We live in the age of data. As the popular statement says, “data is the new oil.” It is a growing resource with high potential. In order to realize this potential, we need to determine how to curate data effectively.

Data curation at its simplest is taking care of resources. It involves the selection, appraisal, storage and dissemination of objects and collections. With regard to data, particularly digital data, it is important to conceptualize and support the whole data lifecycle because without a proper understanding of how data are created and used it is almost impossible to store and preserve it properly. And vice versa – without adequate storage and preservation, data can quickly become obsolete and unusable. Data curation can easily become the responsibility of both everyone and no one. By understanding the data lifecycle, a data curator can alleviate this problem.

The data curator occupies a hybrid role somewhere among the traditional researcher, librarian, early technology adopter and policy maker. Clearly, neither current library and information science curricula nor extra-academic certification programs are ready to prepare people for this complex function. In an attempt to fill this gap, as well as to raise awareness of the significance of the work of such hybrid professionals, the Council on Library and Information Resources (CLIR) has expanded its postdoctoral fellowship in academic libraries to create the Data Curation Fellowship Program (www.clir.org/fellowships/postdoctoral-fellowship-in-academic-libraries-new). The program, which is co-sponsored by the CLIR Digital Library Federation (CLIR/DLF), the Alfred P. Sloan Foundation and participating universities, is aimed at establishing data curation as a profession and providing opportunities for on-the-job training. Recent Ph.D. s already have domain-specific knowledge, as well as expertise in data collection and analysis. During this two-year postdoctoral fellowship, participants are expected to gain experience in data storage, organization, preservation and dissemination. For the 2012-2013 fellowship, there are seven fellows hosted in six institutions: B. Dewayne Branch at Purdue University; Vessela Ensberg at University of California, Los Angeles; Inna Kouper at Indiana University; Ting Wang at Lehigh University; Wei Yang at McMaster University; and Fe Consolacion Sferdean and Natsuko Hayashi Nicholls at University of Michigan.

Inna Kouper is data curation fellow, Data to Insight Center, Indiana University. She can be reached at inkouper<at> indiana.edu.
As a fellow of this program, I participated in a two-week immersion seminar at Bryn Mawr College with other CLIR fellows (from both the data curation and academic libraries programs). The seminar brought library practitioners and researchers together to share our experiences with data and data initiatives and discuss issues facing libraries and academic institutions. The participants came from an astonishingly wide range of domains – from educational research and political sciences to geology and molecular biology. We knew that developing a common language and culture across diverse disciplines would be important, but were not clear where to begin. As someone with experience in libraries and information studies and interest in cross-disciplinary research, I was amazed by some deep disconnects between research areas, researchers and libraries, and libraries and technologies. There is a lot to be done if we want to establish and promote a common culture of data sharing.

In addition to recognizing diversity, participants discussed the changing role of the library and the challenges of developing infrastructure to support data curation. Infrastructure development is not merely a technical/engineering issue – data curators need to be aware of historical context and socio-political issues, as well as the complexities of everyday practice. Because data curation involves many different stakeholders, it is important to ease tension and facilitate consensus among universities, libraries, granting agencies and creative individuals. We also talked about collaborations and interdisciplinarity. Despite the great diversity among fellows and the uncertainties involved in data curation, everyone expressed excitement to work with different kinds of people and materials and do what is called boundary work, that is, connecting people across data types, technologies, disciplines and institutions.

It has been almost two months since the seminar ended and the fellows dispersed to their new places of work. We plan to meet online every month to reconnect, share successes and frustrations, explore potential collaborations and identify more training opportunities. From our first meeting in September, it seems that fellows have been able to engage with the challenges at work and take on the CLIR mission to raise awareness and build capacities for data management and sharing throughout the academy. As CLIR program officer Christa Williford kindly noted, “This cohort of data curation fellows has exceeded our expectations with their level of engagement, problem-solving skills and openness to new ideas. We are learning a great deal from them.”

Along with this promising beginning come many uncertainties. We chose an unusual, trailblazing path in our academic careers. We do not know whether this path will allow us to acquire enough symbolic capital to be recognized as serious researchers. We also do not know whether data curation as a profession will gain high, Ph.D.-required status or whether some of us will have to choose between traditional research and professional careers. We also face the challenges that all academics face – balancing research, teaching and service activities; making an impact; and living a productive life. In one of next year’s RDAP columns, we will address some of those issues and provide a more detailed report about our experiences as hybrid data curation scholars.