

Opening aDORe

RYAN CHUTE

Abstract: In May 2008, the Digital Library Research and Prototyping Team of the LANL Research Library released a new version of the aDORe Archive software package. As was the case with the previous version, it provides the functionality to ingest and store compound digital objects in a combination of XMLtapes (storage of XML-based representations of compound objects) and ARCfiles (storage of constituent datastreams of compound objects), and to access these content objects from their respective storage repository. The new aDORe Archive version additionally provides components that facilitate accessing the multitude of storage repositories as if only one single repository were involved. These new components are the Identifier Locator, the Service Registry, an OAI-PMH-based front-end for batch collecting XML-based representations of compound objects, and an OpenURL-based front-end for retrieving disseminations of constituent datastreams. The aDORe Archive is the result of four years of research and development and its design is inspired by concerns of scale, interoperability and preservation. The resulting architecture is fully standards-based and highly modular. The code base is entirely written in Java and repurposes several existing software components, including OCLC's OAI-PMH and OpenURL packages, the Ockham service registry module, and the Heritrix ARCfile tools. Some core characteristics of the aDORe Archive make it potentially attractive as a long-term storage component to be plugged into other repository solutions: the parallelization of ingestion and dissemination, combined with the distributed storage capabilities address concerns of scalability; the write-once/read-many approach is suitable for content objects that have reached a level of stability in their life-cycle and for preservation use-cases that require maintaining all versions of both the XML-based representations of compound objects as of their constituent datastreams; the concatenation of many content objects into XMLtapes and ARCfiles yields a basic file-based storage solution that is straightforward to manage; the virtualization achieved through the use of non-protocol-based identifiers for all content objects and machine interfaces allows for their straightforward physical relocation; the use of protocol-based interfaces throughout the solution allows for exchanging the underlying technical implementations while keeping the access mechanisms stable over time; the aDORe Archive front-ends provide a single point of access for a potentially extensive collection of content objects. The aDORe Archive solution was taken into production at the Research Library of the Los Alamos National Laboratory (LANL), in September 2007. The LANL deployment hosts over 100,000,000 objects mainly consisting of licensed content from both secondary and primary publishers (e.g., APS, BIOSIS, EI, Elsevier, Thomson Scientific, etc.).

About the author:

Ryan Chute is a member of the Digital Library Research and Prototyping Team in the Research Library at the Los Alamos National Laboratory.