

Scientists and Social Scientists Working in Data Curation: The Inaugural CLIR/DLF Cohorts

Alice Bishop and Christa Williford

In 2015, the U.S. National Research Council (NRC) released [Preparing the Workforce for Digital Curation](#), a report examining the critical need for education and training in digital curation to address the increasing demand for access to and meaningful reuse of digital information. Prepared by the Board on Research Data and Information (BRDI), the report states, “There is no single occupational category for digital curators and no precise mapping between the knowledge and skills needed for digital curation and existing professions, careers, or job titles.” One of its conclusions is that the knowledge and skills required of those engaged in digital curation are highly dynamic and highly interdisciplinary. They encompass an integrated understanding of computing and information science, librarianship, archival practice, and the disciplines and domains generating and using data. The report cites the increase in the number of domain experts learning digital curation as an important development and states, “One promising model for transitioning domain experts into curation work has been developed by the Council on Library and Information Resources (CLIR)/Digital Library Federation (DLF) as an extension of their Postdoctoral Fellowship Program” (71).

Since 2004 CLIR’s Postdoctoral Fellowship Program has recruited, trained, and established cohorts of new PhDs working within the digital environment to help manage, sustain, and generate valuable information in support of higher education. Beginning in 2012 with support from the Alfred P. Sloan Foundation, the program deepened and sharpened its focus on research data curation, a focus that persists to this day. Between 2012 and 2014, CLIR and DLF collaborated to bring 23 scientists and social scientists into data curation fellowships at 20 host institutions in the United States and Canada. The goal was for these fellows to contribute to a more sophisticated understanding of data curation and its often-determining role in the conduct of scientific and social scientific research. What impact did these initial cohorts have? What kinds of ongoing needs and challenges do the fellowships reveal? And what are the implications of their experiences and potential for the stewardship of scholarly research data?

As a follow up to their 2012 CLIR report [The Problem of Data](#), anthropologists Lori Jahnke and Andrew Asher interviewed data curation fellows and their colleagues to assess the impact of the initial three cohorts supported by the Sloan Foundation. In “Ongoing Challenges for Data Curation Support: A Program Assessment of the 2012–2014 CLIR/DLF Postdoctoral Fellowships in Data Curation for the Sciences and Social

Sciences,” they conclude that, “considerable work is still needed to develop a system of data curation that supports the preservation and access of research data while allowing researchers to fulfill their ethical and professional obligations, and an improved understanding of research processes and their variance among disciplines is critical to this work.” Clarifying tools, standards, and workflows for research data curation practice is only part of this work: Jahnke and Asher point out that shifts in academic cultures are also necessary to recognize and reward the contributions of curators who create useful and reusable datasets that are instrumental to the advancement of science. Through creating the CLIR/DLF fellowships in data curation, CLIR aspired to trigger new ways of thinking about divisions of labor and expertise across the academic professions while bringing focused attention to the development of data curation resources and services that would be responsive to the needs of current researchers.

Fellows’ Disciplines

With the addition of the data curation fellowships in 2012, the proportion of scientists and social scientists in CLIR’s Postdoctoral Fellowship Program increased significantly. When the program was founded, fellows came exclusively from training programs in the humanities or, occasionally, from information science. The first social scientists from other fields joined the program in 2009; in 2010, the program welcomed the first fellows trained in natural sciences.

To recruit fellows for the new data curation fellowships, current and former postdoctoral fellows helped solicit candidates through their own networks and communities, while CLIR staff reached out to programs and departments known for data-intensive research to inquire about suitable PhD candidates in the sciences and social sciences. Staff worked closely with host institutions throughout the recruitment and hiring processes.

The candidates hired for the 2012-2014 CLIR/DLF Postdoctoral Fellowships in Data Curation for the Sciences and Social Sciences came from the following 16 disciplines:

Anthropology	2
Archaeology	2
Biology/Biological Sciences	2
Biomedical Informatics	1
Chemistry	1
Comparative Literature	1
Economics	1
Educational Research and Policy Analysis	1
Environmental Studies	1

Geological Sciences	1
History of Consciousness	1
Information Science	4
Neuroscience	1
Political Science	1
Psychology	1
Sociology	2

The high number of candidates in information science suggests a greater awareness of the need for new interdisciplinary work in data curation among iSchool PhDs than among those in the sciences and social sciences. In many disciplines with a strong tradition of postdoctoral research and training, such as in laboratory-based sciences, pursuing a library-based fellowship following the degree was, and still is, unusual. As a result, CLIR continues to invest time and energy in recruiting more disciplinary perspectives into the fellowships, since it is the fellows' academic grounding that helps the program facilitate meaningful, productive connections between scholarly and professional communities of practice.

The attraction of the program for talented but diversely trained young scholars has remained strong since the CLIR/DLF data curation fellowships were created. Of the 15 institutions that hosted fellows through CLIR's program for the first time during the first three years of Sloan Foundation support, 8 have since chosen to host a second fellow, and 4 of these have hosted a third. These numbers suggest that the variety of perspectives the program incorporates is an asset for the ongoing development of research data collections and services.

The value of disciplinary experience for success in a professional role in research data curation can be unpredictable. While hosts identify the specific qualifications they seek in their position descriptions, CLIR and DLF have found that when reviewing the applicant pool, hosts sometimes hire fellows with disciplinary backgrounds that are different from those originally anticipated. That was the case in 2014 when one host institution chose a fellow with a comparative literature degree. The university had to obtain an exemption from its human resources administrative office before hiring the fellow since the original fellowship description had called for a PhD in the social sciences. Regardless of their degrees, many candidates in CLIR's applicant pool stand out not just for their strong disciplinary and methodological skill sets but also for their contextual understanding and desire to make data useful and reusable across disciplines and institutional boundaries.

Fellows' Career Outcomes

Regardless of how fellows come to the program, they generally report that they gained valuable experience that helped them build careers in both academic libraries and on the tenure track. Of the 23 data curation fellows brought into CLIR/DLF fellowships through the Sloan Foundation's support between 2012 and 2014, more than half now hold permanent positions in academic libraries, either at their original host institutions or elsewhere. The majority of the remaining former fellows from these first three cohorts now hold positions in the field of data management and are engaged in closely related work in research and research support, software creation, assessment, or user experience design.

2012-2014 Sloan-Supported Fellows' Career Outcomes (as of April 2019)

Library	56% (13/23)
Tenure track faculty	4% (1/23)
Other non-academic jobs	13% (4/23)
Other academic jobs	13% (3/23)
Other postdoctoral fellowship	4% (1/23)
Unknown	4% (1/23)

The number of these 23 fellows who were hired by their host institutions immediately following the fellowship into permanent roles (39%) is healthy and the regular submission of new job descriptions to CLIR with requests to post specifically to the postdoctoral fellowship community indicates that these fellows share skills and perspectives that are in high demand. However, given the relatively small sample and the uniqueness of the circumstances affecting each fellow's and each host's choices, it is difficult to draw conclusions about which factors are most closely associated with the conversion of data curation fellowships into longer term careers in data curation. CLIR's Postdoctoral Fellowship Program remains a small part of a much larger narrative involving many other organizations, funders, institutions, and individuals. It is a persistent challenge to distinguish between individual fellows' influence on changes in research and research support practices, and the influence of other, larger networks of professionals to which fellows have become connected.

The Larger Data Job Context

As a way to contextualize the earliest CLIR/DLF data curation fellows in the sciences and social sciences within the broader academic library job market, CLIR approached research consultant and current CLIR Program Officer Jodi Reeves Eyre—who was

herself among the 2013–2015 fellows—to undertake an analysis of the responsibilities and qualifications listed in data job descriptions posted by academic libraries between 2013 and 2017, the years during which CLIR’s earliest data curation fellows entered the job market. In her paper, she discusses the ways that the qualifications and perspectives of recent PhDs in the sciences and social sciences may or may not suit the positions as they were described. Her findings indicate that if research and subject knowledge, experience, and skill requirements are the only factors taken into account, then recent PhDs in the sciences and social sciences are strong candidates for data jobs. However, many data jobs listed during this period required a library degree, library experience, or both; these stipulations can pose insurmountable barriers for recent PhDs interested in research data curation. Reeves Eyre’s research also indicates that emphasizing experience in a service or consulting role in data job listings poses additional challenges for recent PhD scientists and social scientists who are more likely to have worked collaboratively in laboratories, in the field, or in the classroom teaching.

What’s Next?

Sustaining the Postdoctoral Fellowship Program long-term has never been a major objective for CLIR. The program was designed to help academic library leaders envision a sustainable future, expanding the pool of talent available to replace those leaders in the coming decades. CLIR intends to remain involved in the program only as long as it genuinely serves these purposes and CLIR’s mission.

Expanding the range of roles and disciplines in recent years has dramatically increased the size of the program, as well as the breadth of stakeholders who provide support to the program and receive its benefits. There is no sign of waning interest, and for the near term CLIR expects to attract healthy numbers of host institutions and candidates. By continuing to cultivate cohorts of differently trained yet like-minded researchers who work together to explore the challenges that the collection, organization, and preservation of data pose for the future of scientific research, CLIR aspires to help build more robust shared information systems, inculcate more deeply interdependent relationships among institutions of higher learning, and develop a more flexible and multifunctional human infrastructure to maintain and strengthen those relationships in the future.