists and curators with digital preservation experience, disciplinary specialists must begin to accept responsibility for helping in this vital scholarly work. One digital curation specialist cannot possibly manage the full and complex range of ... responsibilities. Instead, teams of ... curators with a variety of sub-specialist knowledge, talents and perspectives, working together with a shared understanding of the need for active ... curation, will be required to enable future scholarship in all disciplines.”

Here, we expand upon the concept of professional hybridity

Inna Kouper is a Council on Library and Information Resources (CLIR) data curation postdoctoral fellow. She works at the Data to Insight Center, Indiana University Bloomington. She can be reached at [inkouper@indiana.edu](mailto:inkouper@indiana.edu).

Katherine Akers is a Council on Library and Information Resources (CLIR) postdoctoral fellow. She commenced her fellowship at Emory University Libraries and is now at University of Michigan Libraries. She can be reached at [kgakers@umich.edu](mailto:kgakers@umich.edu).

Matthew Lavin is associate program coordinator, Crossing Boundaries, at St. Lawrence University, Lincoln, Nebraska. He can be reached at [mlavin@stlawu.edu](mailto:mlavin@stlawu.edu).

by describing our projects and experiences as CLIR postdoctoral fellows and by reflecting on how the variety of roles that we inhabit influence digital curation.

**Katherine Akers:** Coming to the library from a neuroscience laboratory, I embrace the “scholar-librarian” mindset while being service-oriented in my work. Through my engagement in a myriad of library projects, including helping to develop research data management services and assess the use of library collections, I am gradually building a framework for understanding academic librarianship. In particular, I am increasingly struck by how my library projects – seemingly disparate at first – are beginning to inform each other. For instance, when interviewing faculty about their research data management practices, my default is to encourage researchers to preserve and share their datasets (all of them!) because it is difficult to predict their potential value. But when analyzing the usage of library databases, journals and books, I sometimes question the value of collecting resources without knowing whether they are subsequently used. I’m realizing that in both cases, knowledge of how existing resources are used should inform future preservation or collection decisions. The links between conducting research and supporting research, as well as the connections beginning to emerge among my separate library projects, add a whole new dimension to my understanding of how scholarly information is created and disseminated.

**Inna Kouper:** For me, data curation primarily concerns two things – cyberinfrastructure and user engagement. I

am primarily involved with the NSF-funded Sustainable Environment, Actionable Data (SEAD) project (<http://sead-data.net>) that aims to create infrastructure to support data curation and management workflows in sustainability science. As a team, we face many architectural and policy decisions daily. How do we support ingest and discovery of heterogeneous datasets? What kind of interface can serve the dual needs of researchers and data managers? I see my role as a curator with social science and information science expertise in making sure that our decisions are well-grounded and that we maintain a balance between availability of resources and user needs. In parallel with building cyberinfrastructure, data curators need to ensure that cyberinfrastructure will be used. Much is being said about how data sharing and preservation leads to more reliable knowledge and improves citations. Recent discussions, including that of Reinhart and Rogoff’s “Growth in a Time of Debt,” demonstrate that data provenance and availability are crucial for the validity, reproducibility and trustworthiness of scientific results. This importance is why my other role is to articulate and disseminate norms and values of data sharing and preservation and to find “data champions” to engage researchers in data exchanges.

**Matthew Lavin:** I am working with other humanists to adapt digital tools and approaches to their disciplines. I’ve spent the year learning about digital scholarship, advancing my technical skills and working on expanding the archive titled “*Death Comes for the Archbishop: A Digital History of the Book*,” which creates an interface to visualize and analyze multiple copies of Willa Cather’s book *Death Comes for the Archbishop*, including inscribed copies, translations, notable editions and issues, and manuscripts. My project takes a bibliographical and book historical approach to digital humanities by (1) establishing a nuanced data model for

bibliographical information, (2) facilitating visualizations of physical book features over time and (3) linking the data structure to high-resolution page scans. Although some scholars may not see this project as being related to data curation, I am approaching the study of book history through its most concrete data – the measurable properties of physical objects. Bibliographers have fastidiously recorded information about books in analog form for hundreds of years, but less has been done to consolidate, share and analyze these data in digital form. Humanists can collectively contribute to data curation when they begin to think about how to articulate data-driven questions, how to create or adapt shareable data and how to leverage this data with renewed rigor.

As these sketches illustrate, digital/data curators take on a variety of responsibilities, including surveying faculty about data management practices, analyzing library data, and building and adapting tools. Moreover, the definition of data differs among curators – from data created in different research domains to information on academic behavior, library collections, services and facilities, and the physical properties of books. Being a hybrid data curator means being open to and flexible about what to curate while maintaining scientific rigor during the curation process. One of the most remarkable aspects of our work is that it makes us open to serendipitous discoveries via the intentional expansion of knowledge bases, cross-fertilization among projects, connections between previous research and current practices, and conceptual thinking about data-research-knowledge ecosystems. Our experiences, over time, are helping to inform CLIR about the roles that subject specialists can play in data management of research data, how postdoctoral expertise complements the work of other academic professionals and the types of training that might benefit students preparing for careers in digital curation. ■