

The University of North Texas Libraries' Portal to Texas History: Archival Challenges and Solutions

Kurt Nordstrom
Programmer
Digital Projects Department
University of North Texas Libraries
(940) 369-7809
knordstrom@library.unt.edu

Cathy Hartman
Head
Digital Projects Department
University of North Texas Libraries
(940) 565-3269
chartman@library.unt.edu

Mark Phillips
Laboratory Manager
Digital Projects Department
University of North Texas Libraries
(940) 369-7809
mphillips@library.unt.edu

ABSTRACT

The UNT Texas History Portal Project strives to balance the goals of accessibility of information and long-term preservation of digital objects. This poster details the system that automates the collection of metadata records to coordinate access to web-viewable files and preservation of archived master files.

Categories and Subject Descriptors

D.2.11 [Software Architectures]

General Terms

Management, Design, Human Factors, Verification.

Keywords

Preservation. Metadata. Libraries. Portal.

Our proposed Portal for Texas History is a multifaceted project which spans many levels of implementation. The two primary areas of focus are a web portal, designed for public access, and an archival system, designed for long-term storage of digital files. Creating a system that uses a single metadata record to coordinate both access and archival storage has presented some design and implementation challenges, which we present here.

The web portal uses an XML schema, in which each record maintains information about one or more files that form a logical object. Examples might be scanned pages of a book, or a scanned poster. The system indexes and searches records by keywords and subject fields. In addition, each XML record contains preservation metadata about one or more corresponding digital master files, stored on a separate archival system.

The challenge comes with finding a way to easily create records on the portal, while maintaining accurate data corresponding to the archived master files. Moreover, automating as much of the data collecting process as possible saves human effort and reduces the possibility of error.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '04, Month 1-2, 2004, City, State, Country.
Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

A custom-designed client program serves as the "glue" between the web portal and the archival system—handling the preprocessing of files, collection of metadata, and the actual uploading of records. While the client program manages most files automatically, exceptionally large files that exceed the available bandwidth may require additional user input.

The simplified diagram below shows the flow of data, from the beginning of the process to the final ingestion. The critical juncture is the creation of master archives and metadata records at a single point. The client software assures that the proper links between the two main systems will be maintained.

