DLF Aquifer Technology/Architecture Working Group
Proposed Activities
Updated version after O'Hare meeting
July 4, 2005

Introduction

This document outlines a proposed set of activities for the DLF Aquifer initiative's Technology/Architecture Working Group, based on discussion within the group via e-mail and conference calls, as well as discussion at the June 28, 2005 face-to-face meeting of Aquifer participants at the Wyndham O'Hare. Activities for phases 1-3 are included, with a natural focus on those in phase 1.

The technical architecture and implementation of Aquifer should follow from the needs of its users and the nature of its collections. Consequently, the Technology/Architecture Working Group's activities are heavily dependent on outputs and plans from the other working groups: Collections, Metadata, and Services. This document should be viewed as a work in progress, continuing to evolve as the requirements and scope of Aquifer become better defined by its participants.

In this document: TWG=Technology/Architecture Working Group, MWG=Metadata Working Group, CWG=Collections Working Group, SWG=Services Working Group, AIG=Aquifer Implementation group

Working Group Charge

The Technology/Architecture Working Group recommends architecture and infrastructure policy to the DLF Aquifer Implementation Group. The working group designs technical solutions for DLF Aquifer Implementation, identifies the resources required to create these solutions and communicates the resource requirements to the DLF Aquifer Director. *(from the Aquifer Business Plan)*

Working Group Membership

Jon Dunn, Indiana University (chair)
David Ackerman, New York University
Eric Celeste, University of Minnesota
Tim Cole, University of Illinois Urbana-Champaign
Jerry Persons, Stanford University
Anthony D. Smith, University of Tennessee
Cory Snavely, University of Michigan
Thornton Staples, University of Virginia

Activities – Phase 1

1. Conduct an environmental scan of relevant projects and technical solutions
Who: Technology/Architecture Working Group members
When: Summer 2005 – ongoing

Description: In order to explore architecture and technology options for Aquifer, we need to have a good understanding of the current library and digital content technology environment. The Technology/Architecture Working Group plans to create and maintain an inventory of related projects, tools, architectures, frameworks, reference models, and protocols that are potentially relevant to Aquifer technical architecture and design. For ease and flexibility of editing, this inventory will be maintained on a Wiki supported by the University of Minnesota.

2. Develop statement of technology/architecture principles
Who: Technology/Architecture Working Group
When: Summer 2005

Description: We propose to develop a short document outlining a set of general principles for architecture development and technology selection in Aquifer to help guide our later work, addressing issues such as use of open-source vs. commercial software, use of open standards, and leveraging the work of existing projects.

3. Develop high-level architecture for DLF Aquifer
Who: Technology/Architecture Working Group
When: Summer 2005 – ongoing

Description: While the requirements for Phases I-III of Aquifer are still evolving, the group feels that it is valuable to develop a preliminary high-level technical architecture diagram and description as a starting point. This will initially be based on the discussion and working group reports from the June 28 face-to-face meeting and will continue to evolve as the technical and functional requirements for Aquifer become better defined.

A member of the TWG will be identified to lead this work, with assistance and feedback from other TWG members, as well as members of the other working groups and AIG. Group members agree it would be helpful to schedule an intensive 1 ½ day face-to-face meeting in late summer to work on architecture definition.

4. Develop requirements document / RFP for DLF Aquifer harvester
Who: Technology/Architecture Working Group, in coordination with other working groups
When: Summer 2005

Description: Per the Aquifer Business Plan and discussions within the Aquifer Implementation Group, one of the primary goals of Phase 1 is to establish a "presence" for Aquifer within the library and higher education community by building on past experiences with OAI metadata harvesting to create a DLF Aquifer harvester and corresponding search/browse interfaces. This harvester will aggregate metadata via OAI-PMH for items in the collections of DLF member

institutions identified by the Collections Working Group according to standards developed by the Metadata Working Group. The search and browse interfaces will provide user-friendly means for searching and browsing item and collection metadata and linking to collection and item content served from DLF member institutions.

While the Aquifer harvesting service will join a growing number of existing OAI harvesters, it will distinguish itself through the quality of metadata harvested, its focus on ease of use, and the coherent collections represented, supporting the needs of scholars in the humanities and social sciences.

We envision that one or more DLF member institutions with expertise in OAI metadata harvesting and user-centered interface development will work on projects to harvest metadata and provide searching and browsing interfaces, based on a functional and technical requirements document developed by a new cross-functional requirements group including members from all Aquifer working groups. These requirements should build upon best practice developed by existing OAI projects. In particular, the scope and requirements for Aquifer Phase 1 overlap significantly with the IMLS grant-funded DLF project to create a prototype second-generation OAI finding system and should make use of the work of that project as much as possible.

The new proposed group will develop exact requirements, but the Technology/Architecture Working Group believes that certain elements are necessary:

- Both searching and browsing capabilities should be provided.
- The search and browse interfaces should be accessible to the public, not just to DLF members.
- A focus on user input, user testing, and usability is essential.
- Services enabling reuse of metadata should be provided, including OAI-PMH data provider, SRW/SRU searching, RSS feeds, and others.
- A minimum metadata standard for participation higher than simple Dublin Core is appropriate.

Other areas that the requirements group will need to address specifically include:

- Required search/browse capabilities – though some details may be left up to implementer(s) based on technical capabilities and results of user testing
- Optional OAI-PMH data provider features required for participation (e.g. OAI sets, inclusion of total result set size in resumption tokens, use of About container, etc.)
- Required subsetting capabilities beyond use of OAI sets, if necessary
- The amount of metadata massaging/normalization required
- Requirements for replication, mirroring, performance and scalability
- Platform requirements to enhance potential future relocatability or replicability of the service

Beyond basic searching and browsing of OAI harvested metadata, we envision a number of activities to be carried out as experiments involving a subset of Aquifer data providers. These experiments could include the inclusion of multiple “actionable URLs” for access to different capabilities of components of a digital object, …

5. Create DLF Aquifer harvester based on OAI-PMH
Who: One or more Aquifer participating institution(s) per requirements from requirements cross-functional subgroup
When: Fall 2005 – Spring 2006 (exact dates TBD)

Description: As described in the previous task, we envision that one or more DLF member institutions will be selected and funded to build and operate a DLF Aquifer harvester based on OAI-PMH metadata harvesting, with interfaces for searching and browsing. Various models are possible. For example, it is quite likely that more than one institution will be involved, e.g. one focusing on metadata harvesting and the other on the public interface for searching and browsing the collections. Institutions with existing expertise and experience in OAI harvesting and user-centered search/browse design and implementation should carry out this work if at all possible, leveraging previous and ongoing work in these areas.

Activities – Phase 2

1. Phase 2 Requirements Definition / RFP Development: Enhancement and Integration
Who: Technology/Architecture Working Group in collaboration with other WGs and AIG
When:

Description: As described in the Aquifer Business Plan, phase 2 is focused on "enhancement," including enhanced content (additional collections, additional formats), enhanced retrieval (enhanced metadata, better search and discovery capabilities), and enhanced services (new tools and capabilities). The Technology/Architecture Working Group will need to continue to evolve the Aquifer architecture to accommodate these enhancements as the other Working Groups continue to elaborate their visions. In addition, building support into Aquifer for the integration of its collections and search and discovery services into local institutional systems—including metasearch and course management—should be a major focus of phase 2, helping to pave the way for deep sharing in phase 3. It should be recognized that models for sharing beyond OAI metadata harvesting will likely be necessary to support the features of phases 2 and 3.

2. Phase 2 Implementation: Enhancement and Integration
Who: One or more Aquifer participant institution(s) per requirements from TWG
When:

Description: Implementation of enhancements as defined by Technology/Architecture Working Group.

Activities – Phase 3

1. Phase 3 Requirements Definition / RFP Development: Deep Sharing
Who: Technology/Architecture Working Group in collaboration with other WGs and AIG
When:

Description: Phase 3 of Aquifer anticipates so-called "deep sharing" of digital objects between institutions, allowing individual scholars or other users to not only work with digital objects in their native delivery environments, but to retrieve the component parts of those objects for use with local tools for display, analysis, and annotation. There are a number of potential models for enabling such sharing by exposing the components and capabilities of digital objects, ranging from harvesting of object packaging metadata in formats such as METS and MPEG-21 DIDL to exposure of "behaviors" of digital objects through Web services-type interfaces. In addition, various supporting services are necessary to enable deep sharing, potentially including cross-domain authentication, authorization, and rights management.

In addition, there is the major question of whether Aquifer only provides deep access to digital content and methods of discovering it for use with users' and institutions' local tool sets for analysis, annotation, visualization, manipulation, etc., or whether it also provides access to the tools themselves.

We will need to define potential technical architectures to support deep sharing and evaluate them against requirements identified by the Services Working Group to see how well they support the services ultimately being envisioned for Aquifer. We should look to other projects that are also exploring issues of deep sharing for ideas, including those funded by NDIIPP.

2. Phase 3 Implementation: Deep Sharing
Who: One or more Aquifer participant institution(s) per requirements from TWG
When:

Description: Implementation of support for deep sharing as defined by Technology/Architecture Working Group.