Recent Developments in the Design, Construction, and Evaluation of Digital Libraries:

Case Studies

Colleen Cool
Graduate School of Library and Information Studies, Queens College, USA

Kwong Bor Ng
Queens College, CUNY, USA

A volume in the Advances in Library and Information Science (ALIS) Book Series

Information Science
REFERENCE

Managing Director: Lindsay Johnston
Editorial Director: Joel Gamon
Book Production Manager: Jennifer Yoder
Publishing Systems Analyst: Adrienne Freeland
Assistant Acquisitions Editor: Kayla Wolfe
Typesetter: Christina Henning
Cover Design: Jason Mull

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Recent Developments in the Design, Construction, and Evaluation of Digital Libraries : Case Studies / Colleen Cool and Kwong bor Ng, Editors.
pages cm.
Includes bibliographical references and index.
Summary: “This book addresses the challenges with digital libraries, as well as describes the recent developments in the design, construction, and evaluation of these libraries in various environments”-- Provided by publisher.
ISBN 978-1-4666-2991-2 (hardcover) -- ISBN (invalid) 978-1-4666-2992-9 (ebook) -- ISBN (invalid) 978-1-4666-2993-6 (print & perpetual access) 1. Digital libraries. 2. Digital libraries--Case studies. 3. Electronic information resources--Case studies. 4. Computer network architectures--Case studies. 5. Libraries--Special collections--Electronic information resources--Evaluation. 6. Libraries--Information technology. I. Cool, Colleen, 1952- editor of compilation. II. Ng, Kwong Bor, editor of compilation.
ZA4080.R43 2013
025.042--dc23

2012040364

This book is published in the IGI Global book series Advances in Library and Information Science (ALIS) Book Series (ISSN: 2326-4136; eISSN: 2326-4144)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 2
Reengineering The Portal to Texas History[SM]: A Case Study

Kathleen Murray
University of North Texas, USA

William Hicks
University of North Texas, USA

Mark Phillips
University of North Texas, USA

Neena Weng
University of North Texas, USA

Dreanna Belden
University of North Texas, USA

ABSTRACT

This case study reports the activities, findings, and lessons learned during a project that replaced the legacy Digital Asset Management (DAM) system of The Portal to Texas History[SM] at the University of North Texas Libraries with an open source system. This unique system decouples the application development framework from the backend infrastructure, effectively relieving the development and growth constraints inherent in the legacy system. In a novel approach for an academic library, genealogists participated in the user-centered, iterative approach used to prototype, develop, and test the user interface. The resulting system promoted productivity gains by enabling programming staff to work in parallel from specialized areas of expertise. A post-project review process identified a number of lessons learned, including the importance of representing the requirements and priorities of internal and external stakeholders. The review process also informed an application development model that may be useful to other digital libraries.

INTRODUCTION

The Portal to Texas History[SM] is a gateway to a significant set of humanities collections within the digital library of the University of North Texas (UNT) Libraries (http://texashistory.unt.edu/). In collaboration with over 200 content partners, the Portal provides access to collections from Texas libraries, museums, archives, historical societies, and private collectors. The Portal archives and provides access to more than 165,000 digital objects, comprising over 2.3 million image files.

DOI: 10.4018/978-1-4666-2991-2.ch002

The range of primary source materials includes maps, books, manuscripts, newspapers, diaries, photographs, and letters. While materials primarily concern the 254 Texas counties, there are items related to most of the states in the USA and to over 40 other countries.

Materials include a fascinating array of rare and invaluable items that document Texas's history, such as handwritten correspondence between Santa Anna and Lorenzo de Zavala from the early 19th century and photograph collections covering seminal events, including the Dallas Police Department's investigation of the 1963 John F. Kennedy assassination and the Texas City Disaster of 1947. Historic newspapers represent a large segment of the materials, with over half a million searchable pages dating from 1829 to the present. This significant collection is not only used in traditional research by scholars and lifelong learners, but also forms the basis for a collaboration between UNT and Stanford University in text mapping research that is exploring new methods for programmatically finding and analyzing meaningful patterns.

Development of the Portal began in 2003 when the UNT Libraries selected a system vendor for its digital library, and, as is typical of many digital libraries, based design decisions largely on the requirements of librarians and what they imagined end users would need. When the reengineering project began in 2008, the number of unique visitors per month had grown from 1,000 in 2004 to over 20,000. This success was accompanied by operational and management challenges, which affected the Portal's content partners, users, and other stakeholders.

Scale issues created by the continuous addition of content and the increased usage required constant attention and distracted the systems team from other areas of development. The underlying data model for both the digital objects and descriptive metadata were limiting the kind of items that could be ingested into the system. Development constraints associated with the underlying technical infrastructure also emerged. In particular, the design and implementation of new features and functions was limited by outdated software and changing vendor priorities. It was often not feasible to make interface design changes without programmatic changes in the supporting infrastructure.

In 2007, a decision was made to replace the legacy asset management system with a digital library infrastructure and framework based largely on modern open source components widely used throughout the world. This approach distinguishes the UNT Libraries within the broader library community, in which libraries generally employ single-vendor, integrated systems for their digital libraries. Replacing the legacy system also presented an opportunity to include users directly in the design process.

The UNT Libraries received a National Leadership Grant from the Institute of Museum and Library Services (LG-06-07-0040-07) for a two-year study (2007-2009) to redesign the Portal's interface. At the outset of the project, an application development model was drafted to guide the project's work. The model employed user-centered design methods within a rapid development framework. The framework required reengineering the underlying technical infrastructure, which was undertaken in conjunction with this project. The model involved three teams within the Digital Libraries Division: (1) system designers and programmers, (2) user interface designers, and (3) user study researchers. The user group involved in the project was genealogical researchers, a significant and growing user group of both libraries and archives. Genealogists participated in needs assessment activities at the start of the project and in usability testing at three points over the course of the project. The goals of the project were:

- Implementation of a rapid development framework within the UNT Libraries, Digital Libraries Division.

- Implementation of an iterative user-centered design process to create a user interface to The Portal to Texas History℠ optimized for genealogists.
- Identification of the information context and information needs of genealogists interacting with The Portal to Texas History℠.
- Creation of an application development model for digital libraries that incorporates user-centered design principles.

The redesigned Portal was launched in two public releases: Release 1 in June of 2009 and Release 2 in October of 2009. Following the second release, project team members participated in a post-project review. From this review, a revised application development model was created. Both the model and the infrastructure resulting from this project have supported the rapid growth of digital collections within the UNT Digital Library and the further enhancement of the user interface.

Objectives

This chapter is a case study of how one large, public university library dealt with growth and technology challenges within its digital library. The first objective is to describe the application development-working model and the specific activities and findings involved in each of the three phases of work within the project: (1) requirements definition, (2) application development, and (3) usability testing. The second objective is to illustrate the value of a post-project review process by including the key lessons learned as well as the final application development model that resulted from the review process. The third objective is to report the project outcomes and impact. In closing, reflections on the experience gained during the project are stated, as well as implications for other digital library application development projects.

BACKGROUND

Digital Asset Management Systems

Most libraries use asset management systems to manage the end-to-end processes that support their digital collections, which may be comprised of multiple object types and formats including photographs, maps, documents, audio recordings, newspapers, video, and books. Digital Asset Management (DAM) systems provide storage for a library's digital objects and their associated metadata, as well as tools to manage and provide access to these assets (Kim, et al., 2007).

Some DAM systems are add-on products that leverage the investment libraries have made in Integrated Library Systems (ILS). These systems interface with an ILS product from the same vendor and generally include tools to import, catalog, edit, store, search, retrieve content and metadata, and generate statistics (Boss, 2009). Examples include Media Management from Innovative Interfaces, Portfolio™ from SirsiDynix®, and DigiTool® from Ex Libris. CONTENTdm® from OCLC® is an example of a DAM system that is not an ILS add-on product but does offer libraries a similar set of tools for the storage, curation, and distribution of their digital collections.

Summarizing from a previous study (Marchionini & Fox, 1999), Marchionini, Plaisant, and Komlodi state that, "most research and development projects to date have been devoted to technology and content" (2003, p. 123). Matuskiak (2006) asserts that most digital library research concentrates on infrastructure and metadata. Early development of DAM systems is consistent with these observations. Historically, development of DAM systems concentrated on providing workflows for content providers, such as libraries, and for persons creating the digital assets, while particularly emphasizing metadata tools. DAM systems have offered little design flexibility to local developers in terms of customization and enhancement of the user interface to meet local requirements (Salo, 2008).

More recent open source DAM-related projects offer increased flexibility for local developers interested in customizing the user interface to meet local requirements (Jansson, 2010; Uzwyshyn, 2008). Open source DAM systems have been created by in-house programming efforts such as the New Zealand Digital Library Project at the University of Waikato, which developed Greenstone, a suite of open source software for building a digital library. Other open source DAM systems have been built through collaborations between libraries and private companies. The Portal to Texas History℠ at UNT Libraries was originally built with Index Data, a software development company.

In the digital repository arena, DuraSpace supports two open source platforms: DSpace, a turnkey institutional repository application, and Fedora, a framework for building digital repositories. Two open source projects, FEZ, developed by the University of Queensland Library as a front end to Fedora, and Manakin, developed by Texas A&M University for DSpace, enable customization of the user interface. Of course, not all libraries have the staff with Web application development expertise to support advanced customization, even with tools such as Manakin (Janssen, 2010). Many libraries continue to require turnkey DAM solutions with a limited set of options for interface customization.

Rapid Application Development

Rapid application development strives to complete development in as few steps as possible and in a relatively short timeframe (e.g., 30-90 days) (Fernandez, Martinez-Prieto, de la Fuente, Vegas, & Adiego, 2008). Software solutions are usually based on standards and strive for simplicity and portability. Because solutions can be designed and developed in an accelerated timeframe, a website, such as The Portal to Texas History℠, can be more responsive to changing needs.

Rapid application development occurs within a Web framework, such as the open source frameworks of Ruby on Rails® and Django®. Web frameworks are configured using modules, components, and tools, which are easily interchangeable, robust, highly cohesive, and loosely coupled. This is sometimes referred to as shared nothing architecture. Components in a Web framework are supported by flourishing active user communities and audited by a large base of users, whereas components developed for niche communities are generally avoided. Within the framework, components at each level are highly scalable to allow costs to be distributed across the framework as increased capacity is needed in particular areas.

Application development within a Web framework tends to follow an established flow in which functional requirements are first identified and generally remain unchanged during the application development process. Developers use prototyping tools to create functional designs based on the requirements. Using an iterative process, the prototypes can be revised in response to user feedback. Marchionini et al. (2003) state that digital library development is well-served by an iterative, process-oriented approach; however, only a few studies (Norberg, Vassiliadis, Ferguson, & Smith, 2005; Van House, Butler, Ogle, & Schiff, 1996) have included an iterative user-centered design approach for creation or redesign of user interfaces to digital library collections. Within a Web development framework, this iterative process between developers and users continues until a final prototype is agreed upon. Subsequently, development and testing of the Web application occurs.

The project reported in this chapter replaced the legacy software development platform for The Portal to Texas History℠, which had a development community with dozens of fellow developers, with an open source Web application framework, whose components have Web communities with thousands or tens of thousands of developers. This

novel approach enabled developers within the UNT Libraries to solve problems by leveraging the solutions and innovations of developers from different disciplines and knowledge domains. The Web development framework made it possible to integrate these solutions and application code into the Portal's user interface, allowing more flexibility in implementing changes in response to user requirements.

Genealogists: A Significant User Community

Many stereotypes have a basis in fact and the view of genealogy as an avocation for those in middle to later life is generally supported in demographic research. A study of New Zealand genealogists revealed that 73% of respondents were female and 87% were aged 51 or over (Kuglin, 2004). The Fullerton Genealogy Study found that (1) 72.2% of respondents were female, (2) the average number of years spent pursuing genealogy was 14, and (3) the average age at the inception of their research was 40 (Drake, 2001).

This user community is poised to grow significantly in the coming years. In 1995 four out of ten Americans were somewhat interested in genealogy (Fulkerson, 1995) and by 2000, 60% of Americans were interested (Gallop-Gerdman, 2000). A 2005 study by Market Strategies, Inc. (MSI) indicated that the percentage of Americans interested in genealogy had grown to 73% (Generations Network, 2005). This trend indicates a growing interest in genealogical research.

Genealogists are also a significant percentage of the users of digital libraries, archives, and record repositories (Boyns, 1999; Cherry & Duff, 2002; Duff & Johnson, 2003). Despite this, only a few studies have investigated genealogists' information seeking behavior (Duff & Johnson, 2003; Fulton, 2009; Kuglin, 2004; Skinner, 2010; Yakel, 2004). The significance of this user group and the fact that it is growing, compel the digital library community to seriously examine this user group to better inform their services and user interfaces. Butterworth (2006) declares that traditional information seeking models such as those put forward by Kuhlthau and Belkin break down when applied to information seekers pursuing personal history. When information seekers do research as a leisure activity, their motivations are significantly different—these users pursue "research undertaken primarily for pleasure" (Butterworth, 2006, p. 2). In a study of amateur genealogists around the world, Fulton (2009) identified leisure as an important information-seeking context.

Taylor (1991) asserts that effective design of information retrieval systems should be informed by an understanding of the unique information use environments of discrete user groups. This project investigated the information needs of genealogists and conducted initial usability tests with genealogists of the existing user interface to The Portal to Texas History℠. The results of these assessments informed the design of a prototype user interface optimized for use by genealogists and developed within a Web application development framework within the UNT Libraries.

THE REENGINEERING PROCESS

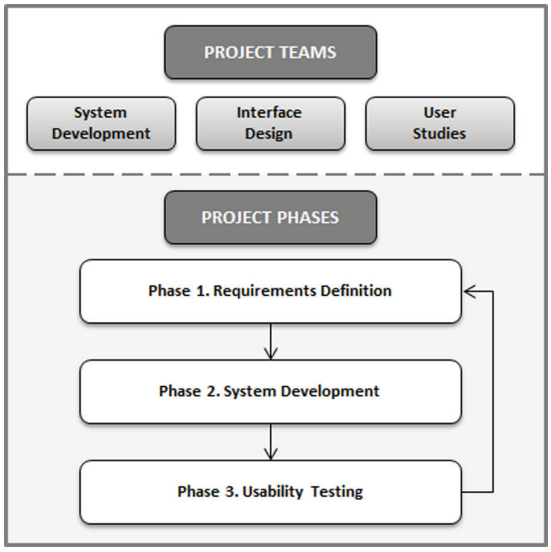
The reengineering project was guided by a draft application development model (Figure 1). Three teams were identified at the start of the project: System Development, Interface Design, and User Studies. Consistent with user-centered design, the three iterative phases of work over the course of the project included: (1) Requirements Definition, (2) Application Development, and (3) Usability Testing. The arrows identified the collaborative interfaces and the handoff points among the teams.

Project Teams

System Development Team

The system development team consisted of two programmers and a system architect. The system architect was experienced in architectural design

Figure 1. Draft application development model



and proficient in data model specification. The programmers had the requisite skill sets to support the backend system components, in particular the search platforms, servers, and the Web development framework. System developers were responsible for configuring, coding, and testing infrastructure components. System developers took great care to identify and specify each component of the Portal system so that the key constraints of the legacy system were removed. The objective was to build a system and a development framework that would enable independent utilization of the expertise of the Libraries' Web interface designers, while at the same time accommodating future growth in both usage and content.

Interface Design Team

The user interface team consisted of two librarians whose singular focus was Web development. Both had experience with best-practices in information architecture, accessibility compliance, and were familiar with the processes of user-centered design. Interface designers created and coded the user interface components. As part of the overall design process, the user interfaces group helped guide the larger development team in their assessment of real world design issues and their application in the current project. After some initial work in needs assessment, the team was tasked with creating visual mock-ups for paper prototyping, and with generating and revising html wireframes. As other teams continued their development efforts, the user interface team transformed the incoming ideas, pseudo-code, wireframes, and requirements documents into the actual source code and presentation logic that would be used for the final deliverable, a working site.

User Studies Team

The user studies team consisted of one researcher, with experience in conducting interviews, focus groups, survey research, and usability testing, as well as a graduate research assistant. The researcher was responsible for all communication and interaction with the genealogists who participated in the project. The team developed all data gathering instruments (i.e., the focus group protocols, questionnaires, quality assurance tests, and usability tests), conducted the focus group discussions, and administered all tests. Subsequent to data gathering the team analyzed the data and reported the findings to the other project teams. These findings directly impacted the design and development activities, as well as measuring overall user satisfaction with the resulting system.

Project Phases

Phase 1: Requirements Definition

Infrastructure and interface requirements were created by the system development and user interface teams. Both teams also generated questions, ideas, and prototypes regarding possible features and functional enhancements to the existing Portal. The user studies team used these to develop a protocol for focus group discussions with members of genealogical societies. In addition to

conducting focus groups, the user studies team (1) analyzed a log of comments submitted by the Portal's users over the previous three years and (2) conducted initial usability tests of the existing Portal with genealogists. After analyzing all of the data collected, the user studies team drafted a set of functional requirements for the user interface. These requirements were refined by members of the three teams and classified into implementation phases or releases.

Phase 2: System Development

The system development team designed and implemented the necessary hardware and software infrastructure to create the rapid development framework. This team created specifications for the framework's overall architecture, components, tools, and workflows. Likewise, system developers created required application specifications and code. The interface design team first translated user, non-functional, and functional requirements into paper prototypes, which were subsequently transformed into HTML mock-ups. These, in turn, informed the final design and its implementation in client-side code.

The system development and interface design teams created and refined workflows to support their separate but inter-dependent design and development work. These two teams also performed internal quality assurance tests. The user studies team conducted Quality Assurance (QA) tests of completed software releases with individuals external to the design and development teams (i.e., library staff, student workers, and users). The system development and interface design teams used the QA test findings to refine each software release prior to its public launch.

Phase 3: Usability Testing

Subsequent to the public launch of each release, the user studies team conducted usability tests with members of genealogical societies. The arrow from Usability Testing to Functional Requirements in Figure 1 indicates a key feedback loop in the work, that is, usability test findings became input for future functional requirements.

PHASE 1: REQUIREMENTS DEFINITION

The activities conducted in phase 1 are described in two sections: (1.1) Needs Assessment and (1.2) System Requirements. The needs assessment includes the activities in which genealogists were directly involved. Section 1.3 lists the major lessons learned in phase 1. These were identified by members of the three project teams who participated in a post-project review.

1.1. Needs Assessment

1.1.1. Identification of Innovative Features

After ample exploration and discussion, more than 15 innovative features, which might have great potential to enhance the user experience, were identified and considered by the team. These included navigation options and displays of metadata, objects, and search results. At the onset of the project, social networking sites were becoming popular on the Web and in keeping with this wider trend, it was anticipated that the new interface would include social features, such as commenting and tagging.

Additionally, a number of open source and freely available widgets and other easily implementable code libraries that could enable users to interact with websites had become available and could be readily adapted to local interface designs. MIT's SIMILE Timeline and the image magnifier code for Zoomify™ were two such examples. From within the world of libraries and booksellers, a number of sites had begun offering faceted displays of search results in order to provide users more choices for filtering and evaluating their results.

In order to assess the interest and usefulness of these new social and interactive functions, mock-ups were created for the focus group discussions conducted by the user studies team. Figure 2 is a mock-up of a possible comment function. Additionally, mock-ups of five object navigation designs were created for the focus group discussions. The options incorporated several of the proposed interactive functions including tabs, widgets, and collapsible menus.

1.1.2. Focus Group Discussions

During February and March of 2008, three focus group discussions were held. Nineteen genealogists from two northeast Texas genealogical societies participated in one of three focus group discussions. Participants were primarily females (84%; n=16) and all were over 50 years of age. On average, participants had been doing genealogical research for 21 years. Of the 19 participants, only three reported having professional genealogical credentials and these three were all members of the Association of Professional Genealogists (APG). However, about 70% reported having memberships and affiliations with local, state, and national genealogical organizations, including the National Genealogical Society, the USGenWeb Project, the TXGenWeb Project, the Hispanic Organization of Genealogy and Research (HOGAR de Dallas), and several local genealogical societies.

The focus group protocol included a semi-structured questionnaire and a slideshow illustrating features of the legacy system interface as well as ideas for new features. Marchionini et al. (2003) noted: "An inherent limitation in directly assessing the human needs for an innovation is the fact that potential users must imagine what the innovation can and will do for them" (p. 123). In order to address this limitation and to stimulate thinking, mock-ups were created by the user interface team to illustrate possible new features for the Portal. The discussions addressed each of the key functional areas within the Portal: (1) Search, both basic and advanced; (2) Browse; (3) Search results; and (4) Object navigation for photographs, maps, and multi-page documents. The discussions were recorded and the audio recordings were subsequently transcribed and analyzed.

In addition to providing input for the functional requirements discussed at the end of this section, the findings identified key information seeking characteristics of genealogical researchers. These include:

1. Genealogists, both hobbyists and professionals, research family histories as well as the historical and cultural context in which the individuals lived. In the former instance, genealogists primarily seek three types of data about major events in individuals' lives: names, locations, and dates. In the latter instance, genealogists seek information that situates individuals within a larger historical context in order to discover insights about the society and culture in which they lived.
2. Most often, genealogists will begin their information search using a surname, although many will start by using advanced search features to add a location (often a county name), as well as a date range.
3. Genealogists include people with a wide range of both research experience and computer skills. Less experienced researchers need visible features to successfully navigate the materials.
4. Some terms commonly used in digital library interfaces are generally misunderstood or not understood by most genealogical researchers. These include metadata, relevance, permalink, fulltext, and creator. Likewise, most genealogists do not understand distinctions among file formats, such as jpeg and tiff.
5. Genealogists often publish their family histories and are attentive to copyright restrictions and the need for permission from copyright holders prior to publishing the materials they discover. Likewise, genealogists are aware of

Figure 2. Example mock-up of a comment function

the importance of citing references for their information sources. However, in practice, many genealogists find this a difficult task and fail to obtain the data elements needed for adequate source citations.

6. Genealogists often wish to see the source objects whose digital surrogates they discover in libraries and archives. Having found items of interest to their research, they will often explore a digital collection more fully and may well visit the library or archive in person at a future date.

1.1.3. Feature Ranking Questionnaire

Participants in the three focus groups and the usability tests completed a questionnaire that measured their interest in specific features that the reengineered Portal might include. The results were tabulated and the features were ranked by the percentage of users indicating they were interested in each feature.

The top two features of interest to all participants were saving items and saving search results. These were followed by adding items to personal favorites and accessing the personal search history for an active session. Compared to the focus group participants, the participants in the usability tests were somewhat more interested in building lists of objects and somewhat less interested in commenting on items and annotating images. The features of less interest to all participants were receiving RSS feeds of search results, rating the historical significance of objects, and commenting on other users' comments.

1.1.4. Overall Design Characteristics

Members of the Digital Libraries Division staff, including system developers, programmers, metadata librarians, interface designers, program coordinators for the Portal, and staff involved in the digitization and preservation of materials, completed a brief questionnaire that included the questions listed below regarding: (1) general information about the Portal, (2) the audience the Portal serves, and (3) the desired perception the Portal should strive to elicit in visitors.

- **General Information**
  - What are the goals and objectives of the site?
  - What aspects of your current site do you like and/or not like?
- **Audience**
  - Describe the audience and typical user(s) of the Portal?
  - How often is the user online?
  - What does he/she generally use the Portal for?

- **Perception**
  - Use a few adjectives to describe how users should perceive the Portal (examples include prestigious, friendly, corporate, fun, forward thinking, innovative, and cutting edge).
  - Is this different than the current image perception?
  - List the URLs of any sites you find compelling and would like us to use in conceptualizing or for reference.
  - What specifically do you like about these sites?

The objective of the questionnaire was to stimulate thinking among internal stakeholders that would identify the Portal's major user groups and inform decisions regarding the overall design characteristics of the interface, as well as desirable features and functions.

1.1.5. Internal Stakeholders Group Discussion

Particularly important to the design effort was achieving a consensus among internal stakeholders regarding priorities for the Portal's overall information architecture, functionality, and design. An additional design goal was to achieve an optimal balance between the needs and priorities of users, content partners, and the library. To these ends a discussion among a group of internal stakeholders was conducted by the interface design team.

The group included current members of the Digital Libraries Division as well as members of the initial team within the UNT Libraries who had identified the need for the Portal, recruited the founding partners, obtained seed funding from external sources, selected the original asset management system, designed the original data model, and launched the first public system in 2004. This enabled a range of experiences and perspectives to be considered in the design of the user interface. Participants shared and discussed websites they particularly liked. Features, navigation, and aesthetic preferences were demonstrated.

Consensual decisions concerning the overall design, functions, and information architecture were identified. For example, the following user groups were identified as the audiences the Portal serves: researchers, historians, genealogists, educators/teachers, K-12 students, lifelong learners, family history enthusiasts, and the general public. However, consensus was reached that the needs and opinions of three expert user groups (researchers, historians, and genealogists) would outweigh the preferences of other groups when conflicting aspects of the design arose. Based on the experience of the design team, decisions based on consensus that are established early in the design process help clarify what is expected of the development team. Additionally, they prevent downstream scope creep that is always costly in terms of time and human resources.

1.1.6. Analysis of User-submitted Comments

Portal visitors submitted 425 comments to library staff via a link on object display pages between October 13, 2005 and January 8, 2008. Content analysis classified these into two groups: comments related to a specific object itself and other types of feedback (e.g., typos and requests for copies). About 60 percent of the comments in the log were classified as comments about objects and about 40 percent as other feedback. Analysis of the comments about objects proved more fruitful in terms of understanding the Portal's users.

The comment group was further classified into three categories: (1) pertaining to people depicted in an object; (2) pertaining to locations depicted in an object; and (3) pertaining to other object content, including descriptive metadata.

Within each of these three categories, comments were further classified as identifications, questions, notes, or errors (e.g., in the content of the descriptive metadata).

The analysis indicated that users were particularly interested in identifying people and locations, although their comments often included dates as well. About half (51%) of the object-specific comments related to people depicted in objects and 28% related to locations depicted. This interest echoed the importance of names, locations, and dates that emerged in the focus group discussions. Further analysis of this group found that user comments identifying people constituted the largest sub-category within the entire log (16%; n=71). Identifying locations ranked third within the entire log (10%; n = 44).

1.1.7. Initial Usability Testing

Consistent with usability testing practices (i.e., Nielsen, 2000), five volunteers from a genealogical society were recruited by a member of the project's advisory board to participate in usability testing of the legacy system. The participants were females between 40 and 70 years of age. On average, they had been doing genealogical research for 14 years. One person was a member of the Association of Professional Genealogists (APG) and two were members of the National Genealogical Society (NGS). Test sessions lasted one hour and a talk-aloud protocol was used.

A test script with four scenarios reflecting the types of research goals that genealogists commonly address was created. These goals were: (1) find an answer to a specific question about a person, (2) test a hypothesis concerning where an ancestor lived around a certain date, (3) find evidence to prove a statement from one ancestor about another ancestor, and (4) investigate a topic related to the history of a certain area. Participants read aloud a scenario and its associated tasks prior to searching or browsing the Portal to find an answer.

Audio, video, and screen recordings of each session were captured using Morae® Recorder software from TechSmith. Each recording was analyzed using Morae® Manager software to identify the duration of each task, key observations and problems, tester prompts, and illustrative video clips. Task completion was rated either 0 (*completed with ease*), 1 (*completed with difficulty*), or 2 (*not completed*).

Morae® Manager was also used to measure the Portal's effectiveness and efficiency in regard to the test scenarios. For the four scenarios tested, the Portal's efficiency, as measured by the average time participants engaged in each scenario, was clearly related to its effectiveness scores. The more difficulty participants encountered in effectively resolving a scenario, the more time they spent attempting to do so. Overall, participants were fairly motivated to resolve the scenarios, even though the records they might commonly use to do so (e.g., census records or burial registers) are not available on the Portal.

Among the findings was important feedback regarding the Portal's home page, which for new users of the system provided no clues regarding the types of information and records contained in the Portal. Other findings supported priorities that emerged in the focus group discussions: (a) name searches are a first priority for genealogists; (b) access to information needed for citations is important; and (c) saving and printing objects is important.

1.2. System Requirements

1.2.1. Functional Requirements

Analysis of the data collected from the various activities of the needs assessment resulted in a draft set of functional requirements for the Portal's user interface. The requirements were categorized into existing or new functions. Existing functions were grouped into the following categories: (1)

search (basic and advanced), (2) browse, (3) search results (list and grid views), (4) metadata, (5) object navigation, and (6) help. Six new functional categories were identified:

1. **Obtain:** Allow users to save and print objects, along with their citations and metadata. Users may order high-resolution prints of images.
2. **Comment:** Allow any user to submit error reports and allow registered users to add comments, view others' comments, and communicate with other registered users.
3. **Register:** Offer users a simple registration process.
4. **Create Lists:** Allow registered users to merge search result lists and create object lists.
5. **View Map:** Allow users to view search results on a map of Texas counties that visually indicates the variance in the number of hits for each county.
6. **View Timeline:** Allow users to view search results on a timeline.

The functional requirements provided a foundation for the redesign of the interface to The Portal to Texas History℠. The requirements were classified by members of the three project teams into one of four development priorities: (1) Implement in Release 1; (2) Implement in Release 2; (3) Consider for Future Releases (i.e., not possible in the project's timeframe); and (4) No Development Planned (i.e., not feasible within the system).

At the time of this classification, existing workflows within the Digital Libraries Division did not include a formal review process for draft functional requirements. One oversight in the review process was the failure to include the Portal's program coordinators. As a result, some requirements that should have been clarified or omitted were not. Additionally, documentation activities (i.e., writing help content, modifying metadata practices, and updating partner and collection profiles) could have been better understood at an earlier date and assigned to appropriate staff members who were not directly involved on one of the three project teams.

A second issue with the draft requirements was with the initial wording of some them. For example, some requirements were redundant and could have been eliminated; other requirements were quite similar and could have been merged. Sometimes requirements included too much design specificity (e.g., "include a print icon on search results pages"). It would have been helpful to have spent more time and resources early in the project to resolve these issues prior to classifying the requirements.

Functional requirements identified for implementation in Release 2 of the Portal were revised in June 2009. The major modification resulted from a decision to defer development of a shared commenting feature, which required development of a registration capability that was beyond the development resources available for the project.

1.2.2. Structural Requirements

As a prerequisite for implementing the rapid development framework, system requirements included specifications and practices to enable implementation and testing of prototype technologies and standards prior to final implementation. The four key specifications are listed below.

1. **Digital Object Manifestations Model:** This is an extensible model that defines what a digital object entity is as well as a consistent naming convention for the possible manifestations of any digital object (e.g., various file formats such as TIFF, JPEG, and PDF for a specific digital object). Based on existing and future content, the object model allowed the development team to create standardized tools for reading and writing digital objects. METS (Metadata Encoding

and Transmission Standard) was used as a serialization format for this object model throughout the system.

2. **Persistent Identifiers:** Archival Resource Keys (ARKs) were implemented as part of the persistent identifier strategy. ARKs provide a globally unique naming system for objects in the digital library. Digital objects within the Portal were mapped to URLs, with ARKs playing a key role in providing logical, hackable, and bookmarkable identifiers for the system. For example, the thumbnail for each object in the system is available by appending/thumbnail/ to the end of the ARK URL, which always returns a thumbnail as in: http://texashistory.unt.edu/ark:/67531/metapth163241/thumbnail/.
3. **Metadata Scheme:** The UNT Libraries uses a locally qualified Dublin Core® metadata format (UNTL[1]) for all collections within its digital library. The UNTL input guidelines and formatting rules were updated to reflect the new data model. Additionally, all controlled vocabularies used within the UNTL metadata scheme, as well as in other parts of the system, were added to a controlled vocabulary management tool. The new metadata model allows for organic change within the scheme by adding qualifiers as needed to the vocabularies management tool. An example would be the addition of new types of dates associated with the lifecycle of an object (e.g., date harvested or date embargoed until).
4. **Edit Application:** Requirements for a new application (edit) to facilitate modification of records in the new system were specified for later development. The edit application provides a dashboard for collection maintainers (e.g., content partners) to interact with their objects in the system. They can search, browse, and limit records to a subset of all their records and edit each of those records through a user-friendly Web-based interface. When saved by the record maintainer, the edit application versions the record. This makes it available for future edits and re-indexes the new record, making any metadata changes available to the public instantly. Many of the record maintainers for the Portal are external partners not affiliated with UNT.

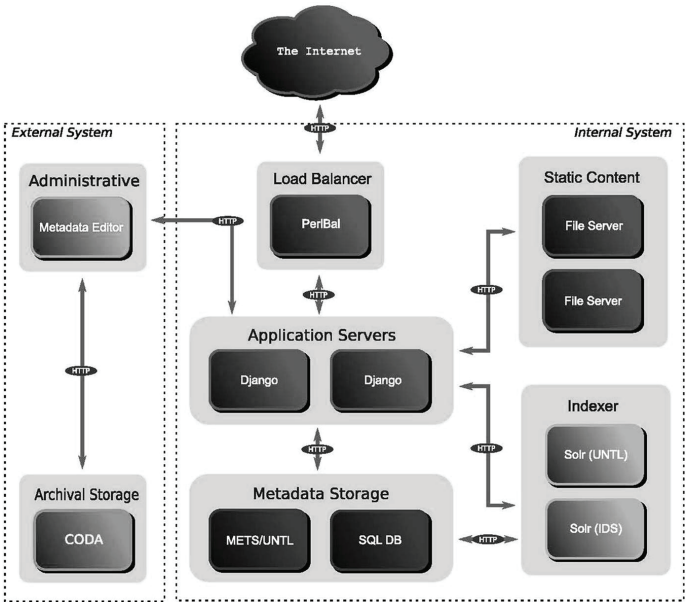
1.2.3. Architectural Requirements

As the number and complexity of digital objects in the Portal grew, it was apparent that key areas within the legacy system needed to be re-engineered to increase flexibility and to accommodate future growth. A major challenge within the legacy system architecture was that one application was responsible for both the access copies of objects as well as the preservation metadata associated with the master objects, which were managed by another system. Additionally, the legacy system made assumptions about the location of files within a traditional file system, which made adding storage resources challenging. The bottom-line result was a bottleneck within the system that limited the amount of new content that could be ingested. In order to address these inherent issues with the legacy system, architectural changes were needed in the core infrastructure. The fundamental change was to move from a monolithic system design to a decoupled, Web services approach to system architecture.

The reengineered system architecture specified two core infrastructure systems: external and internal systems (Figure 3). This architectural design effectively decoupled access services associated with digital library programs, such as The Portal to Texas History℠, from preservation services associated with Libraries' digital repository. Doing so has allowed the UNT Libraries to manage the growth of its storage and processing components in a staggered, systematic manner that was not possible within the legacy system architecture.

The external system includes an administrative component, the *Metadata Editor*, for editing and adjusting records via the *edit* application. It also includes the *Complex Object Digital Archive* or

Figure 3. Core infrastructure systems and components



CODA, which is the archival storage component of the system that supports long-term preservation and curation of the Portal's digital objects. In addition to preservation management of the Portal's digital objects, *CODA* manages the master files for other digital library initiatives within the UNT Libraries.

When possible, open source software and tools were selected, for example, the Ubuntu® Linux operating system, the Apache HTTP Server™, the Python programming language, and mod_python, an Apache™ module that embeds the Python interpreter within the server.

The internal system, named Aubrey, is comprised of the following five core components:

1. **Static Content Storage Servers:** These include the file servers for the Portal's digital objects (e.g., image files and OCR-text files).
2. **Servers for the Open Source Solr Search Platform:** These include indexes at both the digital object level (e.g., a photograph or book) as well as page-level indexes (e.g., pages within a book).
3. **Metadata Storage Servers:** These contain: (1) the METS files for the Portal's digital objects; (2) the UNT Libraries' (UNTL) standard metadata records for digital objects; and (3) the MySQL relational database management system, which includes collection and partner information for the user interfaces.
4. **The Perlbal HTTP Load Balancer:** This is a Perl-based open source application that distributes HTTP requests from the Internet across the Portal's application servers to optimize system performance.
5. **The Application Servers:** These are implemented within the open source Django® Web framework that supports the design and development of the Portal's website. The Django® framework includes the Portals three core applications, search, browse, and contribute, as well as the edit application that supports the Metadata Editor. The framework also includes the ARK component, which is responsible for mapping a digital object's features to the ARK URL presented to the end user. All interactions with the digital

objects are handled through the ARK component, which creates an application where the URL is the API for objects. The Django® framework and the application interfaces to system components are illustrated in Figure 4.

1.3. Requirements Definition: Lessons Learned

1. External stakeholders include the Portal's content partners and funding agencies. Program coordinators are in the best position to represent their interests and needs, which constitute significant input for design and implementation decisions.
2. Internal stakeholders include the Portal's program coordinators and the various departments within the UNT Libraries. Departments need to be informed of changes that might impact their operations. Communicating with them early on would lessen any impacts from changes.
3. User needs should be stated as general cases in requirements documents and broad classes of users, for example, novices and experts, should be identified.
4. It would be advantageous to add a formal review process for draft functional requirements into the workflows.
5. Stakeholders, both inside and outside the library, need to be represented and/or included in the requirements review process and in the process of establishing development priorities.

PHASE 2. SYSTEM DEVELOPMENT

System development activities are described in two sections: (2.1) Prototype Development and (2.2) Application Development. Descriptions of the system architecture and the open source tools are included. Section 2.3 lists the major lessons learned by members of the three project teams during phase 2.

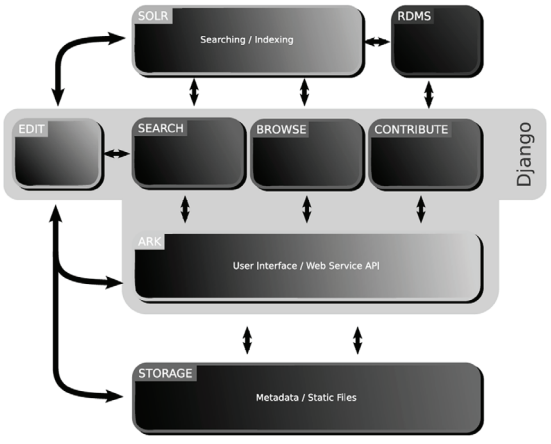
2.1. Prototype Development

2.1.1. Content Migration

A conversion script was written for the migration of digital objects from the proprietary format specified in the content model of the Portal's legacy TKL system to the new METS data model used by the Aubrey system. This conversion transformed the previous descriptive metadata into the new UNTL specification and legacy structural information into the METS specification for the Aubrey system. A number of inconsistencies were encountered in the legacy format, which required a substantial amount of code to be written to test the conversion.

The bulk of the work required in the conversion was in the change from the proprietary XML format for storing metadata and structural information about files (i.e., sequence, page numbers, and OCR-text) to the new METS format. In the legacy system, some content, such as OCR-text, was stored inline whereas in the new system this

Figure 4. Django applications and interfaces



content was stored as static content outside of the METS files. The conversion script was refined during successive migration trials and successful migration of a subset of the Portal's digital objects was completed for use by developers in the prototype system. Prior to public launch of the reengineered Portal, smaller sets of objects than those used in trial migrations were converted in order to further ensure a successful, error-free migration of the Portal's entire content to the new system.

Creating and implementing new backend workflows for moving digital objects in and out of the system took a great deal more time than anticipated. As a consequence, the amount of time and resources required were not adequately estimated. However, these workflows had to be in place prior to beginning development of the user interface, which consequently was delayed.

2.1.2. Prototype System

Since the legacy Portal was a production system, a separate development system was installed for application development. The development environment was defined for multiple developers, each working in different areas of the project. It was decided that separate development, staging, and production environments would be used for the project and they were created using a combination of virtual machines and physical servers. This separation of environments allowed changes to be made in different areas in the architecture that would not interfere with other areas. Additionally, deployment of code to the different environments uncovered bugs based on assumptions of one system versus another.

Servers were installed and configured for the Portal's static content (i.e., digital objects, including image files and OCR-text files). The digital objects were stored using the hierarchical file system specification, Pairtrees for Object Storage. A Pairtree file system enables an application, for example the *search* application within the Portal, to locate digital objects by mapping an identifier string to the file location. As part of Phase 1, Archival Resource Keys (ARKs) were implemented as persistent digital object identifiers within the system. The ARK application within Aubrey maps the identifier strings to specific file locations in the static content storage servers. An example of this mapping occurs in a request for the item *ark:/67531/metapth125323*. The Aubrey system includes the logic to convert this to a file path, *me/ta/pt/h1/25/32/3/metapth125323/*, which is appended to the location of a static content server where all files associated with that object are located.

Servers were also installed and configured for digital object metadata, including both METS and UNTL metadata. These files were again written to disk using the ARK identifiers mapped to the Pairtree file system. The Aubrey application converts the incoming requests for an identifier into a file location where it can access the metadata files. The *edit* application, developed after Release 1 was launched, interacts with the metadata files to both read and write metadata. All changes to the metadata are versioned and stored on the metadata servers in the Pairtree file system.

The metadata files and static content were located on different servers because of their rate of change and expected growth. Metadata files in the Aubrey system are considered *business data* and are part of the Libraries' daily backup procedures, whereas the static content is replicated using other technologies tailored for the high number of files and disk space required.

2.1.3. Search Platform and Indexer

The Solr search platform includes documents (or indexes) that represent the Portal's digital objects at both the object and page levels. A new document schema was specified and implemented to support future user interface features, such as search facets and collection pages. Many of the features of a digital object in the new data model were mapped

to the document schema. This allows quick lookups of features in the index that can be used for system functionality or other types of functions. For example, inclusion of the number of files comprising an object allows statistics about the Portal's content to be generated. Likewise, inclusion of a field designating the access policy for an object allows user interface features to display different information based on an object's access policy. As the project progressed and new requirements were identified through the iterative process, new fields were added to the document schema.

The UNT Digital Library uses an in-house digital object indexer (Irex) which is responsible for the conversion of UNTL and METS records into the format needed by Solr for indexing. A new version of Irex was developed to support the METS and the UNTL metadata models. Additional logic was added to this indexer to improve the experience of end users, for example, removal of the end-of-line hyphenation present in OCR files that splits words across lines. This transformation occurs at index time and is not volatile to the original OCR files. Another example of additional logic used by Irex is the conversion of the various date formats used in an object into standardized machine-readable versions. This is an important process because the Solr platform requires certain formats for sorting and date range searching.

2.1.4. Prototype Designs

Paper prototyping is an essential component of an iterative user-centered design process in that it allows the interface design team to identify design flaws early, to iterate inexpensively, and to adjust aspects of the design rapidly. Additionally, due to their low-fidelity, the paper prototypes present users who test them with a malleable product that is free of aesthetic distractions and largely approachable. Because the product is clearly incomplete, people can be more honest about problems and are apt to think about the items under study rather than adapt to a finished-looking model. As noted by Spool, "real-life representations will likely contain critical flaws. Since the reason we're prototyping is to identify those flaws, we find it essential to draw everything out" (Spool, 2005).

To this end a paper prototype depicting the navigational structure for the redesigned Portal interface was created; although, instead of using paper and pen, the team decided to present cleaner lines drawn with Adobe® Fireworks® since they were intimately familiar with this graphics program. A tree diagram illustrated the top-level functions (i.e., browse, search, help, and about) for the site as well as the navigational flow through the site, for example, from a basic search page to a search results page(s), to the digital object and brief metadata page. The content layouts for specific screens were also drafted as prototypes. These screens included the Portal's home page (Figure 5), the six explore (i.e., browse) pages, basic and advanced searches, and grid, list, and brief views of search results. For individual objects, an *about the object* screen and a *view/read the object* screen were created. Documentation pages for *help*, *FAQ*, and user *guides*, as well as an *about* page for the Portal itself, were also created.

2.1.5. HTML-Based Mock-Ups

HTML-based mock-ups, as a testing medium that more closely resembles the final product than paper prototypes, were deployed as the next interface design step. A non-functional, HTML-based mock-up of the Portal was created using standard Web development tools like Adobe® Dreamweaver® and the Blueprint CSS code library. While not fully implementing all required functionality, the mock-up illustrated possible user interaction flow as well as options for the Portal's overall look and feel, for example, the color choices, image sizes, and fonts. Using a CSS framework ensured that important design aspects such as whitespace, typography, and grid-design were consistent and would be complimentary to the content on pages. The purpose of the mock-ups

Figure 5. Initial paper prototype for the home page

was to visualize all interface related requirements and to expose potential user-interface related issues. As a partially working model, screens within the HTML mock-ups were interlinked and search forms, though not functional, behaved by forwarding correctly through any given scenario.

The goal of the mock-ups was to ensure that the features that would be implemented would satisfy the requirements of both users and internal stakeholders. Unfortunately, genealogists were not included in the prototyping and mock-up activities. In the future, clarification of user requirements would be better achieved with continued involvement of the user community through this stage. However, designers did test several storyline walkthroughs. Likewise, digital library staff members reviewed the mock-ups and provided feedback to the design team, who revised the mock-ups. This pattern of design, review, and revision was repeated until consensus on the designs was achieved.

2.2. Application Development

2.2.1. Collaboration and Management

The system development and user interface teams used two open source systems to manage application development: Apache Subversion® and Trac. Subversion, a project of the Apache Software Foundation, is an open source version control system. And Trac, an open source project of Edgewall Software, is an issue tracking system for software development projects that provides an interface to Subversion. Together these tools facilitated typical software development activities such as check-in and checkout of source code and creating and merging code branches. Trac provided tools for submitting bugs and managing milestones within the project. The project team found both indispensable during the development cycles.

Actual development occurred on two independent testing servers, one primarily devoted to the development of business logic within the Django® application and the other to presentation, which would occur at the browser-level. As the need arose, systems and interface developers created tickets and, upon completion of requested features, checked-in code for updating the other's system. As an example, on individual object navigation screens, there was a requirement for text highlighting of instances of OCR-text that matched a user's search query. The interface designer requested a list of coordinates be returned as an array that could be parsed in JavaScript. While the systems developer built this list in python, the interface developer generated the necessary display logic and, once the code was available, hooked the elements together. Collaborating in this manner allowed each development team a measure of stability in their daily work since the codebase was typically stable.

2.2.2. Software Development

1. **Branding Application:** Requirements for this application were coded and the application was implemented in the prototype system. This application allows a subset of records held by Aubrey to be designated for presentation to an end user community. For example, The Portal to Texas History℠ is a uniquely branded system comprised of a subset of the digital objects held by the UNT Libraries in the Aubrey system. Likewise, the UNT Digital Library is a separately branded system comprised of other digital objects, also held by Aubrey. It is possible for content to be shared among more than one branded system. Each brand has custom variables, which define the layout, style, and text used in its user interface.
2. **ARK Application:** Modifications were made to the ARK application to support the newly implemented Pairtree directory structure. As a request is made to the ARK application, it is mapped to an underlying data structure built from the UNTL and METS metadata records. This application provides a consistent and stable URL structure for the end user and abstracts the filenames and storage platform providing more flexible content management.
3. **Edit Application:** The requirements specified in Phase 1 for the edit application were coded. This application enables an authorized user (or metadata maintainer) to create or edit descriptive metadata associated with a digital object and save the edits, which are automatically re-indexed in the Solr platform. The edit application includes features that allow an authorized user to manage their collections on the Portal. Unfortunately, there were insufficient resources to develop this application in parallel with other essential infrastructure work. It was necessary to suspend the Portal partners' ability to edit their metadata records while the application was developed.
4. **Content Ingest Tools:** Three tools were created and tested: SIPmaker, AIPmaker, and ACPmaker. These tools enable new content additions to the system. The tools are used by the Aubrey system to ingest content of all types into both the Aubrey and CODA systems. There are several ingest servers used during the ingest process and these three tools are used to create archival and access packages which are then stored in either the CODA system or the Aubrey system. Preservation metadata and access derivative generation are handled by these tools.

2.2.3. Interface Development

Guided by the prototypes and mock-ups, the interface design team implemented the Portal's Web interface. Building on the HTML wireframes as an actual skeleton for the site's markup, the team constructed the site using a number of established code libraries and best practices from the larger Web-design community. Using code libraries gave designers a degree of flexibility, in that they provided well tested, community supported foundations upon which to build. In essence, the model that applied to the larger site's framework, using tools such as Django® and Perlbal, could be applied to the design as well.

Having initially employed Blueprint for its CSS framework, the designers employed a more adaptable and robust framework for styling called Compass CSS. Built by Christopher Eppstein, this library expanded on Blueprint's strengths to (a) allow for better semantics within the site's markup and (b) aid in the abstraction of styles into a more programmatic environment using Sass and the Ruby programming language. Because this was code that could be processed, it was written using

DRY (Don't Repeat Yourself) principles and its output could be tested, compressed, and optimized for distribution.

For client-side behaviors, the interface team employed several existing JavaScript libraries, most notably jQuery and OpenLayers. To actually make use of these libraries, several contributed plugins were employed and the interface team wrote several thousand lines of JavaScript to create fully-functional features, such as the page zoomer, text highlighting, and a number of other visual and functional enhancements.

2.2.4. Documentation Creation

Documentation was needed in support of several areas within the reengineered system. Relying on both the functional requirements developed in Phase 1 of the project as well as content design requirements for the Portal interface, an ad hoc team of digital library staff members was established to develop documentation. This documentation related to information about collections and content partners, as well as information regarding internal practices within the digital library, for example, digitization standards, metadata practices, and partnership options.

The metadata for all content partners and collections was updated in a consistent manner that would support ease of future editing and consistent display within the interface. This important work contributed to the dramatic improvement in the visibility of content partners and their contributed materials. Additionally, information regarding options for becoming a content partner and the practices and standards employed for digitization and digital object metadata creation were reviewed and revised in conjunction with Release 1.

A second important documentation area was the user documentation for each of the major functional areas within the Portal. The *Help* feature of the reengineered Portal included specific guides for each major function (e.g., basic and advanced searches), FAQs, and a glossary of terms. Guides for discovering educational materials and genealogical materials were also created. This was a greatly expanded feature within the reengineered Portal and steps were taken to outline the documentation needed and to review any existing documentation prior to creating new content.

Two issues emerged in the process of creating user documentation. The review of existing documentation and the creation of an outline for specific content that needed to be written went smoothly. However, with a team of six individuals creating documentation, inconsistencies in scope, style, and format emerged. Additionally, it became apparent that functional guides could not be completed until the functions themselves were fully operational. Attempting to do so beforehand resulted in unmanageable version control and multiple re-writes.

2.2.5. Quality Assurance (QA) Testing

Subsequent to completion of Release 1 on the development system and prior to its public launch, a Quality Assurance (QA) test was created by the user studies team and completed by members of the UNT digital library staff and student employees. Testers reported an overall success rating of 84% for the 46 tasks. Unsuccessful test results for several tasks were influenced by the limited number of digital objects available in the development system. Test feedback and results informed a set of design and development tasks for implementation. With the objective of optimizing information display and improving the user experience, various solutions were considered, tested, and implemented prior to the public launch of Release 1.

The public version of Release 1 included the whole complement of the Portal's digital objects and a second round of QA testing was conducted in June 2009 after the system was launched. Testers included practicing genealogists in addition to

UNT digital library staff and student employees. Participants reported an overall success rating of 90% for the 37 tasks tested. As with the initial QA test findings, testers' feedback and results informed a set of system design and development tasks.

Prior to the public launch of Release 2, QA testing was conducted with 22 UNT digital library staff members and student employees. With the exception of one task, 73% or more of the testers indicated they successfully completed each task, although some noted issues and problems with some tasks they completed. As before, test feedback informed design changes, which were implemented prior to the public launch of Release 2.

2.2.6. Migration to Production System and Release 1

The production system's infrastructure was configured and the prototype system was replicated and tested prior to the public launch of Release 1 of the new Portal. This infrastructure meets current needs and provides capacity for growth both in terms of throughput and content. From a user perspective, the migration from the legacy system to the Aubrey system was seamless; user requests were automatically forwarded from the legacy system to the new system's ARK-based URLs. It was important to the project team that all existing URLs that users might access continue to lead them to the expected digital object. To complete the testing of the redirect feature, old access logs were used to provide previously requested URLs, which were then fed into the new application. Developers monitored the server logs to make sure that all URLs resolved to the appropriate ARK-based URLs used by the Aubrey system. The code for the redirects was adjusted until all requests were correctly forwarded.

After the initial launch of the system, log files were monitored and additional redirect rules were added as needed. Google Analytics was switched to the new system for continued data collection. Google, bing™, and Yahoo® began including links from the new system within 24 hours of the switch to the new URL structure. A second release of the Portal was launched four months after Release 1.

By staging the required features of the system across two releases, it was possible to deliver a new system to the end user in pieces instead of waiting until the final application had been developed and tested. In the implementation stage of any project, it is challenging to balance the desire to add new functionality with the need to deliver a product at a given time. The hardest part of the process is deciding which features will not be included in one release and will be rolled out in subsequent releases. The use of version control software and tools like Trac are helpful both for tracking bugs as they occur and for tracking feature requests to consider for development in future releases.

2.2.7. Release 2 Implementation

Usability testing of Release 1 of the Portal interface by the user studies team identified additional areas for revisions to the interface, primarily in regard to secondary navigation features. Different design changes were considered, mock-ups were developed, and feedback was solicited from the project team. Once consensus on the design direction was achieved, the changes were implemented, tested, and included in Release 2 of the Portal.

Additionally, two major features were developed for Release 2. The first was the addition to search results pages of a number of facets for limiting search results. Optional designs for the display of facets were investigated and tested. The second feature added a function that allowed users to return to their search results from object display pages.

2.3. System Development: Lessons Learned

1. It is very important to identify the resources required, in terms of people and time, for infrastructure work upon which an application development project is dependent. In

this project, the time and resources required to finalize the migration process from the legacy file format to the new data model were not adequately estimated. This delayed both establishment of the development framework and writing the edit application. In consequence, a decision was made to launch Release 1 prior to developing the edit application, which in turn necessitated a three-month suspension of metadata editing while the application was coded. Better estimation of time and resources would have helped set more realistic stakeholder and partner expectations.

2. Clarification and refinement of user requirements might be better achieved with more user involvement in the prototyping work that follows the assessment of user needs. Ideally, users' involvement would begin with a needs assessment to inform a set of general functional requirements. The requirements would then be translated into mock-ups of user workflows (either paper-based or online), which would be used in usability testing with a set of representative users. The findings of the usability tests would shape the final design requirements.
3. When constructing digital libraries on open source frameworks, development plans should take into account the evolutionary nature of the open source components. Because code libraries and tools come into and fall out of favor with communities of developers over time, the development team needs to consider the ramifications of choosing one library over another. The team might consider: How reliant is the system on any single component? Are there enough developers using it to ensure long-lived support? Regarding the open source project for any tool: Is it being actively developed? Is it stable? Has it fallen out of favor for a different tool? Is the tool well documented? How easily can a new team member learn and begin working with the tool?
4. Specialization allowed technical team members to develop expertise at each of two levels: (1) interface design or system design and (2) specific applications or technology areas within each design area.
5. De-coupling of interface design and development from the underlying system is a key to easily making changes, including upgrades. For example, the look and feel of the portal, along with client-side functionality, could be changed as needed, without impact to the underlying system. Likewise, system changes could be made independently of the user interface.
6. The Subversion system worked well in terms of providing a running log of all changes and informing work ticket assignments for developers. However, shared modules can be problematic. For example, with three developers working in the Django® framework, each one needed to understand who had access to what modules and who needed to be informed of module changes. With even more developers, workflows and additional rules would need to be enforced.
7. Given finite resources, it is prudent for a development organization to follow the design leadership of leading enterprises like Google and amazon.com, which invest heavily in usability testing; following their leadership, in terms of features and design, effectively leverages their investment in testing. The sophistication commercial enterprises bring to the presentation of large masses of information often becomes the intuitive approach that their consumers and digital library users become accustomed to using.

8. Technology changes over the lifetime of a project, as do user expectations. Users adopt new habits through their interactions with leading websites (e.g., Google). This reality particularly impacts user interface design. Several user interaction features, navigation layouts, and underlying technologies changed throughout the course of this project.
9. Quality assurance testing with users prior to publicly releasing a system is one activity that will be incorporated into the workflow for future application development as a result of this project. The testing was both valuable and affordable. Users brought issues to light that internal tests would not have uncovered. For example, developers tend to have a set of searches they use to test; a wider pool of testers will use different queries and therefore uncover different problems. In addition, software development tools often have testing limitations. For example in Django® and Python, internal tests of the search application will not uncover character problems, such as problems with diacritics or ampersands.
10. Including partners and library stakeholder groups in the QA tests would hopefully foster a sense of the Portal as their system. Additionally, they might uncover problems that others did not find.
11. Service interruptions may happen with any major technology implementation or change, such as occurred between June and September 2009 while the edit application was being written. Content partners effectively could not add new metadata records or edit existing ones, and they received inadequate notice of this service outage. Advance notice to stakeholders should make these types of disruptions more acceptable. Ongoing communications with partners and other stakeholders should routinely advise them of system upgrades, expected disruptions, and fairly accurate estimates of completion dates. Internal stakeholders, including program coordinators, need to be cognizant of system development plans and activities that may impact the external stakeholders with whom they interface.
12. The amount of staff time required to develop supporting documentation, such as help guides and FAQs, was also underestimated. The application development workflow needs to include a documentation activity prior to the public launch of a new system or new software release.
13. An ad-hoc group was needed to create the online documentation accessible from the Portal interface. It was noted that while Portal users infrequently access the help guides, they are very helpful reference tools for Portal support staff.

PHASE 3. USABILITY TESTING

Subsequent to both Releases 1 and 2 of the reengineered Portal, the user studies team conducted usability tests with volunteers recruited from three North Texas genealogical societies by members of the project's advisory board. The participants (N=7 for Release 1; N=6 for Release 2) included 10 females and 3 males, ranging from 31-80 years of age. Most participants (85%) were over age 50. Participants had been doing genealogical research an average of 17 years.

The test scripts utilized scenarios that corresponded to the four research goals or information problems that genealogists typically address: (1) questions to answer, (2) hypotheses to test, (3) statements to prove, and (4) topics to investigate. As with the initial usability tests, a talk-aloud protocol was used in each test session and recordings were captured using Morae® software from TechSmith. For each release, the tasks in the test script were targeted to evaluate specific features and functions of the user interface. Morae® soft-

ware was used to analyze each test session and to produce video clips illustrating user behavior.

In the interest of communicating the findings of the Release 1 tests in a timely manner, they were reported in a project team meeting and a number of short videos were produced to illustrate users' behaviors. While the video clips highlighted problems, they did not assign a severity level to the illustrated problems (e.g., critical, moderate, or mild), which would have been useful to developers. Likewise, it would have been useful to have measured task completion.

Findings from the Release 2 usability tests did include a measure of task completion for each of the 42 tasks in the test script. Each task was rated on a 3-point scale: *1* (completed with ease), *2* (completed with difficulty), or *3* (failed to complete). All participants completed 17 (40%) of the 42 tasks with ease. An additional 22 tasks were completed by most users, although some had difficulty doing so. The average completion scores for only three tasks were in the *failed to complete* range (i.e., average score greater than 2). Similar tasks had been identified as problem areas in the Release 1 tests, in particular, tasks that tested users' ability to locate secondary navigation features.

OUTCOMES

In the legacy system, both the addition of storage resources and the development of application enhancements were protracted and time-consuming processes. The reengineered system overcame these constraints. Within the new system architecture, storage servers for both metadata and content can be added as needed, thereby accommodating content growth. Likewise, the Django® framework enables user interface design and development to proceed independently from the underlying system, thereby enabling more agile application development.

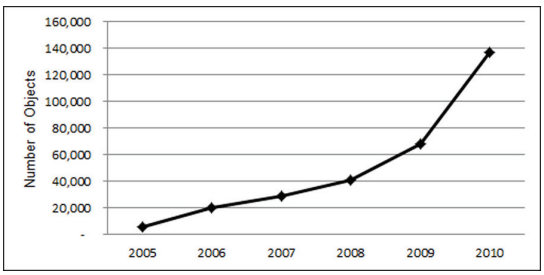
The redesigned Portal was launched in two public releases: Release 1 in June of 2009 and Release 2 in October of 2009. The reengineered system readily scaled to meet the Portal's digital object growth. In 2010, the number of digital objects doubled, from 68,344 to 137,111 (Figure 6).

The infrastructure also proved robust at handling an ever-increasing numbers of visitors. The number of Portal visits per month increased by 59% from June 2009 (21,775), when the new Aubrey system was first released, to January 2010 (34,680), when the project ended. The reengineered system handled this increase in visitors with ease. By January 2011, this number had grown to 45,000 visitors per month.

FUTURE RESEARCH DIRECTIONS

In addition to reengineering The Portal to Texas History℠ to eliminate the constraints of its legacy infrastructure by replacing it with an open source development framework, a major goal of this project was to create a model for digital library application development that included an iterative user-centered design approach. The project was initially guided by a draft application development model (Figure 1), which identified three teams: System Development, Interface Design, and User Studies. The internal post-project review identified areas for improvement in the model and these were incorporated into a revised application development model (Figure 7).

Figure 6. Digital object growth



Perhaps most importantly, the revised model includes a fourth project team: Program Management. Specific digital library programs, such as The Portal to Texas History[SM], are often managed by program coordinators who are not system developers or designers. Program coordinators have responsibilities similar to business account executives. These include marketing products to clients and also providing customer service by conveying clients' needs and issues to the product team. Academic library liaisons have a similar role between the library and the departments and faculty within a university.

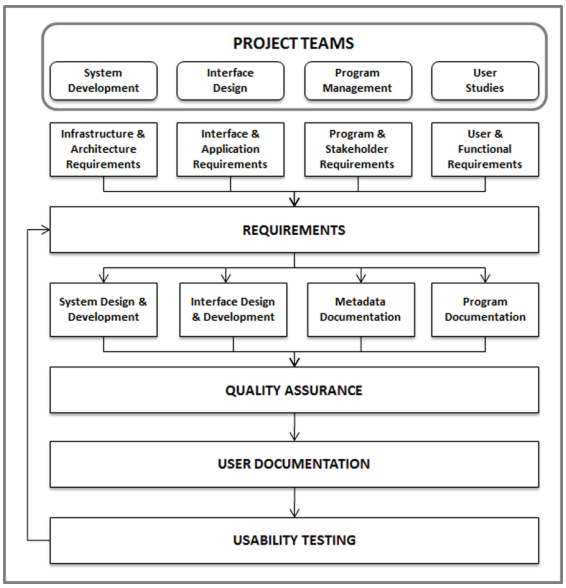
Program coordinators within a digital library are effectively library liaisons for a specific digital library program. At the University of North Texas, program coordinators' clients include significant user groups, both internal and external to the university, as well as content partners and funding agencies. As academic libraries increasingly provide digital access to their unique collections and collaborate with other content providers to develop new digital collections, program coordinators will continue to play critical roles within the library. Identification of the skill set needed for successful program coordination within an academic library is an area for further research.

The revised model also incorporates an understanding that, in the application development process, some requirements are addressed by metadata documentation (e.g., updating input guidelines and formatting rules) and program documentation (e.g., user guides). In order to complete a project on time and within budget, explicit identification of the resources for these activities must be included in project management plans. In the revised model, creation of user documentation occurs after quality assessment of an application is completed and prior to usability testing with external users. In this manner, documentation, in addition to application features and functions, can be included in usability testing. This should allow problems to be corrected prior to publicly launching an application. Creating documentation can be resource-intensive. Understanding to what extent user documentation is needed in a digital library application is an area that could be further investigated to evaluate both its effectiveness and utility for system users and library staff.

With its commitment to open source components throughout the reengineered Portal, the UNT Libraries reflects the acceptance of open systems in the wider academic community. Increasingly, faculties are requiring open access to research publications and funding agencies are requiring open access to research data. As academic libraries adapt their infrastructures and services to address these requirements, it is likely that open source-based systems will become more common in libraries that can provide the necessary technical infrastructure and support. In such libraries, the model resulting from this project supports a relatively low-cost approach to creating and redesigning user interfaces to digital libraries when compared to integrated, tightly-coupled, single-vendor, system solutions.

Applying the model will hopefully promote the design and development of user interfaces that are

Figure 7. Revised application development model



informed by the information needs and contexts of targeted user communities. As group needs and contexts change, responsive interface redesign will be enabled by the development framework. The advantages of open source frameworks for application development include design flexibility in the local library environment and access to a robust community of developers. These benefits are highly desirable for digital libraries seeking to optimize their user interfaces for key user communities. Because digital libraries built by academic libraries often serve user communities beyond their traditional users, models for how to meet user needs are needed. Testing the applicability of the model developed in this project with user communities of other academic libraries is an area for further research.

CONCLUSION

In order to overcome the application development constraints of the legacy asset management system on which The Portal to Texas History℠ was built, the UNT Libraries reengineered the system. To coordinate activities, the project team drafted a working model consisting of three teams: system developers, user interface designers, and user studies researchers. The model incorporated user-centered design methods, involving genealogists in the design process beginning with an assessment of their information needs and continuing through usability testing of the redesigned Portal interface.

Substantial amounts of time and effort were invested in the specification of a new data model and in the migration from the legacy data model to the new model. The model defined a digital object entity and a consistent naming scheme for each manifestation of a digital object within the system. Archival Resource Keys (ARKs) were established and incorporated into persistent identifiers for each object. These building blocks enabled applications within the system to read and write digital objects in a consistent manner and were prerequisites for de-coupling design and development of the user interface from core infrastructure components.

Subsequent to successful migration of a test set of digital objects to the new data model, a development framework comprised largely of open source components was established. This allowed the Libraries' interface design team to implement prototype designs based in part on the findings of the user needs assessment with genealogists and usability testing of the legacy system. Coordination of activities between programmers supporting core infrastructure components and those developing the user interface was facilitated by two open source tools: Subversion®, a project of the Apache Software Foundation, and Trac, a project of Edgewall Software.

The accomplishments of this project are unique in academic digital libraries. This project successfully implemented a digital asset management system comprised largely of open source components, which are supported by a very large and robust community of developers outside of those who typically support this type of system in libraries. Open source digital asset management systems for libraries are more often supported by relatively small library-centric communities.

This project was also atypical of many digital library application development projects in terms of including users who are external to the university in the process. In professional meetings and conferences, librarians often report their inability to involve users in their design and development work and many digital libraries base their interface designs on librarians' requirements. Likewise, few data collection efforts involve users outside of the academic community on the scale accomplished in this project. There are several reasons for this. Many libraries simply lack the resources to involve users. Others have neither the staff with requisite skills nor the necessary technical infrastructure to do so. Often, libraries have no programming

staff at all, forcing them to choose more turnkey digital asset management systems that constrain local modifications to the user interface.

Importantly, the project also demonstrated the value of including internal program coordinators in the application development process, as well as the need for including external stakeholders in quality assurance testing prior to public releases of new and revised interfaces. As the UNT Digital Library's collections grow and become more integral to teaching and research at the university, new digital library programs will be identified. The need for management of these programs is clear so that the requirements of external stakeholders and end users inform application development, operational decisions, and strategic direction.

Reflecting on the success of this project in terms of its original objectives, the constraints of the legacy system were removed and the technical infrastructure has scaled to meet growth requirements. The UNT Libraries have leveraged the infrastructure and development framework engineered for this project to other projects within the university's Digital Library. The rate of collection growth in the UNT Digital Library would not have been possible without the investments made in this project.

REFERENCES

Boss, R. W. (2009). Digital asset management systems. *Tech Notes*. Retrieved August 10, 2011 from http://www.ala.org/ala/mgrps/divs/pla/tools/technotes/DigitalAssetManageme.pdf

Boyns, R. (1999). Archivists and family historians: Local authority record repositories and the family history user group. *Journal of the Society of Archivists*, *20*(1), 61–74. doi:10.1080/003798199103730

Butterworth, R. (2006). Information seeking and retrieval as a leisure activity. In A. Blandford & J. Gow (Eds.), *First International Workshop, DL-CUBA 2006, JCDL 2006, Digital Libraries in the Context of Users' Broader Activities*, (pp. 29-32). Chapel Hill, NC: JCDL. Retrieved August 11, 2011 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.8745&rep=rep1&type=pdf#page=43

Cherry, J., & Duff, W. (2002). Studying digital library users over time: A follow-up survey of early Canadiana online. *Information Research, 7*(2). Retrieved August 16, 2011 from http://informationr.net/ir/7-2/paper123.html

Drake, P. J. (2001). *Findings from the Fullerton genealogy study.* Retrieved August 11, 2011 from http://psych.fullerton.edu/genealogy/

Duff, W. M., & Johnson, C. A. (2003). Where is the list with all the names? Information-seeking behavior of genealogists. *The American Archivist*, *66*(1), 79–95.

Fernandez, J. D., Martinez-Prieto, M. A., de la Fuente, P., Vegas, J., & Adiego, J. (2008). Agile DL: Building a DELOS-conformed digital library using agile software development. In B. Christensen-Dalsgaard, D. Castelli, B. A. Jurik, & J. Lippincott (Eds.), *Research and Advanced Technology for Digital Libraries, 12th European Conference, ECDL 2008*, (pp. 398-399). Berlin, Germany: Springer-Verlag.

Fulton, C. (2009). The pleasure principle: The power of positive affect in information seeking. *Aslib Proceedings*, *61*(3), 245–261. doi:10.1108/00012530910959808

Gallop-Gerdman, G. (2000). We are family. *American Demographics*, *22*(9), 24.

Generations Network. (2005). *Americans' fascination with family history is rapidly growing*. Retrieved August 11, 2011 from http://corporate.ancestry.com/press/press-releases/view/?id=406

Jansson, E. (2010). *DSpace manakin UI: Case study of value and costs*. Retrieved August 11, 2011 from http://hdl.handle.net/10090/12618

Kim, Y., Ahronheim, J., Suzuka, K., King, L. E., Bruell, D., Miller, R., & Johnson, L. (2007). Enterprise digital asset management system pilot: Lessons learned. *Information Technology & Libraries*, *26*(4), 4–16.

Kuglin, H. (2004). *Granny hunting: The information seeking behaviour of genealogists in open access libraries: A quantitative study*. Paper presented at the Lianza Conference 2004. Auckland, Australia. Retrieved April 14, 2010 from http://www.lianza.org.nz/library/files/store_018/Granny_Hunting_Kuglin.pdf

Lancaster, F. W. (1986). Why vocabulary control? In *Vocabulary Control for Information Retrieval* (2nd ed., pp. 1–4). Arlington, VA: Information Resources Press.

Marchionini, G., & Fox, E. (1999). Progress toward digital libraries: Augmentation through integration. *Information Processing & Management*, *35*(3), 219–225.

Marchionini, G., Plaisant, C., & Komlodi, A. (2003). The people in digital libraries: Multifaceted approaches to assessing needs and impact. In Bishop, A., Buttenfield, B., & VanHouse, N. (Eds.), *Digital Library Use: Social Practice in Design and Evaluation* (pp. 119–160). Cambridge, MA: MIT Press.

Matuskiak, K. (2006). Information seeking behavior in digital image collections: A cognitive approach. *Journal of Academic Librarianship*, *32*(5), 479–488. doi:10.1016/j.acalib.2006.05.009

Mills, E. S. (2007). *Evidence explained: Citing history sources from artifacts to cyberspace*. Baltimore, MD: Genealogical Publishing Company.

Nielsen, J. (2000). *Why you only need to test with 5 users*. Retrieved August 12, 2011 from http://www.useit.com/alertbox/20000319.html

Norberg, L. R., Vassiliadis, K., Ferguson, J., & Smith, N. (2005). Sustainable design for multiple audiences: The usability study and iterative redesign of the documenting the American south digital library. *OCLC Systems & Services*, *21*(4), 285–299. doi:10.1108/10650750510629625

Salo, D. (2008). Innkeeper at the roach motel. *Library Trends*, *57*(2), 98–123. doi:10.1353/lib.0.0031

Skinner, J. (2010). Does greater specialization imply greater satisfaction? Amateur genealogists and resource use at the state historical society of Iowa libraries. *Libri: International Journal of Libraries & Information Services*, *60*(1), 27–37.

Spool, J. (2005). *Looking back on 16 years of paper prototyping*. Retrieved August 15, 2011 from http://www.uie.com/articles/looking_back_on_paper_prototyping/

Taylor, M. (2008). *Genealogy research process*. Retrieved August 15, 2011 from http://www.thinkgenealogy.com/wp-content/uploads/genealogy-research-map.pdf

Taylor, R. S. (1991). Information use environments. In Dervin, B., & Voight, M. J. (Eds.), *Progress in Communication Science* (*Vol. 10*). Norwood, NJ: Ablex Publishing.

University of Chicago. (2003). *The Chicago manual of style*. Chicago, IL: University of Chicago Press.

Uzwyshyn, R. (2008). Repurposing open source software for agile digital image library development: The University of West Florida libraries model. *D-Lib Magazine, 14*(9/10). Retrieved August 12, 2011 from http://www.dlib.org/dlib/september08/uzwyshyn/09uzwyshyn.html

Van House. N. A., Butler, M. H., Ogle, V., & Schiff, L. (1996). User-centered iterative design for digital libraries: The cypress experience. *D-Lib Magazine, 2*(2). Retrieved August 11, 2011 from http://www.dlib.org/dlib/february96/02vanhouse.html

Yakel, E. (2004). Seeking information, seeking connections, seeking meaning: Genealogists and family historians. *Information Research, 10*(1). Retrieved August 11, 2011 from http://informationr.net/ir/10-1/paper205.html

ENDNOTES

[1] The UNT Libraries metadata guidelines are available at http://www.library.unt.edu/digitalprojects/metadata/.