

DLF Aquifer Development for Interoperability Across Scholarly Repositories

Framework for Implementation

DLF Aquifer Technology/Architecture Working Group
Working Draft 7
October 20, 2006

About this document

This document describes a framework for the implementation of the technical aspects of the next stages of DLF Aquifer [1] distributed digital library. It aims to name and outline the categories of components, and interactions between those components, that can implement the requirements for interoperability across scholarly repositories. The framework takes the existing DLF Aquifer products and proposes how they can work together with other existing technologies and new components to achieve interoperability.

Overview

The framework is built around a *Metadata Pool*. This contains metadata about digital library collections and items. The metadata is harvested from participant *Repositories* via *Inflow Services*. If the *Repositories* don't provide the interfaces required, then interactions with them are mediated by *Proxies*. The metadata in the pool is then made accessible via *Outflow Services*. Consumers can access metadata directly via the *Outflow Services*. They can also access metadata and content via *Applications* that are built on the *Outflow Services*. Such *Applications* provide additional functionality and customized interfaces to the content of the distributed digital library. *Inflow Services* may validate, normalize, and/or enhance metadata before adding records to the *Metadata Pool*. To do this *Inflow Services* may exploit exposed interfaces of *Enabling Services* and *Outflow Services*. Once *Consumers* have identified items they wish to use, those items are accessed directly from their host *Repository*, or via a *Proxy* or *Application* as appropriate. Various *Enabling Services* are used by components of the framework.

The terms in *Italics* above are component categories in the framework. These are shown in the Top Level View (Figure 1).

Aquifer Scope

A working implementation of the framework will require Aquifer to deploy many components. Some of these have been or will be developed by Aquifer; others are existing technologies. The implementation also assumes the existence of components that are part of the framework but not in scope for the Aquifer implementation. In figure 1 the gray box labeled *Aquifer* shows the scope of the Aquifer implementation. The *Metadata Pool*, *Inflow Services* and *Outflow Services* are core and will be deployed. The component categories that are partially covered by the Aquifer scope box are peripheral (*Applications*, *Enabling Services*, *Proxies*); they may have components that are deployed as part of the implementation. The categories that are not covered by the scope box (*Consumers* and *Repositories*) are out of scope.

Component Categories

The remainder of this document discusses the component categories and, the components that are expected to belong to them. Some of these are show in the Detailed View (Figure 2).

Metadata Pool

The *Metadata Pool* is a store of metadata that has been gathered from participant *Repositories* and possibly enhanced or otherwise altered to conform to Aquifer guidelines or requirements. It consists of item level descriptive metadata in the form of MODS records conforming to Aquifer MODS Guidelines [2], collection level metadata, and actionable URL sets in the form of Asset Action Packages conforming to the Asset Action Schema [3]. The pool stores all of the metadata required for the discovery and use of items in the distributed digital library.

Inflow Services

The *Inflow Services* are components that gather metadata from participant *Repositories* or their *Proxies* and place in the *Metadata Pool*. These components are responsible for ensuring that the metadata conforms to the requirements of the pool.

The main inflow service is a harvester. The harvester in the initial implementation is expected to use OAI-PMH to gather collection level metadata and item level descriptive metadata conforming to the Aquifer MODS guidelines. It will then use the asset action package link in the MODS record to gather the asset action metadata for each item.

Other *Inflow Services* may include the following:

- A Conformance Tester service that reports on the conformance of candidate metadata to the requirements of the metadata pool. This would also be used by the harvester service to ensure that harvested metadata is at least minimally conformant.
- An Enhancer service that augments and normalizes inflowing metadata.
- An Aggregation Analyzer service that determines relations between items and collections.

Outflow Services

The *Outflow Services* are components that provide access to metadata in the pool. These will generally use standard protocols to provide the kinds of access required by various consumers, for example:

- A Harvest service will re-expose harvested metadata as an OAI-PMH data provider.
- A Search service will use SRW/U.
- A Syndicate service will use RSS.
- A Crawl service will use HTTP to expose metadata for crawling by search engine indexers.

Repositories and Proxies

The *Repositories* participating in Aquifer are technically heterogeneous and provide a wide range of collections and items. The *Inflow Services* are responsible for connecting with the *Repositories* or their *Proxies* and obtaining their metadata, in the process ensuring that the metadata conforms to the MODS guidelines and other requirements. The *Repositories* are responsible for exposing their collections via agreed protocols. Initially this will be as OAI-PMH data providers.

Participant *Repositories* are also required to provide access to their content via requests on agreed protocols. Initially this will be via HTTP.

In cases where a participant repository is not able to support the protocols required by the *Inflow Services* or by *Applications* and *Consumers* for accessing content, a proxy service must be developed. This may be done as part of the Aquifer implementation for particular *Repositories*. Possible *Proxies* include a harvest proxy for *Repositories* that don't support the protocol required by the harvester, and an asset action generator proxy for *Repositories* that are unable to expose asset action metadata.

Applications and Consumers

The users of Aquifer *Outflow Services* are *Applications* and *Consumers*. *Applications* are developed on top of *Outflow Services* to provide enhanced functionality and customized interfaces to the Aquifer distributed digital library. Since the *Outflow Services* are built on standard protocols and include actionable references it will be possible for non-Aquifer entities to develop *Applications* that exploit Aquifer collections. Aquifer will develop a portal application that will provide a web interface to browse, search, access and use content from the Aquifer collections.

Aquifer intends to support integration with many kinds of existing services and applications used by scholars. These include:

- Commercial search services such as Google
- Course managements systems such as Sakai
- Citation management software such as RefWorks
- Federated search tools such as SFX

Integration with these *Consumers* will be achieved via the use of standard protocols and by providing access to actionable URLs for semantically predictable links.

Enabling Services

There are many existing and emerging digital library services that will provide a crucial enabling role in the Aquifer implementation. Where necessary these *Enabling Services* will be developed or deployed by Aquifer. Where possible existing services will be used. Examples of such services are:

- A Format Registry service such as the Global Digital Format Registry.
- A Service Registry service such as Ockham.
- A Resolver service.
- Authority Resolver services.

References

1. DLF Aquifer: Bringing collections to light through the Digital Library Federation. <http://www.diglib.org/aquifer/>
2. DLF MODS Implementation Guidelines. http://www.diglib.org/aquifer/DLF_MODS_ImpGuidelines_ver4.pdf
3. DLF Aquifer Asset Actions Experiment: Demonstrating the Value of Actionable URLs. <http://dlib.org/dlib/october06/cole/10cole.html>