

Guest Editor's Notes

The 2008 IASSIST Conference, “Technology of Data: Collection, Communication, Access and Preservation” included a session entitled “Moving Research Data Into and Out of Institutional Repositories” from which several papers emerged. In “Interoperability Between Institutional and Data Repositories: a Pilot Project at MIT”, Katherine McNeill describes a pilot project to enhance study discovery between two repository systems housed in the same institution, DSpace and the Institute for Quantitative Social Science Dataverse Network, by enabling the harvesting and replication of metadata and content across the two systems. In a related project across the pond, Libby Bishop scales this discussion in her description of cross-institutional collection sharing between the University of Leeds and the UK Data Archive in the Timescapes project. Bishop asserts that coordination among multiple agents is likely to be challenging under any circumstances. Challenges magnify when the trajectories of different life cycles, for research projects and for data sharing, are considered. Robin Rice echoes these sentiments in her article on the DISC-UK DataShare Project, a collaboration between the Universities of Edinburgh, Oxford and Southampton and the London School of Economics. Rice provides visual evidence in a compelling diagram of the data sharing continuum based on storage, discovery, and preservation conditions of the digital research materials at each level along the scale -- from the lowly thumb drive to the officious national archive. We see plainly that as one moves up the continuum, more and more human effort and intervention is required to craft the discovery, access, analytic and preservation environment. In other words, data curators matter.

Two other papers tackle these challenges by emphasizing the needs of data producers. Luis Martinez-Urbe introduces the University of Oxford's Scoping Digital Repository Services for Research Data Management project and the findings of a requirement gathering exercise. While the study results reveal researchers' needs and workflows. Martinez-Urbe asserts that the study process itself made an impact on the participants. Study participants reflected on and, as a result, fine-tuned how they work with data, why they create these materials in the first place and were able to articulate reasons for managing these resources the way they do. Similarly, Research Data & Environmental Sciences Librarian, Gail Steinhart, writes about the development of DataStaR, a Data Staging Repository hosted by Cornell University's Albert R. Mann Library. The project developed as a “managed workspace” where researchers contribute datasets they are still actively using in direct response to questions that have to do with sharing in the active research environment, rather than an archival one.

While the authors in this issue describe projects going on in many different places and settings, taken together, these articles address common themes. All address the challenge of scaling data exchange between systems and then between institutions. This raises the perennial question of standards: by what mechanisms will we set them, and how well will we be able to follow them and still accommodate local needs? The importance of aligning repository services with researcher needs is another common thread. Data managers must ask, “how will the active researcher benefit from curation efforts”? The answer may be that benefit is more than finding or accessing a particular resource (yep, I have downloaded the whole thing and all the bits are there), but instead being able to examine this resource in many ways (okay, lets run frequencies, now I want to see it on a map, and let's include some other variables). This is a rich reuse experience, creating a real digital “laboratory.”

Finally, each contributor notes the expanding role of data manager. In its own way, each project described here moves data managers upstream, pre-publication, into the place where research is actively happening. Though all of the articles focus on technological choices and architectures to support research data curation, it is striking to realize that each of these choices emerge from old-fashioned personal, social, and organizational relationships. What we can strive for as data and information managers is to work together as fellow researchers and to be ever curious about how these partnerships and the sharing of information back and forth can be enhanced by thoughtful information and technology design. Some call this the digital plumbing, but I like to think of it as e-gilding.

Gretchen Gano, New York University Libraries