Following the sudden flood of interest in research data, information professionals have played an important role in communicating to the greater research community the advantages, as well as concerns, associated with sharing and reusing research data. The new task of academic librarians is to develop the best practices and methods for how to make data and other valuable forms of research output available to a wider audience. In response to this challenge, UCLA is developing a data registry, which may be one step toward dealing with the data deluge – one that other academic libraries can also explore.

Many different tools are aimed at making digital data and research output discoverable by others. Two significant efforts are institutional repositories and data repositories. While originally designed to store and maintain access to electronic documents, such as article pre-prints or theses and dissertations, many institutional repositories have begun accepting research data as well. In addition, certain disciplines have domain-specific data repositories dedicated to caring for the datasets produced in those fields. Data repositories are a premier tool for discovering data; however, they are usually geared toward the data produced by a single discipline, limiting their scope. Many fields, especially in the humanities, lack the support necessary for managing the digital content they produce.

In light of this need, additional efforts are underway to assist those interested in managing their data. The data registry, for example, has recently gained international recognition as a possible solution. A data registry functions much as a library catalog: it maintains descriptive records of datasets, allowing the end user to discover the data via a surrogate record, rather than containing data themselves. A data registry can also include a link to the location of the data, allowing interested parties to contact the original researcher directly to discuss a possible data exchange. For libraries that are unable to host a data repository, a data registry may be a cheaper and more viable solution to maintaining access to the data produced by campus researchers.

The UCLA library first became involved in such an enterprise in July 2011, when Christine Borgman, professor of information studies, and Todd Grappone of the library received grant funding from the Institute for...
Digital Research and Education (IDRE). The goal of the UCLA Data Registry project was to make a data registry that was both general enough to accommodate multidisciplinary data and specific enough to contain the details necessary to ensure data location and usability. The UCLA library saw this project as a first step toward developing more integrated data management tools.

Project personnel interviewed 20 researchers from disparate fields to determine a) if they would be interested in registering data with the UCLA library and b) what kind of data they would be willing to register. Based on interviews conducted between January 2012 and March 2012, project personnel determined that certain kinds of research projects would benefit from the UCLA Data Registry. For scholars with data stored on UCLA servers, a registry page would provide additional exposure to their data. Some researchers welcomed such supplementary methods for making their data available, because publications only reflect a limited amount of the data collected for a project. Interviewees also acknowledged incentives to registering their data, such as fulfilling funding requirements. In addition, they highlighted motivations that extend beyond data management services, such as the opportunity to establish new collaborations.

Humanists, who typically lack the infrastructure to support the discovery of their non-published work, also noted that they would benefit from registering their projects. However, one of the most interesting interview findings was that, while humanists were among the most interested in registering their work, conversations with them almost always broke down when participants were asked to discuss their “data.” For many of these researchers, curating or sharing the “data” is meaningless without the accompanying software and tools that are created by these projects. This contextual need poses a new challenge for the data registry, as well as data management tools in general.

As of the summer of 2012, the UCLA Data Registry was in its early prototyping and user-testing stage of development. The UCLA library is building this tool with hopes that it will link to current or future tools that build on and add value to its services and functions. In many cases, the most effective tools are not stand-alone services, but instead work in conjunction with other tools to function as a suite of services that collectively form an information infrastructure or a “value chain of scholarship” [1]. Links between journal publications and the datasets on which the articles are based allow an interested user to discover either the dataset or the publication first and then easily access the other components in the chain.

In the long run, the UCLA Data Registry may not be the tool researchers need to help make their research discoverable. The speed at which scholarship and scholarly practices change makes it difficult for academic librarians to predict what kinds of tools will truly benefit research. For this reason, it will be important for UCLA and for other academic libraries to continue collaborating with researchers to implement useful services that do not impose standards or force services where they are not helpful.

---

Resource Mentioned in the Article