

Proposed Activities for Aquifer Metadata Working Group

Liz Milewicz (Emory)
David Reynolds (JHU)
Jenn Riley (Indiana)
Gary Shawver (NYU)
Sarah Shreeves, Chair (UIUC)

Introduction

This document outlines proposed activities for Phases 1-3 (with a particular focus on Phases 1-2) of the Aquifer Initiative. Included are an activity description, the span of time (phases) the activity might take, and a rough outline of the type of resources needed to accomplish the activity. These proposals have been based primarily on the discussions of the Metadata Working Group. The Metadata Working Group welcomes comments and feedback.

Working Group Charge:

The Metadata Working Group recommends metadata policies and best practices to the DLF Aquifer Implementation Group. The working group designs workflows for metadata harvesting, creation and enhancement.

Phase 1 Activities

1. Metadata Profile for Aquifer

Activity Description:

Develop a MODS metadata profile for DLF institutions with cultural heritage and humanities scholarly material based on the DLF/NSDL OAI Best Practices for Shareable Metadata work (see <http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl?MetadataContent>).

The first draft of the profile will include:

- Recommendations for the use of MODS elements;
- Recommendations for controlled vocabulary use;
- Best practices for sharing metadata (beyond the two sections above); and
- Examples of different types of material.

Added in ongoing work will be:

- Appropriate mappings from and to MODS, focusing, in particular on Dublin Core, and working in conjunction with Rebecca Guenther of LC; and
- Guidelines for use of the optional containers in OAI protocol related to metadata.

Span: First draft delivery by August 21, 2005. Will be revised throughout the Aquifer Initiative as needed.

Resources Needed:

The Metadata Working Group along with Bill Landis (CDL) from the OAI Best Practices Group will be responsible for this activity. We will need reviewers and other feedback from Aquifer participants, as well as advice and guidance from LC. We also need to consider how this will be disseminated and maintained.

2. Functional Requirements for Searching and Browsing of Metadata Aggregation

Activity Description:

In collaboration with the Technical Working Group and the Services Working Group, define functional requirements for search/browse interface for harvested metadata and ensure that the metadata profile (described above) includes guidelines to help Aquifer participants meet those requirements.

Span:

Functional requirements development and updating should continue through all phases of the Aquifer initiative, but Phase 1 functional requirements should be completed as soon as possible.

Resources Needed:

Aquifer working groups should work together to develop the functional requirements, but we will need feedback from other Aquifer participants.

3. Incorporation of Collection Description in Item-Level Metadata Aggregation

Activity Description:

The use of collection description metadata within the metadata aggregation can help to provide missing or additional contextual information and can help inform users' activities. See Foulonneau et al for a description of incorporating collection description in searches of item level metadata:

<http://cic.harvest.grainger.uiuc.edu/documents/usingcollectiondescriptions.pdf>.

This activity would require:

- Agreement on a collection description standard (in progress with development of the DLF Collection Registry);
- Participating organizations to include collection descriptions as part of their OAI responses (i.e. within a set description) or a collection level record included in the DLF Collection Registry.
- Inclusion of the collection level metadata in searches (including deciding how to rank the collection level metadata and how to display search results in a meaningful way; should searches always lead to the item level information or should we direct people through collections or other groupings?).

Span:

Phase 1 activity, although upkeep will be continuous.

Resources Needed:

We will need to build on the DLF Collection Registry work and promulgate collection description standards and guidelines for inclusion in OAI responses. We will also need to work with the Technology and Services Working Groups to leverage the collection descriptions in the search.

4) Survey of Metadata Tools and Ongoing Research

Activity Description:

In order to move forward in Phase 2 of Aquifer we need to have a better sense of the tools currently used in DLF institutions for working with metadata and what research into metadata use, enhancement, normalization, mapping, etc is ongoing.

Span:

Phase 1 activity (to be sent out in August 2005), though we may revisit towards the Phase 2.

Resources Needed:

Metadata Working Group can conduct survey (preliminary survey currently being conducted by Liz Milewicz).

Phase 2 Activities

1) Metadata Normalization and Enhancement

Activity Description:

The metadata profile will only go so far to mitigate quality problems in the metadata aggregation. Projects to investigate metadata normalization and enhancement are necessary to build useful services that can leverage the richness of the metadata provided. Below are a number of metadata normalization and enhancement activities that will aid in the development of these services:

- Create and/or improve date normalization algorithms;
- In collaboration with technical group, investigate potential for using authority file cross-references (name, subject) in search/browse system;
- Investigate research systems for automatic metadata improvement (e.g., JHU's automated name authority control
<<http://www.dlib.org/dlib/april01/dilauro/04dilauro.html>>, subject switching languages, etc.) and evaluate their potential for use in Aquifer.

Span:

This can probably begin towards the end of Phase 1 and will continue through Phase 2.

Resources Needed:

Identification of priority services and what metadata normalization/enhancement would best fit with these; metadata expertise to identify most pressing issues and possible resolutions; technical expertise to develop or leverage tools (such as those identified in the survey) to accomplish goals; a suitable collection to use as a test bed. Evaluation and user testing.

2) Investigate methods of obtaining metadata for Aquifer other than OAI

Activity Description:

The Aquifer Initiative should remain flexible enough to use both OAI and other means to either aggregate or federate searches of metadata. This activity will identify what other methods exist, their feasibility for inclusion, and potentially trial implementations.

Span:

Phase 2

Resources Needed:

The Metadata Working Group would need to work with the Technical Working Group for this activity.

3) Developing browsable subject clusters

Activity Description:

One of the seemingly intractable problems of OAI-enabled metadata aggregations is the difficulty in creating useful browse indexes – particularly in subject areas – because of the range of controlled vocabularies (or non-controlled vocabularies in use). This activity would investigate how we might best present browsable groupings of records. This might include:

- Investigating the use of clustering tools (such as those in development at Emory);
- Leveraging what controlled vocabularies exist (mapping?); and
- Comparing groupings based on collections and potentially sets to clustering.

Span:

Phase 2 and 3

Resources Needed:

Metadata and technical expertise; evaluation.

4) Metadata and Deep Sharing

Activity Description:

If we are to move forward to sharing not only metadata, but also content, we need to have a thorough investigation of how metadata and what metadata – descriptive, administrative, technical, structural – will be shared alongside the content.

Span:

Phase 2 (informing Phase 3 activities)

Resources Needed:

Metadata and technical expertise; testers of different methods? This would have to be done in conjunction with the Technical Working Group and the Collections Working Group.

5) Building and/or Customizing Metadata Creation Tools

Activity Description:

Following up on the Phase I survey of metadata tools, we will start assembling a suite of tools to enable Aquifer contributors to create and harvest metadata that fits our profile. This will most likely involve customization and enhancement of existing tools, but might require some completely new tools.

Span:

Phase 2

Resources Needed:

Metadata and technical expertise; software development

Phase 3 Activities

1) User Contributed Metadata

Activity Description

If Aquifer is going to be heavily geared towards “scholars” (who presumably know a lot more about content in DLs than metadata creators do!), we should investigate a system that somehow merges metadata (including text markup) from content experts with that from librarians. This activity would investigate:

- Methods for including user contributed metadata;
- What level of vetting is most appropriate;
- How to display and track user contributed metadata.

Span:

Phase 3

Resources Needed:

Technical and metadata expertise (including mark-up expertise). Testers and evaluation.