

Aquifer TWG Policies and Principles

Purpose

This document is provided by the Aquifer Technology Working Group to clarify assumptions about, and to provide general guidance for, the architecture of systems developed for Aquifer. The principles here are designed to foster collaboration and to balance the ease of participation in Aquifer with the difficulty of advancing access to distributed information. In addition, this document attempts to set some realistic expectations for participants with respect to development effort as well as functionality for users.

General Architectural Requirements

The main goal for this project is to build services that can be used broadly without requiring absolute agreement on how to implement a content management strategy locally. In support of this goal, we make the following assumptions:

- Discovery services are very important; the purpose of these services is to actually find things. Users who find digital resources often want to do something with the resource found immediately, and sometimes want to gather them for later use.
- The Aquifer definition of interoperability, at least initially, will be based on the idea of exposing metadata for arbitrary use, and on providing at least one web-accessible, searchable index of that metadata that allows individual objects to be viewed or retrieved, independently of their type or format. The harvestable metadata will include that needed for discovery as well as that needed for the delivery system to characterize the content found to the user.
- The end-user method of access, therefore, assumes a web browser.
- The architecture should support harvesting for use by members of the consortium and for arbitrary aggregators without any built-in prejudice. In addition, the architecture may support the redistribution of aggregated metadata for other potential uses.
- Varying degrees of interoperability and/or robustness may exist in the system, depending on how well or how completely data contributors follow the recommendations of the Aquifer working groups. Full compliance with recommendations should yield maximum functionality.
- Fine-grained policies about how the data can be used and by whom should be enforceable throughout the life cycle of the exposed information.

Development Considerations

Because participation is to a large extent voluntary, the technical architecture should try to capitalize on existing systems and place as few restrictions on implementors as possible, but needs to place adherence to standards and portability as the highest priorities.

- Adherence to relevant open or publicly-documented standards is a must. Where systems implement standards or parts of standards, they must be *properly implemented*, although not necessarily *completely* implemented.
- Aquifer encourages institutions to leverage their local developments for use with Aquifer to the greatest extent possible, notwithstanding the above.
- Code should be written with an eye toward sharing with other institutions, and substantial consideration should be given to abstraction at appropriate levels and to minimizing operating system or other underlying software dependencies. The ideal technology model to support flexibility and re-use in Aquifer is an interoperable set of lightweight, platform-independent,

standards-based services.

- Developers should give fair attention to scalability; i.e. proof-of-concept models should not be supplied as surrogates for reliable production systems.

Open source versus commercial tools

Open source software is a natural fit for the Aquifer project for a number of reasons:

- The open source community has developed a reputation for providing excellent reference implementations of developing technologies.
- Using open source software minimizes the threat of proprietary influence on functionality and design, and lowers the likelihood of data migration problems in the future.
- Systems built on open source software will be at least cheaper and probably simpler to distribute amongst interested colleagues.

At the same time, we want to encourage participation from all institutions, regardless of a particular technology background.

Therefore, systems built on open source software are highly preferable, and where Aquifer work may make use of commercial systems, we strongly encourage that these systems adhere to relevant standards in the interest of maximizing interoperability.

Replication/redundancy/scalability/performance

For initial work, the focus should be on functionality and not performance. However, all work should be done with an eye toward scalability. In particular:

- OAI-based work in the first two phases should focus on standardization and common practices, so that replication will be easy to achieve, and will produce consistent results for users. This will enable multiple sites to be put on line that harvest the same material and provide roughly the same search experience.
- Work in future stages that may involve distributed technology should focus on evolving standards for those technologies, as well as methodologies for increasing capacity to prevent performance problems as usage grows.
- In general, as architectures introduce interdependencies, the standard principles of avoiding single points of failure should be followed.