Library Curation of Long-tail Science Data

P. Bryan Heidorn

[1]Director, School of Information Resources and Library Science, University of Arizona, 1515 E First St, Tucson, AZ 85719 USA, *Heidorn@u.arizona.edu*

Academic libraries are beginning to take a role in the management of scientific data sets for researchers within major universities. While some forms of data such as some genomic datasets and astronomy data have global or national repositories, other data types have no institutional or disciplinary repositories leaving scientists to curate and preserve their own data. Without training or long-term support, this approach has not been effective for the sustainability of and access to digital data. A number of schools of library and information science have begun formal training programs in data curation. In the United States, the most notable are the University of Arizona, the University of Illinois, and the University of North Carolina. A number of studies have been conducted to identify the skill sets required to manage data repositories within the library context alongside the traditional role of libraries in collecting and disseminating scholarly materials. Some of the standards used in science have been adopted from the library community including, for example, Darwin Core, which was inspired by Dublin Core. Librarians are particularly skilled in highly standardized metadata and data formatting for documents of this type, so training in scientific data management fits relatively easily into existing curricula approved by the American Library Association. Some of these library school graduates work in public or private laboratories but others move to academic libraries. The addition of these newly trained library professionals will greatly expand the services provided by academic libraries and help to ease the potential difficulties some researchers will have meeting data curation expectations now imposed by funding bodies such as the United States National Science Foundation.