

The aDORe Federation Architecture

Herbert Van de Sompel, Ryan Chute, Lyudmila Balakireva

herbertv@lanl.gov , rchute@lanl.gov , ludab@lanl.gov

Digital Library Research & Prototyping Team

Research Library

Los Alamos National Laboratory

Abstract:

The proposition of a “single repository behavior” exposed by a federation consisting of any number of distributed, heterogeneous repositories has been the subject of repository interoperability efforts such as Dienst, NCSTRL, CORDRA, DRIVER, and the Chinese DSpace federation. The desire to federate repositories in such a way typically emerges as a result of the understanding that no single repository hosts all artifacts that are relevant for a specific subject domain, community, or application. Federating can also be a response to challenges of scale. Indeed, the exponential growth in the amount of digital artifacts created on a daily basis, both born-digital, and as a result of massive digitization efforts is forcing architects, engineers and developers involved in creating digital repositories to face the harsh reality that their solutions need to handle an amount of artifacts that is orders of magnitude larger than originally intended. Optimizing, tuning, and tweaking the existing repository infrastructure can initially alleviate performance problems, but eventually limits are reached. At that point, a major redesign of the repository solution is an obvious option. An alternative is to move towards an environment that consists of parallel instances of the existing repository solution and to glue those together into a repository federation that behaves as if it were a single repository.

This paper describes the aDORe repository federation architecture, an outcome of the aDORe research and development effort by the Digital Library Research & Prototyping Team of the Los Alamos National Laboratory (LANL). The architecture is the result of three intersecting drivers. First, a general research interest in repository interoperability as exemplified by the Team’s involvement in standardization efforts such as the ANSI/NISO Z39.88-2004 OpenURL Framework for Context-Sensitive Services (OpenURL), the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), and more recently the Open Archives Initiative Object Re-Use & Exchange effort (OAI-ORE). Second, the Team’s research interest in digital preservation matters illustrated by its involvement in National Digital Infrastructure and Preservation Program (NDIIPP) projects. Third, a concrete need to design and implement a solution for ingesting, storing and accessing the vast and growing scholarly digital collection of the Research Library of the Los Alamos National Laboratory.

The goal of the aDORe federation architecture is to facilitate a uniform manner for client applications to discover and access content objects available in a group of distributed

The aDORe Federation Architecture

Herbert Van de Sompel, Ryan Chute, Lyudmila Balakireva

repositories. This is achieved by means of a 3-Tier architecture illustrated in Figure 1. Tier-3 provides client applications with a single point of access to all content available in the federation, irrespective of the actual location of that content in federated repositories. In order to realize this, the architecture requires all federated repositories to implement the same, minimal set of machine interfaces to make their content accessible. These repository interfaces constitute Tier-1 of the architecture. Moreover, the architecture requires the introduction of a middle Tier, Tier-2, consisting of two shared infrastructure components that keep the books on content objects, repositories, and repository interfaces in the federation. These shared infrastructure components minimally expose one machine interface each. In order to respond to client requests, the federation's single point of access interacts with these interfaces as well as with the interfaces exposed by the content repositories. As a matter of fact, the single point of access to the federation supports exactly the same minimal set of machine interfaces as each federated repository does, effectively making the entire federation behave in the same manner as each individual constituent repository. In principle, this design allows the aDORe federation concepts to be applied recursively, but no experiments have been conducted to date that demonstrate the feasibility of the nested federations idea. The aDORe federation architecture is not concerned with uniform operations to write, update and delete objects in repositories, and considers these the responsibility of constituent repositories of the federation. However, the architecture does ensure that results of these operations can be made apparent to client applications.

This paper will introduce the aDORe Federation's 3-Tier architecture, detailing the core requirements imposed on a repository to become part of a federation, and introducing the components that facilitate exposing an environment consisting of multiple heterogeneous repositories as a single one. The paper will also touch on the aDORe Archive, a specific implementation of the Federation architecture developed at LANL.

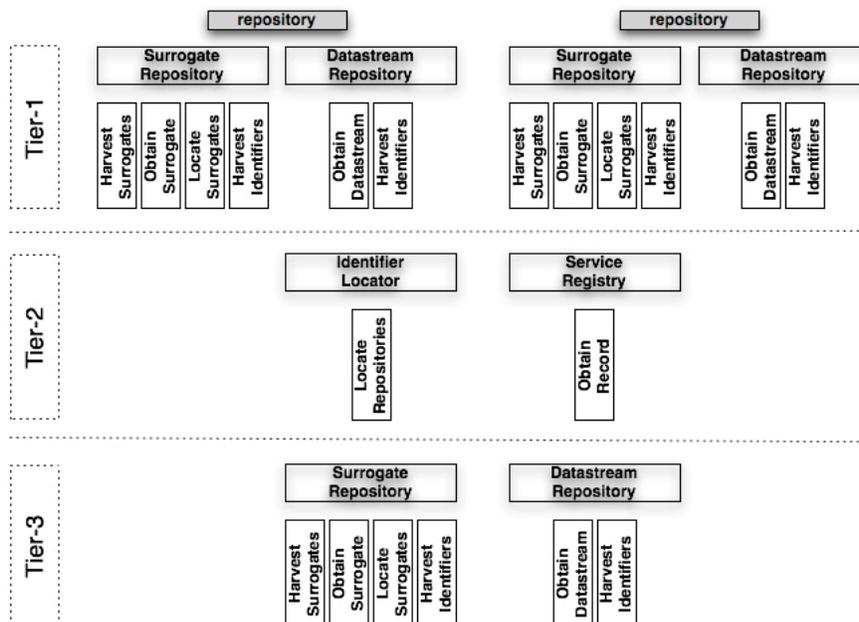


Figure 1: The 3-Tier aDORe Federation Architecture