Giving them what they want: Using Data Curation Profiles to guide Datastar development

Sarah Wright1, Wendy Kozlowski1, Dianne Dietrich1, Huda Khan1, Leslie McIntosh2, Gail Steinhart1

(1) Cornell University Library, (2) Washington University at St. Louis

The Datastar project, which first received NSF funding in 2007, is one of Cornell University Library’s (CUL) efforts in the area of digital data discovery (Steinhart 2010; Dietrich 2010; Khan 2011). Originally envisioned as a “data staging repository,” where researchers could upload data, create minimal metadata, and share data with selected colleagues, Datastar has been reconceived as a data registry to support discovery of datasets. Datastar is now being developed in partnership with Washington University at St. Louis (WUSTL) as a data registry that can be used either as a standalone tool or in conjunction with VIVO, Cornell’s open source semantic web application for research and scholarship networks. These more recent efforts were funded by the U.S. Institute of Museum and Library Services.

With the new focus, we wanted to ensure development decisions were driven by real user needs. To that end, we used the Data Curation Toolkit (<http://datacurationprofiles.org/>) to conduct interviews and create a set of Data Curation Profiles (DCPs) with participants selected at CU and WUSTL (Witt and others 2009). Designed to elicit information needs associated with a data set or collection, the structured interviews provided a general framework for discussing data with researchers and to create a profile covering the stages of the data lifecycle. Participants from a broad range of disciplines were invited to participate, and eight completed interviews.

After evaluation and prioritization of the findings from the interviews, a set of particularly relevant responses emerged. These highly relevant responses had a direct influence on Datastar development. In contrast, some findings from the DCPs were out of scope for the current iteration of Datastar, but they were still interesting because they helped to provide additional details about the ways researchers prefer to interact with their data. Although not all of the findings will result in functionalities in current or later iterations of Datastar, they will almost certainly inform future development. Using the DCPs allowed us to focus Datastar development efforts on addressing real user needs, saving wasted effort. This presentation will discuss both the results of our interviews and the development efforts that were prioritized as a result of our findings.

Dietrich, Dianne. 2010. Metadata management in a data staging repository. Journal of Library Metadata 10 (2) (04): 79-98.

Khan, Huda J., Brian Caruso, Jon Corson-Rikert, Dianne Dietrich, Brian Lowe, and Gail Steinhart. 2011. DataStaR: Using the semantic web approach for data curation. Internation Journal of Digital Curation 6 (2): 209-21.

Steinhart, Gail. 2010. DataStaR: A data staging repository to support the sharing and publication of research data. West Lafayette, IN.

Witt M, Carlson J, Brandt DS, Cragin MH. 2009. Constructing data curation profiles. International Journal of Digital Curation 4(3):93-103.