

Advancing the State of the Art in Distributed Digital Libraries: Accomplishments of and Lessons Learned from the Digital Library Federation Aquifer Metadata Working Group

Jenn Riley and Sarah Shreeves on behalf of the Aquifer Metadata Working Group

April 29, 2009

Accomplishments

From its inception in 2005 through its conclusion in 2009, the Digital Library Federation's Aquifer Metadata Working Group embarked upon a wide variety of activities in support of the Aquifer initiative's goal of advancing the state of the art in distributed digital libraries. Our work began with a focus on setting best practices for distributed digital libraries in general, building on previous work on the challenges to effective metadata aggregation. Beginning in 2007, we additionally tailored some of our activities to the specific needs of the Mellon-funded Aquifer American Social History Online project, which developed a distributed digital library designed to put into production some of the practices recommended by the various Aquifer Working Groups earlier in the initiative. A brief summary of the Aquifer Metadata Working Group's accomplishments follows.

Analyzing current shareability of metadata made available by DLF institutions. In early 2006, the Aquifer Metadata Working Group issued a survey designed to understand the metadata practices of those institutions likely to be the best poised to provide truly shareable metadata (that is, members of the Digital Library Federation) and gather information on what those institutions believe should be the highest priorities for tools needed to support the creation of shareable metadata.

Digital Library Federation / Aquifer Implementation Guidelines for Shareable Metadata. The most public and influential activity of the Aquifer Metadata Working Group was the creation of a set of guidelines for implementing the MODS descriptive metadata format in a way optimized for aggregation of metadata between institutions. Drafts of the *Guidelines* during their development process were widely circulated, and community input strongly influenced the direction our group took with them. The *Guidelines* are relatively aggressive in their recommendations and set the bar high as a goal institutions should strive towards over time. A stable version of the *Guidelines* was released in November 2006 at <http://wiki.dlib.indiana.edu/confluence/download/attachments/24288/DLFMODS_ImplementationGuidelines_Version1-2.pdf>. A revised version with changes suggested by the experiences of the American Social History Online project was released in April 2009.

Supporting Documents for MODS Guidelines. The full MODS *Guidelines* are an imposing document that might be difficult to fully implement for institutions that do are new to MODS. To ease this difficulty, the Metadata Working group developed a series of supporting documents designed to make the implementation of the recommendations in the *Guidelines* less challenging. These supporting documents are a Levels of Adoption framework for the *Guidelines* <<http://wiki.dlib.indiana.edu/confluence/display/DLFAquifer/MODS+Guidelines+Levels+of+Adoption>> prioritizing the recommendations according to the features they allow in a metadata aggregation, an

FAQ for *Guidelines* implementers

<<http://wiki.dlib.indiana.edu/confluence/display/DLFAquifer/FAQ+for+MODS+Guidelines+Implementers>>, and a stylesheet

<<http://wiki.dlib.indiana.edu/confluence/display/DLFAquifer/MARC+to+Aquifer+MODS+XSLT+Stylesheet>> that converts MARC bibliographic records to MODS records that conform to the *Guidelines* as far as possible, based on the Library of Congress' MARC to MODS stylesheet.

Assessment of Metadata Remediation Tools. Although the American Social History Online portal could not put into place more than a minimal metadata remediation processes, the Aquifer Metadata Working Group believed it was important that the Aquifer project as a whole look into and advance the state of the art in tools that aggregators can use to standardize and normalize the metadata they collect. In support of this goal, the Aquifer Metadata Working Group helped to oversee the creation of a report: "Future Directions in Metadata Remediation for Metadata Aggregators." The report was written by staff funded through a grant from the Gladys Krieble Delmas Foundation and is available online from the Digital Library Federation at <<http://www.diglib.org/aquifer/dlf110.pdf>>.

Support for American Social History Online Portal Development. The Metadata Working Group assisted the staff of the Mellon-funded American Social History Online project and other Aquifer project Working Groups with various tasks supporting the development of the project metadata portal. These tasks included analyzing collected metadata using XQuery to plan for portal discovery features, using the information gained from this analysis to consult on the development of data processing guidelines for the project

<<http://wiki.dlib.indiana.edu/confluence/display/DLFAquifer/Data+Processing+for+Aquifer+Records>>, specifying how the OAI-PMH provenance container should be used when re-exposing aggregated records for harvesting, analyzing the MODS import and export feature of the Zotero citation management Firefox plugin, and consulting on the development of specifications for Asset Actions <<http://wiki.dlib.indiana.edu/confluence/display/DLFAquifer/Asset+Action+Project>>.

Support for Development of Tools to Assist Metadata Providers. The Aquifer Metadata Working Group expected from the beginning that it would not be enough to develop guidelines telling implementers how to structure their metadata, but rather that these guidelines must be supplemented by tools that can assist metadata providers with implementing guidelines and best practices. To this end, the Aquifer Metadata Working Group developed a set of checks in the Schematron assertion language that allows current or potential contributors to the American Social History Online portal to assess their conformance to the MODS *Guidelines* through the project's MODS and Asset Action Explorer <<http://ratri.grainger.uiuc.edu/AAX/>>. The Aquifer Metadata Working Group also began work on specifications to assist institutions using the CONTENTdm digital asset management system and the EAD XML language in creating shareable MODS.

Article Summarizing Metadata Working Group Activities. Representatives from the Aquifer Metadata Working Group collaborated on a formal case study describing the work of the group and providing insight into our decision-making processes. This case study was published as:

Riley, Jenn, John Chapman, Sarah Shreeves, Laura Akerman, and William Landis.
 “Promoting Shareability: Metadata Activities of the DLF Aquifer Initiative.” *Journal of Library Metadata*, 8, no. 3 (2008): 221-248.

Lessons Learned and Recommendations for Future Work

Volunteer makeup of the Working Group. The Metadata Working Group, as all Aquifer Working Groups, was made up of volunteers from participating institutions. In general, the Metadata Working Group was very productive and ran into fewer of the problems (lack of time, non-completion of assignments) that many volunteer-based groups face. This may be because institutions were providing in-kind support in the form of staff time and explicitly recognized this time spent on Aquifer as part of the staff member’s responsibilities. The Metadata Working Group also had at the start very concrete tasks – development of the guidelines – which were easy to ‘assign’. It was only in the last months of the Initiative, as support and interest in Aquifer waned, that the Metadata Working Group began to face issues in completion of its work. We conclude from the experience of the Metadata Working Group that all-volunteer efforts can be effective when participants truly believe in a shared overall goal, and that institutional support for these types of activities is a key factor in their success. The work products of the Metadata Working Group likely could have been produced in shorter amounts of time with devoted (rather than volunteer) staff, but identifying individuals with the necessary skills and background to work in these areas as a short-term task would likely be difficult, if not impossible.

Effect of community input on the Guidelines. One of the most important things that the Metadata Working Group did was to actively and consistently solicit community feedback on the MODS *Guidelines*. The Metadata Working Group held several open review sessions at DLF Forums and solicited comment over email. The primary results of these discussions were:

- A stronger and more widely accepted set of guidelines;
- A better understanding of the community needs (and the lack of clear guidelines for the use of MODS); and
- Arguably, wider support of MODS in the digital library community.

Email and in-person feedback periods necessarily slow down development of documents such as the MODS *Guidelines*. The Metadata Working Group attempted to strike a reasonable balance between internally discussing relevant issues with the goal of gaining consensus with timely releases of proposals for community review. Guidelines and best practices work is inherently time consuming, and unfortunately the Metadata Working Group’s experience did not suggest any specific strategies for

combating this reality. We do believe that given the extremely limited time we had available at the end of the Aquifer initiative to revise the MODS *Guidelines* based on feedback from those that used it and assessment reports of the American Social History Online portal, a future similar project might benefit from release of “best-guess” guidelines earlier in the process, to allow more time for actual implementation of them in production systems. Such a model would then provide more real-world information with which to validate assumptions made by the guidelines.

Effect of grant funding on our work. The Mellon Foundation’s funding of the American Social History Online project within the overall framework of the Aquifer initiative changed the nature of the Metadata Working Group’s tasks somewhat. While providing a valuable opportunity to test the recommendations put forth by the Metadata Working Group in practice, the funded project also introduced specific tasks and timetables guided from outside the Working Group that were not present before. These guided tasks and timelines gave the Working Group more focus, but also limited our ability to follow lines of thought to places we might have otherwise gone. We conclude from experience with both of these models that both are needed, and that it is possible for a single group to work in both modes, if perhaps not always simultaneously.

Effect of support from the Digital Library Federation. The Digital Library Federation provided material support for the Aquifer Metadata Working Group throughout its tenure, including support for biweekly conference calls and meeting space and logistical planning in conjunction with Digital Library Federation Forums. While it would not have been impossible to find support for these activities elsewhere, the easy access to these tools allowed Metadata Working Group members to focus more on their specialized tasks and less on administrative. Support for meeting space in particular gave a significant boost to the productivity of the Metadata Working Group. While much of the Group’s work was done over a Wiki, email, and conference calls, the occasional in-person meeting re-invigorated the Group and allowed us to get past particularly thorny issues relatively quickly. The effect of the Digital Library Federation name as assisting the productivity of the Metadata Working Group cannot be underestimated as well. Aligning the Aquifer Metadata Working Group with an *institutional* membership organization such as the Digital Library Federation ensured the individuals serving on the Working Group were participating with the full weight of their home institutions behind them; representing an institutional in addition to an individual commitment to the success of the work. While the Digital Library Federation is currently planning to fold into the Council on Library and Information Resources, we believe it is important that collaborations of institutions continue to prioritize work on distributed digital libraries, and not rely solely on collaborations of individuals to lead the way.

Balancing best case and reality. The connection with the Digital Library Federation helped to ensure the membership of the Metadata Working Group were among those individuals best poised to make effective recommendations prior to external validation of them. This connection also allowed the Metadata Working Group the luxury of assuming a certain degree of technical skill and a commitment to increasing the shareability of metadata by institutions implementing the MODS *Guidelines* and supporting documents. Perhaps predictably, we learned through the course of the Aquifer initiative that

even Digital Library Federation members struggle to find the resources (particularly, time) to focus on improving the shareability of their metadata, especially for legacy collections. We conclude that while more work on guidelines and best practices are needed, the biggest critical need for development in this area is for *tools* that make the creation of shareable metadata easier. It is not enough to tell institutions what they should be doing; they (even Digital Library Federation institutions) need help actually doing it.

Changes in the aggregation landscape. During the existence of the Aquifer initiative, the state of the art in distributed digital libraries changed significantly. The Metadata Working Group's activities focused mainly on providing for effective aggregations of metadata alone, providing advanced functionality without access to the resources themselves. While this is still an important activity, aggregating content in addition to metadata became a much more feasible activity over the four year span of the Aquifer initiative, as seen both in the development of more technical protocols to support this activity and in increasing openness of institutions to the possibility. Future revisions of the Metadata Working Group's *MODS Guidelines*, supporting documents, and tools would benefit from a fresh look with a more content-focused perspective.

Appendix: Aquifer Metadata Working Group Membership

Sarah L. Shreeves (University of Illinois at Urbana-Champaign): 2005-2009; Chair, 2005-2007

Jenn Riley (Indiana University): 2005-2009; Chair, 2007-2009

Laura Akerman (Emory University): 2006-2009

John Chapman (University of Minnesota): 2005-2008

Melanie Feltner-Reichert (University of Tennessee): 2006-2008

Kat Hagedorn (University of Michigan): 2007-2009; ASHO Core Team Liaison, 2006

Bill Landis (California Digital Library/Yale University): 2005-2006, 2007-2009

Tracy Meehleib (Library of Congress): 2006-2009

Elizabeth Milewicz (Emory University): 2005-2006

David Reynolds (Johns Hopkins University): 2005-2009

Gary Shawver (New York University): 2005-2008