The Perspectives of Digital Curators on Building Distributed Repositories

Caryn Wojcik  
Government Records Archivist  
Records Management Services  
Michigan Historical Center  
Department of History, Arts and Libraries  
3405 N. Martin Luther King Jr. Blvd.  
PO Box 30026  
Lansing, Michigan 48909  
wojcikc@michigan.gov

Richard Marciano  
Lead Scientist  
Sustainable Archives & Library Technologies (SALT) laboratory

Chien-Yi Hou  
Digital Preservation Specialist  
Sustainable Archives & Library Technologies (SALT) laboratory

Reagan Moore  
Director of Data and Knowledge Systems  
University of California San Diego/San Diego Supercomputer Center  
9500 Gilman Drive, MC 0505  
La Jolla, CA 92037-0505  
marciano@sdsc.edu  
chienyi@sdsc.edu  
moore@sdsc.edu

Mark Conrad  
Archives Specialist  
Electronic Records Archives (ERA)  
NHE, Room 1540  
The National Archives and Records Administration  
8601 Adelphi Road  
College Park, MD 20740-6001  
mark.conrad@nara.gov

Abstract

The Persistent Archives Testbed project (PAT: 2003-2006) brought together digital curators from libraries, archives, historical societies, scientific data environments, and museums, as well as IT researchers and staff. One of the principal goals of the project was to design a distributed community repository for electronic records management. Each site chose an archival collection for testing on the infrastructure, with the record types varying from web crawls, to e-mail, to images, to voting records, to reports. We implemented a distributed data management system that allowed each institution to separately manage their own records, while using a common software infrastructure. In this architecture, each site controls access and update permissions for their preservation environment independently of the other participants. Each site implemented a different preferred interface for interacting with their archival collections and linking the distributed data to holdings at their institution. Finally, each site tested one or more archival functions - appraisal, accessioning, arrangement, description, preservation, and access – with their archival collection. The ability to manage all of these types of records using
common software infrastructure was one of the significant outcomes of the project. The ability to share a common infrastructure while implementing independent archives is a second major outcome. This paper focuses primarily on the lessons learned and skills needed by the digital curators to automate ingestion, description, and validation of records. The approaches taken by each participating institution are described, as well as the benefits that were achieved by using common infrastructure.