The Issues of Compliances and Interoperability in Integrating Heterogeneous Digital Information Resources: Lessons from Texas History Portal

Daniel Gelaw Alemneh, Cathy Nelson Hartman, and Mark Phillips
Digital Projects Department, University of North Texas, Denton, Texas, U.S.A

Abstract

The University of North Texas (UNT) Libraries received a grant from the Texas Telecommunications Infrastructure Fund Board to create the technical structure for a Web Portal to Texas History. In collaboration with many stakeholders (Texas libraries, museums, teachers, historians, etc.), the portal to Texas history at UNT provides access to a diverse collection of published and unpublished resources (manuscripts, letters, maps, oral histories, photographs, etc.) to researchers, teachers, and young scholars at school, home, office, or anyplace with a computer and Internet connection. Implementing a state-wide system required the creation of a robust application framework for integrating heterogeneous digital information resources from institutions across the state. This paper discusses issues related to portal building and explores compliance and interoperability issues. Based on the UNT Libraries’ Portal to Texas History project experiences, it will attempt to assess the current status and the emerging trends in innovative uses of portal technologies and provide an overall scenario. In context of aggregating a variety of formats, bundling preservation metadata to facilitate long-term access, setting project standards and best practices guidelines, and coordinating a variety of collaborative efforts for participating institutions.

Keywords

Digital Archive, Digital Library, Long-term access, Interoperability, Digital Resource Integration,
The Issues of Compliance and Interoperability in Integrating Heterogeneous Digital Resources: Lessons from the Texas History Portal

Daniel Gelaw Alemneh, Cathy Nelson Hartman, and Mark Phillips
Digital Projects Department, University of North Texas, Denton, Texas, U.S.A

Abstract

The University of North Texas (UNT) Libraries received a grant from the Texas Telecommunications Infrastructure Fund Board to create the technical structure for a Web Portal to Texas History. In collaboration with many stakeholders (Texas libraries, museums, teachers, historians, etc.), the portal to Texas history at UNT provides access to a diverse collection of published and unpublished resources (manuscripts, letters, maps, oral histories, photographs, etc.) to researchers, teachers, and young scholars at school, home, office, or anyplace with a computer and Internet connection. Implementing a statewide system required the creation of a robust application framework for integrating heterogeneous digital information resources from institutions across the state. This paper discusses issues related to portal building and explores compliance and interoperability issues. Based on the UNT Libraries’ Portal to Texas History project experiences, it will attempt to assess the current status and the emerging trends in innovative uses of portal technologies and provide an overall scenario. In context of aggregating a variety of formats, bundling preservation metadata to facilitate long-term access, setting project standards and best practices guidelines, and coordinating a variety of collaborative efforts for participating institutions.

Trends and Developments in Digital Libraries

Digital libraries and supporting technologies have now matured to the point where their contents are incorporating complex and dynamic resources and services. In reviewing the literature of the past few years, there is no shortage of views on the future of digital libraries. Lavoie and Dempsey (2004), Lagoz et al. (2004) and Birmingham et al. (2003), among others, articulated a vision of a digital library environment that resonates with possibilities to create a knowledge management system on the Internet that will enable scholars to navigate through these resources in a standard, intuitive, and consistent way. Many observers (Kumar et al. 2004) and Birmingham et al. (2003), agree that the new scholarly communication systems will inevitably be based on capabilities of interoperable network technology.

However, these innovations may have come at the expense of simplicity, sustainability, and other commonly understood applications in the life cycle management of digital resources. For instance, architectural complexity may have a significant impact on the internal efficiency/effectiveness, preservation, provenance, intra- and/or inter-object interoperability issues.

Introduction

The University of North Texas (UNT) Libraries received a grant from the Texas Telecommunications Infrastructure Fund Board to create the technical structure for a Web Portal to Texas History. In collaboration with many stakeholders (Texas libraries, museums, teachers, historians, etc.), the Portal to Texas History provides access to a diverse collection of published and unpublished resources (manuscripts, letters, maps, oral histories, photographs, etc.) to researchers, teachers, and young scholars in schools, homes, offices, or anyplace with a computer and Internet connection.

The Challenges of Compliance and Interoperability

Digital libraries and supporting technologies must now incorporate complex and dynamic resources and services. The wide variety of application technologies and the growing number of digital library initiatives make the need for collaboration and compliance with standards more critical than ever. Considering the central role of a digital library in scholarly communication, the compliance and interoperability challenge is among the fundamental concerns, and addressing this issue is a key factor to the success of digital library development at large.
Different strategies may be involved for ensuring interoperability both at internal and external levels. As noted by Singh, Gaba, and Pandita (2004) built-in interoperability can be provided for internal layers using the emerging frameworks (such as Semantic Web and Web Services) while it would be difficult in case of external layers, which require various plug-ins, wrappers, and agents.

Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. As described in W3C (2003), Semantic Web is the abstract representation of data on the World Wide Web, based on the Resource Description Framework (RDF) standards and other standards to be defined.

Web services provide a standard means of interoperating between different software applications, running on a variety of platforms and/or frameworks. It uses protocols based on the XML language to describe an operation to execute or data to exchange with another Web service.

In the past there was a perception that Semantic Web and Web Services initiatives are in opposite directions. However, as noted by a number of commentators (Paolucci and Sycara (2003), W3C (2003)), Semantic Web can be used for data integration while Web Services can do the program integration. And both Semantic Web and Web Services can be bridged together and ensure interoperability among various layers.

Recognizing the importance of building and managing flexible and interoperable library portals in this rapidly evolving technology landscape, the University of North Texas’ "Portal to Texas History” project has designed and implemented a robust architecture, with particular emphasis on the issues of compliance and interoperability.

**Texas History Portal Project Background**

The Portal to Texas History provides tools that facilitate single-point access to vast and heterogeneous primary source material about Texas History. In order to fulfill this ambitious goal, the Portal project has been developed in two phases.

Phase-One, which was completed in 2003, accomplished the following four stated objectives:

1. Build a technical structure for the Portal services.
2. Create standards and best practices guidelines required for participation in the Portal project, and design a module to train library and museum staff members at collaborating institutions regarding the digitization processes outlined in the standards and guidelines.
3. Develop a Portal Web site to guide students and researchers in the use of the resources.
4. Create an evaluation plan for the first four years of the project and implement the first year of the evaluation plan.

Phase-Two of the project continues in development, focusing on the following five main objectives:

1. Adding content to the Site, focusing on specific topic areas to support K-12 and college level classes.
2. Coping with controlled vocabularies.
3. Designing a training program for content and metadata ingestion to be used in a test-bed project with member institutions.
4. Developing curriculum guides that fulfill the needs of educators teaching Texas history to schoolchildren.
5. Increasing interoperability options with the addition of OAI (Open Archives Initiatives) harvesting capabilities and participating in Z39.50 compliant federated searching.

Additionally, expanding collaborative partnerships and site evaluation will continue throughout the project.

**Project Architecture**

Implementing a statewide portal system with the magnitude of the aforementioned objectives required the creation of a robust application framework for integrating heterogeneous digital information resources from institutions across the state.
The Architecture should have the flexibility to accommodate the future internal and external layers. Taking many factors into consideration (simplicity, sustainability, versatility, etc.), the Portal to Texas History uses open source software configured to support the diverse digital information resources and services offered by the varied participant institutions. The metadata-based content management software relies heavily on the flexibility of XML file structures. The adaptable nature of the design allows compliance (with a minimum modification) with widely used standards including OAI harvestable metadata protocols and Z39.50 protocols for interoperability related to search and retrieval of information.

The general overview of the Texas History Portal architecture with its major activities and procedures (including heterogeneous contents and metadata creation, archiving, and users’ access interfaces) is graphically illustrated in Figure-2 below.

![Figure 2. The Portal to Texas History General Architecture](image)

**Summary**

While designing and implementing the Portal to be used by historians, researchers, young scholars, and the general public, the Texas History Portal project team constantly reviews the current technology trends, best practices, protocols, and standards. Such an exploration of trends and innovative uses of technologies provides an overall scenario in the area of aggregating a variety of formats, bundling contents and metadata, etc.

Now, having the portal in operation, the focus of future activities will be to ensure the integration of various services, tools, and contents. In this regard, introducing OAI-PMH Federator as a single-point-of-access is one of the top priorities of the next phase of the Texas History Portal project. As part of this effort, evaluation of the multifaceted use of metadata in the portal architecture design is one crucial process that is ongoing.

Recognizing the fact that interoperability occurs at different levels of abstraction, there is a need for more comprehensive evaluative research that takes an in-depth look at the existing portal infrastructure, which is expected to facilitate and assure the deposit, storage, search, retrieval, and preservation of heterogeneous digital resources.

By increasing interoperability options, we will be able to ensure long-term access and facilitate and coordinate a variety of collaborative efforts for participating institutions.

**References**


5. Kurt Maly and Mohammad Zubair, Enhancing Infrastructure for OAI: the DLGrid. D-Lib Magazine, 10 (6), (2004). Available at: [http://www.dlib.org/dlib/june04/06inbrief.html#Sin02](http://www.dlib.org/dlib/june04/06inbrief.html#Sin02) [Sited visited on: October 04, 2004].


Biographies

Daniel Gelaw Alemneh is a doctoral student in Information Science at the UNT. He earned his Master’s Degree in Library and Information Management from the University of Sheffield, UK. Prior to that, he received his BLS degree from Addis Ababa University, Ethiopia. Daniel has been involved in various Digital Projects from their inceptions to their implementations. His work has primarily focused on project managements and development of standards and procedures. Daniel is also an adjunct instructor at the UNT School of Library and Information Sciences.

Cathy Nelson Hartman is Head of the Digital Projects Department at the University of North Texas (UNT) Libraries and a Fellow of the Texas Center for Digital Knowledge. Hartman actively participates in various professional activities at the state and national level, including serving as an ALA Councilor and chairing the Depository Library Council, a federal government advisory board. She was recently elected to the Texas Records Management Interagency Coordinating Council. She publishes and speaks widely and is a successful grant recipient, most recently participating as a partner with the California Digital Library and New York University in a National Digital Information Initiative Program grant to capture and preserve digital information.

Mark Phillips is Manager of the Digital Projects Laboratory at the University of North Texas Libraries. He received his Master’s Degree in Information Science from the UNT in May 2004. Mr. Phillips has also been involved in every level of the portal project and recently presented papers at JCDL and LITA 2004 conferences.