RDAP13 Poster Session Summary

by Jennifer Doty

EDITOR’S SUMMARY

Posters presented at the 2013 Research Data Access & Preservation Summit (RDAP) spanned an array of topics of current interest for research data managers. Helping doctoral candidates archive their research data was the focus of one, stressing the need to provide support early in the researchers’ process. Another described a multi-format data literacy program geared to graduate level researchers. The presentation on one university’s “DataDay” described a training workshop designed to help librarians understand the research data lifecycle and pass on essential points to researchers. Another team of presenters offered a methodical approach to establishing the provenance of research datasets that later appear in other contexts. Several posters addressed data management policies and practices at the institutional level, with one offering model policy language and guidelines and another analyzing current and likely future trends and needs to support archiving research data. The full collection of posters reveals the breadth of issues, progress to date and advances to come for the area.

KEYWORDS

research data sets
data curation
policy
standards

Jennifer Doty is a data management specialist with the Emory University Libraries. Her primary focus is developing services and support for the management and curation of research data at Emory. She can be reached at jennifer.doty<at>emory.edu.

EDITOR'S SUMMARY

With presenters covering topics ranging from provenance to preservation to publication of data, the RDAP13 Poster Session provided a wealth of information on recent trends in research data management. Posters ran the gamut with new explorations into systems and support for accessing, preserving and sharing research data.

The poster presented by David Fearon and Betsy Gunea of the Johns Hopkins University Data Management Services team demonstrated the challenges inherent in supporting previously published data. In a time when many librarians and information professionals are exploring methods to engage with faculty and student researchers, JHU presents a compelling model for working with PhD candidates to archive data related to their publications. The experience posed a challenge in providing support to document and prepare data underlying already published results and reaffirmed the importance of working with researchers earlier in the data collection process to effectively create project metadata.

Data information literacy for graduate students was the focus of the poster from the IMLS-sponsored project represented by librarians from Cornell, Purdue and the Universities of Minnesota and Oregon. By focusing their lens on researchers at the graduate level, the project has developed a comparative model for packaging data literacy instruction in a variety of formats (for example, online course, in-person seminar or workshop, embedded librarian) and applying it in a range of situations with students in scientific disciplines. The team adopted an integrated and tailored approach, and one of their most interesting findings is the usefulness of focusing on the mechanics of data management and the local, immediate needs of a specific research group.

From instructional programming for graduate students, we turn to an
example of participant-driven training for information professionals. The librarians at the University of Colorado Boulder presented a comprehensive overview of their first ever “DataDay” training designed for subject librarians. The program development began with participant input, which resulted in creation of an interactive one-day workshop that incorporated hands-on exercises, panel presentations and informal discussions. Pre- and post-training assessments revealed that workshop participants self-reported having higher levels of understanding of the research data lifecycle and greater confidence in their abilities to incorporate assistance with research data management into their established roles as subject librarians. The success of the initial session is attributed in part to the direct involvement of the participants in designing the workshop training, and future plans include additional assessment of the librarians’ training needs to strengthen this iterative process of acquiring knowledge about research data support.

Institutional data management policies have not yet been consistently implemented or reliably enforced at many higher education institutions. In the wake of the February 2013 Office of Science and Technology (OSTP) memo regarding federally funded research outputs, however, those with plans to raise the issue with campus stakeholders would do well to consult the Association of Southeastern Research Libraries (ASERL) Research Data Coordinating Committee presentation on a “Model for Developing Data Management Policy Language.” A collaborative effort from librarians at several ASERL member institutions, the process included an environmental scan of developments in data management support. The final outcome was model language for establishing institutional policies with clear guidelines to support researchers as they try to meet existing and potential funding agency mandates, as well as institutional expectations, for managing research data.

The challenge of establishing provenance and ensuring reproducibility of research for items lacking adequate metadata was explored using a representative figure from the National Climate Assessment (NCA), which serves as the base for the Global Change Information System. Justin Goldstein et al from the U.S. Global Change Research Program, stepped through their sample process for tracing the lineage of a particular image in the NCA, displayed through three different representations. The complexity of dealing with representations and publications that pull data from multiple sources, especially when not all the data are archived or accessible, can be expected to pose an ongoing challenge for many systems. This measured and methodical approach, firmly grounded in best practices for diagramming the provenance of research data, provides a useful example in delineating the origins of digital objects.

The Cornell University Library used a two-pronged approach to postulate present and future trends for depositing research data in institutional repositories. Wendy Kozlowski et al analyzed usage of eCommons, the Cornell University institutional repository (IR), for existing deposits of items designated as datasets (currently a small fraction of the repository’s holdings)
and conducted interviews with researchers to identify the highest priority data management features and functionality. Their findings were consistent with prior assessments of researchers’ perspectives and include the desire for functions such as citation support, discoverability, versioning, self-service submission and linkages between publications and related datasets. Future developments for eCommons are expected to address those functions that received the highest ranking from researchers. Soliciting input from the users, current and future, to prioritize additional improvement is a commendable approach and an encouraging trend for IR administration.

On the classification and metadata front, the Biodiversity Heritage Library presented a new program using open scientific name data to more completely describe items in their collection. In collaboration with the uBio initiative to develop a comprehensive catalog of biological names, this application of open metadata elements to the digitized books has enabled BHL to organize its collection in a more meaningful manner for biologists and zoologists [1]. It also serves as a useful model for other online collections of digital publications and data to emulate when seeking to incorporate open, collaborative metadata schemas to classify items.

This overview highlights just a handful of the informative and engaging posters on display at the RDAP13 Summit. The variety of projects and research demonstrated in one session bodes well for the advancement of research data access and preservation.

For a complete list of the posters presented at RDAP13 please see the program at www.asis.org/rdap/program/#post.

Slides from a selection of posters (as well as keynote, panel and lightning talk presentations) from the Summit are available on slideshare at www.slideshare.net/asist_org/tag/rdap13.

Resources Mentioned in the Article