

ISBN 1-84544-333-0 ISSN 0737-8831

Volume 23 Number 2 2005

Library Hi Tech

Collaborative digitization programs

Theme Editor: Ken Middleton



www.emeraldinsight.com



Collaborative digitization programs

Theme Editor
Ken Middleton

Access this journal online _____	143
Editorial advisory board _____	144
INTRODUCTION	
Collaborative digitization programs: a multifaceted approach to sustainability <i>Ken Middleton</i> _____	145
THEME ARTICLES	
Development of a portal to Texas history <i>Cathy Nelson Hartman, Dreanna Belden, Nancy K. Reis, Daniel Gelaw Alemneh, Mark Phillips and Doug Dunlop</i> _____	151
Meeting metadata challenges in the consortial environment: metadata coordination for North Carolina Exploring Cultural Heritage Online <i>Katherine Wisser</i> _____	164
Ohio Memory Online Scrapbook: creating a statewide digital library <i>Laurie Gemmill and Angela O'Neal</i> _____	172

CONTENTS

Access this journal electronically

The current and past volumes of this journal are available at:

www.emeraldinsight.com/0737-8831.htm

You can also search more than 100 additional Emerald journals in Emerald Fulltext (www.emeraldinsight.com/ft) and Emerald Management Xtra (www.emeraldinsight.com/emx)

See page following contents for full details of what your access includes.



CONTENTS

continued

Educating future digitizers: the Illinois Digitization Institute's Basics and Beyond digitization training program <i>Amy Lynn Maroso</i> _____	187
Empowerment for digitization: lessons learned from The Making of Modern Michigan <i>Ruth Ann Jones</i> _____	205
Aggregating distributed digital collections in the Mountain West Digital Library with the CONTENTdm™ multi-site server <i>Kenning Arlitsch and Jeff Jonsson</i> _____	220
AlabamaMosaic: sharing Alabama history online <i>Sherida Downer, Sue Medina, Beth Nicol and Aaron Trehub</i> _____	233
The Columbia River Basin Ethnic History Archive: a tri-state online history database and learning center <i>Leslie Wykoff, Laurie Mercier, Trevor Bond and Alan Cornish</i> _____	252
OTHER ARTICLES	
Accessibility and usability of online library databases <i>Ron Stewart, Vivek Narendra and Axel Schmetzke</i> _____	265
The evolution of electronic reference sources <i>Amy S. Van Epps</i> _____	287

Access this journal online

www.emeraldinsight.com/lht.htm



As a subscriber to this journal, you can benefit from instant, electronic access to this title via Emerald Fulltext and Emerald Management Xtra. Your access includes a variety of features that increase the value of your journal subscription.

How to access this journal electronically

To benefit from electronic access to this journal you first need to register via the internet. Registration is simple and full instructions are available online at www.emeraldinsight.com/admin Once registration is completed, your institution will have instant access to all articles through the journal's Table of Contents page at www.emeraldinsight.com/0737-8831.htm More information about the journal is also available at www.emeraldinsight.com/lht.htm

Our liberal institution-wide licence allows everyone within your institution to access your journal electronically, making your subscription more cost-effective. Our web site has been designed to provide you with a comprehensive, simple system that needs only minimum administration. Access is available via IP authentication or username and password.

Key features of Emerald electronic journals

Automatic permission to make up to 25 copies of individual articles

This facility can be used for training purposes, course notes, seminars etc. This only applies to articles of which Emerald owns copyright. For further details visit www.emeraldinsight.com/copyright

Online publishing and archiving

As well as current volumes of the journal, you can also gain access to past volumes on the internet via Emerald Fulltext and Emerald Management Xtra. You can browse or search these databases for relevant articles.

Key readings

This feature provides abstracts of related articles chosen by the journal editor, selected to provide readers with current awareness of interesting articles from other publications in the field.

Reference linking

Direct links from the journal article references to abstracts of the most influential articles cited. Where possible, this link is to the full text of the article.

E-mail an article

Allows users to e-mail links to relevant and interesting articles to another computer for later use, reference or printing purposes.

Emerald structured abstracts

New for 2005, Emerald structured abstracts provide consistent, clear and informative summaries of the content of the articles, allowing faster evaluation of papers.

Additional complimentary services available

Your access includes a variety of features that add to the functionality and value of your journal subscription:

E-mail alert services

These services allow you to be kept up to date with the latest additions to the journal via e-mail, as soon as new material enters the database. Further information about the services available can be found at www.emeraldinsight.com/alerts

Research register

A web-based research forum that provides insider information on research activity world-wide located at www.emeraldinsight.com/researchregister You can also register your research activity here.

User services

Comprehensive librarian and user toolkits have been created to help you get the most from your journal subscription. For further information about what is available visit www.emeraldinsight.com/usagetoolkit

Choice of access

Electronic access to this journal is available via a number of channels. Our web site www.emeraldinsight.com is the recommended means of electronic access, as it provides fully searchable and value added access to the complete content of the journal. However, you can also access and search the article content of this journal through the following journal delivery services:

EBSCOHost Electronic Journals Service
ejournals.ebsco.com

Huber E-Journals
e-journals.hanshuber.com/english/index.htm

Informatics J-Gate
www.j-gate.informindia.co.in

Ingenta
www.ingenta.com

Minerva Electronic Online Services
www.minerva.at

OCLC FirstSearch
www.oclc.org/firstsearch

SilverLinker
www.ovid.com

SwetsWise
www.swetswise.com

TDnet
www.tdnet.com

Emerald Customer Support

For customer support and technical help contact:

E-mail support@emeraldinsight.com

Web www.emeraldinsight.com/customercharter

Tel +44 (0) 1274 785278

Fax +44 (0) 1274 785204

Morell D. Boone

Professor, Interdisciplinary Technology, Eastern Michigan University, Ypsilanti, MI, USA

Michael Buckland

University of California, Berkeley, CA, USA

May Chang

North Carolina State University, Raleigh, North Carolina, USA

Susan Cleyle

Associate University Librarian, QEII Library, Memorial University of Newfoundland, Canada

Timothy W. Cole

Mathematics Librarian and Associate Professor of Library Administration, University of Illinois at Urbana-Champaign, USA

Dr Colin Darch

Centre for Information Literacy, University of Cape Town, South Africa

Professor G.E. Gorman

School of Communications & Information Management, Victoria University of Wellington, New Zealand

Charles Hildreth

Associate Professor, Long Island University, Brookville, NY, USA

Larry A. Kroah

Director, Trenton Free Public Library, NJ, USA

Karen Markey

University of Michigan, Ann Arbor, MI, USA

Joe Matthews

EOS International, Carlsbad, CA, USA

Steve O'Connor

Chief Executive Officer, Caval Collaborative Solutions, Bundoora, Victoria, Australia

Ed Roberts

Head of Information Systems, University of Washington Health Sciences Libraries, USA

Ilene Rockman

Manager, Information Competence Initiative, The California State University, Hayward, CA, USA

Professor Jennifer Rowley

Lecturer, School for Business and Regional Development, University of Wales, Bangor, UK

James Rush

Consultant, PA, USA

Dr Hildegard Schaffler

Head of Serials and Electronic Media, Bavarian State Library, Munich, Germany

Axel Schmetzke

Librarian/Assistant Professor, University of Wisconsin-Stevens Point, WI, USA

Steven Sowards

Head Main Library Reference, Michigan State University, MI, USA

Dr Judith Wusteman

Department of Library and Information Studies, University College Dublin, Ireland

Sandra Yee

Dean of University Libraries, David Adamany Undergraduate Library, Wayne State University, Detroit, MI, USA



INTRODUCTION

Introduction

Collaborative digitization programs: a multifaceted approach to sustainability

145

Ken Middleton

Middle Tennessee State University, Murfreesboro, Tennessee, USA

Received 27 August 2004
Revised 6 January 2005
Accepted 6 February 2005

Abstract

Purpose – To provide an introduction to the eight articles in the issue that deal with collaborative digitization programs.

Design/methodology/approach – This introduction summarizes the articles within the context of sustainability.

Findings – In addition to business planning, efforts to build collaborative networks, empower content providers, and engage end-users are central to developing sustainable programs.

Originality/value – Readers gain a quick overview of some of the key issues that digitization programs face.

Keywords Digital libraries, Training, Empowerment, United States of America

Paper type General review

Introduction

Statewide and regional digitization programs in the USA offer cultural heritage institutions (archives, libraries, and museums) a viable option for digitizing their collections (e.g. photographs, diaries, oral histories, museum objects). These collaborative programs may provide training in digital imaging and metadata standards, access to scanning equipment, and software tools that streamline the creation of metadata records. Most programs also feature a central site for searching across the digital collections of participating institutions. Students, scholars and lifelong learners gain access to a rich source for exploring the history and culture of their state or region.

The Ohio Memory Online Scrapbook (www.ohiomemory.org) illustrates these benefits well. A simple search for “temperance” retrieves digital surrogates of a variety of sources (photographs, temperance society records, letters, pamphlets) from over 30 institutions, including county historical societies and small public libraries. The cost savings and potential educational value of these programs have not gone unnoticed. Counting programs still in planning, groups from at least 40 states are involved in

The author wishes to thank Michael Seadle for offering him the opportunity to serve as a guest editor. He also thanks Jeff Barry, Melvin Davis, and Mayo Taylor for their input and support. Working with his Volunteer Voices colleagues, a group with considerable talents and a collaborative spirit, has prepared him well for editing this issue.



statewide or regional digitization programs (Collaborative digitization projects in the USA www.mtsu.edu/~kmiddlet/stateportals.html).

Eight articles in this issue illustrate diverse approaches to addressing the technical, social, and financial challenges facing collaborative digitization programs. States and regions represented include Texas, North Carolina, Ohio, Illinois, Michigan, the Mountain West (Utah, Nevada), Alabama, and the Columbia River Basin (Idaho, Oregon, Washington State). Although these programs vary in terms of organizational structure, the number and types of participating institutions, and software/hardware options, they share a vision of creating a dynamic and sustainable digitization program. The realization of this vision requires that programs implement sound business planning practices, build strong collaborative networks, empower content providers (i.e. participating institutions), and engage end-users.

Business planning for sustainability

Bishoff and Allen (2004) argue that the business planning approach offers a long-term, strategic perspective that is essential for the development of sustainable digitization programs. The authors include a template that offers cultural heritage institutions a better understanding of the key components of business plans (e.g. the mission statement, market research, promotional strategies, financial planning, product evaluation). A key theme of this valuable guide is that program planners should base their decisions on sound business planning practices instead of assumptions and wishful thinking. For instance, they should conduct market research to determine what current and potential end-users expect from digital collections.

The financial component of business planning becomes increasingly important for digitization planners as they transform grant-based pilot projects into sustainable programs. A financial plan can be used to determine the true costs of digitization and to designate reliable sources of revenue to offset these costs. As part of their overview of the Portal to Texas History, Hartman *et al.* discuss their financial plan for developing multiple revenue streams and improving efficiency. This program will receive revenue by providing digital conversion activities, consulting services, a clearinghouse for licensing digital content, reproductions of historical items, and on-demand printing. In addition to automating part of the metadata creation process, they will improve efficiency by using a variety of software tools that track projects, evaluate customer satisfaction, and create collaborative workspaces.

Building collaborative networks

Some of the most challenging aspects of collaborative digitization programs are social rather than technical. Even at the institutional level, digitization programs often require that staff from different departments work with one another much more closely than they are accustomed. This challenge is compounded when a digitization program includes multiple institutions of various types, sizes, and locations. For example, Bishoff (2004) has detailed the collaborative challenges that librarians and museum professionals face, including the lack of a shared vocabulary and differences in cultures and funding structures. By respecting these differences and insuring that all participants benefit, digitization programs can develop collections that are more diverse, increase the efficiency of the digitization process, and create additional funding options.

Wisser describes just such an inclusive approach in her article about meeting metadata challenges of NC ECHO. She begins by summarizing the results of NC ECHO's face-to-face survey of over 500 institutions. This impressive effort has not only provided project staff with valuable knowledge about how museums and other institution types operate, but has also recognized personnel at these institutions as true partners in NC ECHO. These personnel can help shape NC ECHO by serving on any of its multiple metadata working groups (Dublin Core, Encoded Archival Description, museum metadata, Text Encoding Initiative). Participating institutions benefit from training and consultation services that address their specific metadata needs. This multi-pronged approach promotes the creation of consistent and quality metadata that is so central to access and digital preservation. Equally important to the sustainability of NC ECHO is the sense of community and empowerment among participants that this approach encourages.

Gemmill and O'Neal discuss the direct benefits that Ohio Memory receives from its strong collaborative relationships with six project partners. For instance, OhioLINK provides valuable technical assistance, and houses Ohio Memory's digital objects on its servers. Project staff at the Ohio Historical Society can concentrate on offering content providers a high level of personalized service. As a further incentive, contributing institutions are only required to submit one item. These efforts have achieved remarkable results. Ohio Memory has 330 contributing archives, libraries, and museums.

Empowerment through training and tools

Building a broad-based group of digitization professionals is crucial to sustaining a collaborative digitization program beyond its initial grant-funding period. This is a formidable task because digitization is totally new to so many participating institutions, training can be expensive, and the speed of technological change necessitates continuing education. However, collaborative programs can address these challenges by empowering content providers with flexible training opportunities, support and technical tools.

The Illinois Digitization Institute (IDI) provides an excellent model for addressing the rapidly expanding need for digitization training. Maroso describes how IDI's three levels of training are designed to meet the diverse needs of individuals involved with digitization projects. Inexpensive one-day workshops are ideal for individuals who are new to digitization, whereas IDI's online course provides more detail and offers interactive tools. A third option combines the online modules with hands-on exercises. In addition to arguing for increased training opportunities in the digitization of audio and video, Maroso urges trainers to place a stronger emphasis on the costs and importance of digital longevity. No one should doubt the importance of training to the sustainability of digitization programs after reading this article.

Jones notes that Making of Modern Michigan (MMM) is empowering content providers by combining training with support and appropriate technical tools. Six digitization centers provide scanning equipment, technical advice, and informal training. Project staff developed a metadata entry system that allows participants to enter Dublin Core records. One time-saving feature is that records can be flagged and used as templates for other records.

MMM is also providing a model for making copyright compliance much less of an obstacle for institutions to participate in collaborative digitization programs.

A web-based copyright training video was an instant success, and a copyright clearance service builds on an existing service at Michigan State University. In addition, copyright screening is an integral part of the project's inventory system. By breaking the complexity of copyright law down to a series of yes/no questions, content providers can quickly determine with some degree of confidence whether materials are in the public domain.

Unlike as few as five years ago, a variety of digital content management systems can now facilitate collaborative digitization efforts. Authors in this issue note the particular strengths and limitations of CONTENTdm, Documentum, Encompass, and Index Data. Other programs have selected Digital Library eXtension Service (DLXS) and SiteSearch. Indeed, no single package has cornered the market, and this healthy competition should continue to promote the development of efficient and user-friendly software for content providers and project hosts, as well as elegant interfaces and innovative tools for end-users.

Arlitsch and Jonsson describe the use of the CONTENTdm Multi-Site Server (MSS) as a cost-effective tool for developing the Mountain West Digital Library (MWDL). In this model, six regional digitization centers scan materials from partner institutions to insure digitization uniformity. Partner institutions create and upload metadata to a center's CONTENTdm server. Records from each digitization center server are harvested by the MSS at the University of Utah to create the publicly accessible database. In addition to retaining control of metadata, partner institutions have the option of developing a customized portal to their digital collections. Having already created a quarter of a million digital objects in multiple formats, the MWDL model is highly efficient. In addition, the authors note the success of applying this model to the Western Waters project. Nevertheless, they also note some limitations of the MSS, such as its inability to automatically discover new collections.

Similar to the MWDL's use of CONTENTdm, AlabamaMosaic benefits from multiple installations of Encompass for Digital Access software in the state. Downer *et al.* note that AlabamaMosaic's three digitization centers fulfill most of its digitization needs because each center has different technical capabilities. However, a repository must complete a digital collection plan before personnel can utilize the digitization centers or attend digitization workshops. This requirement contributes to the sustainability of AlabamaMosaic by informing project staff of training and resource needs, empowering institutions to digitize materials themselves, and sometimes forming the basis for grant proposals in support of local digitization projects.

Some institutions may be too remote to take advantage of digitization centers, and others may hesitate to transport their delicate and valuable collections. The Portal to Texas History offers an alternative. After receiving training, staff at participating institutions can borrow a portable digital imaging unit. Wykoff *et al.* offer a second alternative in their overview of The Columbia River Basin Ethnic History Archive. This project utilized a mobile scanning team that visited archives and museums, scanned images, and entered metadata. Although wholesale adoption of this approach is often not feasible, programs that want to encourage participation from small institutions may want to consider offering some flexibility in where items are digitized.

Engaging end-users

The collaborative approach is also essential for developing digital collections that are engaging and relevant to end-users. Beginning with the appointment of historian/educator Laurie Mercier as co-project director, the Columbia River Basin Ethnic History Archive was designed from the ground up with education applications in mind. Rather than having a “lesson plans” link that is buried among other options, they have nicely integrated background materials, research tutorials, and lesson plans into the site. In addition, the public interface is designed to guide the user into the content of the database. Their plans for an online discussion forum for teachers and students, although not fully realized, will add another level of interactivity.

In collaboration with educators, The Portal to Texas History (PHT) is also planning a host of portal services for that will assist teachers in integrating PHT collections into the curriculum. For instance, restricted access sites will provide teachers with tools to design resource sites for specific classes, as well as online forums for students and teachers to discuss assignments. AlabamaMosaic uses Alabama Moments, an existing framework for tying primary sources into the K-12 curriculum, both as a guide for selecting collections to digitize and as a source for subject headings. These examples will hopefully prompt more collaborative programs to give historians and educators an even greater role in planning and development. Participants in Teaching American History, a project that helps teachers integrate primary sources into the K-12 curriculum, are well-suited for this role (www.ed.gov/programs/teachinghistory/index.html).

Statewide digitization programs can make further inroads into the curriculum by collaborating with other projects to create a one-stop site for state history research. AlabamaMosaic is taking one step in this direction by collaborating with another digitization project, *The Encyclopedia of Alabama*. Such a resource could provide contextual information for digital surrogates of primary sources, a feature that is often lacking in digitization projects. An online archive of the back issues of a state’s historical society journal could serve a similar purpose. A federated search of quality primary, secondary, and tertiary sources would combine Google-like simplicity with highly relevant results.

New sources for digitization can also increase a program’s appeal to different audiences. Newspapers in particular have a wide variety of uses for students, historians, and genealogists. Arlitsch and Jonsson highlight the joint efforts of The University of Utah and Brigham Young University to create the first distributed digital newspaper collection in the USA. The National Newspaper Digitization Program, with pilot grants awarded in 2005, will give more programs the opportunity to explore the challenges and rewards of this unique information source (www.neh.gov/grants/guidelines/ndnp.html). Although not covered in this issue, “Ephemeral Cities”, a project of Florida’s Publication of Archival, Library & Museum Materials (PALMM) program, is linking specific locations of buildings on Sanborn Fire Insurance Maps with other relevant digital objects (e.g. photographs, documents, museum objects) (www.uflib.ufl.edu/digital/collections/EphemeralCities/). Students gain an engaging resource for exploring the relationships between place and time.

Individuals who are intimately involved in a digitization program may think that its value is quite obvious. However, there is a danger of thinking, “if we build it, they will come.” Similar to its approach with content providers, Ohio Memory has taken

proactive steps to increase use of the project web site by students and the general public. For instance, they have promoted the use of personalized, annotated online scrapbooks by having well known Ohioans like Jack Hanna and Ohio's First Lady share their scrapbooks on the project web site. The Governor's Challenge is a quiz that links to items in Ohio Memory, and simple games offer opportunities for very young students to interact with the site. A variety of small images on the project home page quickly draws users of all ages into the database without entering a search. By creating such a broad-based group of end-users, project staff can make a convincing case for the continued development of Ohio Memory.

Conclusion

Taken as a whole, these articles provide a multifaceted approach to sustainability. By including archivists, educators, librarians, and museum curators as truly collaborative partners, a digitization program gains expertise, resources, and expanded funding options. By empowering content providers with appropriate tools, training, and support, digitization can become an integral part of the routine work of archives, libraries and museums. By engaging users with powerful, interactive tools, a database of digital objects can be transformed into a must-have learning tool for students of all ages. When combined with efforts to recover costs and improve efficiency, this approach should make digitization programs better candidates for both ongoing and grant-based funding.

References

- Bishoff, L. (2004), "The collaborative imperative: if librarians want to lead in creating the digital future, they need to learn how to work with their colleagues in museums and archives", *Library Journal*, Vol. 129 No. 1, pp. 34-5.
- Bishoff, L. and Allen, N. (2004), *Business Planning for Cultural Heritage Institutions*, Council on Library and Information Resources, Washington, DC, available at: www.clir.org/pubs/reports/pub124/pub124.pdf



THEME ARTICLE

Development of a portal to Texas history

Development of a
portal to Texas
history

Cathy Nelson Hartman, Dreanna Belden, Nancy K. Reis,
Daniel Gelaw Alemneh and Mark Phillips
University of North Texas Libraries, Denton, Texas, USA, and
Doug Dunlop
University of Central Florida Libraries, Orlando, Florida, USA

151

Received 27 August 2004
Revised 29 November 2004
Accepted 17 February 2005

Abstract

Purpose – To help information professionals learn about issues and considerations in portal building.

Design/methodology/approach – The University of North Texas Libraries' Portal to Texas History provides long-term storage and access to digital copies of important original materials illuminating Texas's past. This paper describes the development of the Portal technology and content – presenting objectives, processes, and future plans – and defines the larger goal of facilitating collaboration among resource-holding institutions.

Findings – Practical aspects of creating and populating the portal include development of specifications and standards, construction of an application framework, selection of content, production of metadata, and refinement of user interfaces. Planned future enhancements to the Portal will augment sustainability and provide added value for users. The portal project may also serve as a catalyst for wider collaborative efforts in digitization.

Originality/value – The Portal to Texas History project's experiences described in this paper will inform other stakeholders seeking to develop innovative uses of Portal technologies.

Keywords Portals, Digital storage, Content management, Digital libraries, History, United States of America

Paper type Case study

Introduction

Texas history can thrill us with its bigger-than-life heroes and heroines, champions, charlatans and rogues. Who can forget the bravery of the defenders of the Alamo or the shenanigans of Jim Hogg? Filled with life, Texas history holds the rich potential to enchant school children and captivate learners of all ages.

Imagine – a school child in Pampa, Texas, reads William Barret Travis' last-ditch pleas from the Alamo, breathing new life into the cry, "Victory or death!" A historian in Marshall, Texas, carefully peruses Anson Jones' memoranda and official correspondence as the last president of the Republic of Texas in preparation for publishing a scholarly article. A school teacher in Dallas prepares material for her students covering pioneering African-American Texas Senators, George T. Ruby, Matt Gaines, and Walter Burton. At the University of North Texas Libraries' (UNT) Portal to Texas History, the dream of providing unique online historical content is becoming reality.



Beginnings of the portal

In 2002, the Portal to Texas History existed only as an idea in the minds-eye of a forward-thinking University of North Texas (UNT) librarian who envisioned the portal as an entry point on the web for access to digital copies of important original materials that could illuminate Texas's past. Digital representations of documents, letters, journals, maps, photographs, artifacts, and other materials could be presented to a much larger audience than the original objects, while helping to preserve the materials through reduced handling. In addition, the UNT Libraries hoped to form partnerships with schools, other libraries, museums, publishers, and individuals to provide access to these materials through the portal. UNT could also offer optional services to partners such as hosting their electronic collections on the portal, providing mentoring services for persons beginning first-time digital projects, and coordinating funding for the collaborative project.

A steering committee consisting of librarians, programmers, teachers, subject specialists, and designers gathered information, examined existing software, and decided how the portal could be constructed to serve its constituents. UNT planned to create technology systems and standards for the project with emphasis on the necessity of preserving the site for long-term usage. After preparing technical specifications (http://texashistory.unt.edu/tech_specs.html), UNT secured grant funding, and then selected Index Data Systems of Denmark to implement an open-source software package to structure the portal.

Now in its beta version, the portal is available on the Worldwide web (<http://texashistory.unt.edu/>). As the portal staff members continue to refine its functionality, they are entering text, images, and descriptive metadata. Development of new content with our partners continues to bring exciting new material for users of the portal.

Goals of the portal

The portal project is being developed in two phases. Phase one, which is nearly complete, encompassed four objectives: building a technical structure for the portal services, designing a search interface that embraced both beginning and skilled researchers, creating standards and best practices guidelines – including the design of training modules to educate library and museum staff who lack knowledge of the digitization process, and implementing an evaluation plan.

Objective one. Creating the technical structure and the project services for the Portal to Texas History involved two elements. First, a two-server configuration with expanded system storage and with a sophisticated application framework was planned for the technical structure of the site. The application framework supplies file management capabilities, allows search and retrieval of files, enables harvesting of files from collaborating sites for indexing and preservation purposes, allows Z39.50 interoperability with other servers including the Library of Texas Project (<http://libraryoftexas.org/>), and creates backup copies of all files on a regular basis. Administrative functions allow the project administrator to enable pass-worded access for collaborators at distant sites for purposes of data entry. Second, a portable digital imaging unit, comprising a high-end laptop, a scanner, and a digital camera setup, was purchased to assist with on-site training and scanning activities in libraries and museums. After training, small libraries and museums without access to digital imaging hardware and software will be able to borrow the portable equipment.

Objective two. Project staff created standards and best practices guidelines for the portal project and services, and designed and tested a training module that assists with imparting knowledge contained in the project standards and guidelines to persons in museums and libraries. Project standards include element sets for both descriptive and preservation metadata. The “descriptive” metadata delineate the historical items that are digitized and allow searchers to discover and view the materials contained in the digital collections. The “preservation” metadata maintain reference, context, provenance and fixity information about the files, and technical information about file creation, size, and location, software used to create the file, and preservation activities, which will aid the long-term maintenance and viability of digital files. A standardized element set simplifies data entry and ensures that important information is recorded. Best practices guidelines recommend appropriate processes for digital imaging of various types of materials and artifacts. By following best practices guidelines to preserve long-term access to the electronic files, project participants will help ensure that delicate materials will not require repeated handling and re-scanning in the future.

Objective three. The portal web site embraces both beginning and skilled researchers – guiding them through the process of searching for and retrieving relevant information. In addition, Texas historians will contribute original, supplementary materials to offer guidance for diverse users and to highlight subjects of particular interest.

Objective four. The Portal to Texas History contracted with Dr William E. Moen of the Texas Center for Digital Knowledge to create and implement an outcomes-based evaluation plan for the first phase of the project and, additionally, to write a similar plan for evaluation of the second phase of the project. The evaluation plan examines three objectives: creating the technical structure of the portal; creating standards and best practices; and developing the portal web site to guide researchers. The plan focuses on the milestones required to meet objectives: key activities, output (results), outcomes (possible benefits), indicators, and the data sources or methods used to gather evaluative data. The plan also includes questionnaires for collaborators, digital imaging students, young scholars, and researchers, which assess and measure the effectiveness of the metadata, best practices and standards for digitization, training for the application framework, and usability of the web site. Questionnaires will be implemented at appropriate benchmarks in the project.

Phase two objectives continue in development. The four goals of phase two include construction of an online *Thesaurus for Texas History*, design of a training program for the application framework to be used in a test-bed project with member institutions, establishment of a business plan for cost recovery initiatives, and development of additional content and curriculum guides that fulfill the needs of educators teaching Texas history to schoolchildren. These phase two objectives are discussed in more detail later in the article.

Portal processes

Selection criteria

Considering the burgeoning volume and heterogeneity of information on the web, selection and appraisal of resources for digitization is one of the most difficult tasks in the digital resources management life cycle. Under current criteria, materials selected

for the portal should focus on aspects of Texas history of interest to historians, students and lifelong learners – emphasizing materials that are of key importance to Texas history, as well as meagerly documented topics that could benefit from increased access. Selected items should also provide geographic coverage throughout the State of Texas, hold unique or intrinsic value as historical objects, and not duplicate previously digitized material. Copyright is also an issue, and materials can be used only if copyright has been cleared; i.e. the copyright has expired, or the contributing institution has obtained copyright release or owns the copyright.

Metadata framework

Since its first conceptual description in the *First Monday* peer-reviewed e-journal almost two years ago (Alemneh *et al.*, 2002), the University of North Texas Libraries' (UNT) metadata system has continued to evolve. Considering the very dynamic nature of digital resources, the needs and requirements of resource management systems must change over time. Accordingly, the project team adapts the UNTL metadata system for various uses and applications and continues to add important process functionalities such as a graphical user interface for system administrators, content developers, metadata creators, and end-users. Moreover, as the team gains practical experience in describing objects with metadata, it will provide further definition of the elements, and produce local practice guidelines.

The UNTL metadata element set comprises Dublin Core-based descriptive metadata (<http://dublincore.org>) and detailed technical and preservation information recording how digital resources (text, image, audio, video) are identified, created, formatted, arranged in relevant software applications, and sustained with application of appropriate preservation procedures. User guidelines provide detailed information and assistance to metadata creators. The most current and comprehensive documentation about the UNTL metadata is available on the Worldwide web (http://texashistory.unt.edu/project_documents.htm).

The UNT Libraries are now implementing the metadata system. Among other activities to facilitate implementation, the project team has developed metadata crosswalks. Mapping facilitates large-scale collaboration and fosters interoperability among various metadata schemas. The crosswalks will also enable UNT staff to import metadata records directly from participating institutions in a variety of formats, including harvesting metadata from existing MARC (Machine-Readable Cataloging) records.

Early implementation activities revealed some issues. Balancing the specificity of the information that can be represented in and queried from metadata with the cost of producing records continues to prove challenging. In this regard, UNT is automating several aspects of the metadata creation process, while monitoring emerging technologies for harvesting technical metadata, as well as related initiatives such as the Research Libraries Group (RLG) led project of Automatic Exposure (www.rlg.org/en/page.php?Page_ID=2681). Also at issue is quality control for descriptive metadata to ensure that we satisfy the needs of users, for example, authority control for individual and corporate names and place names. A well-developed thesaurus for the subject area could assist such quality control efforts.

The design of the portal architecture supports the automatic metadata extraction process. Technical information can be captured in the scanning process and harvested

automatically when digital objects are transferred into the archive. For instance, scanning specifications are recorded in TIFF (Tagged Image File Format) headers during the scanning process, and that data is then automatically harvested and recorded in the appropriate metadata fields.

Continuing metadata development and application challenges will be met through monitoring and participating in national and international discussions and standards development. UNT is an active member of various statewide, national and international professional associations and consortia related to digital library development. Portal team members participate in standards building activities such as the PREMIS (Preservation Metadata: Implementation Strategies) group (www.oclc.org/research/projects/pmwg/) sponsored by RLG and OCLC (Online Computer Library Center). Also, working in collaboration with the Texas Center for Digital Knowledge (www.txcdk.org/), the UNT School of Library and Information Sciences (www.unt.edu/slis/), and many other partners in Texas libraries, museums, and archives, the UNT Libraries are experimenting with various strategies to meet the challenges of digital resource management. The UNT Libraries currently provide assistance and guidance for small libraries and museums by identifying standards and “best practices” for digital collection management and by coordinating projects and services that participant institutions need, but cannot develop individually.

Portal architecture

The following sections describe the architecture of the portal from data input through archiving to user access. Figure 1 shows, in a graphical format, the steps involved.

Application framework

The portal is based on open-sourced initiatives such as XML (Extensible Markup Language), PHP (PHP Hypertext Preprocessor), and XSLT (Extensible Style sheet Language Transformations), as well as various libraries for the conversion of images. The foundation of the portal is the metadata record itself. Each metadata record consists of an individual XML file. This XML file uses a schema developed by the University of North Texas Libraries (UNT) Digital Projects Department, based on the UNTL metadata element set (<http://texashistory.unt.edu/guides/help/UNTL-Metadata-Guide.htm>). The schema also introduces some of the program-specific data into the record so that the system is self-contained in these records. After a call has been made from the system to access the information held in the XML record, XSLT is used to transform the metadata record into the desired output format.

Hardware

The portal exists on a Debian distribution of Linux running on two high-end Compaq servers with one acting as a hot backup. Consistent with the project’s commitment to open-source software, the portal runs Apache as the web server. Apache has been adjusted to allow for better support of XML and XSLT for the display of data from the portal. The archival system is currently running on an XP box, but has the ability to run on any platform due to the fact that it has been written in Java. The archival system will allow for the management of the rich digital master files on current or future archival media.

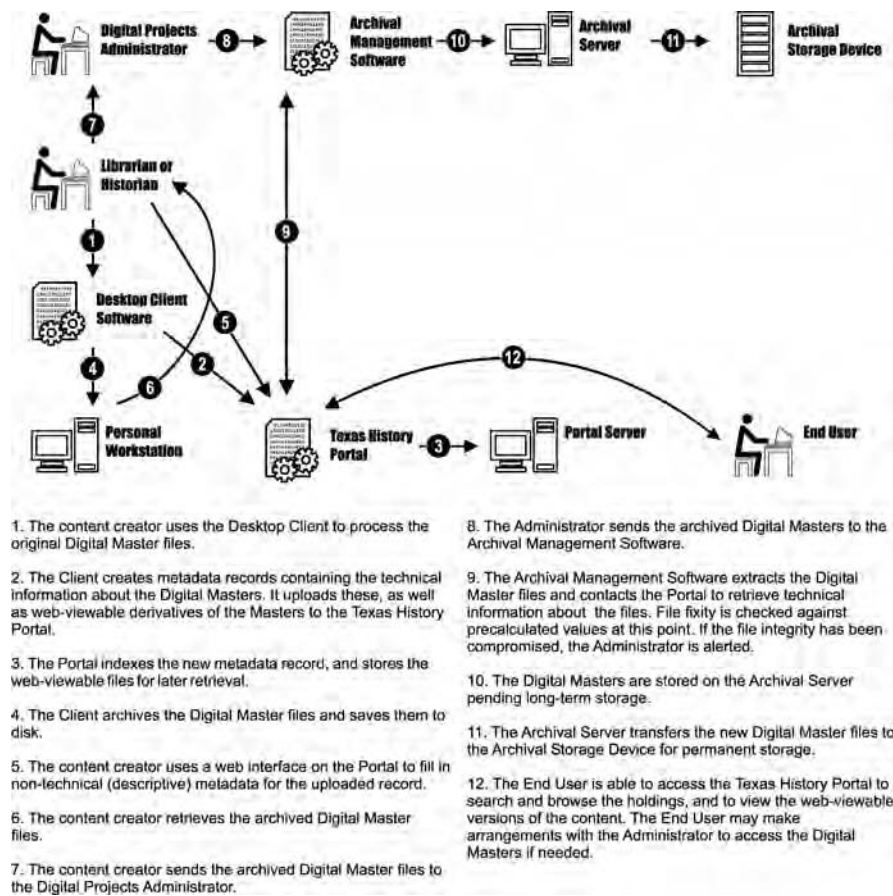


Figure 1.
Texas history portal
architecture

Metadata entry mechanisms

Metadata entry on the portal is facilitated by various administration screens that are created by Xforms, a program designed by Index Data to read an XML schema and output an HTML (Hypertext Markup Language) form based on that schema. The metadata entry form contains sections for both descriptive and preservation metadata, and the contents may be edited as needed.

The portal project staff also created in-house a portal client interface program. Developed with the Imagemagick libraries and written in C, the client allows content creators throughout the state to add information to the portal in a unified way. The client is used in conjunction with the digital master files so that there is a consistency in the collection of technical metadata and 50 information and in the creation of web-viewable derivative images. The client automates these tasks, which are labor intensive for the content creators. The client interface obtains a unique identification number from the portal and creates submission packages of the web-viewable files and metadata records, as well as archival file submission packages. Both the metadata and web-presentable files are sent to the portal through HTTP and the archival files are

sent either by FTP (File Transfer Protocol) or by mail on DVD (Digital Versatile Disc) or Firewire drives.

Digital archive interface

The digital archive system functions as an application separate from the portal, but it uses information contained within the portal to complete its tasks. The archival system takes archival packages created by the client tool and unwraps them for ingestion into the archival system. The files are unpackaged, and then their integrity is verified by comparing them against the fixity information stored in the portal metadata records. Confirmed files are then prepared for ingestion into the storage area of the archive. The files are renamed to ensure uniqueness across the system. The archival software is designed so that it functions independently of the storage mechanism, allowing the system to easily adjust as changes in technology occur. At regular intervals, the archival system checks the fixity of the files on the storage devices and alerts the portal administrators to any problems that might arise.

Public interface

Users access the portal through a web-based interface that allows both browsing and searching. Queries may be phrased using natural language or Boolean operators. Queries initiated from the Young Scholar's page search only the descriptive metadata, while queries from the Researcher's page search both full-text, when available, and metadata. Additional filters control classroom and teacher access to parts of the web site that provide curriculum guides and enable online class discussions.

Benefits to constituents

The Portal to Texas History offers a variety of benefits to its core constituents: students, researchers, educators and teachers, and collaborative partners. Informal surveys conducted by the Portal staff indicate that students of Texas history could benefit from the inclusion of primary source materials in their curriculum. Access to interesting, unique and relevant materials carry learning beyond the strictures inherent in textbooks. Researchers will also benefit from the materials in the portal. Randolph B. Campbell, Regents Professor of History at UNT, and a noted authority on Texas history says:

For more than 30 years, I have researched and written in the field of Texas history, an endeavor that has required more trips to libraries and depositories around the state than I care to count. I also have taught seminars in Texas history for many years, and in every case I had to recognize that my students could not gain access to many important materials unless they took long (this is Texas, after all) and expensive trips, which I could not ask of them. The Portal to Texas History project will bring essential material to the fingertips of scholars and students in a readily usable format (Campbell, 2003).

Plans to include curriculum materials for students will benefit educators and assist them in their efforts in bringing an added-value educational experience to their pupils. Portal partners will also benefit in many ways: training opportunities and guidance on best practices, grant-seeking, and cost-recovery opportunities.

Plans for the future

As the Portal to Texas History increases its content, opportunities for value-added services abound. Some services will enhance the educational value of the available content and other services could assist with site sustainability. Additionally, recognizing the benefits of continued collaboration, the portal project manager is also leading a state-wide effort in conjunction with the Texas State Library and Archives Commission (www.tsl.state.tx.us/) to plan for interoperability and collaboration among already existing digitization efforts for Texan cultural heritage materials. The sections below address future plans related to the development of a business plan to assist with sustainability, creation of curriculum guides to enhance the educational value of the portal content, and continued formation of collaborative relationships.

Business plan for sustainability

Across the nation, digital projects find themselves increasingly challenged by the difficulties of sustainability. It is well established that grant funding and institutional budgets alone are insufficient to provide for the full costs of digitization. In turn, institutions look at models from the for-profit sector in implementing cost-recovery revenue streams to subsidize digitization efforts. The Council of Library and Information Resources publication, *Business Planning for Cultural Heritage Institutions* (Bishoff and Allen, 2004) points out the fallacies often seen in business planning in the non-profit sector, and emphasizes the importance of sound business practices, market studies, and strategic planning. The balance between the traditional mandates for institutions as educators and providers of cultural heritage materials and the goal of cost-recovery will be influenced by an organization's internal and external environment.

In conceptualizing the development of a business plan, the portal partners must consider the full range of resources and staffing that will be required to accomplish revenue producing goals – recognizing that a lack of sound planning can easily lead to failure. A June 2002 paper from the Higher Education Digitisation Service (HEDS) based at the University of Hertfordshire, *Exploring Charging Models for Digital Cultural Heritage* stated that, “on the whole, there is no real business planning and modeling in most institutions, and less than 30 percent of those interviewed had attempted to calculate or quantify the exact costs for creation, storage, or service provision of digital items” (Tanner and Deegan, 2002, p. 16). “There is however a marked lack of clear commercially-led business planning and control in most institutions surveyed . . . charging policies are almost always based on the perceived market value (as defined by what similar organizations are charging) rather than the actual cost of provision” (Tanner and Deegan, 2002, p. 19). The report *Building and Sustaining Digital Collections: Models for Libraries and Museums* (Council on Library Information and Resources, 2001) also emphasized the need to incorporate existing business models without “reinventing the wheel”.

The Portal to Texas History partners are developing a business plan that includes an innovative plural cost-recovery model for multiple revenue streams. The portal operates on a non-profit basis, with free public access to the contents, but cost-recovery projects that can fund equipment upgrades and offset other costs of digitization efforts remain vital. A formal business plan for the project will include possible revenue streams from:

-
- *Digital conversion projects.* In the 2003-2004 academic year, UNT began accepting work for digital conversion projects for outside institutions. Work included scanning, performing optical character recognition, proofreading, and XML mark-up of text. Important statistical and management data collected from these projects will be used to refine process and pricing models and to streamline workflow and management issues. Funds from these projects assisted with hardware and software purchases for UNT's digital laboratory.
 - *Clearinghouse for licensing of digital content.* The partnership plans to serve as a clearinghouse for licensing the content of members interested in participating in this aspect of the plural cost-recovery model. Currently, commercial users researching images for a publication or exhibit must typically contact each institution and work through a confusing maze of differing reproduction licensing agreements, fees and paperwork in responding to the request. Pricing mechanisms would consider copyright, intended use, and service cost.
 - *Reproductions of historical items.* Institutions have traditionally offered copies and reproductions from their collections, but the partnership plans to expand this service with an e-commerce solution. Users will be able to purchase either a high-quality digital image or hardcopy directly from the Portal to Texas History web site. Partners may select services that are in accordance with their institutional culture.
 - *Reprints of public domain books.* The partnership between Historic Pittsburgh and North Haven Press represents a successful model for offering reprints of public domain books. The business plan will explore creating reprints of public domain books from the Portal collection, working either with the University of North Texas Press or another small printing house.
 - *Consulting.* Several universities have approached the University of North Texas Libraries about providing consulting for digital projects unrelated to the Portal to Texas History. Potential clients have an interest in expert guidance, as well as turnkey solutions for digital databases. During the business planning process, this service will receive further study.
 - *On-demand printing.* UNT currently engages in printing posters, maps, presentation materials, and other items for both internal and external customers. Analysis of data collected during the past year for costs of ink, paper, and staffing will enable optimal pricing strategies.

The portal partners plan to incorporate the e-commerce capabilities of the University's Enterprise Information System to provide e-commerce and accounting functions. Other software will be reviewed and selected to provide additional functionality, to include: project tracking, customer relationship management (CRM), enterprise resource planning (ERP), and collaborative groupware tools. Project tracking functions can generate hierarchic project trees and Gantt charts, and monitor pending tasks. E-commerce solutions provide bundle order and invoice management with payment processing. CRM provides a tool and business strategy that focuses on customer satisfaction, comprehensively tracking customer behaviors to better understand, anticipate, and appreciate their needs. ERP packages offer modules covering marketing and sales functions, production, inventory control, procurement,

distribution, finance and accounting. Further plans include the development of a collaborative online workspace for all participants to use in developing membership criteria and business plan objectives. This collaborative workspace will comprise a groupware/virtual environment that allows real-time editing of content.

Should grant funding become available, the portal partners plan to consult with experts on evolving business models and e-commerce solutions to seek their assistance with the development of the portal's business plan. The partners recognize that a subject matter expert intimately involved with cutting-edge business methodologies should bring new and innovative perspectives to the creation of a business plan for sustainability.

Curriculum materials for students

As is true of many states, Texas requires its seventh-grade students to study state history, and many college students also study Texas history. Because history and culture studies textbooks would be enhanced by supplemental primary source materials, thousands of students and their teachers could benefit from the portal services and content. Planned portal services include:

- teacher-only access to model curricula, created by teachers, for integrating portal services and resources into classroom activities and assignments;
- restricted access to specially created class resource sites where teachers, either individually or collaboratively, may place links to assigned materials and other information for the class(es), and students and teachers may then post discussion about the assignment; and
- support for innovative opportunities, such as interactions across school district boundaries for teachers and students that enhance student experiences and learning outcomes in history or culture studies.

An informal survey of educators revealed the need for quick and easy access to primary source materials about well-known events in Texas history, such as the Alamo or the annexation of Texas to the USA. Teachers and school librarians also cited a need for information on the local histories of their areas and to support studies about the Native Americans of Texas, explorers, and other famous historical leaders, while reflecting the broad cultural diversity of all Texans. To support these educational needs, content selections will focus on the desired subject areas, and training sessions will be created to instruct teachers and school librarians on the use of the supporting materials.

With funding from the Texas Telecommunications Infrastructure Fund grants and Microsoft Foundation grants, a strong majority of Texas schools and libraries now possess the technology to access online materials. US Secretary of Education Rod Paige notes that educators have been slow to truly embrace and integrate technology into learning curriculums. He says, "The way we organize schools and provide instruction is essentially the same as it was when our Founding Fathers went to school. Put another way, we still educate our students based on an agricultural timetable, in an industrial setting, yet tell students they live in a digital age" (Paige, 2002). The portal partners believe that the timing is right, and that the need exists for building partnerships among schools, libraries, and museums to make the primary-source

content available on the web and to prepare teachers to integrate the portal services and resources into their classrooms in ways that will benefit student learning.

Continued formation of collaborative relationships

Currently, the Portal to Texas History project's institutional partners include museums, public libraries, and private collections. Recognizing that Texas encompasses a large geographic area with a diverse and rapidly growing population, the project partners believe that many additional institutions could benefit from a wider collaboration. Collaboration within the state of Texas could bring benefits in areas such as:

- *Increasing interoperability among the several existing digital initiatives.* Other state institutions are now sponsoring digitization efforts for Texan heritage materials, and the Texas State Library and Archive Commission's Library of Texas project could serve as a central point of access for a federated search of all such sites to greatly enhance access for users. Current technology can easily facilitate such a federated search among collaborating partners.
- *Creating an online thesaurus for Texas history.* Building a robust thesaurus, and a subsequent controlled vocabulary list, will not only aid in the organization and access of information within the portal, but also provide a valuable tool for students and researchers of Texas history of all ages. A thesaurus will aid in the classification and indexing of information objects for the creation of descriptive metadata records for the portal database. Terms will be generated to represent the broad aspects of Texas history, and as a means of providing subject headings and descriptive terminology. Possible sources for terms could include primary historical sources, existing reference works on Texas history, and the primary information objects contained within the Texas History Portal. Subject experts could provide the structural framework to place the generated terms in, along with choosing contemporary terms that relate to more antiquated terms from earlier literature. Subject experts will also help to balance the needs between the various user groups of the portal, by representing a broad cultural user base, which ranges from school-aged children, to college students, to history buffs and aficionados. A truly collaborative effort will provide a wealth of terms reflecting the depth of Texas history and heritage while enriching the educational life of all users.
- *Sharing resources.* Software, hardware, storage, bandwidth, and grant writers are some examples of resources that could be shared to benefit all members of a wider collaborative organization. For example, expensive scanners for specialized tasks might be purchased by the collaborative for use by all participating institutions, or grant writers in larger institutions might serve as resource persons and readers of draft proposals developed by small libraries and museums that do not have full-time development staff.
- *Offering training opportunities.* Collaborative planning for training to benefit persons in institutions across the state is an effective use of resources and ensures that participants stay abreast of evolving best practices and standards. Training ranging from basic scanning skills to advanced issues related to digital archiving could be provided. For example, tutorials created by one institution to

teach best practices in basic scanning could be used by multiple institutions. Similarly, one workshop related to the Metadata Encoding and Transmission Standard (METS) (www.loc.gov/standards/mets/) could accommodate interested persons from several institutions.

Recognizing the potential for further collaboration, the portal project manager initiated an effort, in partnership with the Texas State Library and Archives Commission, to bring together individuals from institutions across Texas for an initial meeting in May 2004, to talk about wider collaboration (www.library.unt.edu/digitalprojects/texdig/thdi.htm). More than 80 individuals attended this meeting representing public libraries, academic libraries, museums, archives, state agencies, state historical associations and commissions, publishers, and others. Speakers included Marsha Semmel, Director of Strategic Partnerships, Institute of Museum and Library Services, and Liz Bishoff, Vice President for Digital Collection and Preservation Services for OCLC and former director of the Colorado Digitization Program. Their collective wisdom – acquired from years of experience with collaborative projects – provided insight to participants pursuing further collaborations.

In her presentation, Semmel referred to a lecture, “No longer under our control: the nature and role of standards in the 21st Century Library”, in which William E. Moen refers to standards building as “first and foremost a social process” (Moen, 2003, p. 5). Moen listed activities involved in the process of developing and creating a standard as: “Brainstorming, clarifying requirements, collaborative authoring, competitive collaboration, compromising, consensus building, developing proposals, editing, environmental scanning, information sharing/transfer, modeling, negotiating, planning, problem defining, problem resolving, problem solving, setting requirements, socializing, strategizing, visioning, and writing” (Moen, 2003, p. 5). Clearly, collaboration for digitization of cultural heritage materials is also driven by social processes and requires similar activities. With skilled programmers, technology can accommodate almost any collaborative architecture. The pitfalls lie in the social processes to reach agreement on collaborative architectures and then to sustain the agreements through development, implementation, and operation. To do this, benefits to participants must always outweigh the pitfalls.

Conclusion

Overall, the Portal to Texas History initiative strives to build on open source software and existing standards. All images and file structures are designed to support portability and future migration, independent of hardware and software. This flexible design should enable implementation of developing standards such as METS – a schema for encoding descriptive, administrative, and structural metadata for digital objects – and also allow compliance with OAI (Open Archives Initiative) harvestable metadata protocols and Z39.50 protocols for interoperability related to search and retrieval of information.

Brenda Gunn, head of archives and manuscripts at the Center for American History at the University of Texas at Austin, recently noted, “Most historical documents never receive one moment of attention outside of an archives reading room” (Gunn, 2004, p. 15). Materials stay in obscurity because potential users may not realize they exist or may not have the resources for physically visiting the holding institution. At the Portal to Texas History, we hope to transform this reality, enabling users to uncover treasures

of history previously known only to an elite few. Digital libraries enable new styles of learning and scholarship, offering life-long learners access to a wealth of online materials.

Development of a
portal to Texas
history

References

- Alemneh, D.G., Hastings, S.K. and Hartman, C.N. (2002), "A metadata approach to preservation of digital resources: the University of North Texas Libraries' experience", *First Monday*, Vol. 7 No. 8, available at: www.firstmonday.dk/issues/issue7_8/alemneh/index.html (accessed 7 June 2004).
- Bishoff, L. and Allen, N. (2004), *Business Planning for Cultural Heritage Institutions*, Council on Library and Information Resources, Washington, DC, available at: www.clir.org/pubs/reports/pub124/pub124.pdf (accessed June 14, 2004).
- Campbell, R.B. (2003), letter to the Institute of Museum and Library Sciences, Washington, DC.
- Council on Library and Information Resources (2001), *Building and Sustaining Digital Collections: Models for Libraries and Museums*, The Council, Washington, DC, available at: www.clir.org/pubs/reports/pub100/pub100.pdf (accessed June 14, 2004).
- Gunn, B. (2004), "Controversy in the archives: a documentary saga of Davy Crockett's death", *Texas Library Journal*, Vol. 80 No. 1, pp. 10-15.
- Moen, W.E. (2003), "No longer under our control: the nature and role of standards in the 21st century library", Luminary Lectures @ your library™, Library of Congress, Washington, DC, available at: www.loc.gov/rr/program/lectures/moen.html (accessed June 14, 2004).
- Paige, R. (2002), "Introductory letter", *VISIONS 2020: Transforming Education and Training Through Advanced Technologies*, US Department of Commerce, Technology Administration, Washington, DC, available at: www.technology.gov/reports/TechPolicy/2020Visions.pdf (accessed June 14, 2004).
- Tanner, S. and Deegan, M. (2002), "Exploring charging models for digital cultural heritage: digital image resource cost efficiency and income generation compared with analog resources", University of Hertfordshire, Hertfordshire, available from: http://heds.herts.ac.uk/mellon/HEDS_charging_models.pdf (accessed June 14, 2004).



LHT
23,2

164

Received 27 August 2004
Revised 29 November 2004
Accepted 17 February 2005

THEME ARTICLE

Meeting metadata challenges in the consortial environment

Metadata coordination for North Carolina Exploring Cultural Heritage Online

Katherine Wisser

Duke University, Durham, North Carolina, USA

Abstract

Purpose – The purpose of this article is to focus on the approaches to metadata pursued by North Carolina's Exploring Cultural Heritage Online (NC ECHO), a statewide digitization project. Metadata forms the cornerstone to the project and serves as a vehicle to meet the vision of all-inclusive access to the state's unique cultural and historical resources.

Design/methodology/approach – The article begins with a description of the cultural heritage institution landscape in North Carolina and a discussion of that landscape within the framework of metadata challenges. Four distinct but interrelated approaches are then discussed in the context of the project's missions and goals: working groups, "metadata first", training and outreach, and "Seek a Metadata Consultation".

Findings – The described approach demonstrates a commitment to communication, facilitation, and empowerment that is fundamental to the overall mission of NC ECHO: to provide access to the state's cultural heritage materials through an online portal.

Originality/value – The article provides a discussion of the importance of marketing, approval, coordination, and participation in the context of these metadata solutions.

Keywords Portals, Digital storage, Heritage, History, United States of America

Paper type Case study

Introduction

The digital environment has increased pressures on cultural heritage institution communities to create cooperative and collaborative programs. This is evidenced by the growing number of state-wide and regional digital projects [1]. As with many other states, North Carolina Exploring Cultural Heritage Online (NC ECHO) has sought to create a digital project with a state-wide infrastructure through federated relationships with partner institutions (www.ncecho.org/). NC ECHO accomplishes this with a multi-tiered set of goals: a state-wide survey of all cultural heritage institutions, development of an online portal to institutions and their collections, continuing education opportunities for cultural heritage professionals in the state, the development of online tools to aid in digitization efforts, and a grant program aimed at the creation of online content and the acquisition of digitization skills. Metadata forms one pillar of this statewide project. Metadata is fundamental to effective retrieval and use of digital materials, as well as the administration, dissemination, and preservation of those materials in the online environment.



This article focuses on the approaches to metadata pursued by NC ECHO. It begins with a description of the cultural heritage institution landscape which the project serves. This landscape presents several challenges for effective metadata coordination. To meet those challenges, four distinct but interrelated approaches will be discussed: working groups, “metadata first”, training and outreach, and accessibility to expertise through individualized metadata consultations for partner institutions. These approaches demonstrate NC ECHO’s commitment to communication and empowerment, its perceived role as facilitator and disseminator of expertise, and the overall goal of the project, which is to provide access to the state’s cultural heritage materials through an online portal. The article concludes with a discussion of the importance of marketing to assure partner participation and approval in the context of the consortium.

The North Carolina cultural heritage institution landscape: understanding the pond

An understanding of the significance of metadata in the cultural heritage institution landscape would be incomplete without an initial understanding of the institutions themselves. Therefore, this paper begins with a broad discussion of the cultural institution landscape that has been surveyed to this point by the NC ECHO staff. The most striking aspect of this landscape is the sheer diversity of cultural heritage institutions within North Carolina. An examination of the survey data from 535 institutions in July 2004 (projected completion in 2005) indicated that the state’s cultural heritage institution landscape is made up of museums (43 percent), archives (13 percent), library or special collections (23 percent), private institutions (4 percent), and other (17 percent), which includes cultural centers, parks, and registries.

Within these institutions, material dates from 500 BCE to the present day, while 94 percent of institutions surveyed responded that materials from the modern era (nineteenth and twentieth centuries) form the largest parts of their collections. The institutions reported a total of 447,727 linear footage of material, 8,582,534 objects, 650,164 microfilm reels, 41,703 oversized paper items, 24,783 motion picture films, 20,484 video tapes, 102,125 audio tapes, and 7,493 computer media. Institutions also range in staff sizes from solely volunteer organizations to institutions that have over 40 staff members. The majority of institutions reported five or fewer staff members.

The sum total of these statistics demonstrates the diversity that is the hallmark of North Carolina’s cultural heritage institutions. That diversity, while refreshing and exciting, presents several challenges to a state-wide consortium. Institutions range in technological and professional knowledge and capabilities. From equipment to expertise, the whole range of obstacles to achieving open access can be found. Many institutions reported resource limitations, most notably staff issues, infrastructure challenges such as concerns for security, and physical restrictions, including natural disasters, space and equipment.

The visitors to North Carolina’s cultural heritage institutions use a variety of methods to gather information about the material in the institutions’ collections, including postal mail, electronic mail, telephone and fax, but the number of physical visits still far outweighs visits from a distance: a total of almost eight million physical visits per year were recorded by institutions surveyed. While no count of virtual visits to institutions through the world wide web exists, NC ECHO may, in the second phase

of its project, conduct an analysis of its own online effectiveness through visits to its web page and then coordinate with its partner institutions on that front to provide a better picture of the impact of online information to the exposure of cultural heritage materials in North Carolina to its citizens.

Within this landscape, metadata approaches need to be applied to assist institutions with creating access tools for their materials, administrating and managing the materials in their collections, and participating in the larger NC ECHO portal project. In order to ensure that the needs of the entire cultural heritage institution community are addressed through these approaches, metadata solutions need to be constructed through a variety of approaches. These approaches should also work in accordance with the basic goals and objectives of the NC ECHO project: increased access to cultural heritage materials in the state.

Metadata approaches: working groups

The working group approach illustrates two central components of NC ECHO. First, NC ECHO has a small centralized infrastructure, with a current staff of four. Second, working groups rely on the large expert-base from partner institutions. They consist of members from the field who are interested in participating in the formation of best practice guidelines, tools, and making decisions on the application of particular metadata standards. The goal of creating guidelines and tools is to ensure consistency of application and to teach cultural heritage professionals best practices in creating metadata for their materials.

The working group approach has several advantages. Working groups involve the users of systems in the design of those systems, thus increasing their accessibility to potential institutions. This participation also avoids the creation of applications within a vacuum. In addition, with the participation of professionals from a variety of institutions, the metadata solutions are not institution-specific but rather generalized for the diverse institution base of NC ECHO. This produces a wider, more rounded implementation or understanding of metadata, one that is useful for all kinds of institutions. Finally, by soliciting participation in working groups, NC ECHO promotes professional development in the community and gains a stronger endorsement for implementation to the standard. The decentralizing effect of working groups assures that, at least among the members of the working group, professionals have agreed to the importance of a particular standard and its implementation.

Working groups also present some challenges. They can often be slow-moving and ask for participation from professionals already overly busy with their own responsibilities. Meetings are difficult to plan with varying schedules, and travel can be an issue for participants who are located in the extremities of the state. Creative solutions to these challenges include reliance on virtual communication systems, directed requests for participation of key professionals but open membership to the community at large, and an emphasis on the communication of the work that working groups accomplish to engender excitement in the NC ECHO community about the efforts. A little positive reinforcement goes a long way in engaging professionals in external cooperative ventures.

Two working groups illustrate the success of the working group approach. North Carolina Encoded Archival Description (NCEAD) is the longest standing working group, and was formed in tandem with the NC ECHO project in 1998 (www.ncecho.org/). When

NC ECHO hired a metadata coordinator in 2003, NCEAD became part of the NC ECHO project. NCEAD currently has a highly-developed structure, involving an Executive Committee, a Standards Working Group, and a Technical Working Group. The Executive Committee provides overall perspective, including the explication of priorities for the working groups and the review of documentation from NCEAD. The two subordinate working groups deal with the two different but interlocking issues of standards and technology. The initial working group was parceled out this way to allow interest to dictate participation. This structure also keeps the size of the working group meetings to manageable and productive numbers. NCEAD has been working on the standardization and implementation of EAD in North Carolina for over four years now, so this highly-developed structure is a result of its success in recruitment and commitment.

In contrast, North Carolina Dublin Core Working Group (NCDC) is a larger, less developed, and looser working group. Because Dublin Core was adopted as a minimum standard for NC ECHO, the NCDC felt it necessary to provide implementation guidelines. NCDC also developed an online template for the creation of Dublin Core by partner institutions (www.ncecho.org/ncdc/template.html). The working group meets on a semi-annual basis to review, update, and maintain the application of the Dublin Core standard, including information from other metadata working groups focused on more sophisticated standards. The importance of the NCDC working group cannot be overstated. In any search solution, consistent application of Dublin Core is an essential aspect for providing accurate and useful retrieval of information.

Other working groups formed due to demands from the community. These include a museum metadata working group that is currently developing a "Museum Core" content standard, and two new working groups: TEI-NC is set to analyze the application of the Text Encoding Initiative and a Photograph Description group that will be formed in 2004-2005 to interpret the VRA Core standard for NC ECHO. Online oral histories have also surfaced with their own special needs for description and may necessitate a working group solution.

In all working groups, members from partner institutions with skills, experience or expertise are asked to participate. Membership is also open to the larger NC ECHO community for those interested in participating. The metadata coordinator participates in each working group in the same capacity: by convening the groups, by disseminating information regarding the groups including the development of web pages, and by disseminating documentation, facilitating the groups' needs with NC ECHO, and negotiating and balancing the decisions of metadata standards against those of other applications in the NC ECHO project. Negotiating several different metadata solutions in the state, rather than trying to fit all cultural heritage information into one standard, is a challenge that can be overcome with effective communication and consistent goals and objectives.

Metadata approaches: "metadata first"

In creating a program for a state-wide consortium, both knowledge and technological issues need to be addressed. The "metadata first" approach focuses on the creation of quality metadata as a first step. In doing this, the consortium examines existing metadata systems used by institutions. It provides guidance on the application of input standards, controlled vocabularies, and descriptive paradigms. This approach makes

the metadata the primary force, while technological solutions are still being addressed. The “metadata first” approach dictates that the metadata will determine the retrieval capabilities rather than trying to retro-fit the existing metadata into a technological solution.

Given the variability of the technology, NC ECHO operates with two purposes. NC ECHO partner institutions create consistent and quality metadata that will better serve their institution’s participation in a consortial environment while at the same time creating metadata that is robust and in agreement with national standards. The primary objective is for NC ECHO partner institutions to create metadata that will be sustainable for the long-term. It does not presuppose one technological infrastructure; rather, it is system-independent.

As noted in the discussion of the landscape, there is great diversity in the partner institutions. This diversity is reflected as well in the systems that partner institutions use. From museum management systems to online public access catalogs, partner institutions use systems to meet their specific needs. Balancing the variety of systems with the potential leveraging of metadata in an online portal is the next step for the NC ECHO project. By adopting a metadata first approach, NC ECHO provides support to partner institutions while acknowledging that these institutions have invested valuable time and resources into their systems. Therefore, metadata solutions offered by NC ECHO need to conform to those systems. While there is still a great deal of work to be done to see if this kind of metadata approach is tenable in North Carolina, the support garnered by the “metadata first” approach provides a solid foundation for an inclusive solution for all cultural heritage institutions in North Carolina.

Metadata approaches: training and outreach

Closely aligned with the “metadata first” approach is an accompanying education program designed to provide an arsenal of knowledge and skills in consistent metadata applications for as many professionals as possible. These metadata skills include descriptive, analytical, administrative, preservation and structural metadata systems. Metadata skills are among the education needs most in demand for professionals in the field. Many metadata systems are relatively new and are well-suited for continuing education. In addition, providing learning opportunities for NC ECHO professionals presents professional development opportunities that enhance one of North Carolina cultural heritage institutions’ strongest assets: its professionals.

For novices, metadata can be an intimidating topic. Workshop participants anticipate that there will be a decoding of complex concepts. Dismantling the terminology and assuring professionals that metadata is something that they have always done is a fundamental aspect of the education process. What is new is the current technology trends that can take advantage of that work. Bringing these ideas to the cultural heritage professionals forms the foundation for metadata education in the NC ECHO community.

An example of how NC ECHO uses workshops to integrate metadata throughout the North Carolina cultural heritage community is the Encoded Archival Description workshops (www.ncecho.org/conted/continuing_education_template.asp). For the past two years, EAD workshops have been offered in seven different locations throughout the state to sold-out audiences. North Carolina is a geographically dispersed state. There are western, eastern and central regions that form their own communities. Many

partner institutions are limited in staff and resources, making it difficult to support travel or to forego an employee for extended periods to attend workshops held in distant locations. To address these concerns, NC ECHO strives to make all its continuing education opportunities affordable and to bring those workshops to the regions rather than solely hosting them in the central region where NC ECHO is housed. Therefore, EAD workshops have continuously been “on the road”.

Another continuing education program, “workshops to go”, is in the development stages. The “workshops to go” model assumes that an institution would like to host a workshop. The instructor and an NC ECHO representative will travel to the institution, work with registration and marketing to ensure attendance, and “bring the workshop” to the place. This kind of structure demonstrates NC ECHO’s willingness to serve the community for which it is committed. Other metadata training opportunities include a full-day section as part of the Digitization Institute (www.ncecho.org/conted/digitization.asp). As NC ECHO incorporates increasing numbers of metadata schemes into the metadata lexicon, workshops will be held to diffuse those schemes to the NC ECHO community.

Workshops are only one way to provide access to metadata knowledge and skills to the NC ECHO community. NC ECHO gives numerous presentations that focus on metadata fundamentals as well as detailed examinations of metadata systems. In addition, a “Metadata Matters” column has become a regular feature of the NC ECHO newsletter (www.ncecho.org/newsletter.asp). This column informs the NC ECHO community of new and emerging metadata standards, tools developed for the community and available through NC ECHO, and working group achievements. Maintaining open lines of communication and encouraging participation through the column provides another way for the metadata messages to get out.

Overall, metadata systems can only be implemented effectively by partner institutions if the professionals have the knowledge and skills to apply the solutions. Through training, presentations, and open lines of communication, NC ECHO strives to maintain relevance and currency for partner institutions. This assures that all cultural heritage institutions will be active and equal partners in the state-wide community.

Metadata approaches: “seek a metadata consultation”

The significance of providing easy access to metadata expertise cannot be over-estimated. Some institutions require just a sounding board while others benefit from on-site consultations to establish sophisticated metadata programs that will be sustainable and long-lasting. Due to the variety of metadata needs, the project has made metadata consultations available for any partner institution on request. One of the overriding principles of NC ECHO has been the importance of the institutions themselves. This is clearly demonstrated by the support provided by NC ECHO for its staff to do on-site visitations to all institutions as part of the survey and is extended to the metadata portion of the project.

As an institution develops its metadata implementation, it is free to request a metadata consultation by email or phone. The metadata coordinator works with the institution either through electronic mail, over the phone, on-site visits, or, as in many cases, a combination of all of the above. This allows the institution to instigate the metadata consultation, to define its own specific metadata needs in conjunction with its technological sophistication and capabilities, and to learn valuable skills for continuing

success in metadata creation. To advertise these consultations, a “Seek a Metadata Consultation” button is included on the NC ECHO home page and metadata initiative pages. Other efforts to encourage consultations include reminding readers through the “Metadata Matters” newsletter column and encouraging professionals through workshops and presentations. Metadata consultations are growing in frequency as institutions are becoming more aware of the service. It is assumed that the metadata consultation rate will only increase in the next year as more and more institutions are exposed to it.

To date, the metadata coordinator has consulted with a variety of different institutions. The most advantageous type of consultation from a “metadata first” approach is one that involves consultation at the outset. For example, a public library and museum were interested in creating an online portal for the collections of the museum. They were at the beginning stages and were trying to decide on the best approach. They sought a metadata consultation at that point to discuss the importance of descriptive fields for the artifacts in the museum. The size of the museum precluded professional museum staff, so consultations were with a volunteer librarian from a local university. Throughout the consultation, metadata concepts were discussed in a librarian’s framework (MARC cataloging). At the same time, discussions included evaluating workflow and staff time to devote to the project without compromising metadata quality.

Other consultations involve questions regarding the application of NC Dublin Core for records, review of metadata records and more sophisticated metadata problems including the customization of NC ECHO tools developed for NCEAD, on-site training, and facilitation, to highly developed stylesheet additions for EAD. Throughout the consultation process, no issue is too insignificant, no metadata too meager to be considered important and worthy of attention. Confirming the importance of all metadata needs in the NC ECHO community through these consultations directly supports the all-inclusive nature of the project.

Conclusion

For all of these approaches to be successful, and indeed for the project itself to succeed, partner institutions need to feel that they are partners not subsidiaries. The importance of marketing can not be underestimated for any state-wide project, but it bears special consideration when you are asking the staff of these institutions to devote time, energy, and resources to the creation of something that can, on the face of it, be intimidating and frightening. Debunking the metadata myths of technology through training and outreach is one way to create an open and productive environment; communicating metadata information in an informal and non-hostile manner is another.

Participation in metadata initiatives, as seen through working groups, the dissemination of information in training and outreach, and the personalized attention through the consultations places each partner institution on equal footing. At NC ECHO, metadata solutions do not descend from on-high, from a core that is not cognizant of its periphery, from some “black box” that says “just do it”. It comes from the working groups formed from partner institutions and through NC ECHO, succeeds by cooperative effort. These working groups make clear that NC ECHO enjoys the active participation of members from partner institutions. In every single consultation and training opportunity, no matter what the topic, NC ECHO asks for input,

experience, and participation in the decision-making process. Through this dialog the standards become comprehensible, institutions find solutions that are sustainable, and NC ECHO learns how it can support its partners' endeavors. Above all, these strategies represent a "teaching them to fish" approach to metadata in a consortial environment. It stresses empowerment over product in order to embed long-term sustainability for the institutions as well as the project.

These metadata approaches reflect the mission of NC ECHO: to be an all-inclusive resource for North Carolina cultural heritage institutions to improve access to their materials. From the button and aluminum Christmas tree museum to the hallmark collections of rare books and manuscripts, from the sophisticated art collections to the local artists' works displayed in post offices, from the scientific and archaeological research facilities to the historic site plantations of tobacco and local flora and fauna, NC ECHO seeks to address the needs of its partner institutions in the pursuit of access. NC ECHO demonstrates that much evidence of the human record can be found in North Carolina, and metadata is the gateway to those precious artifacts for today and the future.

Note

1. See www.mtsu.edu/~kmiddlet/stateportals.html for a comprehensive list of statewide and regional digitization projects.



LHT
23,2

172

Received 27 August 2004
Revised 29 November 2004
Accepted 17 February 2005

THEME ARTICLE

Ohio Memory Online Scrapbook: creating a statewide digital library

Laurie Gemmill

OCLC Online Computer Library Center, Inc., Dublin, Ohio, USA, and

Angela O'Neal

Ohio Historical Society, Columbus, Ohio, USA

Abstract

Purpose – To provide a model for statewide collaborative digitization projects, based on the development of the Ohio Memory Online Scrapbook.

Design/methodology/approach – Using Ohio Memory as a model, this paper establishes a three-pronged approach to creation of a digital library. Creation of content, development of tools and a focus on outreach were critical to the success of Ohio Memory, a statewide digitization project involving 330 historical societies, libraries and museums throughout Ohio.

Findings – This paper establishes a three-pronged approach to digitization. Creation of content, development of tools and a focus on outreach were critical to the success of Ohio Memory. The collaborative nature of the project is an underlying aspect of Ohio Memory, it led to the creation of a standardized, central point of access for digitized primary source materials relating to Ohio's history.

Practical implications – This paper provides helpful advice to organizations considering digitizing their collections and details a model framework for statewide collaborative projects.

Originality/value – As more organizations establish digital libraries to enhance access to collections, this paper offers an example of a successful project. It also details significant considerations to take into account when building a digital library.

Keywords Digital libraries, History, United States of America

Paper type Case study

Introduction

Primary sources tell the story of Ohio from a first-hand perspective and are invaluable aids for teaching and learning about history. Museums, public libraries, historical societies and other repositories hold historical treasures. The Ohio Memory Online Scrapbook (www.ohiomemory.org/) is a searchable database that provides access to these photographs, documents, artifacts and natural history specimens drawn from all counties of the state so that researchers around the state, and around the world, can use them.

This article outlines the project's history and describes the three-pronged approach that staff used to develop the scrapbook: focusing on content, tools, and outreach. A primary focus on strong content is at the core of the Ohio Memory Online Scrapbook. The site is content-driven, with detailed descriptive metadata that allows users access not only to the digital images, but also the historical significance of the item. While



quality content is of paramount importance, it is also vital to develop tools for users to interact with the items in the scrapbook. Features such as text transcripts, My Scrapbook, and zoom allow users to experience items in the database, sometimes in ways impossible to do with the original items. The development of such tools is the second prong in our approach. The third element of creating a digital library is outreach. Without users, even the best resources will fail. Ohio Memory's focus on collaboration created an instant user base of the more than 330 institutions who participated in the project.

More outreach was necessary, however, to encourage students, educators, reference librarians and the public to use the scrapbook. Presentations, publications, usability testing, and promotional items were critical to the success of Ohio Memory and should not be ignored in developing a digital library. The basis of all three aspects of this approach is collaboration. The collaborative nature of the Ohio Memory project, which includes more than 330 contributing institutions and several project partners, allowed for the creation of a resource that is larger than anything one of the contributors/partners could have created individually. By combining resources, project partners were able to create a powerful resource at a reduced cost. Furthermore, collaboration provides a single site to access collections from hundreds of Ohio's cultural institutions. Standardized content and quality ensure users that Ohio Memory is a reliable source users can count on for accurate information and standardized image quality.

Project background

The project officially began July 1, 2000. Its mission is to commemorate the Buckeye State's bicentennial in 2003; celebrate state and local history; encourage cooperation between archives, historical societies, libraries, museums, and other cultural organizations; and allow the global community to discover and explore Ohio's rich past.

During the first two years, the project staff worked with over 250 cultural repositories throughout Ohio to select, digitize and make available historical materials from pre-history through 1903. Phase two began in October 2002 and concluded December 31, 2003. During that phase work focused on adding historical materials relating to 1903-2003. By 2004, more than 27,000 images were available from 330 participating institutions, including archives, historical societies, libraries, museums and zoos (www.ohiomemory.org/om/participants2.html).

While the Ohio Historical Society is the lead institution in the project, the effort would not be possible without the support of six partner institutions: the Ohio Public Library Information Network, OhioLINK, the Ohio Library Council, INFOhio, the State Library of Ohio, and the Ohio Bicentennial Commission. These organizations provided financial, organizational, and technical assistance critical to the development, maintenance, and promotion of Ohio Memory (Figure 1).

Creating content

The first component in the project was creating content. Early on, several pivotal decisions related to staffing assignments, levels of outreach, and methods of encouraging participation laid the foundation for the project. It was important to set guidelines for participation as soon as possible. Museums, historical societies, archives,



Figure 1.
Ohio Memory home page
(www.ohiomemory.org)

libraries (public, academic and special), and any other repository with historical collections that is open to the public were eligible to participate. While there are many valuable materials held in private hands, staff felt that we would be unable to handle the volume if the project was open to the general public. Also, local museums, archives, libraries, and historical societies have already culled through local materials to acquire historically significant items. Thus an initial sifting would have already taken place, easing the decision making process. However, in the interest of widespread participation, we determined we would make every effort to accept at least one item from each institution that submitted materials.

Ohio Memory was envisioned as a site rich in content. Project staff developed a framework for content development that included selection criteria, subject categories, and a submission process. In an effort to create a balanced, comprehensive history, two staff members, both with master's degrees in history, were hired as content editors to select materials for the site and ensure accuracy.

Historical significance and complementary value were established as primary selection criteria, along with the more technical elements of scanning potential, copyright and ownership.

Historical significance

The primary consideration for inclusion is the value of the materials in terms of historical content. We encouraged submission of collections or items that represent experiences that are either common to all Ohioans or unique to a specific local community, and that reflect one or more of the Ohio Memory subject categories. Staff

found this the most important quality in selection. However, it proved to be the most difficult to describe to participants. Staff encouraged the submission of items that are commonly requested by patrons and considered the treasures of their collections.

Complementary value

The second selection criteria is complementary value. It was important to incorporate a way to include materials representing various geographical areas and time periods and portraying various facets of life in Ohio. Materials were selected that best complemented one another to present a balanced view of state history. They are grouped into 21 subject categories under five major headings: Ohio Government, Citizenship, Ohio's People, Environment and Culture. While no set of categories can provide complete coverage, we found these subject categories to be very beneficial in trying to cover many different topics in Ohio history. The addition of the Environment heading, which includes the categories Plants and Animals, Climate and Weather and Geography and Natural Resources, extended the scope of the project beyond the traditional humanities content, although items in these categories were still selected based on the selection criteria, including historical significance. The subject categories helped define historical significance and gave the team a foundation on which to build the scrapbook. They also forced us to make decisions about content before answering questions from the public. When participants asked about potential submissions, it was helpful to have a handout of subject categories and sample items. Many participants had difficulty identifying items or collections they wanted to submit, so the subject categories provided an excellent place to start.

In addition to the subject category framework, balanced geographic coverage was also important. Accepted items or collections were tracked on a wall map of Ohio, with pins providing a visual representation of areas that are well documented or less fully represented. Individual items or collections were tracked in addition to contributing institutions, as some institutions contributed materials from other parts of the state. Institutions were not limited as to the number of items they could submit. Participation ranged from submission of one or two items, to more than 100 collections.

Scanning potential

The third selection criteria is scanning potential. While many items may have historical significance, not all will digitize well. Typically, any flaws in a document – stains, illegible writing, and smudges – are magnified when digitized. A readable digital image is paramount. Suitability for digitization was evaluated in terms of image quality and physical condition. For example, items that have high contrast and/or a clean, crisp font digitize well and were more likely to be accepted. Inability to digitize the items due to fragile physical condition, size or other constraints was the most common reason for rejection of submitted items.

Time period

In the first phase of the project, only those historical materials created before or related to events that occurred prior to 1903 were accepted, so time period was the fourth criteria. While some institutions were unable to participate in the first phase because they held only twentieth-century collections, focusing on Ohio's early history generally benefited the project. Focusing on this earlier time ensured that the early part of the

state's history is covered. In addition, copyright issues were made simpler by this early focus, and allowed staff to focus on refining the submission and selection processes

Copyright/ownership

The final selection criteria focused on legal issues. The institution needed to hold property rights to the collection or item. The provenance, or history of ownership, had to be documented as thoroughly as possible. Items that violated copyright restrictions were not accepted without the copyright holder's permission. Copyright issues proved to be some of the most difficult to ascertain. Project staff was determined not to violate any copyright, but copyright law is extremely complex and frequently misunderstood. On occasion, difficulty in determining copyright or lack of permission from the copyright holder led to rejection of submissions.

Submission process

In order to participate, cultural heritage repositories sent submission forms to staff for consideration of items. Submission forms could be emailed, submitted online, or sent in the mail via paper forms; submissions were accepted on a quarterly basis during the first phase and for one deadline in the second phase. The project staff debated how much information to require from submitting institutions and decided to ask for the minimum amount to encourage submissions, even though it meant an increased workload for staff. During a series of 12 workshops to encourage participation in the Ohio Memory Project, the subject categories helped to explain to participants what kinds of items we were interested in. Many were surprised to find that we wanted to document Ohio's natural history. Although the categories were broad enough not to limit submissions of any items, they provided participants with a better idea of what we were doing and served as a starting point for them to brainstorm about collections they would like to include.

This framework is also easy to track, allowing us to clearly define another selection criteria, complementary value. Selected items should "best complement collections in the online scrapbook to present a balanced view of state history". When each item or collection was accepted, it would be placed in the appropriate category or categories. We simply made a graph with the numbers of items in each category and immediately determined what broad categories were underrepresented. After the second round of submissions, staff created a Wanted List to distribute to participants. At first, the list included only the broad subject categories. As more material was submitted and categories began to fill, we then focused on specific people, places or events. For example, instead of including the "Transportation" category on the list, "the Winton automobile" was added. Participants were eager to help with the list. It made participants feel more a part of the process and shortened their time in selecting collections to contribute.

To publicize geographic areas where submissions were needed, we created a list of Underrepresented Counties. This list helped us meet our goal of having one item about every county in Ohio by the end of 2001. Once that goal was met, the Underrepresented Counties list was expanded to include "Top Priority" counties with fewer than ten items in Ohio Memory and "Second Priority" counties with fewer than twenty items in the database. As of April 2004, there are only four counties with fewer than ten items in the scrapbook. Like the Wanted List, the Underrepresented Counties list encouraged

participation from institutions that did not want their county on the list and led others in nearby parts of the state to look for items in their own collections that could document neighboring counties.

The Wanted and Underrepresented Counties lists also kept up momentum for the project and made participants feel involved. A monthly update about content focused many submissions and reminded participants that the submission deadline was coming. Without publicity, however, these lists would have been ineffective. Wanted Lists and Underrepresented Counties lists were distributed at Ohio Memory Project Workshops, via listservs, on the project's web site (www.ohiomemory.org/about) and in information packets about the project. Ohio Memory also established a listserv (<http://mail.oplin.org/mailman/listinfo/omp>) hosted by OPLIN (Ohio Public Library Information Network), which served as an effective way to communicate with participants and potential participants. The list's 224 subscribers received project updates, reminders about approaching submission deadlines, and news of web site additions. In addition to the Ohio Memory listserv, monthly or semi-monthly notices were sent to several professional association lists, including the Society of Ohio Archivists, Academic Library Association of Ohio, and the Ohio Public Library Information Network.

Potential participating institutions were frequently unsure of what to contribute. Many institutions have a variety of materials in their collections and wanted more direction about what would be accepted by Ohio Memory. We addressed this in several ways. We created and distributed "Category of the Month" messages that focused on a subject category and provided examples of the kinds of materials we would like to receive. We analyzed submissions as deadlines passed, then created and distributed charts reflecting which subject categories were well represented and which were not. However, personal visits to sites and on-site review of materials with a follow up letter encouraging specific items were the by far the most effective methods for submission of valuable materials. In addition, we found that deadlines were most helpful in encouraging submissions. In a world where everyone has too many tasks to achieve, a deadline is sometimes the only thing that can motivate action.

Digitization

Once materials were accepted, project staff arranged for digitization. Materials could be scanned at the Ohio Historical Center or at the submitting organization, as long as minimum scanning specifications were met (300-400 dots per inch, TIFF files). Higher resolutions were recommended, however, and were used when Ohio Memory staff digitized materials. For example, while photographs were often scanned at 300 dpi, manuscripts were digitized at 600 dpi and slides at 1200 dpi. Archival TIFF files are used to create JPEG derivatives for the web site. Due to contributors' concerns about releasing TIFF files, only JPEG derivatives are available to the public on Ohio Memory. Storage space is not a substantial issue for Ohio Memory due to our partnership with OhioLINK, a consortium of Ohio's college and university libraries and the State Library of Ohio. Through its Digital Media Center, OhioLINK provides virtually unlimited storage space on its servers allowing contributors to submit content of all types.

Maintaining high archival digitization standards proved to be at times an arduous task. Frequently, images would not be digitized according to standards, resulting in rescanning until the proper file format and dpi were achieved. Many participants had

digitized materials for use on the web, but were not familiar with archival standards. Once materials were digitized metadata was applied.

Metadata

Ohio Memory uses the Dublin Core metadata standard (www.dublincore.org/). Dublin Core consists of 15 elements, all of which are optional and may be repeated. The flexibility of Dublin Core allowed staff to develop standards for data values in a way that would incorporate differences between museum objects, photographs, documents, and natural history specimens. As consistency of metadata application is of paramount importance for effective retrieval, the project team developed a set of metadata guidelines to use when creating descriptions for primary source materials. These guidelines were designed to be broad enough to include individual items or collections and flexible enough to describe the various media types included in the project. Project metadata guidelines specify the format of titles, descriptions, dates, place names, creators' names, and other details. Because information provided by institutions varied greatly, staff spent a significant amount of time bringing uniformity to the metadata and ensuring its accuracy (Figure 2).

OhioLINK serves as the host repository for the images that are part of the online scrapbook. The database and images are stored on and served from the OhioLINK servers. Metadata records are first entered into a local Microsoft Access database that also includes all project-related contact information, queries, reports, and forms. Each metadata record is then evaluated, revised and approved. Approved metadata are exported from Access to an XML file for transport purposes. This file is used to

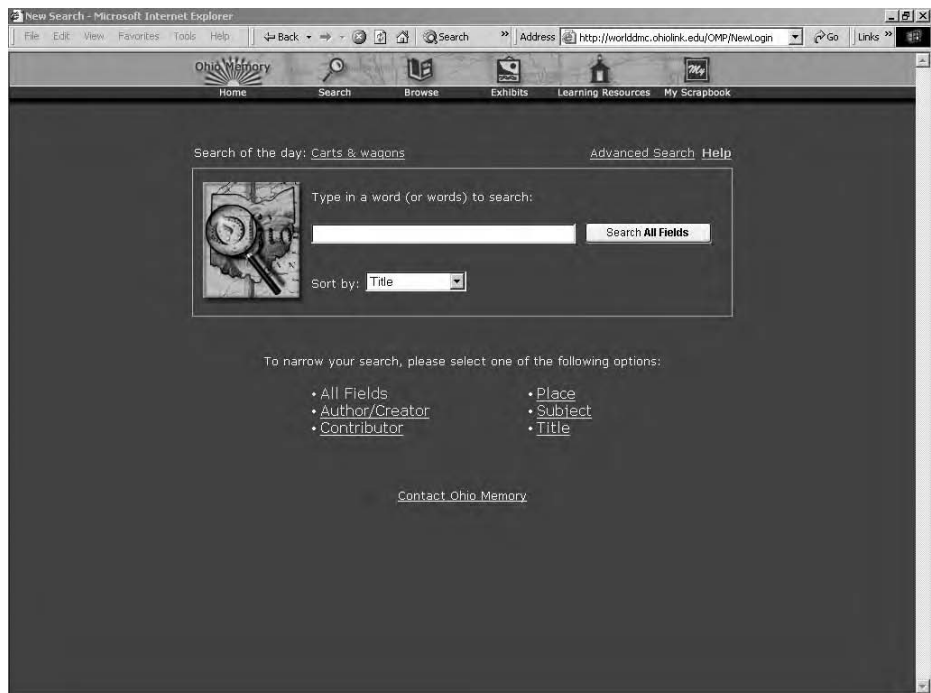


Figure 2.
Ohio Memory search page

register TIFF images to Documentum, along with their corresponding metadata. The Documentum system is hosted remotely on an OhioLINK server. It ingests the registered assets and automatically creates searchable administrative metadata for the image, searchable descriptive metadata (i.e. from the Dublin Core values) and thumbnails for all images.

Proxy scripts run after the asset registration step, creating both a larger jpeg and a zoom proxy for each registered image. New metadata entries in Documentum are copied daily to Oracle tables for use with the Ohio Memory public interface. New entries are evaluated daily for accuracy and completeness. If revisions are needed, modifications are made through Documentum and the changes are published the next day. OhioLINK's original content management software (Bulldog) was purchased by the parent company that created Documentum. OhioLINK continued with Documentum for various technical and interoperability reasons, among which are that:

- Documentum is format and content neutral (it supports images, audio and video);
- has customizable metadata – not tied to MARC or Dublin Core;
- provides support for “groups” of content;
- it is flexible and extensible (has an open Software Developers Kit) and scalable for statewide architecture;
- runs on Unix;
- runs on top of Commercial RDBMS system (some solutions are closed data storage system that don't provide standard SQL access);
- has modular architecture to support specific add-ons for media types (e.g. Video processing/search tools, etc.);
- provides strong support for text documents and XML; and
- has strong workflow development tools.

Workflow

Ohio Memory participants and staff engaged in interaction over a prolonged period, sometimes throughout the entire length of the project. The nature of the interaction was typically complicated and detail oriented: beginning with preliminary conversations about potential submissions, continuing with review of submissions, acceptance, digitization, description, and finally publication online. These interactions typically took place over several months and staff often had as many as 30 contacts with each institution. One mechanism that project staff found helpful was to maintain a database of contacts with individual participants. Each time a staff member spoke with or exchanged correspondence with someone, it was noted in the database. This made it easier to maintain a clear sense of actions and conversations with each institution. Tracking contacts in the database also enabled staff to keep track of questions, and send reminders of deadlines to appropriate contacts. We found this essential when the number of participating institutions grew to over 330.

Searching

An important consideration in metadata creation was how the web site would allow searching. Ohio Memory includes a general key word search but also several

field-specific searches: subject, title, place, author/creator or contributor. Users can limit their search by date, format or contributor. These options allow a variety of access points to effectively find materials in the database. Another feature to pull in users is a Search of the Day. Located on the search page above the search box, the Search of the Day displays a different subject-heading daily. Examples include the Underground Railroad, flight and gardens (Figure 3).

For advanced users, the “Help” file on the search screen details several additional search features. “Fuzzy” searching allows users to search multiple spellings of a word by placing a question mark in front of the search term. For example, a search on “?trylobyte” would return eight hits relating to Ohio’s state fossil, the trilobite. This feature is especially helpful for scientific names, proper nouns and place names.

For users who do not have specific search terms in mind, several ways to browse provide access to the materials. The clickable county map creates instant connections, even for users who initially have no idea what to search for. The map of Ohio’s counties is based on the “Place” field. Users who cannot find their county on the map can also use a dropdown menu. Other ways to browse the online scrapbook are also based on metadata fields: author/creator, contributing institution, format or media type, subject headings and place names (including city and township names). Browse is a simple but effective way to introduce users to Ohio Memory.

Item Description - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Ohio Memory Home Search Browse Exhibits Learning Resources My Scrapbook

« Back to results Previous match Next match »

Save to My Scrapbook Printable View

Ohio Map

Thirteen-year-old Mary Munson drew this map of Ohio circa 1822. It shows Ohio counties as they appeared after the creation of Union County in 1820. Although the northwestern counties of Allen, Crawford, Hancock, Hardin, Henry, Marion, Mercer, Paulding, Putnam, Sandusky, Seneca, Van Wert, Williams, and Wood were also created in 1820, they are not depicted on the map. The land in northwestern Ohio was acquired through the Treaty of Maumee Rapids, which was signed on September 29, 1817 and ratified by the United States Senate on January 4, 1819. Munson refers to the treaty in her notation "Country recently purchased of the Wyandott and other Tribes of Indians." The map measures 13" x 12.25" (33 x 30 cm).

Click for full view
1 image

Date:	1822 circa
Author/Creator:	Munson, Mary
Subjects:	Ohio Government ; American Indians in Ohio ; Geography and Natural Resources ; Maps ; Counties ; Treaties
Places:	Ohio (entire state)
Copyright:	Online access is provided for research purposes.
Contact:	For more information, duplication requests, or to see the item or full collection in its original (non-digital) format, please contact the contributing institution. Contact information is provided below.
Contributor:	Ohio Historical Society
Address:	Archives/Library Division 1000 Nelson Ave.

Figure 3.
Metadata record in Ohio
Memory (Ohio Historical
Society, Map VFM 0051-4)

Development of tools

The second part of the three-pronged evolution of Ohio Memory and a significant element in its growth and success is in the creation of tools to enable users to use the online scrapbook in a myriad of ways. It is important to continue to add new materials but also to envision new methods of using those materials. Each new type of material or each new collection offers unique opportunities to interact with them. In our efforts to develop tools that enhanced the user experience, we drew on our own experience but also made sure to ask users what they wanted. Simple ways to engage new users include rotating thumbnail images on the front page and a “Search of the Day”.

Users rarely know what to expect in a new resource. We encourage them to search to find materials but in today’s world of ever shortening attention spans we need quick, effective methods to draw users in and introduce them to collections. One easy method to highlight collections is to use a series of small changing thumbnails of images on the front page. If someone is interested they can click to find out a little more information and click further to view the entire record and high quality image. This set of images changes so that each time a user returns to the site (or refreshes the front page), they have the chance to view new items. The Search of the Day tool provides another opportunity to introduce new collections to users and provide another point of access.

The needs of users are an essential consideration in developing tools for a digital library. Users want to be able to use tools and experience the materials in different ways. Some users may be interested in reading an entire document no matter how lengthy or difficult to read. Researchers, however, often want to scan a document to see if it contains information relative to their topic. Younger students may have difficulty reading cursive handwriting, particularly from the eighteenth century. To deal with these needs, Ohio Memory includes text transcripts for all handwritten materials. This facilitates reading of difficult handwriting, but it also enables users to do a simple find command (control-F) to search throughout a document. While it can be time consuming to create transcripts, the increased level of access to collections is worthwhile. Volunteers or interns are often available to create transcripts (www.ohiomemory.org/om/transcripts2.html), which are then edited by staff.

Another tool that allows users to experience the items in Ohio Memory is the zoom feature, which enables users to appreciate a small detail in a photograph, view the smallest letter in a piece of correspondence or get a very close up view of a three dimensional object such as a piece of pottery. LizardTech’s MrSid technology is used for the project’s image zooming functionality.

An integral feature of Ohio Memory is the My Scrapbook tool, which allows users to collect images in their own scrapbook, which is accessed by a username and password. No personal information is transmitted. A description field at the top and notes fields next to each thumbnail give users a place to add personal annotations. In keeping with the scrapbook metaphor, users have the opportunity to customize their online scrapbook with different backgrounds and fonts. Users get a URL they can give to others to allow friends, family, teachers or students, access to view the scrapbook but not alter the content. Using this feature, users can engage with the artifacts, photographs, and documents, and make them a part of their own personal history.

The My Scrapbook idea developed from participant workshops and early staff discussions regarding how students, teachers and the public might want to use the resources. The Ohio Memory web developer then worked with OhioLINK staff to

create the application. Testing, however, was critical to success of the feature. The web developer conducted usability testing with approximately 25 representative users of the online scrapbook, among which were advisory group members and staff from participating institutions. Findings were generally positive, but minor adjustments were made in the initial development of the My Scrapbook tool in 2000-2001. In 2004, a substantial upgrade was made to the feature, adding more backgrounds for users to select from and allowing users to control multiple scrapbooks with one username and password. Other innovations suggested by users and accomplished with the My Scrapbook upgrade included the incorporation of check boxes to simplify removing items, an extended description field, and more prominent placement of the URL for the personalized page.

My Scrapbook gives users a reason to return to the site, whether to add to their scrapbooks or share them with others. In many ways it is an empowering tool. In the words of historian Andrew R.L. Cayton, who wrote the introductory essay to Ohio Memory, "You don't need historians any more to tell you what to think. You can do it yourself" (Cayton, 2002). While users can access extensive factual information about an item in the description, the analysis is up to them. Between March 2002 and April 2004 more than 5,000 scrapbooks were created using the My Scrapbook feature. Users create scrapbooks for a variety of reasons: researchers use them to collect items relating to research topics, students use them for class assignments, and many people create an Ohio Memory scrapbook just for fun. The tool allows a wide variety of people from different age groups and with different goals to experience primary source materials relating to Ohio's history.

Exhibits

To encourage users to make their own scrapbooks, Ohio Memory established the Guest Editor program. Periodically, prominent Ohioans are asked to serve as Guest Editors. Ohio's First Lady Hope Taft initiated the program by creating a personal scrapbook with an introduction and notes using the My Scrapbook feature. Guest Editors are encouraged to include some biographical information and a photograph of themselves. Scrapbooks are then published on the on the Exhibit page of the web site. These provide an excellent way to see the personal connection that people have with historical objects and materials and bring in users who may not have heard of site. The program also provides additional publicity, keeps users interested, and builds the foundation for new partnerships.

Staff use the My Scrapbook tool to create a series of simple exhibits. Each month we create a few new scrapbooks on various themes, such as "Back to School", "History of Flight", "Hats", and the "Underground Railroad". These small exhibits highlight various items in the collection and demonstrate the diversity of collections included in the scrapbook. Themes are chosen based on current events, holidays, or suggestions from users. In June 2004, for example, Ohio Memory chose World War II as a theme for its featured scrapbook to coincide with the dedication of the World War II Memorial in Washington, DC. Like the Guest Editor program, featured scrapbooks renew interest in Ohio Memory and, in particular, allow teachers an easily accessible source of primary materials to incorporate into their lessons (Figure 4).

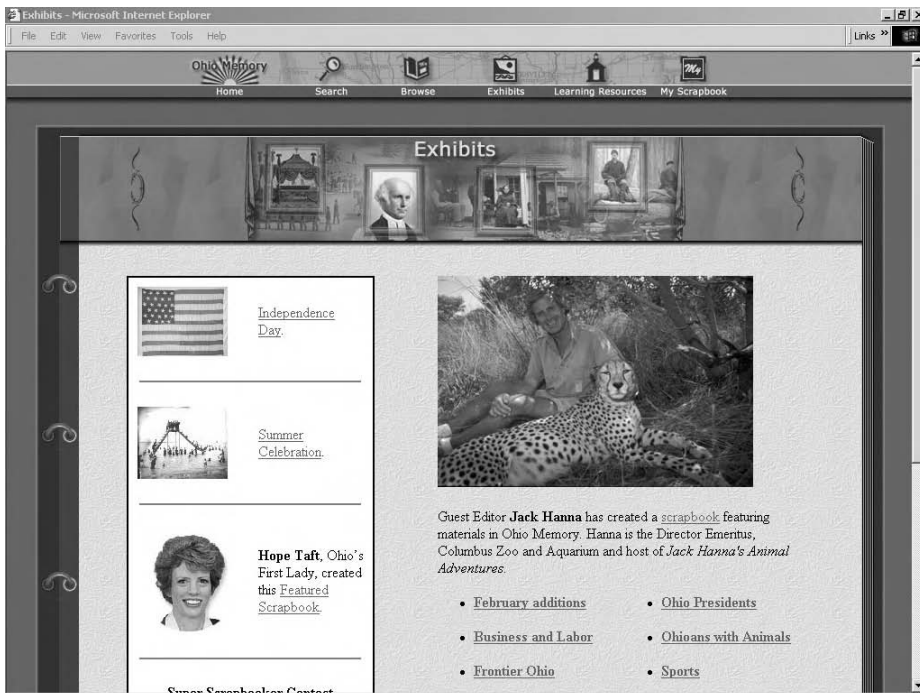


Figure 4.
The Ohio Memory Exhibits page features scrapbooks by guest editors and exhibits on special topics

Educational components

While the tools are imperative to the overall growth and success of the project, it quickly became clear that we also needed to add educational components closely aligned with content standards. Because the Ohio Memory database has reached a critical mass of materials, users cannot easily browse for materials and general searches like “Cleveland” or “Civil War” often return hundreds of hits. Educators in particular want to use primary source materials but rarely have the time to sift through content and create lesson plans around our resources. In an interview titled “Cultural Heritage Institutions Can Provide More Than Raw Material”, Lorna Campbell of the Centre for Educational Technology Interoperability Standards (CETIS) argues, “many teachers and learners do not have the skill to use the raw resources for learning”. She goes on to state that if collections are to be used they will “require some degree of contextualization to provide a meaningful learning experience for the pupils”. Therefore, it is necessary to provide tools such as lesson plans, WebQuests and other educational activities that enable the teachers and school media specialists to incorporate these primary sources with a minimum of effort (von Kasternen, 2003).

Care should be taken to develop tools for various educational levels. To date, Ohio Memory has created several educational activities for adults, students and educators. The “Governor’s Challenge”, a 20-question quiz sponsored by Ohio Governor Bob Taft, is geared toward older students and adults. Since one of Governor Taft’s priorities is education, staff contacted the governor’s office and were given permission to use his name in the feature. Users can test their knowledge of Ohio’s history with questions

such as “Ohio has how many counties?” and “Which two future Presidents served in the 23rd Ohio Volunteer Infantry (OVI) during the Civil War?” The multiple-choice format gives users the option to guess on answers. Once the test is complete, a score is given, along with a ranking such as “Professor”, “Student” or “Alien”.

The Governor’s Challenge is more than a simple quiz, however. All of the choices are linked to items in Ohio Memory. When users are given their “scores”, they also see links to items in Ohio Memory. This increases access to materials in the scrapbook and encourages users to learn more about the correct and the incorrect answers. Links are displayed in a pop-up window to allow users to easily go back to their answers and click on other topics.

Three games were designed for younger audiences or those less comfortable with technology: What Is It? Where Is It? and Who Is It? Each game has a series of images from the scrapbook that users try to identify. To find the answer, they click on the item. This is a very simple way that teachers can incorporate history into a lesson. This model is easily adapted to all items in the database and can be customized for different presentations and topics.

Additional tools, such as lesson plans and WebQuests, will require partnerships with educators and school media specialists. After surveying Ohio teachers and school media specialists, an Ohio Memory study found that teachers are most interested in WebQuests (inquiry-based activities in which most or all of student’s information comes from the web) and lesson plans. The topics they were most interested in included: Ohio Inventors (Thomas Edison, James Ritty, Charles Kettering), settlement, and transportation (canals, railroads, flight). Education may also offer an opportunity for sustainability, either as a source of revenue generation or continued grant funding. Packaged activities based on the Ohio and national Academic Content Standards can be offered to teachers, students and parents for a fee. Revenues from educational tools can then be used to develop more staff and add to the core materials in the database. Educational initiatives, however, are dependent on collaboration and strong partnerships. K-12 teachers, university professors and curriculum coordinators must be involved in the content development process along with project staff. Continued partnerships with established partner organizations, as well as new partnerships, are necessary to guide this process.

Outreach

The final component in the three-pronged approach is outreach. The role of outreach to the success of digital libraries cannot be overstated. Without users even the best sites will fade away. With so many resources and web sites available, it takes a concerted effort to make potential users aware of an excellent resource. Outreach works best when combined with the development of content and tools. With every new collection added and each new tool developed comes an opportunity to tell potential users about it. Some simple methods involve generating press releases and sending out listserv announcements to potential audiences. It is important not just to notify the public about these updates but other information providers such as librarians, archivists, historians and school media specialists.

Presentations to professional associations let professionals know about sites and highlight the content so people know what is there. There is a substantial difference between hearing about something and viewing a feature or seeing a rich resource. The

specifics are what the potential users will remember. Presentations at meetings of state, regional and national organizations such as the Ohio Library Council, American Association of State and Local History, Midwest Archives Conference, SchoolNet, Society of Ohio Archivists, PBS affiliates, Ohio Academy of History, and the Roundtable of Government Documents Librarians all introduced Ohio Memory to new audiences. Articles and news items in organizational newsletters and journals are also a significant source of publicity.

Promotional items can be an excellent outreach tool as well. Once people are familiar with a resource, you need to enable them to remember it. Some simple, relatively inexpensive ways to accomplish this include giving pens, magnets or post-its that include the site URL. Brochures and handouts are also helpful to explain the project and various features. Each feature that is developed provides new opportunities to promote it and the resource. For example, Ohio Memory created a series of bookmarks to promote the new "Governor's Challenge" Ohio history quiz (www.ohiomemory.org/challenge.html). Each of the 12 bookmarks includes an image from the web site and one of the quiz questions. Users must go online to find out the answers and view materials related to the topic.

Creating a poster was one of our most successful promotional efforts. The school age audience was important for us to reach. Staff designed a poster highlighting collections from the web site to encourage use of the web site by teachers and school media specialists. Ohio Memory posters were overwhelming popular with others as well. Thus far, nearly 8,000 posters have been distributed to schools, libraries, museums, and historical societies. Not only are users happy to have an attractive poster, but they have a reminder of our resource hanging on their wall all year long.

Conclusion

Future tools

With more than 27,000 images of primary source materials included in the Ohio Memory Online Scrapbook, the project is approaching a new phase. The Ohio Historical Society (OHS) and OhioLINK are committed to maintaining Ohio Memory and developing more resources for the site. Collections from OHS are added weekly. Several sustainability options are being pursued, including a contract digitization program where cultural heritage repositories can participate in the project on a fee basis.

While staff continue to add new materials, the team's focus has turned to increasing access to the collections included in Ohio Memory and attracting new users. The site receives over six million hits per year, and the number of hits always increases when a new feature or tool is added. For example, in April 2004, when zookeeper Jack Hanna's scrapbook debuted and students were using the site for a "Super Scrapbooker Contest", the site received 774,402 hits, representing more than 14,000 individual users. Hits decreased slightly the following month to around 696,000.

Based on evidence from users who have contacted Ohio Memory, the largest audience groups are teachers, students and researchers. Many have a specific item or topic they are researching. While Ohio Memory is content driven, the vastness of resources available is often lost when users search for one particular topic. Therefore, exhibits are the next logical step. The exhibit will provide context for many of the primary source materials currently in Ohio Memory, while a small number of items

may need to be added to fill in gaps. Collaboration, the key to success with Phase One and Phase Two of Ohio Memory, is also essential in creating successful exhibits.

Conclusion

The three-pronged approach emphasizing content, tools and outreach provides a model for the creation of a successful digital library. An initial focus on creating content results in a quality product. Once standards for the creation of content are established and a critical mass of materials is reached, it is important to develop tools to allow users to easily interact with the content. Ohio Memory's My Scrapbook feature gives users the ability to collect materials from the database in their own personal space. Other simple tools, such as a county map, quiz and identification games, lure users to explore more of the site. Outreach, the third aspect of the three-pronged approach, is crucial. Listserv announcements, publications, and promotional materials introduce users to the web site and get them excited about using it.

Cooperation is necessary in creating a digital library. While it is possible to develop a digital library with the resources of a single organization, the product will always be stronger if partners are involved. By working together, organizations can limit the amount of staff, equipment, and resources necessary to conduct a digitization project. In addition to cost-sharing, partnerships generate new and innovative ideas. Partners are also invested in the digital library from the beginning. They encourage its growth by promoting the project to others and provide a core group of users. The Ohio Historical Society's partnership with the Ohio Public Library Information Network, for example, ensured that Ohio Memory would be used in Ohio's public libraries as well as by the museum community. While digitization can be costly and require significant amounts of staff time, it results in an increased profile for participating institutions, better intellectual control of collections, and the creation of new partnerships.

References

- Cayton, A. (2002), "Introduction to the Ohio Memory Online Scrapbook", available at: www.ohiomemory.org/introduction.html, Columbus, Ohio.
- von Kasteren, J. (2003), "Cultural heritage institutions can provide more than raw material", *DigiCult: Learning Objects from Cultural and Scientific Heritage Resources: Thematic*, No. 4, pp. 10-13.



THEME ARTICLE

Educating future
digitizers

Educating future digitizers

The Illinois Digitization Institute's Basics and Beyond digitization training program

187

Amy Lynn Maroso

University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Received 27 August 2004
Revised 29 November 2004
Accepted 17 February 2005

Abstract

Purpose – The purpose of this paper is to provide a detailed description of the Illinois Digitization Institute's Basics and Beyond digitization training program and to describe how successful the project's different training approaches have been.

Design/methodology/approach – The training consists of one-day workshops, web-based courses, and web-based courses plus a hands-on workshop element. Courses are given at different price points and for different lengths of time. Surveys and quizzes measure how well the participants learn the material.

Findings – As surveys, quiz results, and other data from the courses show, the objective of the project is being accomplished: to present cultural heritage institutions with different types of digitization training to suit their time constraints, budgets, and education needs and produce a new set of professionals who will create successful and long-lasting digitization projects.

Practical implications – The success of the project shows that participants respond well to different training approaches and these different approaches can be implemented to provide cultural heritage institutions with a wide range of digitization learning options suited to their needs. Such methods can also be used for other types of library and non-library training.

Originality/value – This is the first time asynchronous but instructor-led web-based courses have been used for digitization training, and findings indicate that it has been successful. The outcomes of this training can be useful for institutions interested in how well participants respond to this unique style of training.

Keywords Digital libraries, Training, United States of America

Paper type Research paper

Background of the basics and beyond program

In February of 2002, Professors Trevor Jones and Beth Sandore, both from the University Library at the University of Illinois at Urbana-Champaign, published an article on the state of digitization training in Illinois (Jones and Sandore, 2002). They had discovered through surveying Illinois cultural heritage institutions of all sizes that, although 80 percent of cultural heritage institutions in Illinois had digitization equipment, only 15 percent of them had sent staff to digitization training. Less than one-third of the institutions utilized any type of metadata for projects they had completed. These numbers alone proved it was paramount to provide proper digitization training to cultural heritage institutions in Illinois. If one extrapolates the data to other states, it is apparent the need for digitization training extends well beyond the borders of Illinois.



Other research into digitization in libraries and museums had also been done through the Institute of Museum and Library Services (IMLS – available at: www.imls.gov). This extensive report encompasses the national cultural heritage community, not just Illinois; and surveyed museums, public libraries, academic libraries, and State Library Administrative Agencies (SLAAs) to assess their use of technology and the roll digitization plays in their institutions (Institute of Museum and Library Services, 2002).

The report's findings show 78 percent of SLAAs, 34 percent of academic libraries, 25 percent of public libraries, and 32 percent of museums reported digitization activities in the year prior to completing the survey, and all groups stated they were planning digitization projects in the next 12 months and beyond. One of the most notable responses in the IMLS report was what institutions reported as common hindrances to their current or future digitization plans. Although all types of institutions cited "lack of funds to support digitization" as their number one response (which comes as no surprise), "lack of available expertise" was also often mentioned. This is supporting evidence that digitization training can be helpful to institutions across the cultural heritage spectrum. It can also be pointed out that none of the institutions with current or future digitization plans said they "do not see a role for our institution in digitization". So one can see digitization is something institutions are doing and want to continue doing, but need help in learning the process.

The need for digitization education is certainly clear; however, it is futile to create training opportunities if they are not marketable to cultural heritage institutions. To that end, Jones and Sandore (2002) asked survey respondents to indicate in what type of digitization training they would be willing to participate. Over half of respondents (56 percent) stated they would be interested in one-day training, and 19 percent expressed interest in web-based tutorials. These numbers suggest staff at cultural heritage institutions want training which can easily fit into busy schedules and does not require participants to spend too much time away from their other work activities.

Jones and Sandore next looked at the type of digitization training currently offered. At the time there were few options available for people wanting digitization training, and, although those that existed were all nationally-acclaimed, they did not meet the particular needs of potential digitization projects in Illinois. Cornell University offered their Moving Theory into Practice web site (www.library.cornell.edu/preservation/tutorial/) as a free, self-guided tutorial. Cornell also offered a week-long Moving Theory into Practice workshop which took place on the Cornell campus. There is also the Northeast Document Conservation Center's (NEDCC) School for Scanning (www.nedcc.org/), an intensive three-day workshop given at various locations around the country, as well as the Archive Builders' Document Imaging and Document Management workshops in California (www.archivebuilders.com), and Digitization for Cultural Heritage Professionals carried out at Rice University and the University of North Carolina-Chapel Hill (www.ils.unc.edu/DCHP/). The problem with all of these digitization workshops (at least for institutions in Illinois) is they were either self-guided or non-interactive, or they required the time and expense of traveling to a three-days-or-more workshop. One day workshops were given around the country, but they tended to be directed toward regional audiences and not always readily available to participants in Illinois. Obviously, since survey respondents indicated they would like local one-day or web-based training, few suitable options were available to them.

Armed with their findings, Sandore and colleagues applied for, and received, an Institute of Museums and Library Services National Leadership Grant. The grant-funded project sought to provide different digitization training opportunities to cultural heritage institutions in Illinois and, for that matter, to reach beyond the state lines and provide training opportunities for organizations across the country. And with that, the “Basics and Beyond” training program was born.

The idea for *Basics and Beyond* is to give cultural heritage institutions a choice of different digitization training opportunities at different price points and with different time commitments. The training is broken up into three different “tracks” based on the type of training survey respondents said most interested them. Track 1 provides a one-day basic digitization workshop and Tracks 2 and 3 are web-based courses. Track 2 is a three-week on-line course significantly expanding on the information provided in the Track 1 course. Track 3 combines the three week on-line course with an intensive two-day workshop held at the University of Illinois at Urbana-Champaign (UIUC) campus.

Current Basics and Beyond training expands on the work of Trevor Jones and the Illinois Digitization Institute (<http://images.library.uiuc.edu/projects/IDI>), which is part of the Digital Services and Development Unit (<http://images.library.uiuc.edu>) at the University of Illinois Library. Jones originally created a two-week on-line course and several one-day workshops on digitization. Working in conjunction with the Illinois State Library (www.cyberdriveillinois.com/departments/library/home.html), this on-line course was given to recipients of LSTA digitization grants.

The one-day workshops were basic digitization presentations originally designed for staff at institutions such as the UIUC Library and the Indiana Historical SocietyHisHishhh. In June 2003, Amy Lynn Maroso was hired as the project coordinator of the Illinois Digitization Institute’s Basics and Beyond program and took over from Trevor Jones, who chose to return to the museum community. Maroso built on the work already done, expanding the two-week course into a three-week course, creating the content for and presenting the workshops and on-line courses.

A look at the three different training methods

As previously stated, the Basics and Beyond program is comprised of three different training tracks. This approach was specifically designed to give potential participants a variety of options to meet their needs, as digitization knowledge varies widely in the cultural heritage sector and the survey respondents were looking for different types of training. Some institutional participants are seeking very basic fact-finding information whereas those from other institutions might already be starting a digitization project, have funding to start a project, or have completed projects in the past with no training and want to learn the theory behind what they have already practiced. Some institutions do not have the time or budget to allow staff to participate in anything more than a low-cost, one-day workshop, while other organizations can invest more finances and time to take an extended training course.

Track 1

This is the most basic of the training options. Geared specifically to institutions in Illinois and neighboring states, these one-day workshops are given at various locations around the state for a total of 14 workshops during 2003-2005. They run from roughly

9.30 am-3.00 pm at a cost of \$45.00 per person. Based on experience and knowledge from the partnerships with the Illinois State Library (ISL) and the Illinois Heritage Association (IHA) (<http://illinoisheritage.prairienet.org/>), the locations of the workshops reflect where the greatest concentrations of cultural heritage institutions are located within the state. This way, a central location is provided each time to make travel to a workshop easy for those in the area. Announcements about the workshops are sent through ISL e-mail announcements and through flyers distributed to all libraries and museums in the geographical region around each workshop site. Registration is available on-line through the Illinois Digitization Institute's web site.

Each workshop begins with a small pre-workshop quiz which asks attendees to answer several multiple choice and true/false questions on specific digitization topics. The questions cover the information to be presented in the workshop. The same questions are also asked in a post-workshop quiz. Results of the two quizzes can be compared to see what participants already knew and what they learned during the course of the presentations.

The format of the workshop involves two separate presentations. Due to the size of the audiences (averaging 20 participants per workshop) and the amount of material to be covered during the day, there is no time to provide attendees with hands-on opportunities. It would also be difficult to provide hands-on training at most, if not all, of the locations where the workshops are held. Workshops are held in conference rooms at public libraries or museums, which might have access to only one or two pieces of digitization equipment, if any at all.

The main presentation covers seven different topics, including:

- (1) how digitization can work as part of an institution's pre-existing preservation program and what the benefits and drawbacks to digitization are;
- (2) questions institutions must consider when planning a digitization project;
- (3) an overview of what digital images actually are, including a discussion of resolution, bit depth, dynamic range, and their importance to image quality;
- (4) hardware and software needs for projects and what type of equipment is best suited to capture different types of materials;
- (5) current standards and best practices to follow when creating digital images;
- (6) how to store digital images and issues to consider when creating internet access to them; and
- (7) what metadata is and why it is important to create metadata for digital images.

Alyce Scott, Digital Imaging Program Coordinator at the ISL gives a presentation on how Illinois cultural heritage institutions can make use of the Illinois State Library's Illinois Digital Archives (IDA) (<http://eli.sls.lib.il.us/ida/>) to present their images to researchers. IDA hosts access images on Illinois State Library servers and is a searchable database of images about the history of Illinois. Images come to the digital archives from LSTA-funded projects, the Illinois State Library itself, and other Illinois libraries and cultural institutions that might have done a digitization project.

At the conclusion of the day, the floor is open to any questions or comments, followed by the post-workshop quiz and evaluations. A list of additional resources such as books, journal articles, and web sites covering a range of digitization topics (i.e. general project planning guides, scanner review sites, metadata resources, and

digitization help sites) is included in an information packet to help attendees find reliable sources to answer future digitization questions. Some of the main sources recommended in this resource list include: *Handbook for Digital Projects: A Management Tool for Preservation and Access* by the NEDCC, *Moving Theory Into Practice: Digital Imaging for Libraries and Archives* by Anne R. Kenney and Oya Y. Rieger, *Why Digitize?* by Abby Smith, and *Guides to Quality in Visual Resource Imaging* by RLG.

Track 2

Track 2 is a three-week asynchronous on-line course which can be accessed by anyone with an internet connection and web browser. The cost of this course is \$200.00 per person. A total of eight Track 2 courses will be given in 2004-2005 and currently four have already been completed. This type of course was designed in response to Jones' and Sandore's (2002) survey results stating nearly 20 percent of respondents would like to have the option of on-line digitization training. Although self-paced web-based tutorials are already available (e.g. Cornell's *Moving Theory into Practice* and *Creating and Managing Digital Content* by the Canadian Heritage Information Network, www.chin.gc.ca/English/Digital_Content/index.html), Track 2 is designed to be more interactive and structured and most importantly, it is instructor-led.

The course is disseminated using WebCT software (www.webct.com) and is accessed through a secure server on the UIUC campus. WebCT is a tool allowing for the creation of web-based educational environments accessible by anyone with an internet connection. All student interaction with the course content, the instructor, and other students is through a secure web browser. The structure, content, and educational tools used within the course are all configurable by the instructor, and content can be automatically released at given times during the day. Students have access to a live chat feature, an electronic bulletin board, on-line quizzes, and other learning tools.

Track 2 is designed to take the one-day workshops farther while still allowing participants to fit "attendance" requirements into their schedules. Each course runs Monday through Friday for three consecutive weeks. Each day a new section of content is released to the web site. Daily content includes the reading for the day and any associated homework assignment. Students are given an ID and password to access the web site and are expected to log in every day of the course. However, when and how many times per-day they do so is up to each person. By making the course asynchronous, students are given the flexibility to schedule their course time to suit their needs. This also allows people from around the world to participate, since what may be a convenient time to have class in India isn't necessarily a convenient time in Illinois.

Due to the wide geographical audience Track 2 potentially serves, announcements about it are distributed via electronic bulletin boards and listservs for specific cultural heritage groups, such as Cons DistList, Digiline, DigiStates, IMAGELIB, Museum-L, and others. The targeted announcements also tend to be redistributed via those lists' members to additional groups. After sending out the course announcements, inquiries for further information have been received from disparate locations such as Finland, Brazil, India, Wales, Estonia, and Australia.

The course includes 15 days of instruction which closely match the overall topics of Track 1, but considerably expand on them. Track 2 also adds a basic tutorial on digitization-specific Adobe® Photoshop® techniques. The goal of Track 2 is not only to expose participants to the basics of digitization, but to make them feel comfortable enough with the issues involved so they will take the next step and try a small digitization project, or at least plan what items they will digitize and how they would manage a project. The daily Track 2 content schedule is:

- *Day 1:* Learning how to use the course software and how to navigate the course.
- *Day 2:* An overview of digitization and the problems and benefits associated with it that institutions must consider before starting a project.
- *Day 3:* The basics of what digital images are, including discussion on resolution, bit depth, and dynamic range, and how these concepts relate to creating good digital reproductions of analog material.
- *Days 4 and 5:* Planning sustainable digital projects.
- *Day 6:* How to select appropriate materials for digitization.
- *Day 7:* Selecting appropriate equipment for digitization projects based on the items an institution plans to digitize.
- *Days 8 and 9:* Standards and best practices when creating digital images including a look at recommended resolutions and bit depths for various types of items such as text, manuscripts, and photographs, and the challenges associated with digitizing these materials.
- *Day 10:* Best practices in digital conversion and why quality control is so important to creating a good digital product.
- *Days 11 and 12:* Working with Photoshop and options for storing the completed digital images.
- *Day 13:* The need for metadata, including information on current popular metadata schemes, and a closer look at using Dublin Core to make images accessible and provide users with information about images.
- *Day 14:* Delivering images to an audience and image management system options that institutions need to consider when deciding how they will make their images accessible.
- *Day 15:* Issues about digital longevity and course wrap-up.

All the main course content was written by the course instructor; however, each day's reading includes myriad links to other web sites and readings for further information and study. Although many of the topics look very similar to those covered in Track 1, this three-week course moves beyond a presentation-by-speaker format that is completed in a only few hours into a much more in-depth and interactive environment, allowing participants to pose more questions and take a deeper look at each step of a digitization project – from why we digitize to how to deliver images to patrons.

As in the Track 1 workshops, a beginning survey is given to assess the participants' level of knowledge about digitization and how comfortable they are with digitization projects and concepts. Throughout the course students are required to complete several quizzes on course content and these quiz results can be directly compared to some of

the questions asked in the beginning evaluation to see if students are learning key concepts.

Along with the content provided on the web site, each lesson requires participants to do an additional required reading accessed via another site. There are extra suggested readings included which further explain or demonstrate the presented concepts or give examples of digitization projects and outcomes. Students are also required to do assignments which range from short, multiple-choice quizzes to choosing equipment for their potential digitization project based on a small, imaginary budget of US\$1,200.

One of the most important features of the course is the electronic bulletin board. The bulletin board has been very useful in making the course feel more like a cohesive community rather than an incongruent group of scattered individuals. For example, the first day of class students post a brief biography to the class bulletin board, which allows the other participants to learn a little more about their fellow classmates. Some assignments are required to be posted to the bulletin board too. This gives everyone in the course a chance to read the posts and comment on them. There is also a special section of the bulletin board allowing people to post non-assignment related questions, comments, and interesting articles, and this "general discussion and questions" topic receives quite a bit of use.

Track 3

This is the third training option offered by the IDI. Track 3 goes beyond the Track 2 course by providing an on-line portion plus a hands-on workshop allowing participants to actually use equipment to digitize materials. The cost of the course is \$300.00 per person plus travel expenses to the workshop. Four of these courses will be given during 2004-2005, with advertising conducted in the same way as the Track 2 courses. Notice of the course given in July 2004 was also placed in the UIUC summer course schedule.

The balance of class size can be difficult for Track 3. On one hand, a class so small that students do not have enough people with whom to interact and learn from is not beneficial; on the other hand, a very large class does not allow for individual instructor attention for each student. The enrollment for Track 3 is kept to a maximum of 15 people. The July 2004 course had 13 participants. This seems to be the ideal number of students. Not only do they receive individualized attention during the workshop and the on-line section, but the group is still large enough that participants have contact with a greater selection of people from diverse institutions.

This track was designed to allow participants to not only learn the theory behind digitization projects, but to also give them the opportunity to take their training further by putting into practice what they learn through the on-line course. Track 3 students first take a three-week on-line course (covering the same material as Track 2). They then travel to the UIUC campus to participate in an intensive two-day workshop which allows them to use scanners, digital cameras, and other equipment to digitize items. Participants have the option of bringing a few items from their own collection with which to work, or various materials (e.g. bound items, manuscripts, three-dimensional objects, etc.) can be provided for them at the workshop. They create metadata for their items and then upload the images into a CONTENTdm™ (www.contentdm.com) database. Use of Adobe Photoshop Elements 2.0 is covered in a hands-on setting with

participants using the software to do image manipulation such as cropping, rotating, size and resolution adjustment, and batch processing. They are also able to discuss other aspects of digitization, as well as learn about current digitization practices and other digital projects from guest speakers.

In a hands-on setting much more attention can be paid to the practice of image conversion rather than the theory, so this learning opportunity allows students to feel comfortable with using equipment and software, giving them a starting point for their own digitization projects. Track 3 participants tend to be from cultural heritage institutions currently planning a digitization project or actively seeking funding for one, and both the on-line and the workshop information will be directly applicable to them quite soon after completion of the course.

Findings from the training

Findings from completed Track 1 courses

As previously stated, all Track 1 participants are given a brief pre- and post-workshop quiz. Digitization-specific questions are asked, as are questions about what respondents' current level of digitization knowledge is. Of the workshops given to date, 83 percent of participants say they have "none" or "a small amount" digitization knowledge (Figure 1), but 74 percent plan to start a digitization project soon and 22 percent may start one in the near future. Of the 22 percent who were "maybes", the reason for this response was most often due to a lack of funding to carry out a project rather than simply not wanting to do one. As one can see, these workshops are indeed targeting the people who need them.

The question then becomes, "Did the training help?" In the case of the pre- and post-workshops quiz scores, the answer is a resounding, "Yes!" The number of questions missed on the pre- and post-quiz scores is shown in Figure 2.

Once tabulated for all the workshops the data show that, of the same nine questions given on both quizzes, 20 percent of students missed one or two questions on the pre-quiz (none of the participants answered all pre-quiz questions correctly) and 80 percent missed three or more questions. However, on the post-quiz, 86 percent of students missed zero to two questions and only 14 percent missed more than two questions, with no one missing over five post-quiz questions.

On the post-quiz participants were also asked what the importance of metadata is and if they planned to do their project in-house or outsource it and why. All

Track 1: Pre-Workshop Digitization Knowledge

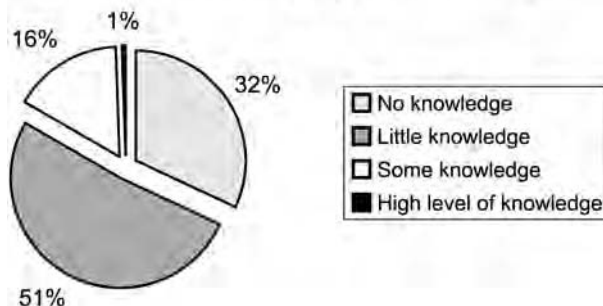


Figure 1.
Digitization knowledge
level of Track 1
participants before
workshops

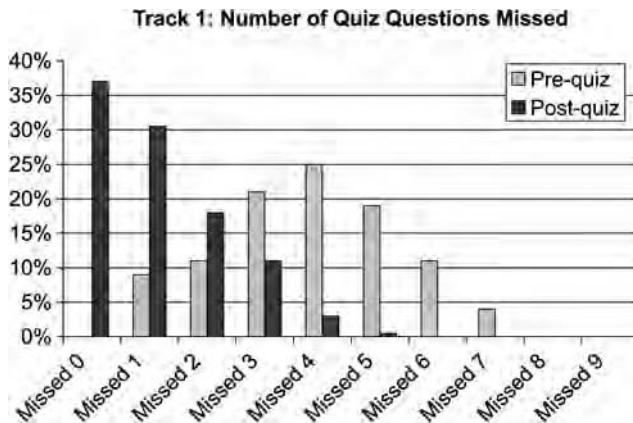


Figure 2.
Number of questions missed by participants in pre- and post-quizzes during Track 1 workshops

participants had a very good grasp of the need for metadata and what its uses are. Responses given to the metadata question included: “To describe the images and help locate them”; “To allow users to learn more about what they’re seeing”; “Increased accessibility”; “To identify and know at what resolution, when, on what equipment, etc. the images were created”; and “ACCESS for both staff and patrons. Metadata also gives you the ability to organize and know what’s available to you.”

Participants were also able to clearly state reasons for wanting to either do their project themselves or to have a vendor do it for them. Most of the participants planning in-house projects stated it is because they want more control over the project and so they will have the skills and equipment for future use. Many are also anxious about sending materials off-site to vendors due to shipping and handling concerns. For those who stated they plan to have a vendor do their digitization, the main reasons were a lack of staff and/or time to do the project or because they cannot afford to purchase the necessary equipment.

The quiz responses show that participants in the one-day workshops learned much from the presentations and were able to assess their own digitization goals and needs and start planning a project for the future. Since the Track 1 workshops are one-day-only, presentation-format sessions, not every aspect of digitization can be covered. As previously stated, the Track 1 workshops give participants broad and basic knowledge of key issues surrounding the digitization process; they are not meant to be a detailed exploration of the field. Since this is the case, there are bound to be questions which arise during the course of the workshop that simply, due to time constraints, cannot be covered. Many participants have questions such as what type of scanner they should buy or what type of digital object management software might be best for their institution. Many have questions about using optical character recognition (OCR) software for text items. Some ask about how to procure grant funding through LSTA grants or other funding means. And many ask about specific costs for projects. All of these questions are particular to individual projects and cannot be covered during the course of the workshop. Participants are encouraged to either contact the workshop instructors after the course or do some research on their own. The extra resources handout provides alternative sources of information for further study.

Findings from completed Track 2 courses: background on the participants

As of July 2004, four on-line courses have been given from which to glean information about the students' digitization needs and what they learned from the course. As with the Track 1 workshops, Track 2 students are asked survey questions prior to the readings to determine their familiarity with digitization. With more time to spend on testing, the pre-course and post-course surveys for Track 2 are much more in-depth than those for the Track 1 workshops. Course participants are asked not only about their current level of digitization knowledge and if they have any projects planned, but also what equipment they might already own, the size of any planned project (how many items they plan to digitize), and what types of items they plan to digitize. Results show many places already own quite a bit of equipment (five people responded that their institutions even own planetary scanners!) and many have sizeable projects they would like to start. However, it is important to note that many participants from the Track 2 courses are from larger organizations and universities and, therefore, have bigger budgets and greater access to equipment. Most Track 1 participants are from small to mid-sized museums, public libraries, and historical societies and so their operating budgets often cannot compare to that of large state agencies or colleges/universities.

As one can tell from Figure 3, there are more participants in the Track 2 courses who have worked on small digitization projects in the past versus the Track 1 workshop participants. Still, 61 percent had no, or only a small amount of, theoretical knowledge about the process, and only 1 percent of respondents had already worked on several projects and felt comfortable with the digitization process.

Once again, it is interesting to see how many of the respondents want to start projects even though they do not have much knowledge of the subject, 73 percent have a project currently planned and nearly 40 percent plan on digitizing 1,000 or more items (Figure 4).

When asked about equipment, only two people stated their institution owned no equipment of any kind. The overwhelming majority owns a flatbed scanner (46 total) or digital camera (37 total) and many institutions own both. Several institutions own even more specialized equipment such as film/slide scanners, microfilm scanners, or planetary scanners (Figure 5).

Not only do many institutions already have equipment with which to digitize, the surveys show the collections people want to digitize are quite varied. Although the

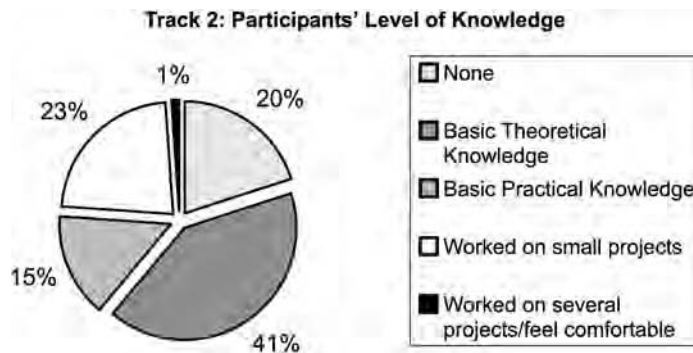


Figure 3.
Digitization knowledge level of Track 2 participants before start of course

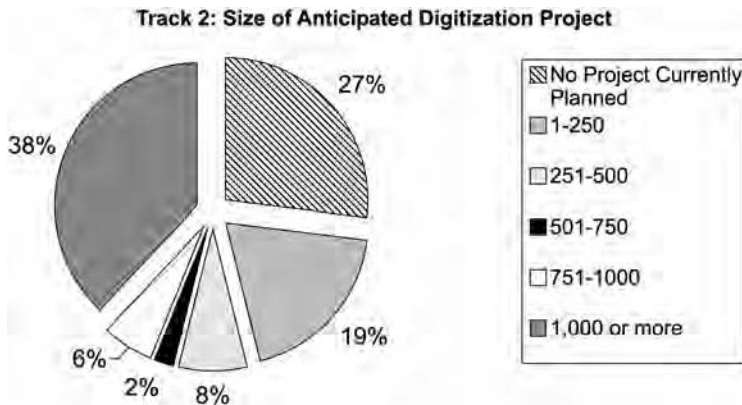


Figure 4. Size of participants' anticipated digitization projects by number of items they plan to digitize

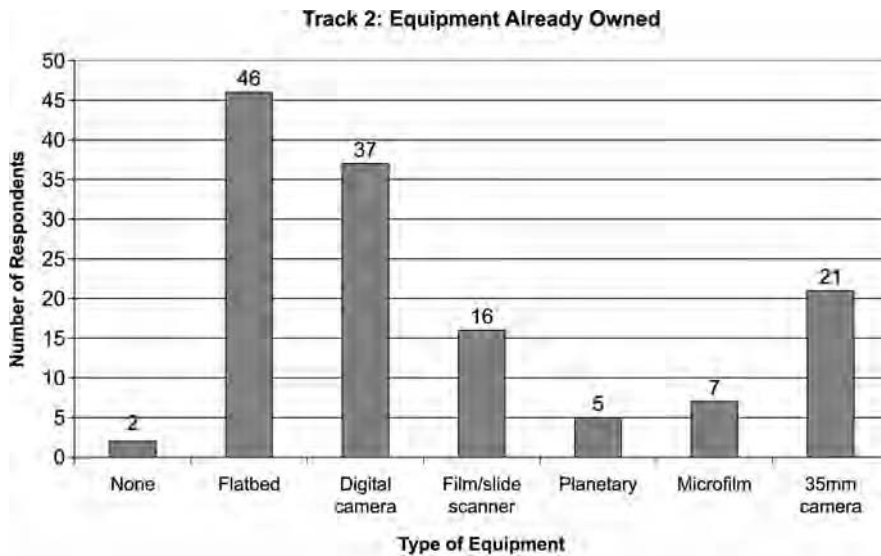


Figure 5. Type of equipment already owned by Track 2 participants' institutions

most common items are text, photos, slides/film, and manuscripts (Figure 6), there is also a sizable number who want to tackle audio and video as well (16 and 12 respondents, respectively).

What participants learned. In addition to asking about the Track 2 students' digitization backgrounds and projects they have planned, they were also asked several questions to determine what their level of digitization knowledge was before the course commenced. These questions were a sampling of future quiz questions that students would be completing later in the course after reading specific topics. By comparing the beginning survey answers with the scores of the later quizzes, one can tell if the participants learned the key points stressed in the course.

The number of Track 2 students indicating they had some knowledge of or experience with digitization projects was greater than in the Track 1 workshops. This

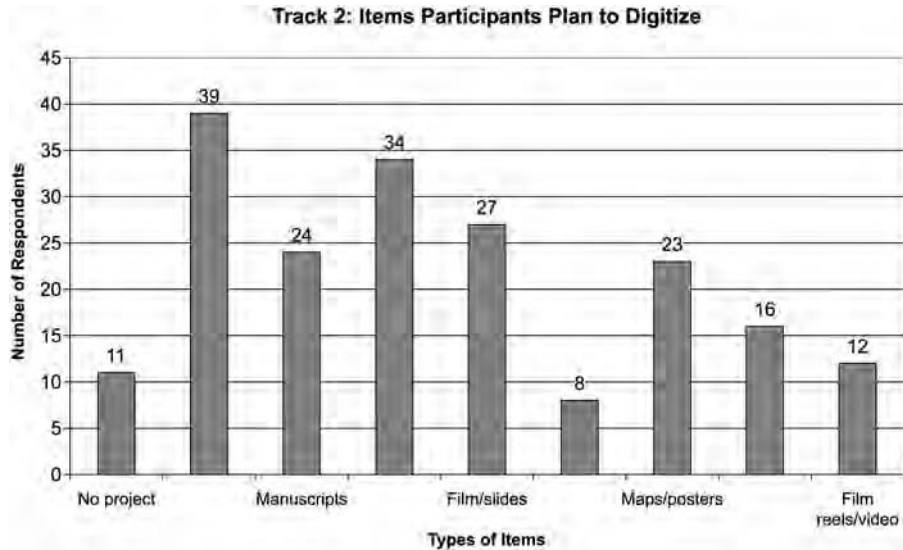


Figure 6.
Types of items Track 2
participants plan to
digitize

should indicate that Track 2 respondents would fare better on a pre-course quiz than Track 1 respondents, and the data correlate with this assumption. Of the ten quiz questions on the beginning survey, Track 2 respondents answered correctly 59 percent of the time.

However, no matter how well respondents did on the pre-quiz, the larger question is whether the scores for the quizzes they took after doing assigned readings were higher? Unlike the Track 1 participants who did not have the opportunity to review materials before taking the quiz, the Track 2 participants could go back to the readings as many times as they wished before taking the quizzes or even while they were taking the quizzes (*à la* an “open book” exam). There are four topic-specific quizzes during the on-line course corresponding to the following course sections: digital image basics, standards and best practices for digital projects, image capture and quality control, and Photoshop techniques and image storage. Quizzes are all completed and graded on-line and include a variety of different types of questions from true/false to short answer. Once the quizzes have been graded, students can go back and review questions they missed to further reinforce their understanding of the course material.

Looking at the aggregate results for each quiz, it is obvious the Track 2 students did extremely well overall. As shown in Figure 7, the average quiz scores for each course were normally well above 90 percent, and many reached or exceeded scores of over 95 percent per specific quiz. Even though the on-line students knew more about digitization before they started the course, these data prove they did much better on the quizzes when presented with an opportunity to review readings before taking each quiz. It also shows the additional information given in the longer, detailed, three-week on-line course is beneficial to the students’ understanding of the topics.

Going beyond quiz scores. While quiz scores are an important measure of how well participants did in the course and how much information they learned about the given topics, standardized score results do not always tell the whole learning story. The *Basics and Beyond* Track 2 digitization course is unique in the fact that although

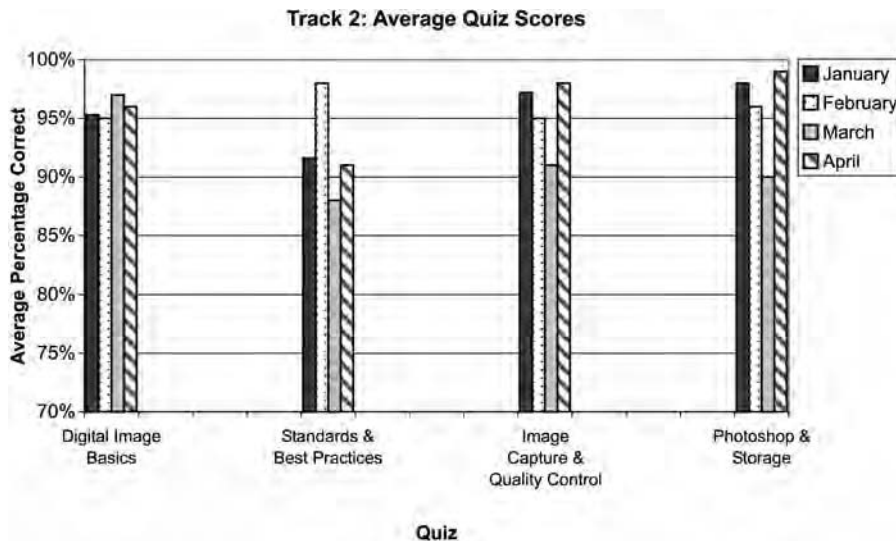


Figure 7.
Average quiz scores for all
Track 2 courses on each
quiz

participants are not physically in the same location or even doing readings at the same time of day, there is still an opportunity to interact and pose questions and discussion topics to the instructor and class and have people respond to questions and comments.

The WebCT electronic bulletin board feature allows both the students and the instructor to communicate with one another and the rest of the course participants. Several Track 2 assignments require students to post their completed work to the class bulletin board and the instructor encourages everyone to comment on the work of others. By requiring three of the first four assignments to be posted the bulletin board, participants are encouraged to share ideas, comments, and questions with one another from the very beginning rather than e-mailing questions and comments to only the instructor.

The first assignment requires participants to post a brief biography so everyone can learn a little about everyone else on both a professional and personal level. Assignments posted to the bulletin board also include what students feel are some of the benefits and drawbacks of digitization and what's known as the "Reality Check" assignment – which asks student to post how they are feeling about their own digitization goals after reading the course information on project planning. Student also post Dublin Core-based metadata they create for a specific item and the equipment purchases they make for an assignment called "Go Shopping," where each participant has a small, fictional budget of US\$1,200 to spend on equipment they will need for their planned project.

But the class bulletin board goes beyond simply being a convenient vehicle for submitting and commenting on homework. It also allows course participants to pose other questions and delve into topics that might not otherwise fit into or may be too specific to include in general course materials. There is a section of the bulletin board for general discussion and questions where anyone is allowed to post questions and comments about digitization or even questions and comments not related to the subject at all. In fact, the April 2004 course had so many gourmards and wine aficionados in it

a new topic was added to the bulletin board called “Culinary Corner” for people to post recipes and favorite wines!

This ability to go beyond just submitting homework assignments creates a much more dynamic environment. For instance, when a new posting to a digitization-specific listserv or web site is made available that might apply to a group’s or even a particular student’s digitization project, that information can be posted to the class bulletin board and discussed. Or participants can pose questions about specific products and materials such as “Is anyone else scanning maps?” or “I’m considering CONTENTdm™ for my image management system. Has anyone used it before and can tell me about it?” or “Can anyone recommend a vendor in my area?” Articles related to digitization (even if only tangentially) can also be posted and commented on and used to point out specific instances where digitization was used to benefit a certain institution or collection. The latest articles about advances in digitization-related industries, such as Blu-ray Disc technology (www.blu-ray.com), might not warrant discussion within the readings but may still be of interest to some people. This type of information can be posted so those who would like to read it can, without making it required reading for everyone. Likewise, upcoming digitization-related conferences and other meetings can be advertised by course members.

The class bulletin board works for digitizers of all skill levels since any type of question, from the most basic to the most advanced, can be posted for the instructor’s and other students’ comments. In one particular course, a student posted a question about how to see if her institution’s scanner would work with their current computer operating system. She didn’t have the scanner’s manual, and since the scanner was an older model she was concerned there would be compatibility problems. The instructor found a PDF version of the appropriate owner’s manual on-line and the problem was solved. In another course, there was a participant involved in a digitization team and her institution could not decide on which image management database to use. So she posted the question to see what other people are using. Several suggestions were proffered along with offers for help when she needed to set up the database.

Adding to the asynchronous interaction that was accomplished with the electronic bulletin board, Track 2 students also have the option of participating in a live chat session called “Ask the Experts”. This live forum is conducted through the WebCT software and brings the course participants together with experts in the field of digitization from UIUC, the Illinois State Library, the Illinois Heritage Association, and other locations. The live chat is held for at least one hour on the last day of each course and allows students to log in and pose questions in a real-time forum. Questions in the live chat are just as varied as questions posed on the electronic bulletin board. Many people want to discuss specific software or ask about their own projects and how best to go about starting them (e.g. whether it would be more cost-effective to outsource microfilm digitization or to purchase a microfilm scanner, or what type of scanner would work best for rare materials). Since all of these questions can be answered in real time, the panel of experts has the opportunity to easily follow up on questions and explore topics in a non-static environment.

Findings from completed Track 3 training

One Track 3 course has been given as of July 2004. The course had 13 students. Results of the on-line portion of the course were very similar to Track 2 results, with many of

the participants having some knowledge of digitization before they started the course. All participants had projects in mind and wanted to take the next step and learn about the process from both the theoretical and practical sides.

After completing the on-line portion of the course, the students came to the UIUC campus for the two-day workshop. The following schedule was set for the two-day workshop:

(1) *Day 1:*

- welcome and discussion of everyone's digital projects;
- presentation on preservation/conservation in digital projects;
- using 10x scale loupes to find the smallest significant character in an item to determine scanning resolution;
- hands-on use of scanner, digital camera, and large-format scanner to digitize items;
- demonstration of planetary scanner; and
- hands-on image manipulation with Adobe Photoshop Elements 2.0.

(2) *Day 2:*

- presentation on LSTA grant-funded map digitization project;
- presentation about the Open Archives Initiative;
- presentation on XML;
- hands-on creation of metadata using the Dublin Core schema;
- hands-on entering of images and metadata into CONTENTdm image management database; and
- evaluations and wrap-up.

Part of the intention of the two-day workshop was to have participants create a mini-collection from start to finish, using what they learned in the on-line portion of the course to create a good final digital project. Students were also exposed to different perspectives on digitization (such as preservation and its relation to the digitization process) and different concepts (the Open Archives Initiative and XML) they might come across when researching and completing their own projects.

An informal focus group was conducted during lunch on the second day of the workshop and evaluations were handed out at the end of the workshop. The responses from both will be used to help structure future Track 3 workshops. Many participants responded that they were glad to have had the opportunity to work hands-on with Photoshop Elements and CONTENTdm, since these are software packages they might be purchasing in the future. Many also thought it was very informative to hear about other institutions' digitization projects and to have an opportunity to meet other people new to digitization.

Two notable comments from the focus group were that participants wanted the class bulletin board (from the on-line section of the course) and the on-line community of new digitizers to remain active after completion of the course. Couple those statements with participants' remarks that they were glad to have an opportunity to meet with people also starting digitization projects, and it is evident there is a desire for new digitizers to communicate with one another. This is being accomplished through

listservs such as IMAGELIB and DigiStates, among others, but a community of people new to digitization might also be advantageous. Those less familiar with the subject many times perceive the established digitization forums as “master classes” for the well-informed, and that might be a bit intimidating. An additional forum would allow people new to the field to discuss projects, planning, and other topics in a community of their peers without fear of asking basic questions.

In addition to what was recorded from the focus groups, students were also asked how much of what they learned would be applied to their digitization projects. Of the participants, 75 percent said they would use 100 – 75 percent of what they learned from the course, and 25 percent said they would use 74 – 50 percent of what they learned (Figure 8). From these figures it is apparent the information provided in the on-line course and the workshop is useful to those participating. It is anticipated that future Track 3 courses will net the same positive results.

What’s next for digitization training?

From these data one can see the need for digitization training and that many institutions are ready to plan and create digital projects but still have the desire to learn more about the process before they start.

Conclusions

There is no doubt the need for effective training is out there, and is necessary around the country and around the world. There are many very good digitization education programs offered in a variety of formats, but now the question has to arise of “What else can we do to prepare institutions to successfully carry out these projects?” There are several steps that should be taken to further improve digitization training and to reach even wider audiences.

First, the training needs to go beyond just targeting digital imaging (i.e. the conversion of materials like books, manuscripts, photographs, and slides) and provide specific training for the conversion of audio and video. Most of the digitization information currently available (the Illinois Digitization Institute’s included) is heavily skewed to print and photographic materials. Even the general guides and tutorials for planning and implementing digitization projects, such as *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*

Track 3: Amount of information that will be applied

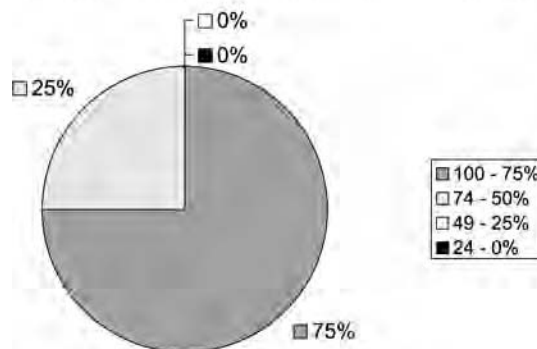


Figure 8.
Amount of information learned from Track 3 that respondents will apply in their digitization projects

(www.nyu.edu/its/humanities/ninchguide/), the NEDCC's *Handbook for Digital Projects: A Management Tool for Preservation and Access*, or Cornell's *Moving Theory into Practice* do not cover audio and video conversion as deeply as they cover converting images. However, a significant number of Track 2 students stated they want to convert audio (16 respondents) and video (12 respondents). As more people enter the digitization arena more will want to convert items such as oral histories stored on magnetic media or Betamax tapes. One must also consider that end-user demand for digital audio and video is also increasing at a rapid pace. Who knows how much longer it will be before DVDs make VHS tapes completely obsolete?

It is also time to broaden the discussion of digital longevity. Large institutions are well aware of the short shelf life of digital objects, but making sure the “up-and-coming” group of digitizers – small- to-mid-sized institutions – is alert to this future challenge is paramount. And not only is it vital to make them aware of the reality of obsolescence and what this means to the cost and permanence of their digital projects, but to help them plan for it. Through the workshops and on-line courses we see time and time again that smaller institutions are not considering the up-keep of their digital images, and that the cost of upkeep must figure into their permanent budgets to keep viable information they worked so hard to create. They may not yet realize that digital objects aren't like books and “digital documents last forever – or five years, whichever comes first” (Rothenberg, 1997). Fortunately, Cornell has received grant funding from the National Endowment for the Humanities to carry out a new series of one-week workshops and an on-line tutorial called *Digital Preservation Management: Implementing Short-term Strategies for Long-term Solutions* (www.library.cornell.edu/iris/tutorial/dpm/). The NEDCC also covers the topic of digital longevity in its three-day School for Scanning workshop, and other digitization guides touch on the subject as well.

Finally, there is a need for permanent funding for national digitization training or feasible profit-making training models which encompass a range of different training options. Many of the different programs available now are funded through grants, and we all know that grant money eventually runs out. As is the case of Basics and Beyond, which has a set number of workshops and courses it will give during a specific time period, grant-funded programs carry out only a finite number of training programs. On-going instruction opportunities are available from organizations such as the NEDCC and Archive Builders, however; the variety of different options currently offered is limited. Digitization is not going away any time soon and, in fact, it will only increase in popularity within the cultural heritage sector as more users demand material on-line. The quality (good or bad) of a digital project does not simply reflect on the institution which produced it; it reflects on the field as a whole. A concerted effort needs to be made to make different types of instruction available to all cultural heritage institutions – no matter how large or how small. This means offering on-going education in different locations, at different cost points, with the use of different technologies, and encompassing the different skill levels of participants. Most importantly, this type of training needs to be made consistently available; not just when funding permits it.

Considering the meteoric rise of digitization and technology as a whole over the past 15 years, we can only assume there is no end in sight to what we will be able to digitally provide our patrons. Users of our institutions are becoming more and more

computer-savvy, and we will soon have on us generations that have never known life without the internet. Keeping up with the demand for on-line access to our cultural heritage materials will only become more challenging. We are taking steps in the right direction to produce “master digitizers” with the current methods of training, such as the on-line courses and workshops given by the Illinois Digitization Institute, that provide institutions with the knowledge and resources they need to complete successful projects. Data show we are targeting the right people for training and that much is being learned by participants in the workshops and courses. Overall, it is believed the project has been, and will continue to be, very successful in its goal of offering digitization instruction in both traditional (workshop) and a novel (interactive on-line courses) formats. Hopefully, this type of training can be continued and further expanded on to encompass even more diverse methods and locations.

References

- Institute of Museum and Library Services (2002), *Status of Technology and Digitization in the Nation's Museums and Libraries*, Institute of Museum and Library Services, Washington, DC, available at: www.ims.gov/Reports/TechReports/2002Report.pdf (accessed 5 August 2004).
- Jones, T. and Sandore, B. (2002), “We don't know the first this about digitization: assessing the need for digitization training in Illinois” *RLG DigiNews*, Vol. 6 No. 1, available at: www.rlg.org/legacy/preserv/diginews/diginews6-1.html#feature1 (accessed 5 August 2004).
- Rothenberg, J. (1997), “Digital information lasts forever – or five years, whichever comes first”, *RAND Video V-079*, Leiden.



THEME ARTICLE

Empowerment
for digitization

Empowerment for digitization: lessons learned from The Making of Modern Michigan

205

Ruth Ann Jones

*Digital and Multimedia Center, Michigan State University Libraries,
East Lansing, Michigan, USA*

Received 27 August 2004
Revised 29 November 2004
Accepted 17 February 2005

Abstract

Purpose – The purpose of this paper is to examine the planning and implementation of a project (The Making of Modern Michigan) which was undertaken to develop a digital collection on Michigan history and to provide training, equipment and technical resources needed by staff in small and medium-sized Michigan libraries to begin their first digitization projects.

Design/methodology/approach – Using a case-study approach, The Making of Modern Michigan is placed in the larger context of academic-library digitization efforts since the mid-1990s. The needs of the target audience are reviewed, and the services and technical resources provided to them are described with special attention to the innovative aspects of the project: tools for copyright status profiling and metadata entry.

Findings – The article discusses the advantages and disadvantages of The Making of Modern Michigan's decentralized approach to managing a collaborative project involving more than 50 institutions.

Practical implications – The library profession may benefit significantly when staff from small and medium-sized institutions are offered the opportunity to build their skills and participate in a collaborative project in a meaningful way; however, the time required to provide training, coordination and oversight is much greater than anticipated when The Making of Modern Michigan was originally planned.

Originality/value – As a collaborative project involving small libraries and focusing on materials of regional and local interest, The Making of Modern Michigan represents an emerging trend in the field of library digitization. The practical lessons reported here will benefit those planning or implementing similar projects.

Keywords Digital libraries, History, Project planning, Training, United States of America

Paper type Case study

It is a truism – but perhaps important to recall – that early digitization projects in North American libraries tended to originate in the largest, most well-funded institutions. Within its first 12 months of publication (July 1995 to June 1996), *D-Lib Magazine* announced such venerable projects as the Duke Papyrus Archive (www.dlib.org/dlib/november95/11clips.html#papyrus) the Berkeley Digital Library SunSITE (www.dlib.org/dlib/february96/ucb/02tenant.html) and the Alexandria Digital Library Project (www.dlib.org/dlib/march96/03contents.html). World-class collections of broadly significant primary documents in history, the arts and humanities attracted much of the attention and funding for early digitization efforts.



A decade later, smaller institutions are digitizing collections of regional and local interest as well. Key to these efforts is the diffusion of digitization experience from the largest libraries to those with fewer staff and resources. The Making of Modern Michigan (MMM) project was designed to facilitate this process among Michigan libraries.

The MMM was established in 2002 with funding from the Institute of Museum and Library Services. The proposal was sponsored by the Digitization Committee of ATLAS, the “Action Team for Library Advancement Statewide.” ATLAS was an initiative of the Library of Michigan, which contracted with the Michigan Library Consortium to develop a statewide information delivery service. Michigan State University (MSU) is the administrative home for MMM, with institutions throughout the state serving as partners.

MMM has two goals: empowerment and content development. The project seeks to provide the training, equipment, and technical resources needed by staff in small and medium-sized libraries to begin their first digitization projects. Taken together, their digitized materials will then form an online collection on Michigan history, which can be searched and browsed through a single web site.

Planning the project

In the course of developing the initial grant proposal, the ATLAS Digitization Committee surveyed libraries in Michigan to determine the level of interest in digitizing local history materials and what type of support library staff felt they needed to begin. Access to equipment and training in digitization techniques and standards emerged as two of the highest priorities. ATLAS members identified several other needs: training and assistance in dealing with copyright issues and with metadata creation; centralized, searchable access to the collection; and arrangements for long-term secure storage of the master copies of digital files. Unsurprisingly, given the tight budgets of most libraries, many indicated that they would be more likely to consider undertaking a digitization project if funding were available.

The survey also revealed that libraries were generally not enthusiastic about any approach that involved sending collections offsite to be digitized by another institution, for fear that delays would cause high-use material to be unavailable for long periods. And, it was recognized that any collection on Michigan history that excluded the twentieth century and the development of the auto industry would be incomplete at best. Therefore, copyright issues needed to be carefully addressed.

With these priorities in mind, the ATLAS Digitization Committee developed a plan for a statewide project with the following elements:

- Regional digitization centers in six locations throughout Michigan, putting one center within a two-hour drive for most libraries in the state. The regional centers provide participants in their area with training and access to both basic and specialized digitization equipment.
- Training in four key areas: project management, copyright issues related to digitization, standards for preservation-quality scanning, and creating metadata records in Dublin Core format.
- Assistance and support in requesting permissions from copyright holders, provided through the Copyright Permissions Unit of the MSU Libraries’ Digital and Multimedia Center.

- Software tools to be developed at MSU, which would facilitate the entry of metadata records in Dublin Core format by participating libraries.
- Long-term secure storage for the master copies of digital files to be provided by MSU and by Wayne State University (also a partner in the grant), both of which have elaborate systems in place to guarantee data security.
- Small incentive grants, intended to partly compensate for the staff time required to attend training and to plan and carry out a digitization project.

In several respects, the MMM plan follows the model of the Colorado Digitization Program (www.cdpheritage.org/), particularly the use of regional centers, provision of software tools for metadata, and the general emphasis on training staff at small libraries to plan and carry out their own digitization projects.

Other elements, such as the assistance with copyright issues, represent new developments. The MMM project has also investigated offering print-on-demand services to participants. Once a work has been scanned, shelf copies can be produced at relatively low cost, which saves some additional handling of fragile artifacts. Reproductions of historic titles could also be produced in small print runs for library gift shops and other fundraising purposes.

Putting the plan into action

MMM was funded by IMLS for a two-year period beginning October 1, 2002, and work began that fall. The various activities divided neatly into two categories: incorporating MMM into existing work processes where possible, and developing new work processes for everything else.

Incorporating MMM into existing work processes

Several significant aspects of the MMM plan fell into this first, easier category. These included the incentive grants, copyright assistance, and long-term storage.

Administration of the incentive grants was handled by the Library of Michigan (LoM) as its contribution to the project. Like other state libraries, LoM administers the distribution of LSTA (Library Services Technology Act) funds through a grant application process; the MMM incentive grants could thus be advertised through LoM's normal channels, and the disbursement of funds managed by LoM's accounting department. The LSTA staff at LoM were also responsible for assembling a peer review committee to evaluate the applications, and consulted with the rest of the MMM management team on appropriate criteria.

Copyright assistance for MMM participants takes two forms. Two members of the management team for MMM serve as the copyright specialists for their institutions and are able to provide information about whether materials fall into the public domain. For items needing permission, participants may ask for assistance from the MSU Libraries' Copyright Permissions Unit, which handles permissions for material used in the university's online courses. The staff in this unit have experience in locating successor companies, searching for out-of-town addresses, and similar tasks. In practice, the majority of MMM participants have preferred to contact copyright holders (often local residents) on their own, using standard permission letters provided on the MMM web site. However, the MSU copyright staff have been able to assist in a handful of cases.

Finally, long-term storage for digitized collections is handled by the MSU Libraries and the Wayne State University Library. Planning for storage needs is an ongoing task which both libraries engage in already. The increase in storage capacity made necessary by MMM is only a fraction of the increase needed to handle the expansion of our own digitized collections.

Development work

The more critical work on MMM, of course, is the development of new things: new forms of training, new workflows, new software tools. This developmental work has been concentrated in three areas: copyright training, managing site architecture issues, and metadata creation.

Copyright issues have been a major focus for MMM. Most large library digitization projects (including several of MSU's own efforts) purposely concentrate on works published in the United States before 1923, since these are in the public domain. Works published after 1923 are often simply excluded from consideration. Unpublished works have their own set of protections and thus may be seen as problematic for library digitization projects.

By contrast, the MMM project actively seeks to include works which reflect not only Michigan's historic involvement in logging, mining and fur trading – activities which characterized the eighteenth and nineteenth centuries – but also Michigan's central role in the development of the auto industry. In addition, the project recognizes that small public libraries have a strong interest in digitizing their rich local history collections, material for which their users may feel a personal connection. These collections often contain significant numbers of unpublished works.

Building a digital collection that includes copyrighted works imposes significant requirements for educating staff at the libraries that are contributing to the collection. MMM staff can assist with individual items, but most of the work has to be assumed by the participating library. And, although copyright issues are given significant attention in the day-long training sessions attended by all participants, two other methods were developed to reinforce the information. The first was to make copyright screening an integral part of the MMM "inventory" system; the second was to produce an inexpensive video to repeat the information covered in training.

Copyright screening system

The MMM inventory system was designed to address several needs. The first was to control the assignment of filenames to digital objects created for MMM. The second was to allow libraries to browse the proposed digital collections of other participants to prevent duplication of work. And, the third was to provide an easy method of determining the copyright status for works that MMM participants wish to digitize. All of these functions are handled through a password-protected web site, which serves as the interface with a MySQL database (Figure 1).

When a new item is entered in the MMM inventory, users are asked to supply basic identifying information. Then, the system asks a series of questions that determine whether the item is in the public domain. The questions are based on the summary of U.S. copyright law given in the document "When US Works Pass Into the Public Domain", (www.unc.edu/~unclng/public-d.htm) by Laura N. Gasaway, director of the law library and professor of law at the University of North Carolina.

THE MAKING OF MODERN MICHIGAN COLLECTION INVENTORY SYSTEM

Empowerment
for digitization

[My Collections](#) | [Info](#) | [Administrator](#) | [Browse All Participants' Collections](#) | [Metadata Entry](#) | [MMM Public Site](#) | [Logout](#)

You are logged in as:

Participant's Name: **Michigan State University** | Username: **JA**

MY COLLECTIONS [Add Collection]

[CONTACT](#) [REPORT](#)

Click a Collection below to view its Subcollections. You can also Add a Collection.

Collection Name	Collection Code	Number of SubCollections
1. Interviews with Michigan Supreme Court Justices	02	1
2. Michigan Cookbooks	03	1
3. The Speculum	01	1

Figure 1.
The Collection Inventory System allows MMM participants to register collections, subcollections, and items to be digitized, and to determine the copyright status of items proposed for digitization

MMM staff developed a flow chart which represents a reconceptualization of Gasaway's summary. Instead of a series of categories based on date of publication and publication status – five categories in all – the flow chart necessarily expresses the situation as a series of yes/no choices, the answers to which can be collected from the user through the interface (Figure 2).

The initial question is “Was the work published or created in the USA?” If no, the user is advised that the work may be protected, because its status will depend on the copyright laws of the country of origin. Given the Michigan focus of MMM, of course, virtually all items are of US origin (Figure 3). If yes, the user is then asked whether the work is published or unpublished. The answers given determine what questions will follow. They may include the format of the material (print or non-print), date of creation, and date of the author's death (Figure 4). At the end, the user is told either that the item is in the public domain, or that permission may be required. In most cases, “may be required” really means “required”, but there are a few rare exceptions that are deliberately ignored by the system so as not to slow the entry of information (Figure 5).

This copyright screening process has both advantages and disadvantages. By boiling down the complexities of copyright law to a series of yes/no questions, it allows MMM participants to get a reasonably accurate answer without knowing a great deal about the law. On the downside, it can be repetitive if a particular collection contains many items where the answers to the copyright questions are identical, such as a collection of photos from a single photographer. For future phases of the project, a “copyright bypass” may be added to the system.

The copyright screening system's greatest contribution may be the constant reinforcement it provides about the complexity of US copyright law. On the whole, MMM participants are very much aware that copyright restrictions exist, and they want to be in compliance with the law. However, most entered the project with incorrect assumptions about the specific provisions of the law: believing that the 1923 threshold applies to unpublished as well as published materials, for example, or that anonymous works are unprotected, or that print and nonprint materials have the same

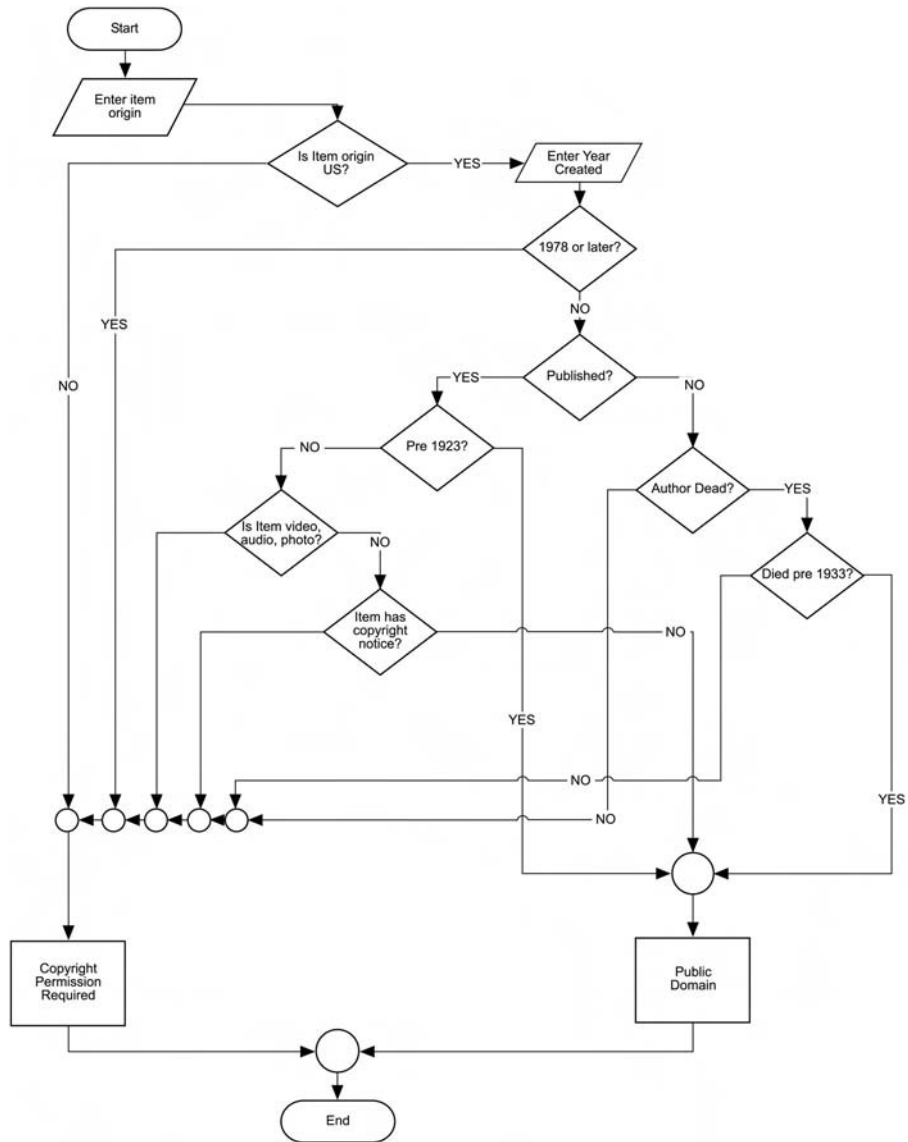


Figure 2.
A flow chart for
determining whether
works published in the
USA are in the public
domain

protections. Although the copyright system's primary function is to determine copyright status rather than to teach the specifics of the law, alert users will learn from the exposure it provides to the variety of factors determining copyright status.

Copyright training video

A later effort to supplement traditional, face-to-face training on copyright issues was the production of a short instructional video, which is now available on the MMM web

The Making of Modern Michigan - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://mmm.lib.msu.edu/inventory/addinfo.cfm?type=item&subcollection=117&Cc

THE MAKING OF MODERN MICHIGAN COLLECTION INVENTORY SYSTEM

My Collections | Info | Administrator | Browse All Participants' Collections | Metadata Entry | MMM Public Site | Logout

You are logged in as:
Participant's Name: **Michigan State University** | Username: **JA**

MY COLLECTIONS /
Collection: Michigan Cookbooks Test II | Collection Code: 07 [Add Subcollection] [Edit Collection Name] / **CONTACT** **REPORT**
SubCollection: something [Add Item] [Edit Subcollection Name]

ID Code: **JA07a008**
Provisional Title:

Copyright Information
Was this item **published** or **created** (if not Published) in the U.S.?
 Yes
 No

Done

Figure 3.
The first issue in
determining copyright
status is the country of
origin

site (<http://mmm.lib.msu.edu/html/videos.html>). The initial production covers published items and a second video, still being edited, will address unpublished materials.

The two six-minute videos were scripted and produced by MMM staff at MSU and filmed in-house with a hand-held digital camera. An MSU staff member from the Copyright Permissions Unit served as the on-air host, introducing the essential concepts of copyright law and illustrating them with examples of material from MSU's Special Collections Division (Figure 6).

Producing the copyright video was one of the most interesting challenges presented by MMM, especially explaining copyright concepts through a visual medium. Although MMM staff were available who had extensive experience in library instruction, filming a talk on copyright would not have made an interesting teaching tool. Instead, the focus was on library materials that represented particular categories of publications. These were interspersed with scenes of student workers at work digitizing print and audio materials. Even familiar scenes in the library were used to provide appropriate – and very low-budget – visuals. At one point, a voiceover explains that materials given away free to the public are considered “published”. On screen this is dramatized by a shot of a student walking up to a handout rack and walking away with several brochures.

The initial video has been surprisingly popular, considering its on-the-fly production methods and the relatively grainy image quality delivered to users with restricted bandwidth. Within weeks of its release, OCLC added a link to the video to the

The Making of Modern Michigan - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://mmm.lib.msu.edu/inventory/addinfo.cfm?type=item&collection=70&subcoll...

THE MAKING OF MODERN MICHIGAN
COLLECTION INVENTORY SYSTEM

My Collections | Info | Administrator | Browse All Participants' Collections | Metadata Entry | MMM Public Site | Logout

You are logged in as:
Participant's Name: **Michigan State University** | Username: **JA**

MY COLLECTIONS / CONTACT REPORT

Collection: Michigan Cookbooks Test III Collection Code: 07 [Add Subcollection] [Edit Collection Name] /

SubCollection: something [Add Item] [Edit Subcollection Name]

ID Code: **JA07a008**

Provisional Title: Copyright status example

Is this item **published** or **unpublished**?
 Published
 Unpublished

Year that item was **published** or **created** (Estimate if necessary):

Done

Figure 4.
The next issue is whether
the work is published or
unpublished

“Online Copyright Resource Kit” provided by its Digitization and Preservation Online Resource Center (<http://digitalcooperative.oclc.org/copyright/>).

Managing site architecture issues

The next major planning challenge for MMM was setting standards for issues that would affect site architecture. Chief among these was the issue of filenames. The filenaming scheme was designed to fill a number of requirements related to project management and site architecture planning. To understand these goals, it is important to make a distinction between “digital files” and “digital objects.”

A “digital file” is a single file in a format such as tif or jpg for image files, wav or mp3 for audio files, etc. A “digital object” is the level at which an item is catalogued; it may be made up of one file (a photograph) or many (pages of a book). Obviously, the web site displaying the MMM collection needs to be able to locate all the digital files associated with a particular digital object. MMM filenames are made up of the following parts. For files which represent an entire digital object, such as a photographic image, a representative filename would be *JA03d007.tif*

- the first two characters, JA, are a two-letter code indicating the library digitizing the material; this allows up to 676 participants to be accommodated (26 x 26).
- the next two characters, 03, are a collection code; up to 99 collections can be established for each participating library.

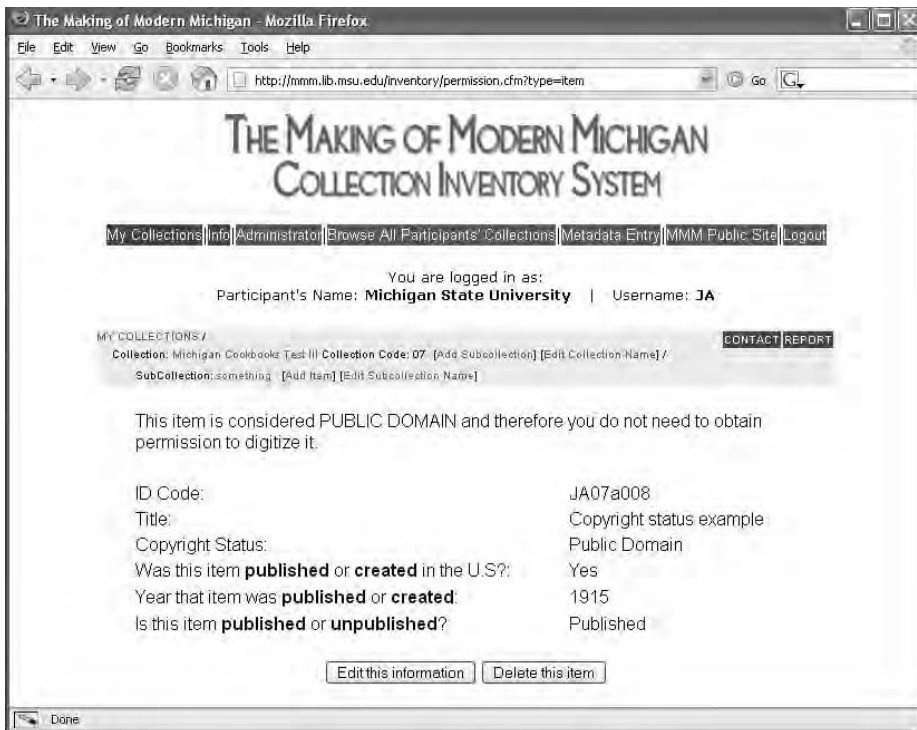


Figure 5.
The status “public domain” means the item may be digitized without permission

- the next character, d, stands for a subcollection; up to 26 sub-collections can be established under any one collection. A subcollection code “a” is required as a placeholder if no subcollection divisions are desired.
- the last three characters, 007, indicate the digital object within the collection; up to 999 items can be listed under a single sub-collection.

For files which represent only part of a digital object, such as image files of the pages in a book, a representative filename would be JA03d007p005.tif. The last three characters indicate that this file is the fifth “part” of the digital object “JA03d007”.

The filenaming scheme accomplishes a number of goals. First, simplicity is served by relating filenames to the digital object’s identifying number in the Dublin Core record. The MMM inventory system assigns the identifying number for a digital object when it is “created” or registered in the collection, the necessary step before the copyright screening. This number is then carried into the metadata entry system (described below) as the < identifier > element in a Dublin Core record.

Second, filenames are unique across the entire collection. This has two benefits. One is that there is no duplication of names, even for materials never intended to reside in the same directory; this means less chance of accidental loss of data by overwriting files. The other is that the filename can be used (without the extension) as a unique identifying number for the digital object. This allows MMM to make use of Digital



(a)



(b)

Figure 6.
An MSU staff member served as the narrator for the copyright training video. Items from the library's Special Collections were used to illustrate points of the law

Object Identifiers, an ISBN-like system for identifying digital content (www.doi.org/). Since DOI prefixes cost \$1,000 each, making one prefix serve the entire collection is obviously a financial consideration.

Third, the filenames encode key information about the digital object: which library contributed the material and where it fits in their collection structure. This hierarchical arrangement of collections and subcollections can then be mirrored by the MMM directory structure. Item “JA03d007.jpg” should be found using the network path `//mmm/JA/03/d/`.

This characteristic was meant to anticipate two situations. “Misplaced” materials (stored in an incorrect directory) would be easy to spot and relocate. And, it provides a way for the MMM search interface to locate materials on other servers, for participants who want to host their own collections. The base URL for the participant’s server would be substituted for the MMM URL, but the directories on the participant’s server would follow the same structure.

Finally, the filenames indicate when an individual file is only part of a digital object made up of multiple files, and which object that is.

Search interface

The hierarchical arrangement of the MMM collection – with contributing libraries and collections and subcollections within those libraries – is clearly evident in the MMM search interface when users choose to browse by institution. Keyword searching is also offered, or users may browse by subject heading. Like the copyright screening system, the search interface was designed by programming staff at the MSU Libraries, using ColdFusion as the programming language and a MySQL database (Figure 7).

There are advantages and disadvantages to developing a user interface in-house rather than using commercial software. Development work can be slow, impacted by other institutional priorities; programmers may be pushing the envelope of their current capabilities, so that many problems are a learning experience rather than a quick fix. On the other hand, a system designed in-house is a system designed to meet the precise needs of a given set of users. And, the experience gained from the development work serves to deepen the expertise of library staff – an outcome consistent with MMM’s emphasis on empowerment and learning.

Metadata entry system

The final major development effort for MMM was designing a webform to allow participants to create Dublin Core records to describe their digital objects. This was a stated goal of the original grant proposal and considered a high priority by everyone involved in providing training to MMM participants.

There are several similar tools available on the web, the most widely-known probably being the Dublin Core Metadata Template provided by the Nordic Metadata Project (www.lub.lu.se/cgi-bin/nmdc.pl). These tools are generally quite user-friendly, but they have a significant limitation in that every record is created from scratch, with each field being keyed in by the user. In the context of the MMM project, where there may be dozens or even hundreds of similar metadata records for items within the same collection, this seemed a serious disadvantage.

To accommodate this, the MMM metadata entry system was designed to allow users to flag a particular metadata record as a “template” and re-use the information it contains as the basis for another record (Figure 8).

enter metadata - Mozilla Firefox
http://mmm.lib.msu.edu/inventory/dcredit.cfm

THE MAKING OF MODERN MICHIGAN METADATA ENTRY SYSTEM

My Collections | Info | Administrator | Browse All Participants' Collections | Metadata Entry | MMM Public Site | Logout

You are logged in as:
Participant's Name: **Michigan State University** | Username: **JA**

Dublin Core Metadata Entry Page

** Required fields are in red. [?] Description of Dublin Core Element.

These elements identify and describe the original artifact being digitized (the book, pamphlet, postcard, diary, letter, etc.). Not all the fields will be applicable to all artifacts.

Title:

Creator 1:

Creator 2:

Creator 3:

Contributor 1:

Contributor 2:

Contributor 3:

Source:

Original Date 1:

Original Date 2:

Acceptable formats:	
Year Only:	1982
Year-month:	1982-12
Year-month-day:	1982-12-01
Date range:	1911-1918
Approximate date:	c.1926
Date not available:	Unknown

Figure 7. The MMM search interface allows browsing by institution and subject heading, and keyword searching through most elements of the Dublin Core record

Results

As of July 2004, the MMM project had been underway for 22 months of the initial two-year grant period.

Regional digitization centers were established at six Michigan libraries throughout the state: Superiorland Library Cooperative (upper peninsula); Traverse Area District Library (northern lower peninsula); Western Michigan University (western lower peninsula); Michigan State University (central lower peninsula); and Wayne State University and University of Detroit-Mercy (both in the Detroit metro area in southeast Michigan).



Figure 8.
The MMM Metadata system is a simple webform that allows users to enter Dublin Core records from any computer with an internet connection

An application process for digitization incentive grants was coordinated by the Library of Michigan in winter-spring 2003. A total of 43 applications were received and 39 grants were awarded. Some 12 additional libraries have subsequently joined the project and many more are exploring participation.

The collection inventory system and metadata entry system were designed, tested, and put into production in Winter-Spring 2003.

Training was provided to all participants in Summer-Fall 2003, covering project management, copyright issues, use of the MMM inventory and metadata entry systems, and guidelines for Dublin Core records.

The copyright video was scripted and filmed in early 2004.

A secure FTP site was established in early 2004 for participants to deliver their digitized collections to MSU.

A prototype search system was developed in Spring 2004, to allow users to search the MMM collection by keyword and browse by subject heading or institution.

By August 2004, approximately 5,000 metadata records and digital files had been delivered and were scheduled to be made available on the MMM web site during the following months.

Lessons learned

There are certain challenges inherent in the dual goals of the MMM project: building a useable digital collection while working primarily with novice digitizers. The two goals do not have to be mutually exclusive, but combining them does require careful planning and a great deal of managerial effort.

A digital collection made up of materials from many different contributors must somehow impose a degree of standardization in several areas. Copyright restrictions must be observed, which means that all participants must share at least a basic understanding of the law. Filenames must follow a consistent pattern, if only to avoid duplication; the user interface must know what file formats will be found in the collection. It is likely that digitization and metadata will have to meet certain specifications to satisfy funding requirements.

At the same time, a project cannot be truly collaborative without some degree of decentralization. If one institution controls all aspects of the project, it may be much easier to achieve consistency and standardization, but the other partners will not be participating in a meaningful way.

One way this has played out in the context of MMM is in training. Orientation to the project and training on essential skills has been carried out at several levels. The MMM management team provided training in copyright, technical standards, metadata creation, and project management to the staff of the regional digitization centers. The regional center staff, in turn, provided training to the individual participants. In many cases, the staff who attended MMM training were then responsible for directing the work of others at their institution. Critical information was, therefore, being passed through at least one group on its way to another.

An exception to this was the group of participants working with MSU, who were trained by an MMM staff member who served simultaneously on the management team (the top level in the training scheme) and as coordinator of the regional digitization center at MSU (the middle level in the training scheme).

Can this type of communication process be successful? It is difficult to judge, but one imprecise measure is the degree to which the first group of participants were able to complete their projects on time and to specifications. Of the ten participants working with MSU, eight were able to complete their project on time and to specifications (80 percent). The on-time, to-specification completion rate for the other groups ranged from zero (zero out of three projects) to 71 percent (five out of seven projects.) The MSU participant group contained the same mix of novice digitizers and slightly experienced digitizers, but had the advantage of receiving their training directly from the MMM management instead of through an intermediary. In some cases, this may have resulted in a clearer understanding of expectations and requirements.

This is not to say that a more centralized structure for MMM would be preferable. The reason for working with novice digitizers is to educate them about this type of work, and the same principle applies to the institutions serving as regional centers. Some of these partners had significant experience of their own with digitization projects; some had the advantage of an established leadership position in their region of the state, but less hands-on experience with digitization. Centralizing all training and

management in one institution would not serve to deepen the skills and experience of staff from those libraries new to the digitization field.

Frequent communication and evaluation of work-in-progress are necessary to bridge this gap, however. In hindsight, it seems likely that there would have been less confusion about certain issues (particularly copyright, technical specifications, and filenaming) if participants had been asked to provide samples of their work as each stage of the project was begun. Fortunately, this is a lesson that can be applied immediately as MMM takes on new participants.

Another unexpected outcome to the project was the divergence between which tasks were expected to present the greatest challenge for participants and which tasks actually did. When the original MMM proposal was written and when project implementation was planned by the management team, metadata creation was expected to be the most difficult area for most participants. Much effort went into creating a user-friendly webform with many examples to follow. In incentive grant applications and in face-to-face training, the participants themselves expressed concern about metadata creation more than any other digitization task.

By contrast, the MMM management team did not anticipate significant difficulty in having participants follow the project's minimum standards for preservation-quality scanning: specific resolutions, color depths, and file formats for various types of materials. The standards were covered in training and repeated in numerous points in the various documentation supplied to participants, including forms they were required to complete to obtain incentive grant funds.

As participants began to deliver their completed projects, however, it became evident that our predictions were incorrect. Few participants actually reported difficulty in creating their metadata and in fact, the majority of the metadata submitted was complete, thorough, detailed, and used the Dublin Core elements correctly. On the other hand, problems with technical specifications of submitted images were much more frequent.

It is always possible to improve training, and the technical specifications will be emphasized more strongly with future participants. Nonetheless, the relative familiarity – or unfamiliarity – of the task may have been a factor too. During training sessions, librarians schooled in traditional MARC cataloging frequently expressed hesitation about working with Dublin Core. Once they began creating metadata, their questions were insightful and demonstrated a solid understanding of the relevant concepts. By contrast, many participants had no previous hands-on experience with digital imaging, and concepts such as color depth and compressed vs. uncompressed files proved confusing to some.

Despite these difficulties, interest in MMM remains strong. Presentations at state and local conferences continue to draw new potential participants, many of whom are willing to take part even without incentive grant funding to defray their expenses. Many participants who have completed their initial projects are now planning future contributions. The regional centers are in place, ready to train, advise, and provide equipment to new users. And, knowledge and experience have deepened, not only among the grassroots participants, but among project managers as well. This may be MMM's most lasting contribution to the library profession.



LHT
23,2

220

Received 27 August 2004
Revised 29 November 2004
Accepted 17 February 2005

THEME ARTICLE

Aggregating distributed digital collections in the Mountain West Digital Library with the CONTENTdm™ multi-site server

Kenning Arlitsch and Jeff Jonsson
University of Utah, Salt Lake City, Utah, USA

Abstract

Purpose – The purpose of this paper is to describe the creation of the Mountain West Digital Library (MWDL), a cooperative regional program distributed throughout Utah and Nevada. Its metadata are aggregated at a single web site. Six digitization centers at the largest universities in both states digitize their own collections and support partner institutions in their geographic regions. Each center runs a CONTENTdm server, and an aggregating server at the University of Utah harvests metadata into a single searchable index. Local control and identity of collections are hallmarks of the MWDL.

Design/methodology/approach – This paper describes the technical structure of the MWDL, focusing on the multi-site server aggregating software from DiMeMa Inc.

Findings – The MWDL was the first cooperative digital project to implement this software, and inspired the same infrastructure for a nine-state project known as the Western Waters Digital Library. In 2005 the MWDL will also become the first in the nation to aggregate distributed digital newspaper collections.

Practical implications – Describes the creation and aggregation of a distributed regional digital library with digital asset management software that is already in use at over 200 institutions in the US. The solutions implemented in the MWDL offer a mechanism for large and small institutions to work together in a cost-effective manner.

Originality/value – Examines the benefits and problems associated with creating a regional digital library.

Keywords Digital libraries, Software, Information retrieval, United States of America

Paper type Technical

Background

The Mountain West Digital Library (<http://mwdl.org>) was established in early 2002 after a pilot project between the University of Utah's Marriott Library and the Utah State Historical Society (USHS) proved the concept of remote digitization support using CONTENTdm software. After scanning a collection of glass plate negatives for the USHS, the Library created collection space on its CONTENTdm server and gave the USHS client-side CONTENTdm Acquisition Station software. The USHS imported the digital images into the Acquisition Station, added metadata, and then uploaded the images to the Library's CONTENTdm server. The USHS built its own web site, with dynamic links into the CONTENTdm server to search and browse the images in the collection. To a user visiting the digital Shipler Commercial Photographers Collection



(<http://history.utah.gov/Photos/C275/>) it appears as if they never leave the USHS server, though all images are actually served from the Library at the University of Utah.

The pilot project worked so well that the initial 400-image contract between the Library and the USHS turned into a 10,000-plus image contract. The Library charged a fee for every glass plate scanned, generating enough revenue to purchase two high-end flatbed scanners and hire part-time technicians to do the scanning work. The USHS paid for the initial work with a grant from the Library Services and Technology Act (LSTA) and the larger contract with a grant from the National Endowment for the Humanities (NEH). The success of this initial project laid the ground work for the idea of a state-wide digital library.

University of Utah discussions with DiMeMa Inc., the creators of CONTENTdm, revealed that the company was developing code that was designed to harvest metadata from distributed CONTENTdm servers to form a searchable index with links to the images on the remote servers. This architecture could then support the searching of a larger “virtual” set of collections state wide while still maintaining local initiative and control. The parameters were in place for the next stage.

The Utah digital landscape in 2001

In 2001 the largest academic institutions in Utah were just beginning to develop digital imaging projects. Brigham Young University was most advanced and was leading a collaborative effort with the University of Utah and several other institutions to digitize the Overland Trails Collection (<http://memory.loc.gov/ammem/award99/upbhtml/overhome.html>) for the American Memory Project. Southern Utah University was in the midst of creating a multimedia digital collection known as Voices of the Colorado Plateau (<http://archive.li.suu.edu/voices/>) And Utah State University had begun to digitize some photographs and a collection of Jack London book covers. None of the institutions owned strongly established digital asset management software.

Utah academic library consortium

The Utah Academic Library Consortium (UALC, www.ualc.net) comprises the libraries of the 14 institutions of higher learning in Utah, as well as the State Library and the Utah Technical Centers. Nevada academic libraries are official members of the UALC. UALC has enjoyed a remarkable cooperative spirit, and its achievements include the purchase of a state-wide integrated library system, a federated state-wide catalog search (www.ualc.net), coordinated collection development, a shared database-purchase agreement, and a state-wide document delivery service known as Utah Article Delivery (UTAD, <http://www.lib.utah.edu/ualc/jour.html>).

MWDL proposal

Following the success of the pilot project between the University of Utah and the State Historical Society the UALC Digitization Committee submitted a proposal for state-wide digitization to the UALC Council. The proposal outlined a plan to establish regional digitization centers at the following universities in Utah: Utah State University (USU); University of Utah (UU); Brigham Young University (BYU); and Southern Utah University (SUU). By participating in the project each center agreed to

create and contribute their own digitized collections as well as support partners who wished to create digital collections, but who lacked the resources to establish their own digitization infrastructure. Partner institutions were generally understood to be public libraries, museums, historical societies, other colleges and universities, and other cultural heritage institutions. Partners were expected to pay a digitization center to scan their materials, thereby helping to offset the cost of scanning equipment, personnel, and server space. Suggested pricing ranged from \$50 for 35 mm slides to \$4.00 for print photographs, and \$6.00 each for large-format glass plate negatives.

The proposal included the following goals:

- build a digital collection reflecting the history of the region;
- include resources from regional cultural and educational institutions;
- offer local control and low-cost digitization to partners;
- standardize metadata for interoperability; and
- make content accessible to all internet users.

The proposal requested \$100,000 from UALC Council, mainly for purchase of hardware and software. Recipients of the proposed funds pledged to provide additional hardware as needed and to maintain and upgrade software licenses. Since the University of Utah and BYU had already established digitization centers most of the proposed funding was aimed at USU and SUU.

- (1) *CONTENTdm software for USU and SUU:*
 - $\$10,000 \times 2 = \$20,000$ (level 2 licenses).
- (2) *Flatbed scanners and computers for USU and SUU:*
 - $\$20,000 \times 2 = \$40,000$.
- (3) *Multi-Site Server hardware and license:*
 - \$10,000.
- (4) *Part-time hourly scanning technicians:*
 - $\$7,000 \times 4 = \$28,000$.
- (5) *Training:*
 - \$2,000.

Full funding was awarded by unanimous vote of the Council in January 2002, and by May all four centers were up and running. In September the first aggregation of collections occurred and a MWDL web site was built for searching and browsing. The two largest universities in Nevada – University of Nevada-Las Vegas (UNLV) and University of Nevada-Reno (UNR) – joined the MWDL in 2003 (Figure 1).

Library services and technology act

Partner institutions in the MWDL rely on grants and other funding sources to digitize their collections at one of the regional centers. Much of this funding has come from the Library Services and Technology Act (LSTA), monies from the Institute of Museum and Library Services (IMLS) which the 50 states distribute through locally administered grant programs. The LSTA Council in the state of Utah recognized early on that the MWDL could be effective in helping to maximize grant funds



Figure 1.
Mountain West Digital
Library web site

awarded for digitization projects. Council resolved to require participation in the MWDL for all digitization proposals to assure that digital collections were created according to national standards, and that they would become part of the state-wide collection. The cooperation of LSTA Council has helped tremendously to make the MWDL a success.

Technical structure of the MWDL

Figure 2 shows the technical structure of the Mountain West Digital Library as it was first envisioned for the state of Utah. The ovals represent the regional digitization centers, each running a CONTENTdm server on the platform of its choice (Linux, UNIX Solaris, or Windows). Each center supports multiple partners by providing scanning services, Acquisition Station software, training, and support. The multi-site server, running at the University of Utah harvests metadata from each of the centers, and users search the aggregated index through the MWDL web site (Figure 3).

Multi-site server

The CONTENTdm Multi-Site Server (MSS) quickly and efficiently harvests the index information from multiple CONTENTdm Servers. It presents a unified, aggregated search capability of all harvested collections, displaying a thumbnail and index data for each resulting record. Once a user clicks a thumbnail, he/she is instantly connected with the item using its unique URL from the host server.

CONTENTdm is an XML database, storing its index data in a monolithic tagged description file, with an offset file used to locate records within the larger file. The

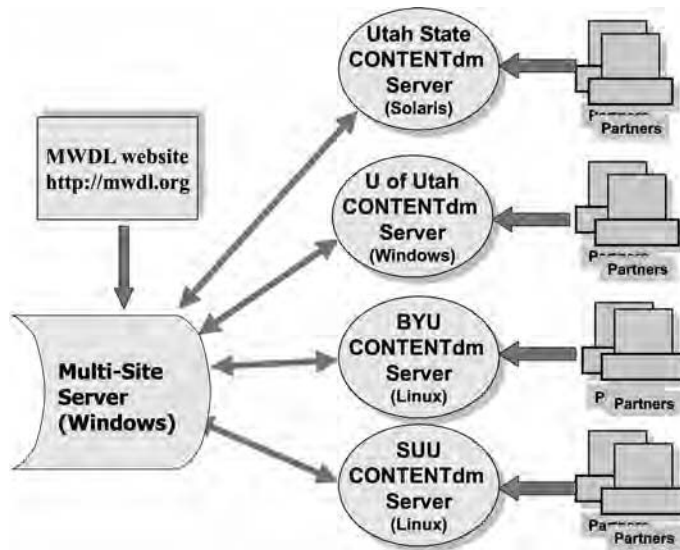


Figure 2.
Structure of the MWDL

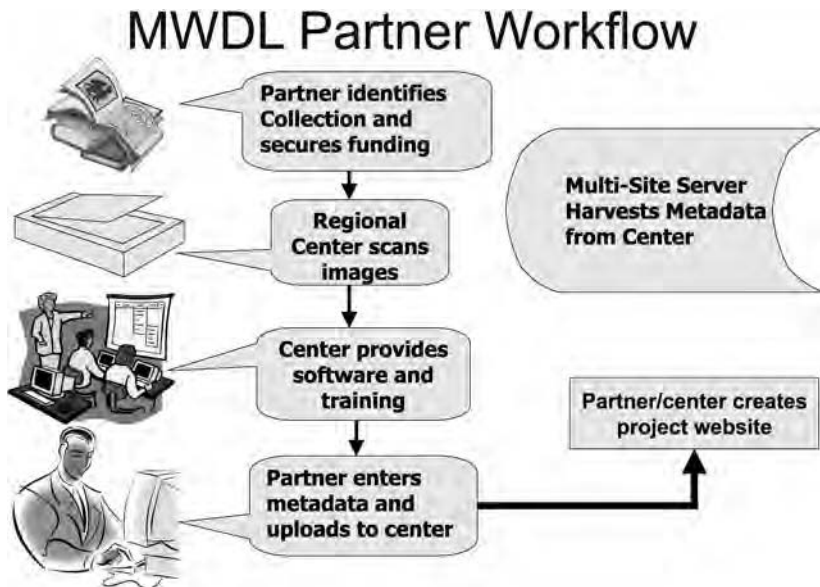


Figure 3.
Digitization workflow for
MWDL partner
institutions

images and thumbnails are stored in a separate directory within the collection directory, which are exposed to the worldwide web as uniquely addressable items. DiMeMa Inc. chose to build their software this way in order to maximize both extensibility and flexibility. The Multi-Site Server is an example of the flexibility of this back-end database software.

The Multi-Site Server was delivered by DiMeMa Inc. in September 2002. It was installed on a Windows server[1] at the University of Utah and harvesting began. Regional centers identified collections on their CONTENTdm servers for harvesting, providing the server address and the names of the collection directories. Currently the MSS uses command line configuration utilities to add collections for harvesting, and to execute the harvesting itself.

Because of its design, the MSS can run on a small, low-cost server, and still deliver good performance. The MSS harvests the indexed metadata files from the remote CONTENTdm servers using WGET (www.gnu.org/software/wget/wget.html). It downloads only the current index and offset files from the remote sites and stores them locally on the MSS, where it offers the indexes for searching at a central location. This lightweight data harvesting method is, in effect, invisible to the remote sites. The collection index files use little disk space, and take only seconds to download from the remote servers at LAN speeds. The MWDL database currently hosts 290,000 records, and the database files for the whole repository only use 2.5 GB of disk space (Figure 4).

A search at the MWDL web site queries the MSS indexes, producing a local result. The search results page is generated by the MSS, and since the thumbnail images are not stored on the MSS itself, they are pulled, in real-time, from the remote CONTENTdm servers. A click on a thumbnail will instantly redirect the browser to the item on the remote server. If a remote CONTENTdm server is unresponsive at the point of search, the images from that server simply become unavailable, but the results from other servers are unimpaired (Figure 5).

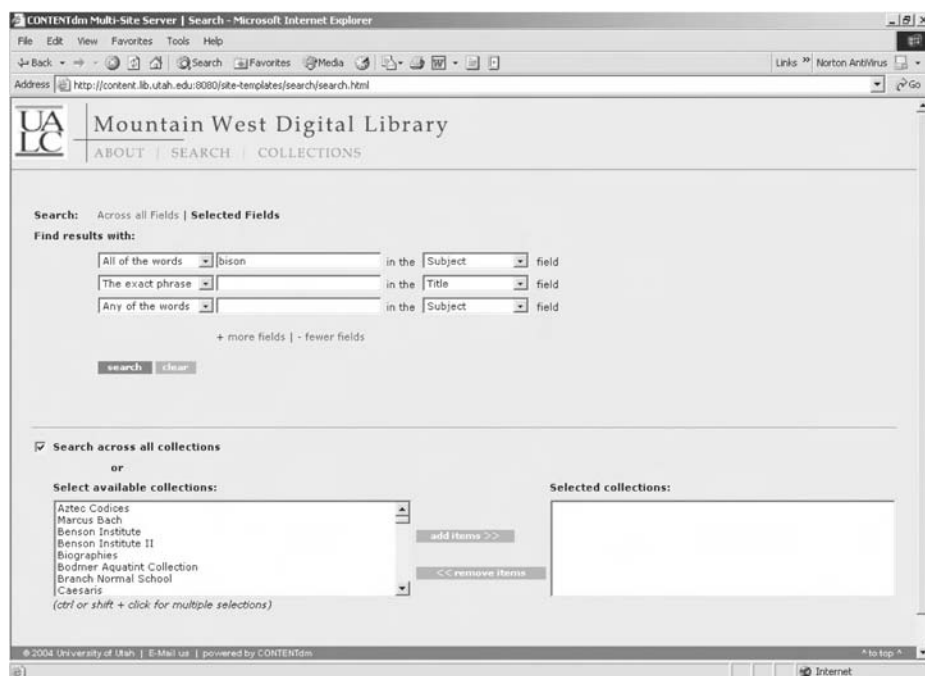


Figure 4.
The main search interface to the Multi-Site Server, as customized for the MWDL

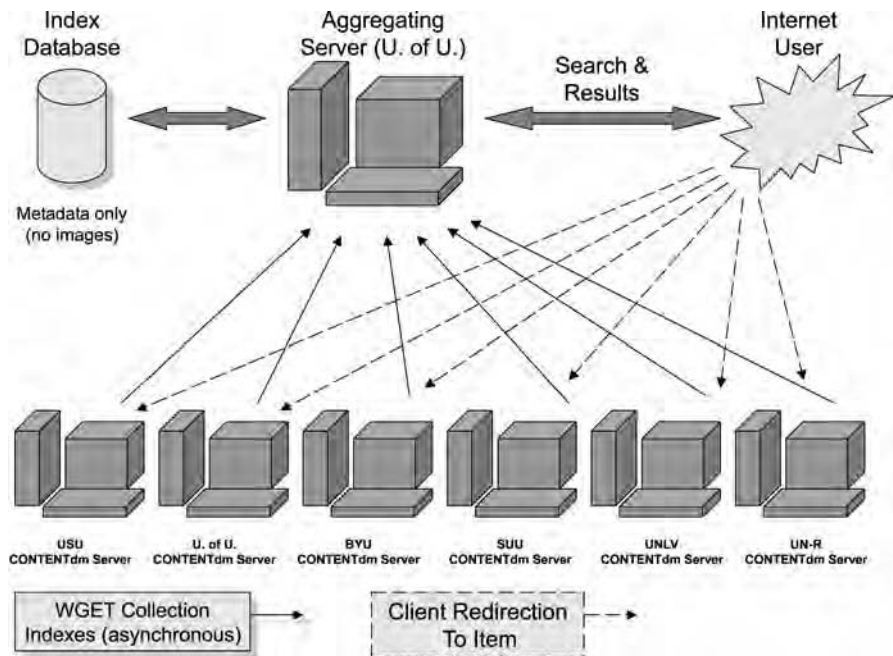


Figure 5.
CONTENTdm multi-site
server harvesting model

The Multi-Site Server includes a subset of the Custom Queries and Results wizard from the CONTENTdm software. This wizard generates the HTML code needed to create search boxes, drop-down lists, or predefined queries for collection web sites. Varying results views can also be specified in the query, including the following options:

- thumbnail images only;
- titles only;
- bibliographic view (up to five metadata fields may be displayed for each hit); and
- grid view (up to five metadata fields, including thumbnail image, may be displayed for each hit).

The MSS provides only the initial search results template (Figure 6). The regional center where the larger display image resides provides the local display template, allowing customization by collection and institution. The web front-end for the MSS can be built to suit the needs of the project (Figure 7).

It is important to note that the Multi-Site Server is only one alternative for aggregating searches across CONTENTdm databases. CONTENTdm software is fully compatible with the Open Archives Initiative (OAI, www.openarchives.org/), and in technical terms, registering the individual servers of the MWDL with one of the existing major OAI harvesters instead of with the MSS is certainly an option[2]. But a regional aggregation that focused on materials found in the MWDL was most desirable for this project, so registering with an existing OAI harvester to offer searching that included materials from other projects was not considered a good choice. (Future plans do include providing MWDL data to an OAI harvester.)

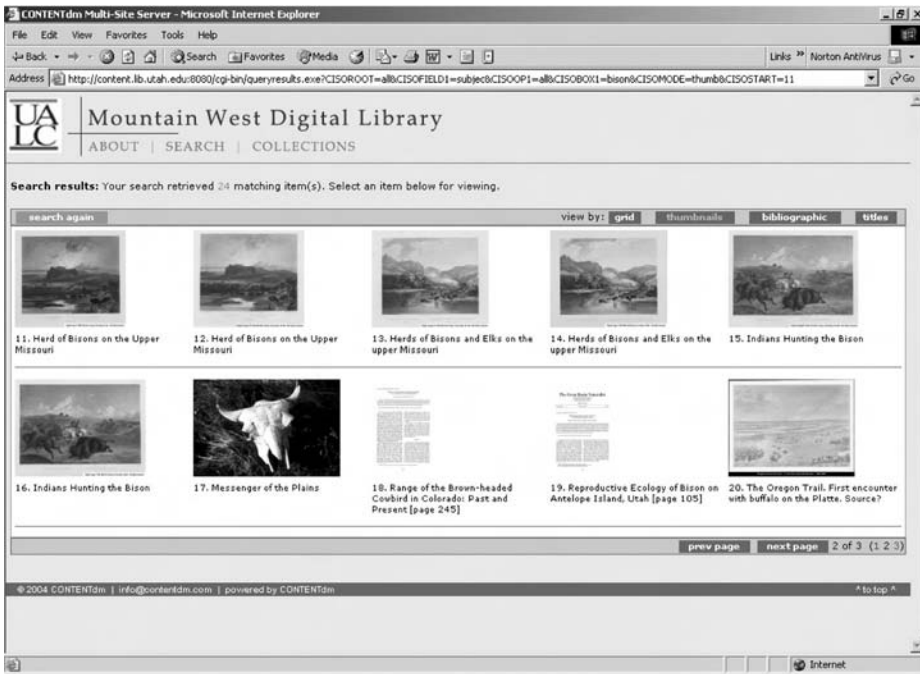


Figure 6. The search results screen, showing the thumbnail images view



Figure 7. Search result being displayed from local site, using customized template

Local development of an OAI harvester was also a possibility, but it is the out-of-the-box readiness of the CONTENTdm MSS that makes it so attractive. The MSS includes a powerful user search interface, plus the customizable, predefined search and results tools described above, that make it easy to build links, drop-down lists, and search boxes for a web site.

Yet another option would have been to purchase or build a federated search system based on Z39.50 queries. CONTENTdm is not natively Z39.50-compliant, but in 2003 the University of Utah developed open source add-on software called ZContent (www.lib.utah.edu/digital/ZContent.html) ZContent is a Perl module developed to convert Z39.50 queries to a CONTENTdm query. It returns results in MARC or XML format.

These options, while decreasing the software costs associated with the MSS, would have created the need for in-house development and maintenance of a front-end application. The Multi-Site Server was the most cost-effective method for achieving the goals of the Mountain West Digital Library.

The Multi-Site Server software installs on a Windows server in just a few minutes. Setting up collections for harvesting takes minimal time using a command-line interface. The MSS for the MWDL was largely pre-configured by DiMeMa as a test case, but a similar project, the Western Waters Digital Library (WWDL, www.westernwater.org) also utilizes the MSS. It took under three hours to completely install the MSS software on an existing Windows 2000 server and define and harvest ten collections from the first four institutions participating in the WWDL.

Metadata

The MWDL follows Dublin Core metadata standards as further defined by the Western States Dublin Core Metadata Standards document (http://www.cdpheritage.org/westertrails/wt_bpmetadata.html). Utah representatives participated in the development of the document, which was led by Liz Bishoff as part of the IMLS-funded Western Trails project (<http://cdpheritage.org/westertrails/index.html>). The document generally follows Dublin Core definitions and specifies field mappings to default Dublin Core whenever field labels are changed for local use. These mappings are crucial for the MWDL because the MSS searches across multiple CONTENTdm collections using default Dublin Core fields.

Participation in aggregated digital projects naturally requires additional care in adherence to metadata standards. Usage reflecting local needs or biases can become serious problems when those metadata are harvested into larger projects and do not match the usages at other institutions. The most important metadata fields for aggregations where the purpose is mainly to provide a single search interface are those that most users will want to search.

Current conditions

By early 2005 the MWDL aggregated collection had reached nearly 300,000 digital objects of every format. Photographs, documents, books, maps, art prints, audio, and video clips are all represented. The MWDL supports partners in the form of colleges, public libraries, historical societies, and museums ranging from Logan, in northern Utah to Las Vegas, and to Reno, Nevada.

In March of 2004 a second generation beta version of the Multi-Site Server was installed to match the new developments in version 3.6 of CONTENTdm

(<http://contentdm.com>). Updated results and display templates, a new custom queries and results wizard, and MSS Open Archives Initiative data provider abilities were all part of the new version.

Why the MWDL works

Cooperation and a respect for the differences in participants' missions, institutional cultures, and funding structures are required to make digital consortia work (Bishoff, 2004). The MWDL model has been successful for several reasons related to the architecture of the project that help address these differences.

Local control

Partners who contribute materials to the MWDL retain control and ownership of their original materials and the digital versions. Partners of the University of Utah are granted administrative rights to their collections on the center's CONTENTdm server, allowing them to edit metadata as needed. They are also trained to apply metadata prior to initial upload, thereby retaining control of its application and contributing to the digitization process.

Identity

The concept of retaining identity cannot be overestimated. The flexible nature of the MWDL means the aggregated collections also allow individual creativity and design without duplicating server/storage resources. Each partner is free to construct its own web site portals into its collections, and results and display templates allow that web site design and identity to be retained even after a user has left the site and entered the CONTENTdm server. Some of the best examples of sites that have retained their identity are the Utah State Historical Society (<http://history.utah.gov/Photos/C275/>) and the Topaz Museum (<http://www.topazmuseum.org/archive.html>). These institutions house their own web sites and have built search and browse queries into the University of Utah's CONTENTdm server using the Custom Queries and Results wizard.

Low cost

The scanning fees charged by the centers include server space, training, and the Acquisition Station software. The cost savings to partners in terms of hardware, software, and personnel is enormous. Long-term storage of high-resolution files is also available for an additional fee if requested.

Digitization uniformity

Because the digitization is done at the centers it occurs in a uniform manner, using high-end equipment and standards. Adoption of the Western States Dublin Core Metadata Standards, as developed by the Western Trails project, has also insured uniformity and interoperability across collections. Beyond the required metadata fields, all institutions are free to implement whatever additional fields they feel are necessary for their collections.

Revenue for regional centers

The scanning fee generates a revenue stream and helps the MWDL to be somewhat self-sustaining. Revenue at the University of Utah has funded new scanners, servers, and part-time scanning technicians.

Concerns

Creation of the MWDL has not been without problems. In any cooperative project there are concerns about willingness, agreement, and equal participation. Following are some specific concerns experienced in the MWDL.

Funding and staffing

Aside from the initial investment of the UALC, and the LSTA grants to fund individual projects, the MWDL has thus far received no additional funding. There is no staff devoted exclusively to the project; participation has been voluntary and in addition to regular jobs. This has impacted the speed of development of the project.

Participation

Not all regional centers have participated equally. Some have supported no partners to date, for reasons that have included personnel issues and initiative.

Notification

The CONTENTdm MSS currently has no ability to discover new collections automatically, and there is no way to harvest all collections from a server without knowing their directory names. Centers are therefore required to notify the University of Utah that new digital projects have come online.

Training and support

Partners who are supported by a regional center are dependent on training and continued support from that center. The CONTENTdm Acquisition Station software is not difficult to use, and an afternoon of training is usually sufficient to get started, but basic computer skills are a requirement. The diversity of institutions in the MWDL naturally means there is a diversity of technical expertise, and remote support can be challenging.

Data transfers

CONTENTdm offers a Full Resolution Manager tool, allowing collection administrators to keep track of archival digital files, whether they choose to store them online or offline. University of Utah has encouraged its partners to store their archival files on the University's server, which requires a data transfer via FTP. Delivering high-resolution scanned images to partners for import into the Acquisition Station, and then uploading them back via FTP is not always feasible due to variations in internet reliability and the large size of archival files. After experimenting with CD and DVD we have decided that the most reliable method for the high-resolution images is to mail external hard drives.

OAI policy issue

In 2004 we convinced DiMeMa Inc. (against their better judgment) to build OAI-compliance into the MSS. We thought that it would greatly simplify things if we could provide data to OAI harvesters at the aggregated level, but of course the danger is that redundant files from the MWDL could find their way onto OAI harvesters if individual institutions were also registering their CONTENTdm servers. In late 2004 the UALC Digitization Committee agreed to leave OAI data provider responsibilities with the participating institutions, and we have thus left disabled the OAI capability of the MSS.

Other projects modeled after the MWDL*Western Waters Digital Library*

The Western Waters Digital Library (WWDL), as mentioned earlier, is an IMLS-funded project to develop a digital library of water information resources for the western states. Led by the Greater Western Library Alliance (GWLA), a consortium of 30 research libraries in the Midwest and West, the WWDL is modeled after the infrastructure of the Mountain West Digital Library.

Of the 30 GWLA institutions, 12 are actively participating in the two-year project, which focuses on four major river basins: Platte, Rio Grande, Colorado, and Columbia. As in the MWDL each institution runs a CONTENTdm server on the platform of their choice, and a Multi-Site Server installed at the University of Utah harvests metadata from the participants.

The grant began in November 2003, and by February 2004 the new WWDL Multi-Site Server had harvested existing digital collections at four of the participant libraries, quickly bringing the WWDL online with nearly 30,000 digital objects.

Utah digital newspapers

Over the past two years the University of Utah has developed an extensive newspaper digitization program, again employing CONTENTdm as the digital asset management tool (Herbert and Arlitsch, 2003). There are already more than 200,000 pages posted, and with current grants from IMLS and LSTA we expect to reach more than 400,000 pages by late 2005.

As a partner to the University of Utah in the IMLS grant, BYU is separately digitizing 40,000 pages of the Deseret News and posting them on their own CONTENTdm server. During the early part of 2005 aggregation of the distributed collections with the MSS will begin. We believe this will be the first aggregation of a distributed digital newspaper collection in the nation.

Because of the zoning of individual articles the newspaper digitization process creates a much greater number of files than the actual pages counted in the project. The 200,000 newspaper pages on the University of Utah's CONTENTdm server translate to over two million individual files. For this reason we have been leery of simply aggregating them into the existing MWDL; the sheer number of files would overwhelm most searches with newspaper results. The aggregation plan calls for the installation of a separate MSS, which will aggregate only newspaper collections. Users of the MWDL will be able to select whether to include newspapers in the general search, or whether to search them separately. And the aggregated search will also be available from the Utah Digital Newspapers web site.

Conclusion

There is always more work that can be done. The MWDL web site can be improved to offer guided research rather than simply search and browse functions. Establishing a dedicated staff to manage the project, and securing a direct source of funding to support the MWDL would help enormously.

In its two years of existence the MWDL has grown to nearly three hundred thousand digital objects, and continued growth is expected. We have established a viable technical infrastructure that will serve us well into the foreseeable future, and the success and cooperative nature of the MWDL positions us well for future grant funding. It is the shared vision, cooperation, and enthusiasm of the members of the UALC that has made this project possible, and as long as that spirit endures the MWDL will continue to prosper.

Notes

1. The Multi-Site Server, like the CONTENTdm software, runs on three different platforms: Linux; Solaris; and Windows.
2. Michigan's OAISTER and the OAI harvester at UIUC are two examples of major existing OAI harvesters.

References

- Bishoff, L. (2004), "The collaboration imperative", *Library Journal*, Vol. 129 No. 1, p. 34.
- Herbert, J. and Artlitsch, K. (2003), "Digitalnewspapers.org: the Utah digital newspapers program", *The Serials Librarian*, Vol. 47 No. 1, p. 2.



THEME ARTICLE

AlabamaMosaic

AlabamaMosaic: sharing Alabama history online

Sherida Downer

Auburn University Libraries, Auburn University, Alabama, USA

Sue Medina

Network of Alabama Academic Libraries, Montgomery, Alabama, USA, and

Beth Nicol and Aaron Trehub

Auburn University Libraries, Auburn University, Alabama, USA

233

Received 27 August 2004
Revised 29 November 2004
Accepted 17 February 2005

Abstract

Purpose – This paper aims to summarize a three-year federal grant from the Institute of Museum and Library Services to create a statewide digital collection.

Design/methodology/approach – This paper describes the Cornerstone Project: its background, the processes used to determine appropriate standards, guidelines, and best practices for digital collections, the technology selected for the digital production centers and the web portal, and its plans for the future. The project is a collaborative statewide initiative to make unique historical treasures from Alabama's archives, libraries, museums, and other repositories electronically accessible to Alabama residents as well as students and scholars around the world.

Findings – The project has demonstrated a successful model for statewide collaboration among all types of repositories. NAAL member institutions are helping other repositories create digital collections of unique Alabama historical resources.

Practical implications – The Cornerstone Project offers practical procedures that can be replicated by any consortia considering distributed digital collection building.

Originality/value – The project offers training to help librarians, archivists, and other staff successfully plan, implement, and complete digitization projects. Three shared digital production facilities at the Alabama Department of Archives and History (ADAH), Auburn University, and The University of Alabama serve as training centers and assist with local digitization efforts.

Keywords Digital libraries, Collections management, Information operations, United States of America

Paper type Case study

The Network of Alabama Academic Libraries (NAAL), a consortium of Alabama's public and private four-year colleges and universities, was established 1984 to coordinate academic library resource sharing. The academic institutions in NAAL are fully committed to education, research and service. Their "open door" policies help ensure that every citizen has access to the rich information resources housed in their libraries. All NAAL members facilitate statewide access to their print collections by supporting interlibrary lending at no cost to Alabama public libraries.

An early adopter of networking technology, NAAL embraced electronic databases for cooperative collection development, becoming one of the first statewide consortia to license the newly available Expanded Academic ASAP for all its members. The libraries' "open door" policies resulted in walk-in community users discovering various



databases in academic libraries. When these users sought, sometimes aggressively, remote access which was denied to anyone not an “authorized user”, NAAL envisioned a “virtual library” to serve all Alabamians. The Alabama Virtual Library (AVL) marked a NAAL initiative to foster new partnerships among public school systems, colleges, and public libraries to use technology to strengthen information resources available for all Alabamians. State funding for the AVL represented a “new” library, one not physically owned by an individual library or educational sector but rather a consolidation of resources to benefit all citizens. Since its initial funding in FY2000, the AVL has emerged as an essential portal for access to information for K-12, the higher education community, and the public.

While open door access for community users, AVL databases, and liberal interlibrary lending are extremely valuable information resources, they do not fully serve the information needs of all citizens or the state’s research and economic development agendas. Commercial databases, marketed for national markets, rarely include material of regional or local interest. These products, most often created by aggregating secondary source materials as it is published, offer few full text historical materials older than the final quarter of the last century. Despite the richness of Alabama’s electronic collections, searches of commercial databases do not as a rule yield much information about Alabama’s history and culture. Furthermore, liberal interlibrary lending policies rarely extend to Alabamiana collections.

Planning for a digital collection

Nearly all NAAL libraries, along with other repositories such as archives and museums, maintain collections of unique materials relating to the history of the state or a research priority of their institutions. In 1998, NAAL’s chairman, Dr Stella Bentley, university librarian for Auburn University, appointed a Digital Content Committee to propose a role for NAAL in contributing primary sources to the “library commons”. Important factors for any policy were avoiding duplication of effort, avoiding the adoption of disparate digital systems, and encouraging the exchange of information content across the broadest possible audience. Because this committee was working at the same time NAAL was seeking funding for the AVL, it was critical that any digital collection become part of this new “portal”.

NAAL recognized that making primary sources accessible would be more difficult than sharing circulating print collections. Many items in special collections are already in great demand, but their value or fragile condition restricts use. Many more are inaccessible because of the absence of adequate descriptions and finding aids. Poor environmental conditions at some sites with concomitant deterioration could mean irrevocable loss of historical artifacts. In Alabama, tornados, hurricanes, or floods are threats that can destroy collections.

The existing large special collections raised questions about how best to begin a digital collection. The Committee proposed that any digital collection plan must seek to protect valuable artifacts, reduce or avoid damage caused by handling fragile items, provide secondary access to the intellectual content, and extend access to users who cannot physically visit the library. The Committee struggled with adopting appropriate selection criteria to assure a manageable scope, assuring participation by all types of repositories, and contributing to the mission of the AVL. Tracey Berezansky, Assistant Director for Government Records for the Alabama Department

of Archives and History (ADAH), suggesting using Alabama Moments in American History as a guide for subject selection for digitizing materials.

Alabama Moments in American History: Supplemental Teaching Materials for High School Teachers grew out of concern by state historians over changes in history instruction in Alabama public schools (www.alsde.edu). This curriculum eliminated a required one-semester Alabama history course. The new course, offered over two years for 10 and 11 grade students, stresses US and world history along with civic responsibility. While historians applauded the new course of study, many were concerned about eliminating the separate Alabama history course. Including Alabama history in the new US history curriculum was hampered by the lack of resources for teachers, who would still be using standard US history texts containing little information about Alabama. Few if any of these teachers have an extensive background in Alabama history or access to the resources needed to integrate Alabama history into their class assignments. To enrich the new curriculum, Alabama historians produced supplementary materials organized under the heading Alabama Moments (www.alabamamoments.state.al.us/) (Figure 1).

Each online topic contains a brief summary, a more detailed explanation, a bibliography, primary source documents to enrich classroom presentations, and suggested enrichment activities such as class projects, research topics, field trips, videos, or web sites. *Alabama Moments* offered an ideal framework for developing a statewide collaborative digital collection.

The Digital Content Committee also recommended that NAAL emphasize digitizing collections related to the Civil Rights era in Alabama. Many Alabama libraries have received gifts from “front line” participants of the Civil Rights movement; and unifying related materials “virtually” would be a special contribution to scholarship on this subject.

The Cornerstone Project: October 2001 to September 2004

A hallmark of NAAL has been collaboration to strengthen institutional capability to anticipate and manage an increasingly complex education and information environment. Changes in information delivery have been carefully managed by NAAL members to encourage and extend access to global information. Professional practices and procedures for the acquisition, maintenance, and delivery of information have been driven by advances in technology for information storage and delivery as well as by changes in user expectations and behaviors. Alabama libraries have had great success in sharing electronic resources and circulating print materials, but less success in providing access to primary sources. Also, libraries, museums, and other repositories have not engaged in systematic, planned efforts to facilitate user access to related materials held across many locations – a crucial factor in promoting scholarship about the struggle for Civil Rights.

In this environment, NAAL forged The Cornerstone Project: Building the Foundation for Sharing Unique Treasures to make historical materials held by Alabama’s archives, libraries, museums, and other repositories electronically accessible. This online collection would enable Alabamians to explore the images, diaries, letters, and other special artifacts that document their history – all from the convenience of a desktop computer. Further, the collection would contribute to scholarly research while also making historical collections accessible to a new audience

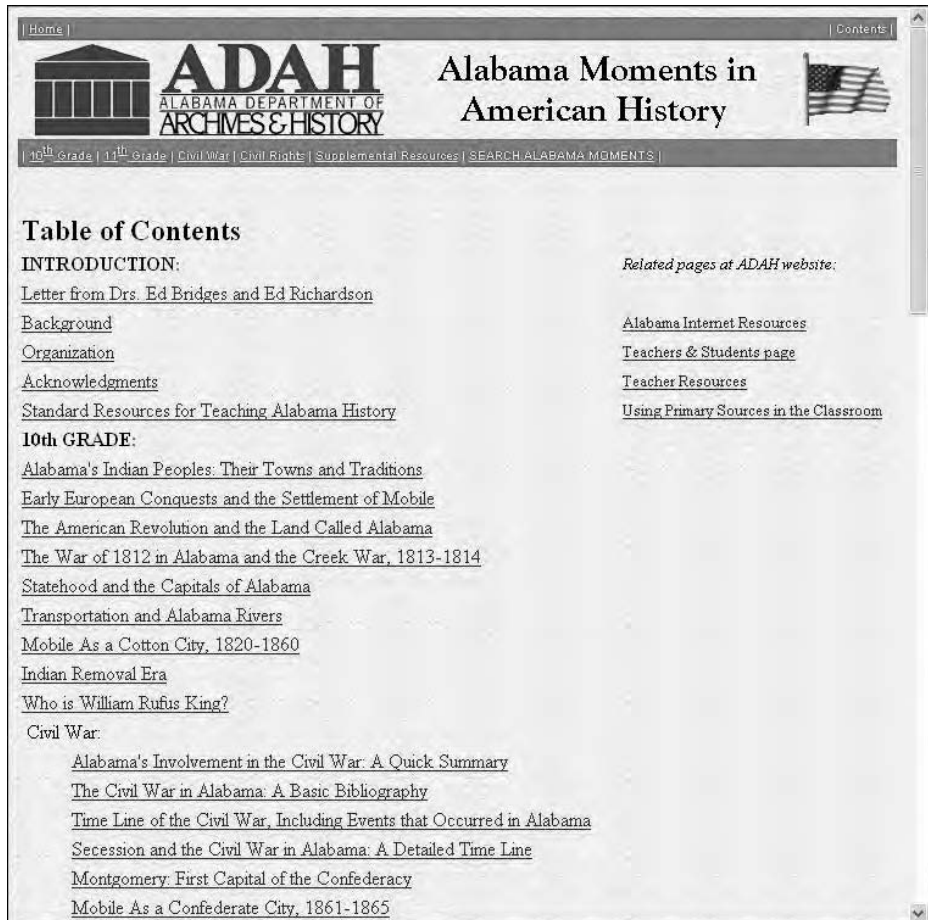


Figure 1.
“AlabamaMoments in
American History”
curriculum enrichment
web site

comprising scholars in other countries, elementary school students in other states exploring Alabama history, and industry leaders tracking industrial and economic information. Equally important, the digital surrogates of the unique items in Alabama’s special collections would eliminate barriers to use while protecting the original materials.

Alabama is typical of many states in having different state agencies providing leadership for its library sectors. The Alabama Public Library Service (the state library agency), Alabama Department of Archives and History (state archives), Department of Postsecondary Education (two-year colleges), and NAAL (a consortium of academic and other research libraries based in the Alabama Commission on Higher Education) have cooperative relationships for a number of statewide programs. Also, strong relationships exist through the AVL with the State Department of Education and local K-12 school systems to promote use of the virtual library resources by students and teachers. Under the aegis of the Cornerstone Project, these agencies are beginning to work together on expanding the availability of digital information for instruction and

research. The Cornerstone Project is one part of a long-term vision to encourage libraries and other repositories to integrate digitization of materials into their usual and routine tasks. Its goal is to effect systemic change in the way libraries and other repositories develop their collections and make resources available. As a result of this project, scholars, students and other researchers will find a web-based collection of digital surrogates for unique documents, images, recordings, maps, realia, and multimedia held by Alabama repositories. This collection will unite surrogates for materials that are physically separated from one another and from users. It will open opportunities to “mine” and analyze information in ways not previously possible. Students, teachers, and other researchers will gain an unparalleled opportunity to efficiently search widely dispersed, significant collections of information resources. The project will integrate searching the intellectual content of formats such as photographs and recordings with the searching of traditional print materials such as books and journals.

Funds to match the IMLS grant are provided as in-kind services by the Alabama Department of Archives and History, Auburn University, The University of Alabama, and NAAL. Services include use of equipment, space, and – most important – the availability of staff to undertake project tasks. NAAL received notice on September 26, 2001 that the Cornerstone Project had been awarded a National Leadership Grant to begin on October 1, 2001. However, the attacks of September 11 2001 delayed many of the activities planned for that fall and winter. The project never overcame this delay and subsequently requested and was approved for an extension through September 2004 to ensure its successful completion.

Phase 1: gathering information

One of the most pressing needs was strengthening our knowledge about digitization programs. To expand our knowledge, librarians from Auburn University Libraries (AUL) visited Cornell University to discuss their digital program and identify appropriate best practices to follow. They learned about the Cornell Institute for Digital Collections, Cornell’s participation in the Making of America Project, the Saving America’s Treasures project in which anti-slavery pamphlets were digitized for preservation and access, the Core Historical Literature of Agriculture (to which Auburn contributed records and holdings locations) and Core Historical Literature of Home Economics. They discussed ENCompass development and how content is organized for ENCompass, reviewed imaging standards, and learned how Luna Imaging software was being used and how metadata were added to scanned images. This site visit assisted the Cornerstone Project with planning its processes to select, scan, identify, and organize digitized materials.

Librarians from The University of Alabama visited the University of Southern Mississippi Special Collections department to learn about its “virtual reading room”. This program had successfully obtained copyright permission to digitize and make available the University’s collection of editorial cartoons created by members of the American Association of Editorial Cartoonists and for the Civil Rights in Mississippi Digital Archive, an exceptional collection of oral histories. The University of Alabama team was also interested in the department’s staffing and workflow.

Auburn University and The University of Alabama also sent several staff members to the School for Scanning course offered by the Northeast Document Conservation

Center. This highly regarded week-long curriculum covers project management, copyright and other legal issues, selection of content for digitization, standards, quality control, costs, collaborative project models, text and image digitization, metadata, and digital longevity and preservation.

An invaluable asset for the Project was Liz Bishoff, then-director of the Colorado Digitization Project. At a statewide meeting of interested repositories in April 2002, Ms. Bishoff reviewed the issues the Project would need to address. She helped the project planners organize anticipated tasks into manageable assignments, each with an appointed “working group” charged with completing the “essentials”. The working groups became training sessions in their own right, as members worked hard to understand all the concepts, standards, practices, and technology that would support the digital collection.

Finally, to broaden training opportunities beyond the NAAL membership, the Project contracted with the Southeast Library Network (SOLINET) to offer its highly regarded workshops – Introduction to Digital Programs, Hands-on Digitizing, and MetaData Standards – in Alabama.

Phase 2: establishing working groups

Three working groups were formed in April 2002 to survey repositories and develop a model for digital collection plans, evaluate and recommend metadata standards, and recommend hardware and software for the web site.

1. Directory survey and digital collection plans

One Project goal was for every repository to complete a directory entry and then a digital collection plan. The survey would identify possible participants and determine their available or needed capabilities for: staffing; facilities; internet presence (existing web sites or need to create new web sites); hardware and software, data storage; facilities; and, other infrastructure needs. The digital collection plan would identify characteristics (print, images, sound, video, artifacts, etc.) and physical condition of source materials selected for digitization and assess these materials’ value to history topics in Alabama Moments. Dwayne Cox, Head of Special Collections for Auburn University, chaired a working group to develop a survey instrument. The survey form was mounted online at the Alabama Virtual Library. An announcement with a Project brochure was mailed to repositories encouraging them to respond online to be listed in the directory of repositories owning Alabama history materials. Response was low, and a follow-up letter included a survey form to be returned by mail. This resulted in only a few more responses. The project planners discussed the low response rate and determined that the survey form and letter failed to instill a sense of excitement about the project. Until the project could “show and tell” the online collection through a web presence, it would be difficult to generate interest in a directory. Thus, it was decided to delay creating a directory of repositories until after the project’s web site became available.

During her consulting with the project, Ms Bishoff created an outline for a digital collection plan. Mr Cox supervised two graduate students in history/archival studies who worked with repositories to test this outline and develop “model” collection development plans. Model plans were developed for an academic library, a county archives, and a public library. A fourth model will be developed for a museum. The

Project planners will likely create this plan because the original archival studies students have finished their studies.

NAAL believes a digital collection plan is an essential element for a successful digital project. Although many possible participants (especially NAAL academic libraries, the largest public libraries, and museums) are known to hold relevant materials for digitizing, most are not prepared to initiate a digitization program. The process of developing long-term and strategic plans helps staff focus on manageable projects. Preliminary testing of the collection plan outline found that most Alabama repositories need help with the basic description of collections, creation of finding aids to hasten identification of specific items appropriate for digitization, skill development, and assistance with project planning. In keeping with the recommendations of the *Handbook for Digital Projects*, the project stresses good management, with each repository setting clearly defined goals, writing a long-term plan for achieving its goals, and developing a management strategy.

A completed digital collection plan is required for a repository to be named a “project participant”. This enables project planners to determine what assistance and resources the repository may need from the project. It also “opens the door” for the repository to send staff to project workshops in scanning, creating metadata, and complying with copyright; request workshops to be offered in the local community, and utilize the training and digitization services of the digital production centers. For some repositories, the plan becomes the basis for developing proposals for additional funding to support local digital projects.

2. Metadata standards

The Metadata Standards Working Group, chaired by Sterling Coleman, then Digital Resources Librarian for Auburn University, was charged with identifying standards and best practices for tagging digital files for storage and retrieval. This group investigated practices used by libraries, museums, and archives for a variety of digital projects. It was determined that repositories with professional staff already had high standards for cataloging their materials, although most had not yet adopted a standard for their own digital surrogates. A major project concern became cataloging digital files for items held by repositories such as local history societies, museums, or archives. These repositories generally do not employ professional librarians or archivists, and many are staffed by volunteers.

After much discussion, the working group recommended the adoption of unqualified Dublin Core Metadata (<http://dublincore.org/>). This standard offers fifteen data elements to describe a digital file. Elements can be repeated and qualified as needed to fully describe items from the local collection. The group agreed that the project would encourage professionally-staffed repositories to adopt the appropriate standard to enable their own users to search records for both their bibliographic catalog and digital files. This recommendation assumes these facilities will use all elements of Dublin Core to fully describe a digital surrogate or use MARC which can be crosswalked to Dublin Core. The project will train staff from other repositories to create metadata using all 15 Dublin Core elements. However, the Project recognizes that some repositories may have difficulty with metadata creation, so only four elements are required to submit a record: contributing repository, title, URL for location, and one subject heading. This decision was reached after much debate and

did not enjoy unanimous agreement. In the end, however, project planners decided that it was more important to include repositories from throughout the state, even if they contribute only “short records”, than to require the higher level of description preferred by the working group.

3. Hardware and software

The Hardware and Software Working Group, chaired by Richard Wood, Dean of Libraries, University of South Alabama, was charged with recommending hardware and software for the project. Because the central site would be the public presence as well as the basic infrastructure for the Project, it needed a search engine capable of searching disparate systems and returning an integrated results set, a digital file management system to organize digital files for repositories that are too small to have their own web servers, a work management system to facilitate metadata creation for participants who are not professional catalogers, and space for “web exhibits” to enrich Alabama Moments and other topics.

To advance the Project’s ambitious goals, the working group members reviewed online sites for digital collections functionality and began contacting vendors with products that seemed appropriate for the project’s goals. The group developed a set of criteria with elements in five categories: Entry/front ware, search engine (queries, retrieval results, display, and navigation), digital file management system/database, metadata input and management, and miscellaneous system functions. To help the group with evaluation, each vendor was asked to respond to questions about capabilities of the software and the company. Some 18 of these related to criteria which would be “required” of any project software:

- (1) Can the interface be customized?
- (2) Does the company offer design assistance?
- (3) Does the company offer hardware assistance?
- (4) Does the product support Z39.50?
- (5) Does the product support XML?
- (6) Does the product Encoded Archival Description?
- (7) Does the product Text Encoded Initiative?
- (8) Does the product support Dublin Core?
- (9) Does the product support Open Access Initiative?
- (10) Does the product support MARC?
- (11) Does the product support Unicode?
- (12) Describe the company’s support for maintenance.
- (13) Describe the company’s support for updates.
- (14) Describe the company’s support for training and ongoing support.
- (15) Does the product provide for a central repository for digital files?
- (16) Does the product support searching/access to external digital repositories?
- (17) What report capabilities are supported; what reports are available?
- (18) Does the product support audio-video streaming?

The remaining were “highly desired” features:

- Is the product web-based?
- Does the product include bulk record importers?
- Does the product provide for normalization (vocabulary, field names)?
- Does the product support Open URL?
- Does the product support user browsing?

After much deliberation, the working group recommended ENCompass Solutions, offered by Endeavor Information Systems, Inc. for its full-featured applications, open systems connectivity, and ease-of-use for the public and library staff (<http://encompass.endinfosys.com/>). The group anticipated that the large number of Endeavor systems installed in Alabama would help shorten the technology learning curve and promote speedier implementation of the Project database.

The Cornerstone Project had planned to seek state funding for FY2004 to support the central site. However, declining state revenues for FY2003 rendered this impossible. Fortunately, NAAL members have a strong history of cooperation and a firm commitment to succeeding. In this case, Auburn University Libraries (AUL) offered to host and manage the Cornerstone web server at its own expense until such time the State could fund the project. Accepting this generous offer enabled use of an AUL open contract for Endeavor products and services. This allowed the Project to move forward more quickly without having to bid the server and software. It also reduced training time because Auburn was already installing ENCompass for its own digital collections.

The central site: technical details

The central ENCompass installation at Auburn University currently resides on a two-processor Sun 280 server computer with two gigabytes of RAM and 72 gigabytes of disk storage. The server runs Solaris 9 as its operating system, Oracle 9i as its database management system, and Apache 2 and Tomcat as its web server software.

The public interface is a web page. There are no special requirements for access beyond freely available plug-ins such as Acrobat Reader for PDF files; QuickTime, RealPlayer, or Windows Media Player for audio and video, and the ability to enable both cookies and javascript. Because of the quality of the digitized images, the web display is designed for no less than 800 by 600 pixels.

The current version of ENCompass has some shortcomings for this statewide project. To use the metadata template, the contributing library must locally install client software. An upgrade to web-based access to the metadata template has been announced, but not yet released. External scripts and applications must be used for compound documents to present such materials as pamphlets, college yearbooks, and other multi-file objects in a useful and coherent fashion. Also, the project must create its own interfaces to present images using vector-based graphic formats such as JPEG2000 and MrSid.

Over the course of meeting its charge, the Hardware and Software Working Group learned a great deal about digital library software. It gradually became clear that the group was requesting operational features and functionalities that had not yet been put into production by any vendor. Consequently, the group invited Endeavor to become a

development partner to provide a “turn-key” digital collection solution. Endeavor agreed to assist in integrating “third-party” software – for example, software that facilitates the navigation of complex, multi-element files (e.g. compound-document software, for use with pamphlets, monographs, and articles), or software that allows the project managers to evaluate the level of use (e.g. web statistics software).

ENCompass provides a framework for accessing, categorizing, and organizing electronic content. Several distinct components function to provide seamless simultaneous searching across disparate collections. There are three sub-systems: ENCompass for Digital Content, which categorizes, organizes, manages, and searches local digital collections; ENCompass for Resource Access, which performs federated searching of remote and local resources; and LinkFinderPlus, an OpenURL Link Resolver. One advantage of ENCompass is the ability of these components to facilitate a single search across geographically dispersed collections and return a consolidated set of records for digital resources.

ENCompass for Digital Content categorizes, organizes, manages, and searches local digital collections. These functions form the basic building blocks for individual sites to contribute their own digital collections to the Alabama history database. Any Cornerstone Project participant can have its own “space” on the project’s central server, called an ENCompass Repository. Each ENCompass Repository supports metadata using a single Document Type Definition (DTD). Most participants will use a DTD mapped to Unqualified Dublin Core. Once assigned an ENCompass Repository, each participating institution can create its metadata using the ENCompass Dublin Core template and export locally-created metadata to its ENCompass Repository at the central site. A single ENCompass Repository may contain the contributor’s metadata for locally-held digital objects along with metadata records for objects stored on the central server. Digital objects can include documents, photographs, sound recordings, text documents, web pages, and so forth. Additional space can be allocated to a contributor’s ENCompass Repository if needed.

Project participants may be assigned additional ENCompass Repositories if they need to use other DTDs. This enables contributors to utilize additional metadata fields and alternate DTDs to accommodate the specific needs of their local digital collections. ENCompass currently supports Encoded Archival Description and Text Encoded Initiative materials.

One issue for multiple repositories sharing a virtual digital collection is assigning each institution its own workspace name (also known as an ENCompass Repository name). Cornerstone Project participants are assigned a workspace name which is constructed using the institution’s OCLC symbol, if available. All Alabama academic and public libraries have OCLC symbols for a statewide resource sharing group; similar workspace names are constructed for institutions that do not have an OCLC symbol. This workspace name is not displayed to the public; instead, the software displays each institution’s descriptive and publicly recognizable name.

ENCompass for Resource Access adds the ability to search remote “non-ENCompass” resources. The ENCompass connection is proprietary and connects to “like” software such as Voyager and other Z39.50-compliant Online Public Access Catalogs (OPACs) which utilize the MARC record type, resources for which there are a Hypertext Search Engine, and resources made available using

XML-gateway technology. External resources can include MARC records in OPACs, citation databases, digital collections, and web sites.

The ENCompass federated searching function enables users to search digital collections individually, select and search multiple collections, or search all collections. These collections may contain metadata only or metadata plus the digital objects. A strength of the ENCompass software is its ability to connect seamlessly to these collections whether they are stored on the project server or in remote repositories. Multiple ENCompass installations within Alabama make it possible to establish permanent, seamless connections between collections hosted on local ENCompass servers and the project server in Auburn. This capability allows the project to avoid the time and expense of moving digital files and their associated metadata into the central server.

The search engine returns merged and relevance-ranked result sets. For digital files stored on the central project server, the user can view a thumbnail image created “on the fly” by the software. For collections stored remotely, a link facilitates viewing the image.

For non-ENCompass sites, the Project has experimented with importing the local repository’s metadata into the central ENCompass server for searching. To date, this has been accomplished by having the local repository export metadata with object locations (the crucial Dublin Core data element is *Identifier*, the URL which points to the digital object) included, transforming the data into the appropriate format and importing into ENCompass using a bulk load process. CONTENTdm, a widely-used digital file management system, supports exporting collection metadata as tab-delimited or XML-formatted files. Other possible external sources include data created in Access or Excel and exported as XML or other database programs that can create these types of file extracts.

The Cornerstone Project is focusing on providing digital files for public access; it is not a storage facility for master files. Repositories may store only web-ready presentation images or objects on the project server. Each participating institution must accept responsibility for creating, storing, and maintaining its own high-quality digital master files. Currently, the project accepts GIF or JPEG images (maximum resolution: 150 pixels per inch, in color or greyscale) which have been optimized for web presentation. Other formats (e.g. Luna Insight, Mr. Sid, JPEG2000) require third-party applications for display, and these are not currently supported by the central site. ENCompass can present digital formats such as those for audio, video and text, but these may require the end-user to install plug-ins for the web browser.

Finally, the Project will make metadata available as XML files to metadata harvesters, in accordance with the latest version of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH, www.openarchives.org/OAI/openarchivesprotocol.html). Given the Project’s geographical focus, the most appropriate OAI harvester would be the server for the AmericanSouth project at Emory University in Atlanta, Georgia (www.americansouth.org/). In the words of its creators, “AmericanSouth is an on-going project undertaken at Emory University in collaboration with scholars from around the nation for the study of Southern culture. It seeks to improve access to digital resources on Southern histories and cultures by indexing metadata concerning research collections and by providing a discovery mechanism for primary resources relevant to this focused subject domain”. The AmericanSouth project is supported by a grant from the Andrew W. Mellon Foundation.

Digital production centers

Alabamians have a deep appreciation for the history of their state, evident in the many small archives and museums that can be found throughout the state. Several individuals own remarkable personal collections of historical documents and photographs that would significantly enrich a digital Alabamiana collection. In planning the digital project, planners recognized that much of the content will draw initially on the collections at the Alabama Department of Archives and History, Auburn University, University of Alabama, and Birmingham Public Library. Long traditions of collecting Alabamiana have enabled these repositories to accumulate historically significant collections. However, there are also rich treasures dispersed throughout the state in repositories that lack the capability to undertake major digital projects.

To help create the initial digital collection, the IMLS grant established two new digital production centers at Auburn University and University of Alabama; these centers joined an existing center at the Alabama Department of Archives and History. Together, these digital production centers are contributing 15,000 digital surrogates over the duration of the IMLS project – materials selected primarily from their own collections but also contributed by other repositories. They agreed to assist other repositories with planning and to help support local digital projects. As participants are accepted as Cornerstone contributors, they are invited and encouraged to visit one of the digital production centers to learn more about digitization, especially best practices for scanning, workflow, and quality control. Through September 2004, which is the end of IMLS funding, this includes some limited scanning services for Alabama repositories with significant Alabamiana materials to contribute to the database.

Initially, the project planned to create two identical centers at Auburn University and The University of Alabama. This would have enabled trainers to use either center interchangeably. However, after some discussion, the project coordinators agreed that different configurations would be even more useful, allowing the centers to collectively support a greater variety of scanning needs.

The center at the Alabama Department of Archives and History primarily uses a PULNiX MFCS-50 H/S Digital Overhead (planetary) System. This fixed-array device is capable of scanning 6 bit (16 greyscale levels) and 8-bit (256 greyscale levels) and bitonal images. System outputs are TIFF and Bitmap files. Bound, stapled, and oversized materials up to 22 by 34 inches can be scanned easily. Archives selected a planetary scanner because fragile items might be harmed by placing them on flatbed. Also many bound items cannot be opened flat without damage. Unfortunately, the PulNix is no longer manufactured. Archives also uses a flatbed MicroTek ScanMaker 9800 XL that will scan 12 by 16 inches area. A transparency adapter allows the scanner to be used for 35mm slides and some film.

The Auburn University digital production center supports high volume digitizing using flatbed scanners as well as imaging for oversize materials such as large color posters, architectural sketches, detailed maps, and drawings. Its Epson 1640XL flatbed scanners offer 1,600 by 3,200 dots per inch (dpi) scan resolution and 42-bit color depth for documents up to 12.2 by 17.2 inches. Some of the scanners are also equipped with automated document feeders which can handle up to 30 single page sheets. To digitize oversize materials, Auburn uses a Contex Cougar color scanner with 425 dpi optical resolution. Hardware-interpolated resolutions up to 2,400 dpi are possible. The Context

Cougar has a scanning width of 36 inches but accepts flat documents up to 43 inches wide with unlimited length. The scanner supports 24-bit color scanning, eight-bit greytone with 256 gray levels, and one-bit black-and-white imaging.

The University of Alabama digital production center selected a Better Light Super 6K-2 scan back camera on a stand to generate high resolution, first generation digital images. These cameras are capable of color, black and white, and infrared photographs of extremely high resolution. It offers 6,000 by 8,000 pixels in native CCD resolution with 24-bit RGB (274 MB in 48-bit RGB) or 9,000 by 12,000 pixels in Enhanced Resolution™. The camera can capture a large format image with all pixels in true RGB color, not software-interpolated color. With Enhanced Resolution™, the Super 6K-2 offers the capability to increase resolution in several additional increments up to 150 percent of the previous pixel dimensions and to create file sizes up to 2.25 times larger than the current models. Unlike post-capture interpolation, the enhanced resolution settings record additional image information during the scan for sharper, more detailed images. Using this camera with a copy stand assures high-quality images of often poor-quality originals and protects fragile materials during handling.

Phase 3: going public with the AlabamaMosaic web site

The project's technical infrastructure and guidelines were essentially in place by late 2003. At that point, focus shifted from building infrastructure and management process to creating the project's public presentation. This shift has concentrated on three areas: developing a public web presence with rich content, providing training around the state, and attracting new participants.

The first step in this shift was the design of an attractive, easy-to-navigate web site that would contain basic information about the project and allow users to browse the various ENCompass repositories and search for digitized materials submitted by participating institutions. To forge a closer public identity to Alabama, the web site is named AlabamaMosaic instead of Cornerstone and can be visited at <http://www.alabamamosaic.org>. This site was initiated for the public by Dr. Robert Martin, IMLS director, during the centennial Alabama Library Association conference in Montgomery in April 2004. In September 2004, when federal funding for the Cornerstone Project ceases, AlabamaMosaic will supersede the project title, Cornerstone, to become the program name (Figure 2).

The site is still a work in progress. To date, some 15 institutions in Alabama have submitted collection and digitization plans to the Project. Five of these – the Alabama State Council on the Arts, Auburn University, Huntsville Madison County Public Library, Air University at Maxwell-Gunter Air Force Base, and University of Alabama Libraries – have begun adding content to the site through the ENCompass workflow. Other participants are training staff and organizing their internal processes to begin adding digital files to the collection. When the federal grant ends in September 2004, top priorities for NAAL will be attracting and supporting new participants and adding new content to the AlabamaMosaic collections.

Content and organization

The AlabamaMosaic site designers at Auburn University decided to organize materials by general subject area. Specifically, the site uses the 14 subject categories

Interested in joining AlabamaMosaic?
Click here to contact the Project Coordinator

Partnering Institutions

- Alabama Department of Archives & History
- Alabama State Council on the Arts
- Auburn University
- Birmingham Public Library
- City of Mobile Municipal Archives
- Huntsville Madison County Public Library
- Jacksonville State University
- Maxwell-Gunter Air Force Base Air University
- Spring Hill College
- University of Alabama
- University of Alabama at Birmingham
- University of South Alabama

Welcome to AlabamaMosaic

AlabamaMosaic.org is the Web site for **The Cornerstone Project**. The Cornerstone Project is a collaborative initiative funded by a grant from the Institute of Museum and Library Services (IMLS) and administered by the Network of Alabama Academic Libraries (NAAL).

Its purpose is to make unique historical treasures from Alabama's archives, libraries, museums, and other repositories electronically accessible to Alabama residents as well as scholars throughout the world.

[More information on the Cornerstone Project...](#)

The Collections

AlabamaMosaic is a work in progress. A number of collections are available for browsing and searching. More collections will be added in the near future.

You can:

[Browse the collections](#)

Or:

Search the collections for specific subjects (examples: "sacred harp", "agriculture")

Acknowledgments

AlabamaMosaic would like to acknowledge the [following contributors...](#)

Auburn University Digital Library | 231 Mell Street | Auburn University, AL 36849-5606 | (334) 844-1731 | diglib@auburn.edu

Figure 2.
The AlabamaMosaic
web site

that appear on the web site of *The New Georgia Encyclopedia*, another state-specific digital project (www.georgiaencyclopedia.org/nge/Home.jsp). These categories are:

- (1) The Arts;
- (2) Business and industry;
- (3) Cities and counties;
- (4) Education;
- (5) Folklife;
- (6) Government and politics;
- (7) History and archaeology;
- (8) Land and resources;
- (9) Literature;
- (10) Media;

-
- (11) Religion;
 - (12) Science and medicine;
 - (13) Sports and recreation; and
 - (14) Transportation.

In ENCompass, *objects* (i.e. individual digitized items) and *containers* (i.e. groupings of digital objects) can be linked from the institutional ENCompass Repositories to subject categories, thereby making them more easily searchable and visible to users. Items can be linked to any number of subject categories as appropriate. They can also be un-linked by the originating repository or by the central site manager (for further editing or enhancement, for example).

Providing adequate subject access to materials in AlabamaMosaic is a program concern and a problematical issue. All project participants are urged to use Library of Congress Subject Headings (LCSH), but this may not be a realistic expectation for repositories without professional staff. To support assigning appropriate subjects, two lists of subject headings were compiled specifically for use with AlabamaMosaic. The first consists of chapter headings and article titles derived from *The Encyclopedia of Southern Culture* (Wilson and Ferris, 1989), in alphabetical order. The second list consists of headings from *Alabama Moments*, also in alphabetical order. The Project planners are aware that some Cornerstone participants (especially public libraries and local historical societies) may not want to use subject headings from controlled vocabularies in their metadata at all, but at least the lists of terms are available if they do (Figure 3).

It should be stressed that these content-related questions – subject categories, subject headings, recommended depth of indexing, etc. – are all open to discussion and modification. Similarly, the Project is in the process of setting up a working group to guide the future design and development of the web site, with representation from each of the coordinating institutions.

Training opportunities

NAAL and its member institutions are committed to supporting training to insure that digital collections meet national standards and best practices. To make it easier for repositories to know and follow these practices, the Cornerstone Project has developed four basic workshops on: planning digital collections, scanning, creating metadata, and copyright.

Planning digital collections

NAAL Director, Sue Medina, offers a basic planning digital collections workshop. It introduces digital collection development and projects and covers many of the issues addressed throughout this article. This planning workshop has been presented at the Baldwin County Public Library System, Athens State University, Florence-Lauderdale Public Library; Sheffield Public Library; Talladega College, University of Montevallo; University of South Alabama, University of Alabama, and Jacksonville State University. Each host site included representatives from the cultural institutions in its service area.


Full Record Display
Help | Exit

[Browse/Search](#) | [Find Resources](#)

You did a basic search for "jones"

[view relationships](#)
Next

The Andrew Jones Family on Their Farm

Caption: "The Andrew Jones family, who own and operate their own farm. Jones has twelve children and sends three (Lottie Bell, Sarah, and Bessie Mae) to school." Date: July 8, 1925.

Place: Alabama

Date: 1925-07-08

Creator: Unknown

LCSH: African Americans -- Alabama -- Pictorial works; African Americans -- Alabama -- Social life and customs -- 20th century; African Americans -- Education -- Alabama -- History; African Americans -- Land tenure -- Southern States -- History

Cornerstone: Agriculture

Cornerstone: Agricultural Extension Services

Cornerstone: Alabama

Cornerstone: Black Life

Cornerstone: Education, Black

Cornerstone: Family, Black

Cornerstone: Farm Buildings



[Click to enlarge](#)

Figure 3.
Part of a sample AlabamaMosaic record in ENCompass, showing thumbnail image, metadata, and subject headings

Primer on scanning

Todd Jennings and Beth Nicol of Auburn University developed a workshop on basic scanning techniques and principles. Built around a single scanner and workstation, the workshop concentrates on scanning techniques, presents participants with a set of digital files, and leads them through various image editing functions using standard image-editing software (Adobe Photoshop) (www.georgiaencyclopedia.org/nge/Home.jsp).

The Cornerstone Project has struggled with determining appropriate scanning guidelines and best practices for its diverse participants. Archives, some museums, and academic libraries with special collections generally intend to use high scanning resolutions and pixel depth to support archival preservation as well as use needs. These sites generally save an unaltered master TIFF, which is stored under archival

conditions. They retain a second master for presentation. This first derivative, generally a JPEG file, is saved as a working master and is used to create the presentation JPEG for public access (and, if needed, the thumbnail GIF). Other participants have indicated that they plan to create and store only JPEG files. Their digital files are primarily to promote remote access to their collections. They are unconcerned that they might need to re-scan originals sometime in the future. Further, they often lack the capability to store master TIFFs under archival conditions.

The Project recommendation is that each repository determines the scanning resolution that will best capture the original for subsequent presentation in digital format. The primary determining factors are quality of the original and anticipated needs of users. For example, high contrast black and white documents using modern typeface will require the lowest resolution. If these modern documents are widely available, then users may have access through multiple locations and the online “look and feel” may not be critical. However, faded documents written by hand with uneven surface pressure may require much higher resolution. If these documents are unique and the surrogate needs to be a faithful replication of the original, then the presentation file may need to preserve as much fidelity as possible.

Creating metadata

The Project uses the Dublin Core Metadata Initiative as the basic source for identifying and defining the 15 basic data elements (<http://dublincore.org/>). This is supplemented by the *Western States Dublin Core Metadata Best Practices* (www.cdphheritage.org/resource/metadata/wsdcmbp/index.html). Patty Pilkerton, Authority Control Librarian at the Mervyn H. Sterne Library, University of Alabama at Birmingham, developed the Project workshop, “Creating Metadata”, to train staff from local repositories in basic Dublin Core and project requirements. The workshop uses the ENCompass workform so that participants can begin adding metadata to the Project database immediately. To date, Creating Metadata has been presented at Auburn University, at the Alabama Library Association meeting in Montgomery, and Jacksonville State University.

Copyright

The Project presented its copyright workshop at the Alabama Library Association April 2004. Unfortunately, the attorney who developed the presentation recently accepted a position outside Alabama and is no longer available. An ambitious schedule is planned to offer these workshops throughout the state during 2005. The one difficulty is finding a presenter for the Copyright workshop.

Attracting new participants

Under the leadership of NAAL, the Project is campaigning to identify repositories of suitable materials and enlist their participation in AlabamaMosaic. These collections may be based at colleges, universities, or libraries, or they may reside in local historical associations, museums, or even private hands. The expectation – the hope – is that, as the project gains momentum and publicity, it will attract the attention of prospective participants and contributors. The Project is determined to broaden participation beyond libraries and is especially pleased by the participation of the Alabama Council on the Arts, a statewide arts program, and the Southeastern Indian Research Center,

based in Jacksonville State University. Their examples hopefully will encourage other cultural entities to seek to add their digital materials to AlabamaMosaic.

The Project is especially interested in preserving and making publicly available collections of materials on the history of the civil rights movement in Alabama. The history of the USA is incomplete without the civil rights movement, and the civil rights movement is practically synonymous with Alabama. From the Civil War, Reconstruction, and Jim Crow through the Great Depression, the New Deal, and the Second World War, to the first bus boycotts, protests, and marches of the 1950s and 1960s, Alabama was at the epicenter of the most important and far-reaching transformation in America's history. In the space of less than a hundred years, the state moved from being the cradle of the Confederacy to (as one chronicler has called it) "the cradle of freedom" (Gaillard, 2004). Unfortunately, some of the most important materials on the civil rights movement are currently housed in unsatisfactory conditions. Other materials, we must presume, remain undiscovered, undescribed, and unavailable to students, scholars, and citizens. We know that behind the famous names – Rosa Parks, Fred Shuttlesworth, Martin Luther King, Ralph Abernathy, George Wallace, Bull Connor – there are hundreds of other individuals whose stories, although just as compelling, remain untold and unknown to any but their family and friends. AlabamaMosaic can rescue these stories from obscurity and restore them to the national narrative.

Some groundwork for this effort has already been laid. "Transforming America: Alabama's Role in the Civil Rights Movement" is a joint venture of the Auburn University Libraries, the Auburn University Libraries Special Collections and Archives Department, and other institutions that have contributed metadata to the repository through a web-based workform (www.transformingamerica.org/index.html). Begun in 2000-2001, the project was suspended in the past year until a new director of library technology could be hired. The advent of AlabamaMosaic will provide a platform for accessing the Transforming America project and encouraging its future development.

The future

The Cornerstone/AlabamaMosaic Project has made remarkable progress in its three short years of existence. Still, it faces several difficult challenges in the very near future.

The first challenge is financial. The IMLS grant supporting the project through its design, inception, and public launch will end in September 2004. At that point, the project will have to move from soft money to permanent in-kind support, or develop another grant to assist with funding. In short, it will have to live up to the goals announced at the beginning of this article: namely, "to encourage libraries and other repositories to integrate digitization of materials into their usual and routine tasks" and "to effect systemic change in the way libraries and other repositories develop their collections and make resources available". For better or worse, reliance on grants and other forms of external financial support has become a quasi-permanent feature of the academic library landscape. It remains to be seen whether this project can buck that trend.

The second challenge is forging closer relationships between the Project's coordinating institutions – the Alabama Department of Archives and History, Auburn

University, and The University of Alabama – and the state’s historically black colleges and universities (HBCUs). These include three universities, Alabama A&M University, Alabama State University, and Tuskegee University, and five colleges: Concordia College, Miles College, Oakwood College, Stillman College, and Talladega College. In addition to addressing longstanding disparities in resources, the Project will provide a way for the HBCUs to collaborate and encourage access to their collections, while still maintaining control over the precious information, objects and materials in diverse formats held by those institutions. In our view, this makes collaboration all the more necessary and valuable. Indeed, the Project provides an excellent opportunity to achieve a stronger collaboration between historically white and historically black institutions in Alabama.

The third challenge is one already alluded to above: that is, simply getting institutions to participate and contribute materials regularly to the AlabamaMosaic collection. Giving workshops and presentations throughout the state should help raise the Project’s profile and attract new participants. So should enhancements that have been promised for the next release of ENCompass – in particular, the replacement of the ENCompass client with a web-based metadata workflow. We hope that this will make it easier for institutions with smaller collections to participate.

Along with the challenges; there are also opportunities. One of these is a nascent collaboration between the Cornerstone Project/AlabamaMosaic and *The Encyclopedia of Alabama*, another digitization project of Auburn University and the Alabama Humanities Foundation that is just getting underway (www.encyclopediaofalabama.org/). The *Transforming America* Project mentioned previously has the potential to become a much larger project involving the development of a digital collection from the Civil War through the Civil Rights era, and extending beyond the boundaries of the state of Alabama to include collections throughout the Southeast. Another opportunity is involving students at the School of Library and Information Studies at The University of Alabama in the design and management of the AlabamaMosaic web site, perhaps on a practicum basis. Not only would this help distribute the costs of AlabamaMosaic, it would give the students valuable experience with an established digital library project.

Alabama has a uniquely rich history. The Cornerstone Project and AlabamaMosaic are already making parts of that history accessible to students and scholars throughout the state and around the world. This is a long-term project, requiring a long-term commitment from participating institutions. Given the high level of interest that we have already achieved, we are confident in predicting that this commitment will be forthcoming.

References

- Gaillard, F. (2004), *Cradle of Freedom: Alabama and the Movement that Changed America*, University of Alabama Press, Tuscaloosa, AL.
- Wilson, C.R. and Ferris, W. (Eds.) (1989), *Encyclopedia of Southern Culture*, University of North Carolina Press, Chapel Hill, NC.



LHT
23,2

252

Received 27 August 2004
Revised 29 November 2004
Accepted 17 February 2005

THEME ARTICLE

The Columbia River Basin Ethnic History Archive

A tri-state online history database and learning center

Leslie Wykoff and Laurie Mercier

Washington State University Vancouver, Vancouver, Washington, USA, and

Trevor Bond and Alan Cornish

*The Washington State University Libraries, Washington State University,
Pullman, Washington, USA*

Abstract

Purpose – This paper aims to describe the three state digital archive collaboration between the Washington, Oregon, and Idaho state historical societies, and Washington State University, a multi-campus university.

Design/methodology/approach – The paper discusses the three major aspects of the IMLS Leadership project: the process of the museum/library collaboration, the technical aspects of the building, the digital archive and the metadata scheme, and the effort to engage the archive user with the past through interactive tutorials.

Findings – Building a contextually coherent digital archive from numerous geographically remote collections is possible to do given current meeting technologies and funding for travel. Project staff and partners need to be explicit and forthcoming about institutional and professional interests involved in collaborative projects. User focus groups found the site to be very engaging.

Originality/value – The paper describes the collaboration process, technically how the archive was built, and the development of the history tutorials and discussion board.

Keywords Libraries, Museums, Digital storage, Audio media, Oral history, United States of America

Paper type Case study

Introduction

The Columbia River Basin Ethnic History Archive (CRBEHA) project received an Institute for Museum and Library Services (IMLS) leadership grant in September 2001. The grant was titled: Engaging the Public with Digital Primary Sources: A Tri-State Online History Database and Learning Center. The institutions involved in the project were Washington State University Vancouver, Washington State University in Pullman, the Oregon Historical Society (OHS) in Portland, the Washington State Historical Society (WSHS) in Olympia, and the Idaho State Historical Society (ISHS) in Boise.

The project's goal was to identify, digitize, and describe library and museum collections concerning non-Native American ethnic groups of the Columbia River Basin, and to stimulate public use of and discussion of those sources. The collaborative project developed a digital archive that highlights the hidden ethnic histories of the



region, provides tutorials on how to research and interpret library and museum collections, and encourages public discussion of critical historical and contemporary issues concerning ethnic history. Although many institutions have embraced digital archives to make their collections more accessible, few have joined in multi-state efforts to combine resources concerning a specific topic to explore the medium's teaching potential. This is the first of a series of projects envisioned by these collaborating partners in the Pacific Northwest.

Along with the digital archive, the CRBEHA web site includes a tutorial module and a discussion board module (CRBEHA web site, available at: www.vancouver.wsu.edu/crbeha/). The tutorials provide the web site's users the tools necessary to become their own historians and encourage more sophisticated use and analysis of library and museum collections. Using the archive as the "research focus", the tutorials include junior high and high-school level lesson plans, articles written by project historians about how to analyze documents, oral history interviews, photographs and artifacts, and illustrations by project librarians on how to search the CRBEHA database and a library's history database. The discussion module is a moderated threaded discussion board that links historical research to contemporary issues about Columbia River Basin ethnic groups.

Collaboration

The IMLS award was in a category called "library museum collaboration", therefore the actual work of collaboration was an important process for the CRBEHA partners to study. What actually is the definition of collaboration? Michael Scharge (1995), in *No More Teams! Mastering the Dynamics of Creative Collaboration* states that:

... collaboration is the process of *shared creation*: two or more individuals with complimentary skills interacting to create a shared understanding that none had previously possessed or could have come to on their own.

Scharge also makes the point that "the linguistic root of communication is the Latin verb *communicare* – which does not mean "to communicate" but "to share". The CRBEHA partners and project directors discussed the meaning of collaboration at our earliest face-to-face meetings as we wrote the IMLS grant proposal. Our core understanding of "collaboration" for the CRBEHA project was that each partner and project director, as institutional representatives and as professionals, had specific reasons why we wanted to participate in the project and goals we wished to attain through the project.

It was very important in the CRBEHA collaboration that everyone's goals were understood to be of equal importance. The Oregon Historical Society was developing new emphases on partnerships, wanted to make their collections more visible, needed to digitize more of its collections, and was increasing its educational activities. The Washington State Historical Society hoped to demonstrate the educational potential of digital archives. The Idaho State Historical Society wanted to share its collections with a broader public and hoped to learn more technical skills in the process. The Washington State University Libraries Manuscripts, Archives, and Special Collections department wished to continue to expand its online collections. Laurie Mercier, co-project director and history professor at WSU Vancouver, wanted to create a learning space where primary sources were accessible along with background

materials and tutorials to help people learn to do good history research. Leslie Wykoff, co-project director and library and information technology director at WSU Vancouver, wanted WSU Vancouver to become a recognized archival site for southwest Washington historical materials. Throughout work on the CRBEHA two-year project, meetings (conference calls and videoconferencing) were held to brief partners on project progress, and to check in the collaborative effort. MaryAnn Campbell at OHS, Linda Morton-Keithly and Jody Ochoa at ISHS, and Gary Schalliol at the WSHS and their staffs participated in the original IMLS proposal design, nominated collections for inclusion in the CRBEHA archive, and organized their organizations' spaces and staff to work with the traveling project team.

The IMLS grant allowed the co- = project directors to hire staff and also to use specified Washington State University colleagues for the project. Robert Schimelpfening, CRBEHA Research Associate, is a doctoral student in religion with interests in history and librarianship. Mario Compean, Research Coordinator, is a doctoral student in American History with a specialty in Pacific Northwest ethnic history. Jeannette Altman, Principal Artist/Designer, is interested in turning very technically complex online projects into beautiful web sites. These staff members carried out the daily work of the project from selecting and digitizing materials to populating the database, from adding the metadata and writing tutorials to adding watermarks, from designing the web site to writing partner reports and lesson plans. The project was kept on track by meeting minutes and grant administration provided by Jane Sanders at Washington State University Vancouver.

Other Washington State University faculty members provided their expertise to the project team as in kind contributions to the grant. Nicole Campbell, reference librarian at WSU Vancouver, conducted CRBEHA web site usability focus groups for local ethnic group members, senior citizens, and teachers. Rick Sawyer, professor of education at WSU Vancouver with an interest in qualitative research, conducted the project's evaluations. Trevor Bond, the Special Collections Librarian in Manuscripts, Archives, and Special Collections, and Al Cornish, Systems Librarian, both at the Washington State University Libraries in Pullman, provided essential technical expertise and support, materials for the CRBEHA collections, and access to the CONTENTdm software and server.

The project directors extended the invitation to collaborate to all members of the CRBEHA project team and faculty colleagues. All the project's team members shared the professional goals each of us hoped to achieve by working on the project. There was a distinct climate of mutual respect for one another's goals among all the team members. One team member said to the project evaluator that project team members and partners were "flexible and cooperative. It felt a little like heading off into a mysterious place, like there was some risk, but everyone seemed rather up for the game".

Much collaborative work between the partners and the project team occurred in the fall and winter of 2001-2002 when both the material selection process and the metadata field names were determined. During this time, following the guidelines for nominating materials for scanning from the *Handbook for Digital Projects* (Sitts, 2000) the museum partners agreed to provide general inventories of their Columbia River Basin ethnic group holdings, noting specific collection strengths. The inventories were shared at a partners' meeting where the ethnic groups that were strongest in terms of both variety

of material and depth of collection between all the museums were nominated for inclusion in the archive. Once the specific ethnic group collections were nominated, a timeline for museum visits by CRBEHA project staff was established. Robert Schimelpfenig and Mario Compean, the traveling scanning and metadata crew, selected materials from each of the nominated collections at the partner museums. Another important collaboration was determining the metadata scheme. CONTENTdm uses the Dublin Core standard for the metadata. The WSU Vancouver project team members worked closely with the WSU Pullman project partners to draft a proposal on metadata field names for a multi-institutional, multi-material type digital archive for all CRBEHA partners to consider. The scheme was unanimously approved by all partners.

Throughout the CRBEHA project careful attention was paid to the collaborative process. Using Schrage's definition of collaboration being a "process of shared creation", collaboration for CRBEHA meant that each member of the project team, including the project partners, were all on the management team. All talents, all conceptualizations, all contributions were important for the project's development. Meeting minutes became critical for recalling our decisions since there was such a complexity of inter-related professional interests involved. This collaboration was not always easy because all of us are independent-minded academics, librarians, administrators, artists, philosophers, and historians. The CRBEHA project goals as outlined in the IMLS proposal were very useful in helping to focus our efforts. And, as collaboration thinkers might expect, each of us, with our individual intensely-held professional goals, was simply compelled to create our vision of an outstanding online ethnic history archive and learning resource.

Technical aspects

The OCLC/DiMeMa CONTENTdm multimedia software suite is used to support object and metadata management for the CRBEHA project. CONTENTdm supports the decentralized development and central housing of collection data, both of which matched well to the needs of this project. Importantly, the CONTENTdm toolset has been successfully used in several multi-institutional projects, including the Mountain West Digital Library (Arlitsch and Zick, 2002) and in Louisiana Library Network (LOUIS) member projects. To support collection building, the CONTENTdm Acquisition Station client software package was loaded onto a workstation at the WSU Vancouver Library. A CONTENTdm collection was created for each participating institution. At the central server located at Washington State University in Pullman, institutional collections (e.g. Oregon Historical Society, Idaho Historical Society) were created and populated with objects and metadata. Manuscripts, Archives, and Special Collections (MASC) at WSU in Pullman also contributed two existing CONTENTdm collections to the project.

To scan items selected for the project and to collect metadata, the traveling scanning and metadata team visited each of the participating institutions. During these visits to the Oregon, Washington, and Idaho Historical Societies, the team scanned objects with a Microtek 9600XL ScanMaker or employed a Nikon D1x Camera to photograph materials. The descriptive metadata was entered on site into a Microsoft Excel file on a Dell Latitude Lap top. Using the Acquisition Station software, this Excel file (as a tab-delimited output) and the digitized object files were imported into CONTENTdm

using its batch import capability. Collection objects and metadata were then transmitted from the Acquisition Station client at WSU Vancouver to the server supporting the project, which is housed at WSU Pullman. By using this model (a mobile scanning and metadata team complete with their own hardware), the project team ensured that collections could be contributed to the CRBEHA project from repositories without an established technical infrastructure. Even though the partners did not actually scan and describe the own collections, a critical component of the success of this methodology was the realization during the planning stages of the proposal that participating institutions would nevertheless be required to devote considerable staff time to establish workspace for the visiting crew and to select and retrieve materials, as well as to answer questions. To alleviate these staffing demands on the already inadequately staffed partners, funds for temporary employees were provided to them in the project's budget.

The CRBEHA search page (www.vancouver.wsu.edu/crbeha/browse.htm) provides an integrated search and retrieval interface for all six project collections (see Figure 1). To enable access to the CONTENTdm collections, project staff used the CONTENTdm Query Builder application to generate search

BROWSE THE ARCHIVE

Welcome to the Columbia River Basin Ethnic History Archive. We have several ways to search our database:

Search by ethnic group:

[African Americans](#) [Basque Americans](#) [Chinese Americans](#)
[German Americans](#) [Italian Americans](#) [Japanese Americans](#)
[Jewish Americans](#) [Mexican Americans](#) [Russian Americans](#)
[Native Americans](#) [Other Ethnic Groups](#)

Each of the above pages utilizes Flash. Get the Flash Player [here](#).
Audio files require Real One Player, available for download [here](#).

Search entire database by:

Keywords: GO

[View our list of Subject Headings](#)

Or use one of these predefined searches:

Search by Historical Society: GO
OR

Oregon Historical Society
Washington State Historical Society

Search by Material Type: GO
OR

Search by Date: GO
OR

Search by Subject: GO

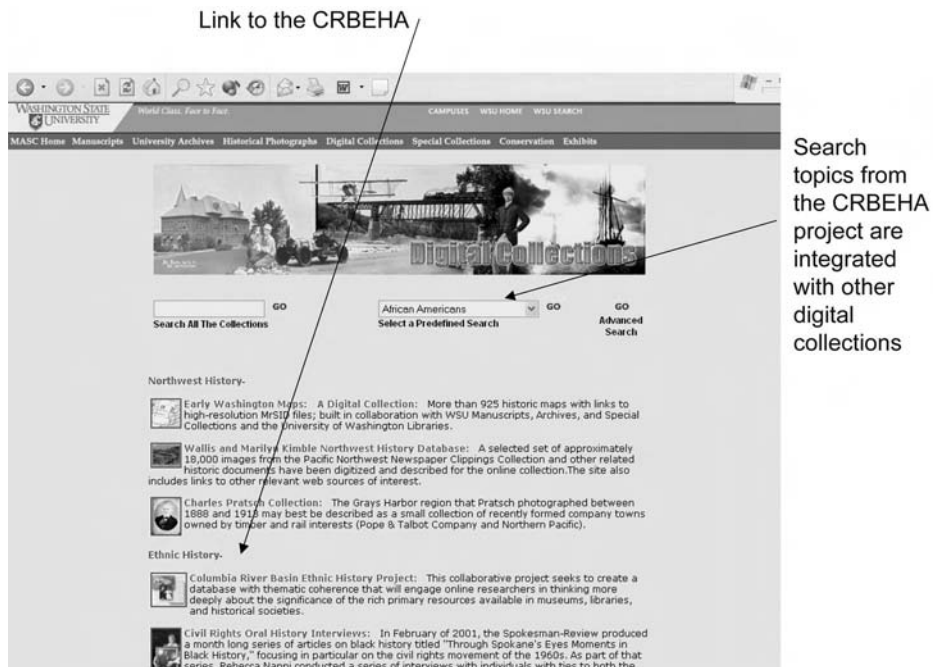
Need search tips or strategies? Visit our [Tutorial](#) section!

Figure 1. CRBEHA main search page. Search options include keywords and controlled vocabulary searches within ethnic groups, and also searches by museum collection, by date, or by material type

query strings. These query strings or database commands were then embedded into HTML documents that have the same look and feel as other CRBEHA web documents (Bond and Cornish, 2002). So, the user who enters keywords on the main search page is actually searching across six individual collections. By creating individual institutional collections, participants, such as the Oregon Historical Society, can also link directly to their own collections. This framework will allow for the addition of new collections from future partners.

All of the collections that form the CRBEHA web site are fully integrated in the Digital Collections page hosted by MASC at WSU Pullman. Visitors to the WSU Digital Collections site will find a link to the CRBEHA project under the heading Ethnic History, or can search across all of the collections hosted on the WSU Pullman Server. Topics covered in the CRBEHA site, such as African Americans and other groups, are listed in a drop-down list of pre-defined searches along with topics relating to the other WSU digital collections (see Figure 2).

An important advantage of CONTENTdm is that it supports collection dissemination using the Open Archives Initiative-Protocol for Metadata Harvesting. Using OAI-PMH, metadata for the CONTENTdm collections that make up the CRBEHA project is exposed so that an OAI Service Provider can harvest the metadata and integrate it with collection information from other institutions into a subject repository. The harvested information includes descriptive information and pointers to the original objects (images, streamed sound files) residing on a WSU Libraries server. For the CRBEHA project, metadata has been harvested into several repositories, including the Gateway to Cultural Heritage Materials housed at the University of



Search topics from the CRBEHA project are integrated with other digital collections

Figure 2.
The Washington State University Libraries Digital Collections. Link to the Columbia River Basin Ethnic History Archive. CRBEHA is part of the larger WSU digital collection

Illinois. By sharing the descriptive metadata of the CRBEHA project, users searching other sites, such as the Gateway to Cultural Heritage Materials[1], may discover the resources included in the CRBEHA project in addition to similar resources from libraries, museums, and other cultural heritage repositories from around the country. Sharing metadata in this way ensures another layer of back-up by distributing project data off-site. A further benefit for long-term sustainability of the CRBEHA project is the demonstration that all metadata can be successfully migrated to future systems independent of proprietary software (see Figure 3).

In addition to image files, oral history interviews were accessible from the CRBEHA site. To support this, oral history interviews were converted from analog formats and encoded into MP3 files. The digital audio files were then mounted on a Helix Universal Server streaming server, which assigned each audio file a unique URL (Bond, 2004). To catalog the streaming audio files, project staff entered the audio URLs to the appropriate CONTENTdm collection by selecting the import URL feature in CONTENTdm's acquisition station. This process launches a CONTENTdm metadata template so that project staff can add descriptive metadata that describes the audio file. Alan Cornish, Systems Librarian, also entered additional metadata into the Helix Universal Server. This information displays in the RealAudio player as the users listen to an interview.

Though users must download and install a RealAudio player, the use of the Helix Universal Server provides significant functionality including the ability for users to

The screenshot shows a Microsoft Internet Explorer browser window titled "UIUC Digital Gateway to Cultural Heritage Materials - Microsoft Internet Explorer". The address bar contains the URL: <http://nergal.grainger.uiuc.edu/cgi/bib/bib-idx?type=simple;sid=4e1af1b05990f0d0ffd82dce3c1fd2;xc=1;rgn6=identifier;q1=hispanic;rgn1=ent>. The main content area displays a record with the following metadata:

- Title:** Rena, senoreena, ma foreign lady
- Author/Artist:** E.K. [clay pot symbol]
- Subject/Description:** women | Songs with piano | Singers Hawthorne, Nellie | Songs | Society and Culture--Women | Discrimination--Hispanic
- Online Access:** [View Item](#)
- Collection:** Duke University Historic American Sheet Music Collection

Below this record, a RealAudio player is embedded, showing a play button, a progress bar at 0:01 / 6:11, and a volume control. The record continues with:

- Title:** Sample audio clip from an interview with Antonio Hernandez Rodriguez, November 1991
- Author/Artist:** Interviewee: Rodriguez, Antonio (Tony) Hernandez
 Interviewer: McDaniel, Patricia
- Subject/Description:** Mexican Americans
 Nampa (Idaho), Canyon County
 Social life and customs
 Manners and customs
 Celebrations
 Holidays
 Cultural relations
 Oral History
 Rodriguez, Antonio (Tony) Hernandez
- Online Access:** [View Item](#)
- Collection:** Washington State University Libraries

Next record:

- Title:** California as it is & as it may be; or, A guide to the gold region
- Subject/Description:** Felix Paul Wierzbicki (1815-1860) left his native Poland after participating in the doomed revolution of 1830. He made his way to America where he received a medical degree and practiced in Providence, Rhode Island. When the Mexican War broke out, Wierzbicki enlisted in the Army and was sent to California. Wierzbicki left the Army shortly after reaching the West and practiced medicine until the di...
- Online Access:** [View Item](#)
- Collection:** Perseus Digital Library

Final record:

- Title:** In San Domingo, 'heath stary skies
- Author/Artist:** Barbelle, Albert W., 1888-1957
- Subject/Description:** women | costume | Songs with piano | Music | Clothing & dress | Society and Culture--Women | Fashion and Costume | Legacies of Racism and Discrimination--Hispanic
- Online Access:** [View Item](#)
- Collection:** Duke University Historic American Sheet Music Collection

The browser's status bar at the bottom shows "Discussions not available on http://nergal.grainger.uiuc.edu/".

Figure 3.
A CRBEHA record
("Sample audio clip from
an interview . . .") within
the UIUC Cultural
Heritage Repository, an
OAI Service Provider site

listen to and navigate through extremely long interviews without actually downloading the file and the ability to fast forward, pause, and rewind the streaming files using the RealOne Player (RealNetworks Inc., 2002). The Helix Universal Server can also sense the speed of the user's connection and stream the selected file accordingly. Some participants in the CRBEHA project also liked being able to limit the ability of users to download the streaming files without formally requesting permission to do so.

Metadata scheme creation

In January 2002, Trevor Bond, Robert Schimelpfenig, and Leslie Wykoff drafted a data dictionary (or data schema) that outlined the use of the Dublin Core metadata in the project. This basic guide included three elements. The first element listed the field name as displayed in the public interface. The second element showed the corresponding Dublin Core field and included notes indicating whether the field would be searchable and hyperlinked. The third element of the data dictionary contained notes and examples for applying the descriptive information. After a period of review, the project adapted this data template and agreed on the display order of the Dublin Core elements. The metadata scheme was unanimously agreed on by the project's partners.

The Dublin Core identifier field was repeated so that identifiers used by partner institutions that describe the physical objects would be maintained in addition to those identifiers for the digital files. Given the collaborative nature of the project, participants were eager to have clear rights statements and repository information. One somewhat strange choice in the metadata was the treatment of circa dates. CONTENTdm does not elegantly support the searching of date ranges so we adopted a "work around" used by colleagues at the University of Washington. We included a date field that displayed to the public but was non-searchable. To allow for the retrieval of these circa dates, we added a second date field that did not display, but was searchable. For example if someone searched the collections for photographs from 1935, an image listed as ca. 1935 would display since the "hidden", but searchable date field included a string of dates within a decade of the circa date or 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939, 1940. As CONTENTdm continues to develop new features, we will revisit some of the metadata decisions, such as the treatment of circa dates.

For the subject terms, the Thesaurus for Graphic Materials (TGM) published by the Library of Congress served as the base controlled vocabulary (Library of Congress, 1995). In addition to the TGM vocabulary, the project staff created a controlled list of local terms that reflected current usage for describing ethnic groups, e.g. Basque Americans. A link – located just below the keyword search box – is provided that lists all the subject terms used in the project. This list also includes the locally created terms in bold (CRBEHA subject terms, available at: www.vancouver.wsu.edu/crbeha/subheads.htm).

Engaging database users with the past

From the outset of planning the project, we sought to make this web site more interactive than scores of digital archives inhabiting the internet. In recent years, many institutions joined the digital age by scanning and posting their collections, but without exploring the pedagogical possibilities of the medium's interactive dynamism.

As Carl Smith (1998) noted in the American Historical Association's newsletter, many web sites "seduce the senses without engaging the mind". While digitizing collections and making them more accessible through the web has been a laudatory goal, we wanted to push this accessibility further to more deeply engage users with the rich historical sources that the database would highlight. To paraphrase the historian Carl Becker, we wanted to help make "every [person] his/her own historian". To accomplish this, we created interactive tutorials that lead users through the steps of how to locate library secondary sources and archival primary sources, how to evaluate and interpret primary sources, and how to incorporate and teach library and museum resources in the classroom. By focusing on the ethnic history archive and creating an online discussion board, this process also implements another goal of the IMLS project: to increase public awareness of the region's diverse ethnic groups and their histories.

Users are introduced to the ethnic history of the Columbia River Basin in the historic overviews, which appear for each ethnic group in the "Browse the Archive" section. Combining selected photographs, written materials, oral history interviews, and artifacts from four major regional repositories enabled us to create a significant online database about an important and popular topic. This collaborative effort also allowed us to incorporate new information that the CRBEHA provided in original essays. Bibliographies, which suggest where to find more information and take users to the existing literature about Pacific Northwest ethnic groups, follow each overview. In addition to providing general background information about each ethnic group, the overviews highlight a variety of database items in order to pull the reader into exploring the collection (CRBEHA historical overview, available at: www.vancouver.wsu.edu/crbeha/ja/ja.htm). For example, in the Japanese American section, tantalizing images of a handbill from the Yasui Brothers' business, a photo of the Japanese baseball club of Nampa, and a letter to the Ed Tanaka family in Pocatello describing the residual postwar hostility towards Japanese Americans returning to their homes in the Puyallup Valley, invite users to click on the image and learn more about the particular source (see Figure 4).

The second major section of the web site, "Tutorials and Lesson Plans", is designed to deepen user engagement with the digital database. Here, users are introduced to the categories of primary and secondary sources and may select one of six tutorials under "Locating Materials" or "Interpreting Materials". Under Locating Materials are tutorials on searching the CRBEHA archive, using simple examples of Boolean operators. Also included here is a more complex strategy for searching a Pacific Northwest ethnic history question in the library database *America: History and Life*.

In tutorials that explain how to interpret written documents, oral history interviews, and photographs, users are encouraged to "read" these sources with a critical eye and consider factors and questions that go beneath the surface appearance. The tutorials attempt to problematize these sources by scrutinizing what and how museums and libraries collect. In researching ethnic history, reports produced by officials and observers outside the cultural group need to be examined in an appropriate context. For example, in the tutorial about interpreting written documents, users are asked to examine WPA and state of Washington reports on African American communities and folkways in Washington during the late 1930s, and a "Memory Book" produced by a black church in Hanford during the Second World War. After studying the documents, the user returns to the tutorial (www.vancouver.wsu.edu/crbeha/tutorials/int_text).


JAPANESE AMERICANS IN THE COLUMBIA RIVER BASIN

HISTORICAL OVERVIEW

[First Arrivals and Their Labors](#)
[Establishing Communities](#)
[Resisting Discrimination](#)
[Japanese American Associations and Culture](#)
[Japanese Americans and World War II](#)
[Recovering Community and Remembering History in the Postwar Era](#)
[Bibliography](#)

First Arrivals and Their Labors

Japanese immigrants first came to the Pacific Northwest in the 1880s, when federal legislation that excluded further Chinese immigration created demands for new immigrant labor. Railroads in particular recruited Issei –or first generation immigrants--from Hawaii and Japan. Thousands of Japanese workers helped construct the Great Northern, Northern Pacific, Oregon Short Line and other railroads in the Columbia River Basin. By 1907, the Japanese comprised about 40 percent of Oregon’s total railroad labor force. These workers commanded higher wages from railroad companies as the sugar beet industry began competing for their labor.



Interior view of a Japanese grocery in Portland, Oregon, 1919.

Japanese in larger cities like Portland provided rooming houses, restaurants, stores, social contacts, and employment services that helped new immigrants get established in the region. Shintaro Takaki came to Portland to sell Japanese goods to Chinese merchants and by 1889 had started a restaurant in the city. Takaki soon became a labor contractor and helped make Portland a center for distributing immigrant workers to fish canneries, farms, sawmills, and railroads throughout the Pacific Northwest. The

Figure 4.
Historical overviews introduce the user to each ethnic group and provide images from and links to specific items in the collection

htm) to think about a series of questions that probe the nature of the sources and ask for tentative interpretations and conclusions that might be drawn about African American communities in Washington during this time period (see Figure 5).

In the oral history tutorial, users are guided to examine a series of interviews, both aural and written excerpts, with Mexican Americans of the Snake River Valley of Idaho. Users return to the tutorial (www.vancouver.wsu.edu/crbeha/tutorials/int_oh.htm) to evaluate the differences between oral and written forms of the interviews; to explore the theme of cultural traditions and determine what connections and conclusions, if any, can be made about the origin, importance, and persistence of some of these traditions in various communities; and to interpret the dynamics of the interview as a creation at a particular time and place and with a particular interviewer.

Educators and historians have increasingly recognized the value of using primary sources to teach critical thinking and the ways historical evidence is used in formulating ideas about past events. A series of structured lesson plans for teachers of grades 6-12 address four curriculum topics – the history of African Americans in the Columbia River Basin, immigration and migration, immigration and settlement, and ethnic culture and identity – which model the kinds of lessons that teachers might prepare about a specific ethnic group or more comparative theme. Each lesson plan includes activities that take students directly to CRBEHA primary sources and encourage analysis and interpretation of those sources. For example, in one activity students read and critically analyze two oral history interviews selected by the teacher, and then evaluate the photographs of early African American settlers in the region.


TUTORIALS & LESSON PLANS

Locating Materials


[Using The Columbia River Basin Ethnic History Archives](#)

[Using History Databases and the Internet](#)


Interpreting Materials



[Interpreting Photographs](#)




[Interpreting Documents](#)



[Interpreting Artifacts](#)

COLUMBIA RIVER BASIN ETHNIC HISTORY ARCHIVES



TUTORIALS

Introduction to Primary and Secondary Sources

Historians use a variety of sources to try to make sense of the past. To begin their research, they read widely on their subject in the secondary literature, collect available primary information about the topic, and then select portions of their research materials that in particular illuminate the topic or research question. They sift through their notes, interpret the material in light of their experience and knowledge, and present their findings to a larger audience in books, articles, exhibits, films, web sites, and performances.

Historians usually group sources into two broad categories: primary and secondary. Think of secondary sources as "second-hand," that is, they are interpretations of topics by scholars and others, somewhat removed or distant from the subject. Secondary sources include books, journals, magazine or newspaper articles, dissertations, reports, and so on. Some of these sources at times can be considered primary sources. For example, a reporter's eyewitness account of a 1968 civil rights demonstration could be considered a primary source if published at the time of the event. A retrospective newspaper account of the same event, but published in 1998 on the thirtieth anniversary of the demonstration, would be considered a secondary source. The later account would be filtered through the decades and provide additional context, information, or biases not apparent in the more immediate 1968 report.

Primary sources are the "stuff" that make history, or the history that historians come to understand through these available sources. Primary sources are the "first-hand" observations or accounts of actual events.

Figure 5. The second major module of the CRBEHA web site, the tutorial section, instructs users how to "become a historian", or how to interpret documents, photographs, oral history interviews, and artifacts

Following these activities, students then examine two additional sources of their choosing. This process not only introduces them to new historical content, but also involves them in situating the sources in an appropriate historical context. In addition, each lesson plan relates content and activities to the Essential Learning requirements for social studies curriculum in Oregon, Idaho, and Washington. As a result of a teacher focus group created in the summer of 2003 to evaluate the effectiveness of the web site, one teacher remarked:

I learned what a great resource this site is for me and my students. It reinforced my desire to use more primary documents in the classroom . . . Something that I get frustrated with [in teaching] is the lack of textbook information about minorities in the Northwest. The books just don't tell their stories . . . This site allows me to have my students become historians and study the experience of minorities first hand.

Focus group members reported that the site not only allowed students to dig deeper into history, but also gave teachers a meaningful way to introduce discussion about diversity within a classroom.

Finally, the "Discussion Forum" module of the web site is the least realized module to date. We hope this interactive forum will provide teachers, students, and the general public a place to talk about discoveries made or dilemmas posed by items in the archive. We want to launch stimulating discussions about various aspects of ethnic history, including ethnic identities and race relations, work and labor, immigration and migration, discrimination and civil rights, and family life, religion, and social customs.

Ideally, this could be an online humanities program, a place for scholars and the general public to interact and create a community of learners, and a place where members of various ethnic groups might share their stories, information, and recollections. Employing a WSU-developed threaded discussion software program, the forum encourages discussions whereby participants can respond to one another's comments in a hierarchical or layered manner. Because of its accessible format, teachers can and have used the discussion board for encouraging students to post observations and interpretations of the materials they encounter in the digital archive.

Our inability to strengthen the discussion portion of the web site reflects a larger problem for institutions with limited resources to expand the educational possibilities of their digital archives. Initially, we had hoped we could employ graduate students to develop, stimulate use of and moderate the discussions. But a lack of funding and time has stalled this effort for the time being. Nonetheless, we hope that teachers will make use of the site and that special projects may be developed in the future to expand the general public's consideration, analysis, and discussion of ethnic history primary sources and issues. As one museum professional participating in a focus group noted:

The discussion boards will likely evolve as there is more participation. The contributions of users may provide interesting first-person accounts to supplement the prepared histories, thereby enlivening the site with current experiences and attitudes.

All of these educational components of the CRBEHA will stimulate more thoughtful consideration and use of the rich primary resources of the partnering institutions. We hope that by increasing the visibility of these collections and the subjects they illuminate, along with the growing public and school interest in the history and cultures of the peoples of the region, the project will expand collecting, preservation, and educational activities in this area. Moreover, our successful collaboration has motivated all the partners to agree to work again on a project that will combine our resources to illuminate a broad, central topic in regional history that has been "hidden" or underrepresented in the documentary and archival record.

Note

1. Gateway to Cultural Heritage Materials, available at: <http://nerval.grainger.uiuc.edu/cgi/b/bib/bib-idx>. Additional sites that have harvested metadata from the CBEHA site include Denmark's Electronic Research Library, available at: http://preprints.cvt.dk/cgi-bin/egw_sindap/1240/screen.tcl/name=find-a&service=prep&lang=eng; and The University of Michigan's Digital Library OAIster project, available at: <http://oaister.umdl.umich.edu/o/oaister/>.

References

- Arlitsch, K. and Zick, G. (2002), "Digital collection cross-server sharing: a new integrated model and technology for building and accessing digital libraries at multiple locations in Utah", Project Briefing at the Coalition for Networked Information Task Force Meeting, December 5, 2002, available at: www.cni.org/tfms/2002b.fall/abstracts/PB-Digital-ArlitschZick.html
- Bond, T. (2004), "Streaming audio from African American oral history collections", *OCLC Systems and Services*, Vol. 20 No. 1, pp. 15-23.

Bond, T. and Cornish, A. (2002), "Digitizing special collections using the CONTENTdm software suite at Washington State University Libraries", *Microform & Imaging Review*, Vol. 31 No. 1, pp. 31-6.

Library of Congress (1995), *Library of Congress Thesaurus for Graphic Materials*, Library of Congress, Washington, DC.

RealNetworks Inc. (2002), *RealServer Administration Guide RealServer 8*, RealNetworks Inc., Seattle, WA.

Schrage, M. (1995), *No More Teams! Mastering the Dynamic of Creative Collaboration*, (updated, rev. ed. of Shared Minds), Currency Paperback, New York, NY.

Sitts, M.K. (Ed.) (2000), *Handbook for Digital Projects: A Management Tool for Preservation and Access*, Northeast Document Conservation Center, Andover, MA.

Smith, C. (1998), "Can you do serious history on the web?", *Perspectives*, Vol. 36 No. 2, p. 5.



OTHER ARTICLE

Online library
databases

Accessibility and usability of online library databases

Ron Stewart and Vivek Narendra

Oregon State University, Corvallis, Oregon, USA, and

Axel Schmetzke

University of Wisconsin-Stevens Point, Stevens Point, Wisconsin, USA

265

Received 10 August 2004
Revised 16 August 2004
Accepted 1 December 2004

Abstract

Purpose – The Americans with Disabilities Act (ADA) mandates that all library services, programs and activities be accessible to people with disabilities. Web-based indexes and databases have become a major staple in today's academic libraries. This study aims to investigate the accessibility and usability of these online resources for people using adaptive software to interface with computers.

Design/methodology/approach – Database interfaces were evaluated with two different methods: Selected screens of the interfaces were checked for compliance with major Section 508 and web Content Accessibility Guidelines components; and involving minimally trained subjects to carry out routine search-related tasks, the interfaces were tested for functionality when used along with adaptive software commonly employed by people with print disabilities. Content accessibility within documents was judged based on document format.

Findings – The results indicate that, while most indexes and databases are now largely compliant with common accessibility standards and permit the performance of common search tasks, their actual user-friendliness for people with disabilities tends to be low.

Research limitations/implementations – Future research could benefit from closer attention to the degree of difficulty involved in performing search tasks and to the accessibility of document content.

Practical implications – Usability information, provided for each of the 37 evaluated databases, helps librarians make better informed procurement decisions, offers guidance to those vendors willing to design more inclusive products, and alerts those who assist users of screen readers or audio browsers to the potential difficulties involved in searching specific library databases.

Originality/value – This is the first published study that systematically explores the accessibility of the major library databases for people with disabilities.

Keywords Online databases, Disabilities, Indexing, Information networks, Worldwide web, United States of America

Paper type Research paper

Introduction

Students and instructors in higher education are expected to retrieve, use and apply information in their chosen field of study. In today's technology-rich education environment, much of this information is found in, or with the help of, web-based indexes and databases. General-subject databases that provide access to articles in magazines, newspapers and journals – such as Ebscohost's Academic Search and MasterFILE, Wilson's OmniFile and Readers' Guide, and Proquest's Newspapers – have become the bread-and-butter of lower-level undergraduate research.



Subject-specific online indexes, which have largely replaced their paper-based predecessors, are bibliographic tools essential for higher-level course work and faculty research. (For convenience, the term “databases” will be used in this article in its broadest sense, referring to both indexes and full-text databases.)

The migration of library indexes and databases to the online environment has created unprecedented opportunities for people with certain disabilities – people with “print disabilities”, who, because of lack of sight, cognitive disabilities (such as dyslexia) or insufficient motor control, cannot independently access printed works. Assistive input and output devices, such as modified computer keyboards, audio web browsers and screen readers with speech or Braille output, provide access to electronic text. However, the extent to which these new opportunities become realized depends, for a good part, on the design of the web environment. From the perspective of accessibility, design matters as much in the online world as it does in the physical world. Even the most advanced assistive devices cannot overcome the barriers associated with inaccessible design.

Features that make web pages accessible include, among others, alternative text (“ALT tags”) to describe the content or function of images, image map hot-spots, and animations; meaningful text for hyperlinks; proper mark-up for structural components, such as headers and list items; consistent page organization; verbal summaries of graphs and charts; and the selection of meaningful titles for frames. Two documents, both of which are considered authoritative, provide comprehensive guidance for the design of accessible web sites: the Web Content Accessibility Guidelines (WCAG) issued by the Worldwide web’s Web Accessibility Initiative (W3C/WAI) and the federal Access Board standards issued under Section 508 of the Rehabilitation Act of 1964, as amended in 1998 (www.w3.org/TR/WAI-WEBCONTENT/, www.access-board.gov/sec508/508standards.htm).

During the past five years, numerous articles, published in the education and library literature, have emphasized the need for accessible web page design and generated at least some awareness about this issue. However, ample anecdotal evidence suggests that, while librarians may acknowledge the need for accessibility with regard to their libraries’ own web pages, they rarely raise this issue when selecting electronic indexes and databases for procurement from outside vendors. One significant reason for this neglect is insufficient information about the competing products when it comes to the question of accessibility and usability. The vast majority of librarians have neither sufficient expertise with assistive technology nor the time to acquire it. They thus depend on others to provide this information.

It would be reasonable to expect that the database producers themselves would provide information about their products’ accessibility. While some database vendors have begun to post such information on their web sites, most have not done so. Those who do, typically focus on the conformance of their products to certain accessible-design standards, such as the federal Access Board standards (Section 508) (www.access-board.gov/sec508/508standards.htm, http://library.uwsp.edu/aschmetz/Accessible/websurveys.htm#vendor_info). However, none of the vendors report how their products perform when tested with actual screen readers or other assistive devices used by people with disabilities to interface with their computer. A recent survey by Byerley and Chambers (2003) confirms these observations. While increasingly, vendors pay attention to compliance with accessibility standards, only three of the eleven companies that responded to the survey reported that they conduct usability tests involving people with disabilities. The survey also revealed that most

database companies do not train their sales representatives to respond to accessibility-related questions. Clearly, librarians turning to vendors for input on which to base an informed decision are prone to be left empty-handed.

This study seeks to overcome the dearth of information on database usability for people with disabilities. More comprehensive and up-to-date than previous evaluation efforts, it provides information that should be of interest to both database developers and librarians. While the former may become inspired to make their products more accessible, the latter will be able to include accessibility among their criteria for selecting, or deselecting, indexes and databases.

Literature review

As already indicated earlier, the number of books, articles and web publications addressing the accessibility of web pages for persons with disabilities has increased steadily since the mid-1990 s. The gradually growing awareness that the Americans with Disabilities Act (ADA) also applies to the online environment, the revision of Section 508 of the Rehabilitation Act and heightened concerns within academia about the problems faced by individuals with disabilities may have attributed to this increase. Before the advent of the world wide web (WWW), most of the literature in the domain of accessibility focused on adaptive technologies (Lazzaro, 1993) and the human factors that are to be considered when designing these technologies (King, 1999). As the web has grown more popular, it soon became clear that poor design of web pages can seriously hamper access to page or site content. Today, several books, numerous articles and a plethora of web resources offer everything from quick tips, broad recommendations and highly detailed technical advice on how to design accessible web sites (e.g. Paciello, 2000; Thatcher *et al.*, 2002; Slatin and Rush, 2003; Schmetzke, 2004). Even the *Chronicle of Higher Education*, much respected and widely read in all academic circles, devoted several articles to the accessibility barriers faced by students with disabilities in the online environment (Young, 1998; Foster, 2001; Carlson, 2004).

Some six years ago, researchers began collecting data on web page accessibility in the post-secondary environment (<http://library.uwsp.edu/aschmetz/Accessible/websurveys.htm>). Among the first to do so, Rowland and Smith (1999) collected accessibility data from a random sample of colleges, universities and online learning institutions from all 50 states ($n = 400$). These researchers found only 22 percent of the home pages to be accessible. Summarizing the data pertaining to library-related web sites, Schmetzke (2002, 2005) reported a rather wide range in findings: At the library web sites evaluated between 1999 and 2002, 19 to 75 percent of the web pages were found to be free of major accessibility problems; among the main campus library web sites of the University of Wisconsin, the average number of errors per page varied between 1.3 and 6.1. While web accessibility had improved at some institutions, a comparison of 2000 and 2002 data that were collected from the same nationwide sample consisting of the web sites of 24 mostly larger academic libraries showed that, on the average, the percentage of barrier-free pages had actually slightly declined – from 59 to 53 percent (Schmetzke, 2003). A break-down of the web sites into those that had undergone a major redesign during the period in question, and those that did not, revealed that accessibility at the former had drastically declined (from 47 to 24 percent), whereas accessibility at the latter had improved considerably (from 68 to 81 percent). Apparently, web designers involved in the complete redesign of their web

sites tended not to pay attention to accessibility. That the occasion of a redesign can be an opportunity for improving accessibility was shown by a British study involving web accessibility audits at 11 higher-education web sites, at least some of which included “gateways” to information provided by libraries. Of the six sites that underwent a major redesign, all showed significant improvements with regard to the problems revealed during the audit (Sloan *et al.*, 2002).

Perhaps most remarkable is the fact that the web pages of accredited schools of library and information science (SLIS) – those institutions that train the next generation of librarians – tended to be particularly inaccessible. In 2002, only 30 percent of the SLIS pages (at US campuses) were free of barriers. Accessibility was barely higher, at 36 percent, at Canadian library schools (Schmetzke, 2003). The author reasons that, quite likely, these figures reflect wide-spread unawareness about the need for accessible design not only among the SLIS web designers but also among the library school faculty and staff, who hire the designers and give them direction.

With very few exceptions, web-page accessibility studies relied exclusively on the automated checking capabilities of Bobby, a tool originally created by the Center for Applied Special Technology (CAST) to help people evaluate the accessibility of their web pages. Based on the authors’ own experience at Oregon State University (OSU), such automated checking leaves at least 30 percent of significant access criteria unevaluated. Furthermore, as Schmetzke (2003) points out, the exclusive use of Bobby’s automated function for evaluative purposes is problematic since it produces some falsely-positive and falsely-negative results. For example, when encountering an image, Bobby will not report an error as long as some alternative text is provided – no matter how meaningless or nondescript this text may be; and pages that, at the very beginning, provide a “text-only version” link may not pass Bobby’s muster because it continues checking the graphics version. Even more vexing is Bobby’s inability to examine the functional impact of the errors found at a given site. From a functionality perspective, missing alternative text (or “ALT tags”) for a bullet, to name just one example, is very different from a missing text description for a crucial piece of graphical content (see also Diaper and Worman, 2003).

Until recently, information about the accessibility of web-based library resources, other than library web pages, has been scarce. Prior to 2002, little had been published in this area. Then, in 2002, *Library Hi Tech* came out with two special issues that included accessibility studies on selected web-based library information resources. Five studies (Bowman, 2002; Byerley and Chambers, 2002, 2003; Horwath, 2002; McCord *et al.*, 2002) examined the accessibility of online indexes and databases, including those with broad-range coverage, such as Ebscohost and Proquest Research Library, and those that are subject-specific, such as Medline and Cancerlit; one study (Coonin, 2002) focused on the accessibility of electronic journals.

Schmetzke (2002) summarized the findings in the above studies using a five-point scale ranging from “very accessible” to “absolutely inaccessible” – with accessibility “mildly reduced”, “significantly reduced” and “severely reduced” occupying the in-between points: Of the 14 resources investigated by Bowman (2002); Byerley and Chambers (2002, 2003); Horwath (2002); McCord *et al.* (2002), the majority (ten) fell in the “mildly reduced accessibility” category, while three were best described as having “significantly reduced accessibility”. Extreme cases were rare: Whereas only one index/database interface was described as “very accessible”, none of the resources fell into the “severely reduced

accessibility” or “absolutely inaccessible” categories. Three resources – Ebscohost, OCLC FirstSearch and Infotrac (Gale) – were evaluated by more than one research team. The findings for each of these three databases were close, but not identical – with those of Riley’s (2002) study suggesting slightly lower degrees of accessibility. Byerley and Chambers (2002, 2003) concluded their article with the observation that accessibility does not necessarily equate usability – that web-based products can be accessible in terms of Section 508 compliance without being user-friendly. While compliance with Section 508 is important, knowing “how screen readers render information is another important piece of the accessibility puzzle that must not be ignored”.

Coonin (2002), who examined the accessibility of e-journals of 11 providers found, with one exception, some major accessibility errors in the “basic access pages” (introductory, basic search or result pages) of all e-journals. Lack of ALT tags for images was the most frequent problem. Four e-journal providers failed to include ALT tags for some of the image-map hot spots. Another four vendors neglected to include titles for all the frames. Coonin also took notice of the format in which the articles themselves were provided. Six of the providers offered an accessible HTML format option for their e-journals. Of the five providers who did not offer this option consistently, three utilized a text-based PDF format, which, for the most part, is an accessible product. All of JSTOR’s, and some of Project Muse’s e-journal articles, were (at the time) provided in image-only PDF files.

The methods used by the above contributors to evaluate the accessibility of database interfaces varied from study to study. Two major approaches were used – sometimes in combination: compliance checking and functionality testing employing assistive technology. For the former, the authors selected major search-related screens and then checked their design for compliance with the Web Content Accessibility Guidelines – either “manually” or with the aid of Bobby’s automated checking function. For functionality testing, the authors selected some basic search-related tasks and then observed the problems encountered when attempting to perform these tasks using assistive technology (typically screen readers, such as Jaws and Window-Eyes). Three author teams, the members of which were all sighted, did these searches themselves; two research teams collaborated with sight-impaired users already familiar with the use of assistive devices.

In early 2003, not long after the publication of the two *Library Hi Tech* special-theme issues, Stewart (2003) published the results of an investigation, conducted in November 2002, that exceeded the scope of all previously reported studies: Using two audio browsers and three different screen readers, he evaluated all 36 vendor interfaces located on the web site of the OSU Library for usability by people with print disabilities. He discovered that, of the 36 vendor products evaluated, two (6 percent) were totally unusable, nine (25 percent) had significant accessibility problems, and only three (8 percent) provided a fully usable interface. In light of the findings of earlier, more informal investigations at OSU, he concluded that despite continuous improvements in the overall accessibility of these products, profound issues continue to exist with regard to their usability by persons with disabilities. In addition to usability testing, Stewart also checked the interfaces for compliance with OSU’s Web Accessibility Standards, a hybrid composed of major Section 508 standards and components contained in the W3C/WAI’s Web Content Accessibility Guidelines (www.w3.org/TR/WAI-WEBCONTENT/, www.access-board.gov/sec508/508standards.htm). Comparing his 2002 findings with similar unpublished compliance data collected in 1999, Stewart observed a “complete reversal in the inaccessibility of

the online library databases from approximately 95 percent inaccessible to 95 percent accessible". Like Byerley, he cautioned that accessibility, defined in terms of conformance to certain accessible design standards, does not automatically result in usability. As an example, he pointed to SilverPlatter's database interface which, for the most part, conformed to the federal Access Board standards (Section 508), but which was so poorly designed that it was extremely difficult to use with an audio browser.

Research scope and focus

This study follows the footsteps of Stewart's (2003) study. Unlike the patchwork of investigations previously conducted by other researchers, it is not only larger in scale but also more consistent in that it employs the same standards, tasks and tools to evaluate the accessibility of the various indexes and databases. More recent than the data collected by Stewart (2003) in November 2002, it provides librarians, vendors and assistive technology support staff with much needed current information about product accessibility.

Data for this study were collected between November, 2003, and January, 2004. All of the major research databases found on the OSU Library web site were included, with the addition of commonly used databases provided by three other companies: Emerald, ProQuest and Wilson, the products of which were accessed from the University of Wisconsin-Stevens Point Library web site. All in all, 37 different vendor interfaces providing access to over 120 databases were evaluated. An initial pilot investigation revealed no difference in accessibility among the various databases offered by the same vendor. Problems with searching and retrieving documents were found to be specific to a vendor's interface, not the particular database queried. In light of these findings, the study was limited to just one arbitrarily chosen database from each vendor.

For databases to be accessible for users of assistive computer technology, they need to be designed in accordance with certain accessibility standards. However, mere compliance with such standards is not enough. As was indicated earlier, an interface that conforms to certain accessible design standards is not necessarily easy to use with assistive technology. (The situation is not different for users without disabilities: While, in theory, they have access to the full content residing on a given site, poor design may render its use difficult.) This study looks at both aspects of interface accessibility: compliance with certain design standards and functional usability for those employing assistive technology. In addition, but to a more limited extent, this study also concerns itself with the format of the documents found in full-text databases. In summary, this study focuses on the following research questions.

- RQ1.* Do the database interfaces conform to recognized standards of accessible design?
- RQ2.* Are users of assistive computer technology able to perform basic search-related tasks (no matter how much effort this may require)?
- RQ3.* Which major difficulties are encountered by individuals who interact with the database interfaces through assistive technology?
- RQ4.* Is the content of the retrieved full-text documents provided in an accessible format?

Evaluation procedures

The evaluation was conducted on platforms found in most of OSU's Windows-based workstations with public access: Pentium 4 computer systems running Windows 2000 Pro or Windows XP Pro. The databases were evaluated from a network based connection, using Internet Explorer 6.0. At the time of the evaluation, all of the latest service packs had been installed on the workstations.

Two major evaluation procedures were employed in this study. In connection with the first research question, the simple and advanced search screens, as well as the screen displaying the retrieved results, were checked "manually" (i.e. without the help of an automated tool, such as Bobby) by checking the respective screens and, where necessary, the underlying HTML code for compliance with selected benchmarks contained in the 2002 draft of OSU's Web Accessibility Standards. These standards were composed of major Section 508 and Web Content Accessibility Guidelines components, requiring features such as alternative tags for images, short-cut links to pass over repeating navigation elements, and properly labeled tables. A complete list of these standards with brief descriptions is included in Appendix 1 (first section).

For functionality, or usability testing (*RQ2*), adaptive software products designed primarily for blindness were selected since it is in this area of disability that both database users and designers encounter the greatest challenge. A total of five products, which represent the leading Windows-based adaptive software, were chosen: two audio browsers, Connect Outloud 2.0 and Home Page Reader 3.0, and three screen readers, JAWS for Windows 4.50, Window Eyes Pro 4.2, and Supernova Pro 5.1. The versions used reflected the most current updates available on the OSU campus. The adaptive software was used in the default configuration, without any modification. This configuration was selected to reflect the typical situation in OSU's campus computer labs, where custom configuration or user modification of the installed software is not permitted. While functionality could have been increased in most of these products through the enabling of scripts and the customization of certain settings, unmodified use provided conditions that are likely to exist on many, if not most, of the nation's campuses.

The usability evaluation involved library research tasks typically undertaken by OSU's undergraduate students. These tasks, which were selected by OSU librarians from course assignments containing library-research components, required the evaluators to impose search limits, to enter a search term and submit the respective query (in both the basic and advanced search mode), and to browse through the query results (see Appendix 1, "Usability" section). The search tasks were considered accomplished if the evaluators managed, in some way, to get them done – regardless of the effort involved.

Particular interface-specific problems that were observed during the standards compliance checking or during the functionality testing (*RQ3*) were carefully documented. A brief summary of these observations is provided in Appendix 2.

Six hired students, all of whom were sighted, were trained by the main author to function as evaluators for both the standard compliance checking and the functionality testing. During the course of the training, these students gained a beginner's level of familiarity with the various adaptive software products, a level that is representative of the entry-level skills of the typical student with a disability enrolled at OSU. (Less than 5 percent of OSU students with disabilities have full mastery of their adaptive devices.) The work of the student evaluators was closely monitored through frequent spot

checking by two of the authors. Any major anomalies or failures were re-evaluated by the main author. A case in point is the evaluation of the SciFinder Scholar Access interface. After student evaluators reported a complete incompatibility with the assistive technology they were using, the main author verified this finding by testing this interface with several different assistive programs.

Content accessibility within a document (*RQ4*) was judged by the document's format. Text-based formats – no matter whether straight ASCII, HTML, RTF, or text-based PDF – were deemed accessible. Image-based formats, such as graphics-only PDF, GIF or TIF, were considered inaccessible.

Findings

Table I provides an overview of interface compliance with the selected web accessibility standards (*RQ1*). Product-specific information on this particular evaluation aspect is provided in Appendix 3 (top section), with further comments included in Appendix 2.

Table I reveals a rather mixed picture. With two exceptions, Micromedex and Scifinder Scholar Access, all databases were technologically compatible with the assistive technology employed, at least at a basic level. All but three databases (92 percent) allowed for navigation with keyboard strokes alone. Almost all vendors provided a standard layout that was both logical and consistent throughout their respective sites. Yet, navigation was not necessarily easy. About a third of the interfaces could not be navigated without special instructions. Only 11 percent of them provided links permitting users to skip over repetitive navigation components. Descriptive link text, as well as properly labeled dialog boxes and tables, were frequently neglected by vendors. Even though ALT tags are among the most obvious and the most easily implemented accessible design features, only 60 percent of the

Standard	Conformance (%)
Access to site (from on- and off-campus) possible	95
Ability to navigate sites without special instruction ^a	69
Site construction (logic and consistency) good/ok ^a	89
Standard layout throughout the site good/ok ^a	94
Meaningful ALT tags for all images and image-based elements ^a	66
Keyboard only navigation possible ^a	91
Skip-over-navigation links provided ^a	15
Link text descriptive ^a	60
No pop-up windows ^a	80
Dialog boxes labeled ^a	71
Tables properly labeled ^a	52
Usable with assistive technology (in default configuration)	95
Fully usable with scripting turned off ^a	71

Table I.
Percentage of database interfaces compliant with selected web accessibility standards

Notes: ^a Percentages reflect only those interfaces in which the specified feature was available; the figures do not reflect the fact that Micromedex and Scifinder Scholar Access could not be accessed at all with the assistive technology employed in this study.

Source: OSU's Web Accessibility Standards (2002 draft)

interfaces used them meaningfully and consistently. In some cases, ALT-tag related problems had drastic consequences for users, as for example in the case of Emerald's interface, in which the button-type images for adding to, and deleting from, the marked list were given the same ALT tag ("mark"). Of the databases, 29 percent were rendered unusable when scripts were disabled.

Table II shows how successful users of assistive technology were when performing selected core search functions in the 35 interfaces that were technologically compatible with screen reader software (i.e. all but Micromedex and SciFinder Scholar Access). Evaluators were able to perform the basic search functions – entering a query in the search field, submitting a query and limiting query parameters – in the vast majority (91 to 100 percent) of the databases. While limiting the query posed the most frequent problems, it affected only nine percent of the interfaces. Surprisingly, the percentages were even higher for advanced searches; the evaluators succeeded in entering and limiting a query in all the databases, and when they tried to submit a query they encountered problems in only three percent of the interfaces. Without exception, the query results (list of citations matching the search terms) were returned in an accessible format, and they were free of any extraneous coding that could have impeded their usability.

Readers must keep in mind that the above figures do not say anything about the ease with which specific search tasks could be performed. The data merely indicate whether the evaluators were able, somehow, to accomplish the tasks. The comments in Appendix 2 hint at the many difficulties that users of assistive technology have to grapple with. For example, in ABC-CLIO users could find a field for entering their search terms, but it was not until they received actual search results that they knew for sure that they had indeed stumbled on the correct input box. Web of Science products required users to navigate through a series of pages before they could actually submit a query. Repeated system errors and crashes, caused by compatibility problems with the screen readers, added further complications. SilverPlatter's frame-rich design suffered from poor linkage between the pages. Further confounded by the extremely complex page layout, site navigation in SilverPlatter, which provides access to some of the most commonly used databases, was extremely cumbersome. Like several other products, SilverPlatter also lacked full compatibility with the assistive technology used by the evaluators, who – in some instances – had to exit the site and then to restart their audio software in order to continue their research.

Criteria	Conformance (%)
<i>Usability</i>	
Basic search (where available)	
Entering query topics	97
Submitting query	100
Limiting query parameters	91
<i>Advanced search (where available)</i>	
Entering query topics	100
Submitting query	97
Limiting query parameters	100
Use of returned search results	100
Absence of extraneous data	100

Table II.
Percentage of database interfaces (not including the technologically incompatible Mircomedex and Scifinder Scholar Access interfaces) permitting users of assistive technology to meet selected usability criteria

Searches in 91 percent of the full-text database lead to articles provided in a primarily text-based format. However, sporadic checks of HTML-formatted documents revealed numerous instances of image-based components without proper ALT tags.

Among the vendors that provided image-based documents, JSTOR deserves special mentioning. JSTOR functions as an electronic archive for journal back issues – issues that, for the most part, have never existed in digital form. Faced with the task of digitizing masses of documents, they chose a compromise solution that was economically feasible, contractually permissible, and true to the authentic look-and-feel of the original print articles: By adding graphics-based TIFF files to their output options, the company provided a format that is easier to convert into text with current optical character recognition software (www.jstor.org/about/accessibility.html).

Discussion

At the surface, the findings include both expected and unexpected aspects. As previous studies already suggested, while complete inaccessibility is the exception, not the rule, few interfaces are perfectly flawless. While the vast majority of interfaces do well with regard to some accessible design standards (basic compatibility with assistive technology, provision for navigation by keyboard only, logical and consistent site construction, and standard lay-out throughout the site), most sites contain some features affecting their usability for people with disabilities. For example, one-third of the interfaces lacked meaningful ALT tags for all images and image-based elements, 40 percent were in need of descriptive link text, and 85 percent did not include mechanisms that would permit users to bypass repeatedly occurring navigation and page elements.

Based on the findings in previous studies, the cumulative effect of the various accessibility barriers could have led one to expect a significant detrimental impact on the interface functionality. *Prima facie*, this does not appear to be the case: Even in cases in which conformity to accessible design standards was relatively poor, the data often indicate that basic search tasks could be successfully performed nonetheless. There is a likely explanation for this apparent contradiction: In this study, the bar for performing a search-related task successfully was set extremely low. A task was recorded as having been successfully conducted, regardless of how cumbersome the process was – as long as the evaluator managed, in some fashion or another, to perform the specified task. The data thus do not reflect the many cases in which the performance of the task took on the character of a fishing expedition, where trial and error prevailed and confirmation of the moves as correct did not occur until some results were pulled up from the database's depth. Functional usability, as measured in this study, is thus not to be equated with user-friendliness.

With an eye to future studies of this kind, the question needs to be asked whether this study's approach toward evaluating usability is worth repeating. Does it make much sense to break down the overall search process into small isolated tasks and then to record the ultimate success or failure for each of these in mere binary terms? The experience gained in this study suggests going a different route: At a minimum, success for each task should not be treated as an all-or-nothing value but evaluated on a scale ranging from "very easy" to "impossible". Better yet, and more useful, would be the careful observation of the search process and the identification of the particular barriers encountered during each of the specified tasks. As documented in Appendix 2, the identification of specific barriers was part of this study; however, no systematic

attempt was made to associate a particular barrier with a particular search aspect. While general awareness about the partial lack of ALT tags in Bowker's interface, for example, is certainly of some utility for both its designers and professionals working with assistive technology users, knowing the particular instance or place where this problem occurs would be even more helpful – both for those seeking to correct the problem and for those planning instructions on how to use the interface.

A second self-critical point must be added. This study seems to suggest that most full-text database vendors (91 percent) furnish articles in a more-or-less accessible text-based format. While similar findings were reported by other researchers, a more careful collection of the data pertaining to document format may reveal a more differentiated picture, and possibly lead to a different conclusion – for the following three reasons: First, not all text-base document formats are accessible. Just like HTML-based documents, text-only PDF files may or may not be accessible with screen readers – depending, among others, on whether they were created with the proper tags that identify their structural elements (Slatin and Rush, 2003, Chapter 12). In order to gain a true picture of the accessibility of documents, even those that are text-based need to get closely scrutinized. Second, it is quite conceivable, if not likely, that document formats (and thus their accessibility) vary among the different databases offered by the same vendor. While vendors typically develop their own database interfaces, the different databases they offer may have been procured from different sources, each favoring a different format for their full-text documents. Thus, with regard to the accessibility of a vendor's documents, one cannot generalize the finding gained from examining just one of its products. Third, vendors aggregating their own full-text databases are at the mercy of the contributing journal publishers, who may provide the journal content in an accessible or inaccessible format. Since aggregators pull their content from numerous publishers, there is likely to be found variability in document format even within a given database. In order to arrive at a realistic picture about full-text accessibility within a given database, the investigation needs to involve a broader sample of its documents.

As argued above, future studies should take into account possible variability with regard to document format. This issue is not merely of academic interest. Accessibility advocates need to know at which point articles, most of which originate as word-processed documents on their authors' personal computers, are converted into inaccessible products so that pressure to provide accessible alternatives can be directed to the right parties. Strategically, it does not make much sense to complain to a vendor about the inaccessible format of a specific journal if it is the publisher who makes only this format available.

Conclusion

While this study is not the first to examine the accessibility of online library databases, it is the first print-published study that employs the same evaluation procedure to a large set of commercial interfaces. More recent than any of the other published data, the findings should be useful for a number of groups: The data in Appendix 3, combined with the specific comments in Appendix 2, may guide database and interface designers in the development of more accessible products. Librarians caring about accessibility have information enabling them to make better procurement decisions and to challenge vendors through specific questions about their products' functionality. Finally, the findings of this study may alert professionals who help

users of assistive technology to the potential difficulties involved in searching specific library databases.

As is the case with other computer-related products, the design of library database interfaces is in a permanent state of flux. For the sake of currency, studies like this one will have to be repeated periodically.

In addition to the improvements already suggested in the discussion section, future studies should also include the interfaces of federated search products, such as Exlibris' MetaLib and Endeavor's Encompass. As these products, with their meta-search functionality, gain in popularity, they will become the major gateways to online library information. Evaluating, monitoring and promoting their accessibility are therefore of utmost importance. If these tools are usable with assistive technology, design barriers in the native interfaces provided by the individual database vendors and aggregators can be circumvented. Due to their costliness, the procurement of federated search systems tends to involve a request for proposals (RFP). Librarians concerned about accessibility need to ensure that accessibility is addressed in the RFP. Competing vendors should be asked to provide documentation not only about their products' conformance to Section 508 standards, or some other comparable set, but also about their actual usability with assistive technology in common use by people with disabilities. It would not be unreasonable at all to ask vendors for a demonstration of this functionality.

References

- Bowman, V. (2002), "Reading between the lines: an evaluation of WindowEyes screen reader as a reference tool for teaching and learning", *Library Hi Tech*, Vol. 20 No. 2, pp. 162-8.
- Byerley, S.L. and Chambers, M.B. (2002), "Accessibility and usability of web-based library databases for non-visual users", *Library Hi Tech*, Vol. 20 No. 2, pp. 169-78.
- Byerley, S.L. and Chambers, M.B. (2003), "Accessibility of web-based library databases: the vendors' perspectives", *Library Hi Tech*, Vol. 21 No. 3, pp. 347-57.
- Carlson, S. (2004), "Left out online", *Chronicle of Higher Education*, Vol. 50 No. 40, pp. A23-5, available at: <http://chronicle.com/free/v50/i40/40a02301.htm> (accessed 11 July, 2004).
- Coonin, B. (2002), "Establishing accessibility for e-journals: a suggested approach", *Library Hi Tech*, Vol. 20 No. 2, pp. 207-13.
- Diaper, D. and Worman, L. (2003), "Two falls out of three in the automated accessibility assessment of Worldwide web sites: A-Prompt v. Bobby", paper presented at the 17th Annual Human-Computer Interaction Conference (HCI 2003), Bath, September 8-12, 2003.
- Foster, A.L. (2001), "Making web sites work for people with disabilities", *Chronicle of Higher Education*, Vol. 47 No. 21, pp. A30-2.
- Horwath, J. (2002), "Evaluating opportunities for expanded information access: a study of the accessibility of four online databases", *Library Hi Tech*, Vol. 20 No. 2, pp. 199-206.
- King, W. (1999), *Assistive Technology Essential Human Factors*, Allyn and Bacon, Boston, MA.
- Lazzaro, J.J. (1993), *Adaptive Technologies for Learning and Work Environments*, American Library Association, Chicago, IL.
- McCord, S.K., Frederiksen, L. and Campbell, N. (2002), "An accessibility assessment of selected web-based health information resources", *Library Hi Tech*, Vol. 20 No. 2, pp. 188-98.
- OSU (2002), *OSU's Web Accessibility Standards*, (draft), Oregon State University, Oregon.
- Paciello, M. (2000), *Web Accessibility for People with Disabilities*, CMP Books, Lawrence, KS.

-
- Riley, C.A. (2002), "Libraries, aggregator databases, screen readers and clients with disabilities", *Library Hi Tech*, Vol. 20 No. 2, pp. 179-87.
- Rowland, C. and Smith, T. (1999), "Web site accessibility", *The Power of Independence*, Outreach Division, Center for Persons with Disabilities, Utah State University, Logan, UT, Summer Issue.
- Schmetzke, A. (2002), "The accessibility of online library resources for people with print disabilities: research and strategies for change", in Miesenberger, K., Klaus, J. and Zagler, W. (Eds), *Computers Helping People with Special Needs*, Springer, Berlin, pp. 390-7.
- Schmetzke, A. (2003), "Web accessibility at university libraries and library schools: 2002 follow-up study", in Hricko, M. (Ed.), *Design and Implementation of Web-enabled Teaching Tools*, Idea Group Publishing, Hershey, PA, pp. 145-89.
- Schmetzke, A. (2004), "Accessible web page design: resources", available at: http://library.uwsp.edu/aschmetz/Accessible/pub_resources.htm (accessed 11 July 2004).
- Schmetzke, A. (2005), "Digitization of library information and its accessibility for people with disabilities", in Khosrow-Pour, M. (Ed.), *Encyclopedia of Information Science and Technology*, Vol. 2, Idea Group Reference, Hershey, PA, pp. 880-5.
- Slatin, J.M. and Rush, S. (2003), *Maximum Accessibility. Making Your Own Web Site More Usable For Everyone*, Addison-Wesley, Boston, MA.
- Sloan, D., Gregor, P., Booth, P. and Gibson, L. (2002), "Auditing accessibility of UK higher education web sites", *Interacting with Computers*, Vol. 14 No. 4, pp. 313-25.
- Stewart, R. (2003), "Accessibility of online databases: a usability study of research databases", Technology Access Program, Oregon State University, Corvallis, OR, available at: <http://tap.oregonstate.edu/research/ahg.htm> (accessed 11 July 2004).
- Thatcher, J., Waddell, C.D., Henry, S.L., Swierenga, S., Urban, M.D., Burks, M., Regan, B. and Bohman, P. (2002), *Constructing Accessible Web Sites*, Glasshouse, Birmingham.
- Young, J.R. (1998), "For students with disabilities, the web can be like a classroom without a ramp", *Chronicle of Higher Education*, Vol. 44 No. 27, p. A31.

Appendix 1. Accessibility evaluation criteria

Compliance with selected accessibility standards

Site accessibility and usability criteria. The site can be accessed both through the campus gateway and through a remote location using assistive technologies.

Ability to navigate site: The site can be navigated using the Assistive Technology being used. [Difficult – if the site can be navigated, but if it is complex and special instructions are required. Simple – if the site can be easily navigated using TAB and Arrow standard navigation keys.]

Site construction. The site is constructed in a consistent and logical manner

Standard layout. The pages within the site are standardized and navigation and presentation elements are consistently replicated.

ALT tags. Image and image-based elements have appropriate text associations that briefly describe the element.

Keyboard only navigation. The site and all its component parts can be fully accessed using only a keyboard.

Skip nav link. Repeated navigation and page elements can be passed over using a shortcut link.

Descriptive link text. Links on pages have appropriate text descriptions that clearly identify their function.

Pop-up windows. The site generates a secondary web page or scrip based presentation window.

Tables properly labeled. Data tables and table-based pages are properly labeled.
Usable with assistive technology. The site can be used with appropriate adaptive and assistive technologies.

No effect with scripting turned off. The site loses functionality or its presentation properties change if scripts are disabled.

Usability

Note: In this study, the basic and advanced search tasks were considered accomplished if the evaluators managed, in some way, to get them done. As explained in the main article text, a “yes” in Table II and Appendix 3 does not reveal anything about the difficulties involved in accomplishing the respective task.

(1) *Basic search:*

- Entering query topics – can query topics be entered in the appropriate dialog boxes?
- Submitting query – can a query be submitted to the database system?
- Limiting query parameters – can the query parameters be limited or expanded?

(2) *Advanced search:*

- Entering query topics – can query topics be entered in the appropriate dialog boxes?
- Submitting query – can a query be submitted to the database system?
- Limiting query parameters – can the query parameters be limited or expanded?

Returned information. Are the query results returned in an accessible and usable form?

Absence of extra data. Do the query results include no extraneous data or code that distracts from the information proper?

Accessible format. Are the query results returned in a text-based format, such as RTF, HTML, or text-based PDF?

Appendix 2. Specific evaluation comments

(1) *ABC-CLIO:*

- poor link descriptions;
- skip-over navigation links not provided;
- initial query buttons not properly labeled;
- link and dialog boxes combined and confusing; and
- site cannot be accessed if scripting is turned off.

(2) *Bowker:*

- ALT tags not always descriptive, or some of them are missing;
- non-essential page elements interfere with effective use and are not fully accessible or properly labeled; and
- site is unusable if scripts are turned off.

(3) *Brown University women’s writing project:*

- advanced search not possible with scripts turned off; site does inform user of issue.

(4) *Cambridge:*

- complex page structure may make site difficult for novice users;
- site is not structured using a logical tabbing order;
- in basic search mode you have to navigate back to initiate a search;

-
- links not always properly labeled;
 - individual records in query presentation are not clearly separated, but they can be determined by navigating through the query results; and
 - advanced search function is easier to use than basic search function.
- (5) *EBSCOhost*:
- Site does not clearly indicate which search mode you are in;
 - Can not select the specific search mode until you are into the search entry dialog.
 - The site layout and structure may be confusing to a novice user.
- (6) *HRAF Inc.*:
- quick search function is easy to pass over; and
 - site assistance is exceptionally good.
- (7) *Emerald*:
- skip-over navigation links not provided; and
 - alt tags not always descriptive, or some of them are missing.
- (8) *Engineering Information*:
- in basic search, query can not be entered without a mouse click on the text entry field; and
 - advanced search is fully accessible.
- (9) *First Search*:
- site has a tendency to loop and prevents moving back out of site to gateway; and
 - site defaults to advanced search mode on loading.
- (10) *ICPSR/University of Michigan*:
- site usability is degraded due to graphics layout.
- (11) *IEEE*:
- quick search requires scripts to be turned on;
 - query results are not presented if scripts are turned off; and
 - documents are returned in a variety of formats, PDF files are in PDF image format.
- (12) *Ingenta*:
- site requires that an individual user account be set up, a process that requires scripts to be turned on.
- (13) *IOP Axiom*:
- a separate “ada” compliant site must be used which is not as functional or responsive as the java-based main site;
 - connection problems with “ada” site that never seemed to occur with default site;
 - page does not read properly but is still usable; and
 - query results are provided in both HTML format and RTF.
- (14) *Illinois Researcher Information Service (IRIS)*:
- text-only site is much more usable; and
 - the telnet version was not accessible.

- (15) *JSTOR*:
 - very complex page selection process in basic search mode;
 - returned documents are scanned images and not accessible; and
 - TIFF format option – while not text-based, this format is easier to convert into text with OCR software than other graphics formats.
- (16) *LexisNexis Academic Universe*:
 - turning scripts off provides a more usable and informative site.
- (17) *Marcive*:
 - very usable plain HTML-based site.
- (18) *MathSciNet*:
 - no skip-over navigation links;
 - Micromedex;
 - requires a separate program; and
 - unable to install or use with AT.
- (19) *News Bank*:
 - site is very poorly constructed, but it is usable.
- (20) *NISC Biblioline*:
 - use of scripts makes site difficult to use with audio browsers;
 - pages are frame-based; and
 - with scripts turned off, pages require repeated refreshing.
- (21) *Online Computer Library Center (OCLC)*:
 - no skip-over navigation links.
- (22) *OID Technologies*:
 - complex site structure will require extensive user instruction;
 - site construction and button placement are counterintuitive;
 - poor element labeling makes navigation cumbersome; and
 - repeated AT software crashes when retrieving documents.
- (23) *Patent database*:
 - very extensive menus in some dialog boxes.
- (24) *Project Muse*:
 - not all images have alt tags;
 - no skip navigation links; and
 - some features of the site are unusable with scripting turned off.
- (25) *Proquest*:
 - text only site is much more usable.
- (26) *Pub Med*:
 - element descriptions and labels are confusing; and
 - highly recommend using the “text-only” site.

- (27) *RDS*:
- poor site construction site severely affects usability; and
 - improperly structured code results in numerous page errors.
- (28) *RLG's Eureka*:
- document retrieval requires frequent refreshing of page;
 - slowness of site affects usability when scripts are turned off; and
 - advanced search is more usable with scripting turned off.
- (29) *Roth's Publishing*:
- extensive navigation required to access query results; and
 - extensive dialog boxes and menu selections made advanced search very cumbersome.
- (30) *Science Online*:
- poor element labeling makes site very difficult to use;
 - site coding errors are distracting; and
 - site is extremely complex.
- (31) *SciFinder Scholar Access*:
- requires that a separate program be installed on computer; and
 - program interface is not usable with AT.
- (32) *SilverPlatter*:
- site launches a new browser with each operation;
 - pages are frame based and not properly linked;
 - very difficult to use site with audio browser;
 - extensive navigation required to get to query; and
 - site caused repeated failure in audio browsers.
- (33) *Standard & Poors*:
- site is constructed of improperly linked frames; and
 - once proper frame is located query is straight forward.
- (34) *University Microfilms Inc. (UMI)*:
- text-only site is much more usable;
 - accessibility help is provided in text site; and
 - page load errors required repeated refreshing of queries.
- (35) *Web of Science*:
- navigation required through several pages to get to search page;
 - search links are located at end of page;
 - site is extremely difficult to use;
 - site is unusable with scripting turned off; and
 - site caused repeated crashes of AT software.
- (36) *Wilson Web*:
- needs javascript support by the browser;
 - frames design is inaccessible with lynx; and
 - site is unusable when scripts are turned off.

Appendix 3. Accessibility evaluation findings

	ABC-CLIO	Bowker	Brown Univ.	Cambridge	EBSCOhost	eHRAF	Emerald	Engineering I.
Compliance with Selected Accessibility Standards								
Access to Site	YES	YES	YES	YES	YES	YES	YES	YES
Ability to Navigate Site	VERY DIFFICULT	SIMPLE	SIMPLE	DIFFICULT	SIMPLE	SIMPLE	SIMPLE	SIMPLE
Site Construction	GOOD	OK	OK	POOR	OK	OK	OK	OK
Standard Layout	GOOD	OK	OK	POOR	OK	OK	OK	OK
Alt Tags	PARTIAL	PARTIAL	YES	YES	YES	YES	PARTIAL, CONFUSING	NO
Keyboard Only Navigation	YES	PARTIAL	YES	YES	YES	YES	YES	NO
Skip Nav Link	NO	NO	NO	NO	NO	NO	NO	NO
DescriptiveText Links	SOME	SOME	YES	SOME	YES, BUT CONFUSING	NO	NO	YES
Pop Up Windows	YES	NO	NO	NO	NO	NO	NO	NO
Labeling of Dialog Boxes	POOR	SOME	YES	YES	YES	NO	YES	NO
Tables Properly Labeled	NO	NO	NO	NO	NO	YES	YES	NO
Usable with Assistive Tech.	YES	YES	YES	YES	YES	YES	YES	YES
Usable with Scripting Turned Off	NO	NO	YES	NO	YES	YES	YES	YES
Usability								
Basic Search								
Entering Query Topics	YES	YES	YES	YES	YES	YES	YES	NO
Submitting Query	YES	YES	YES	YES, BUT DIFFICULT	YES	YES	YES	YES
Limiting Query Parameters	YES	YES	YES	YES	YES	N/A	YES	YES
Advanced Search								
Entering Query Topics	YES	YES	YES	YES	YES	YES	YES	YES
Submitting Query	YES	YES	YES	YES, BUT DIFFICULT	YES	YES	YES	YES
Limiting Query Parameters	YES	YES	YES	YES	YES	YES	YES	YES
Returned Information Usable	YES	YES	YES	YES	YES	YES	YES	YES
Absence of Extraneous Data	YES	YES	YES	YES	YES	YES	YES	YES
Accessible Format (text-based)	N/A	N/A	YES	N/A	YES	YES	YES	N/A

(continued)

Figure A1.

	FirstSearch	ICPSR	IEEE	INGENTA	IOP Axiom	IRIS	JSTOR	LexisNexis
<u>Compliance with Selected Accessibility Standards</u>								
Access to Site	YES	YES	YES	YES	YES	YES	YES	YES
Ability to Navigate Site	SIMPLE	SIMPLE	SIMPLE	SIMPLE	SIMPLE	SIMPLE	SIMPLE	SIMPLE
Site Construction	OK	GOOD	OK	OK	OK	OK	OK	OK
Standard Layout	OK	GOOD	OK	OK	OK	OK	OK	OK
All Tags	YES	YES	YES	YES	YES	YES	YES	NO
Keyboard Only Navigation	YES	YES	YES	YES	YES	YES	YES	YES
SKIP Nav Link	YES	YES	NO	NO	NO	NO	YES	YES
Descriptive Text Links	YES	NO	YES	YES	YES	YES	NO	YES
Pop Up Windows	NO	NO	YES	NO	NO	NO	NO	NO
Labeling of Dialog Boxes	NO	NO	YES	NO	YES	YES	YES	YES
Tables Properly Labeled	YES	NO	NO	NO	YES	YES	YES	NO
Usable with Assistive Tech.	YES	YES	YES	YES	YES	YES	YES	YES
Usable with Scripting Turned Off	YES	YES	NO	NO	YES	YES	YES	YES
<u>Usability</u>								
Basic Search								
Entering Query Topics	YES	YES	YES	YES	YES	N/A	YES	YES
Submitting Query	YES	YES	YES	YES	YES	N/A	YES	YES
Limiting Query Parameters	YES	YES	YES	YES	YES	N/A	YES	NO
Advanced Search								
Entering Query Topics	YES	YES	YES	YES	YES	YES	YES	YES
Submitting Query	YES	YES	YES	YES	YES	YES	YES	YES
Limiting Query Parameters	YES	YES	YES	YES	YES	YES	YES	YES
Returned Information Usable	YES	YES	YES	YES	YES	YES	YES	YES
Absence of Extraneous Data	YES	YES	YES	YES	YES	YES	YES	YES
<u>Accessible Format (text-based)</u>	YES	YES	VARIES	N/A	YES	YES	NO	YES

(continued)

Figure A1.

Compliance with Selected Accessibility Standards	MARCIVE	MathSciNet	Micromedex	NewsBank	NISC Biblioline	OCLC	OVID	Patent Database
Access to Site	YES	YES	UNUSABLE WITH ASSISTIVE TECHNOLOGY (AT); AT CRASHES AS SOON AS DATABASE ACCESS IS ATTEMPTED	YES	YES	YES	YES	YES
Ability to Navigate Site	SIMPLE	SIMPLE		SIMPLE	VERY DIFFICULT	SIMPLE	SIMPLE	SIMPLE
Site Construction	OK	OK		POOR	OK	OK	OK	OK
Standard Layout	OK	OK		OK	OK	OK	OK	OK
Alt Tags	NO	YES		PARTIAL	YES	YES	YES	YES
Keyboard Only Navigation	YES	YES		YES	YES	YES	YES	YES
Skip Nav Link	N/A	NO		NO	NO	NO	NO	SOME
Descriptive Text Links	N/A	NO		NO	NO	NO	YES	YES
Pop Up Windows	NO	NO		NO	NO	NO	NO	NO
Labeling of Dialog Boxes	NO	YES		YES	YES	YES	NO	YES
Tables Properly Labeled	YES	NO		NO	YES	YES	YES	YES
Usable with Assistive Tech.	YES	YES		YES	YES	YES	YES	YES
Usable with Scripting Turned Off	YES	YES		YES	NO	YES	YES	YES
Usability								
Basic Search								
Entering Query Topics	YES	YES		YES	YES	YES	YES	YES
Submitting Query	YES	YES		YES	YES	YES	YES	YES
Limiting Query Parameters	YES	YES		NO	YES	YES	NO	YES
Advanced Search								
Entering Query Topics	YES	YES		YES	YES	YES	YES	YES
Submitting Query	YES	YES		YES	YES	YES	YES	YES
Limiting Query Parameters	YES	YES		YES	YES	YES	NO	YES
Returned Information Usable	YES	YES		YES	YES	YES	YES	YES
Absence of Extraneous Data	YES	YES		YES	YES	YES	YES	YES
Accessible Format (text-based)	N/A	N/A		YES	N/A	YES	N/A	YES

(continued)

Figure A1.

	Project Muse	Proquest	PubMed	RDS	RLG's Eureka	Roth's Publ.	Science Online	SciFinder S. A.
<u>Compliance with Selected Accessibility Standards</u>								
Access to Site	YES	YES	YES	YES	YES	YES	YES	UNUSABLE DUE TO MULTIPLE ACCESSIBILITY PROBLEMS WITH INSTALLATION AND USE OF PROPRIETARY SOFTWARE
Ability to Navigate Site	SIMPLE	VERY DIFFICULT	DIFFICULT	DIFFICULT	SIMPLE	SIMPLE	DIFFICULT	
Site Construction	OK	Good	OK	OK	OK	OK	OK	
Standard Layout	OK	OK	OK	OK	OK	OK	OK	
Alt Tags	PARTIAL	PARTIAL	ONE MISSING ALT TAG	NO	NO	YES	YES	
Keyboard Only Navigation	YES	YES	YES	NO	YES	YES	YES	
Skip Nav Link	NO	YES	NO	NO	NO	NO	NO	
Descriptive Text Links	NO	NO	YES	YES	YES	YES	YES	
Pop Up Windows	YES--HELP PAGES	YES--HELP PAGES	NO	NO	NO	NO	NO	
Labeling of Dialog Boxes	YES	YES	YES	YES	YES	NO	YES	
Tables Properly Labeled	YES	YES	YES	NO	N/A	YES	NO	
Usable with Assistive Tech.	YES	YES	YES	YES	YES	YES	YES	
Usable with Scripting Turned Off	PARTIAL	YES	YES	YES	SITE SLOWS	YES	YES	
<u>Usability</u>								
Basic Search								
Entering Query Topics	YES	YES	YES	YES	YES	YES	YES	
Submitting Query	YES	YES	YES	YES	YES	YES	YES	
Limiting Query Parameters	YES	YES	YES	YES	YES	YES	YES	
Advanced Search								
Entering Query Topics	YES	YES	N/A	N/A	YES	YES	YES	
Submitting Query	YES	YES	N/A	N/A	YES	YES	YES	
Limiting Query Parameters	YES	YES	N/A	N/A	YES	YES	YES	
Returned Information Usable	YES	YES	YES	YES	YES	YES	YES	
Absence of Extraneous Data	YES	YES	YES	YES	YES	YES	YES	
Accessible Format (text-based)	YES	YES	N/A	YES	N/A	YES	YES	

(continued)

Figure A1.

Figure A1.

	SilverPlatter	Standard and Poor's	UMI	Web of Science	Wilson Web
Compliance with Selected Accessibility Standards					
Access to Site	YES	YES	YES	YES	YES
Ability to Navigate Site	VERY DIFFICULT	DIFFICULT	SIMPLE	DIFFICULT	QUITE DIFFICULT
Site Construction	OK	OK	OK	POOR	POOR
Standard Layout	OK	NO	OK	OK	OK
Alt Tags	YES	YES	YES	YES	YES
Keyboard Only Navigation	YES	YES	YES	YES	YES
Skip Nav Link	NO	NO	NO	NO	NO
Descriptive Text Links	YES	YES	YES	YES	YES
Pop Up Windows	YES	NO	YES	NO	YES
Labeling of Dialog Boxes	YES	YES	YES	YES	YES
Tables Properly Labeled	NO	YES	YES	NO	YES
Usable with Assistive Tech.	YES	YES	YES	YES	YES
Usable with Scripting Turned Off	NO	NO	YES	NO	YES
Usability					
Basic Search					
Entering Query Topics	YES	YES	YES	YES	YES
Submitting Query	YES	YES	YES	YES	YES
Limiting Query Parameters	YES	YES	YES	YES	YES
Advanced Search					
Entering Query Topics	YES	YES	YES	YES	YES
Submitting Query	YES	YES	YES	YES	YES
Limiting Query Parameters	YES	YES	YES	YES	YES
Returned Information Usable	YES	YES	YES	YES	YES
Absence of Extraneous Data	YES	YES	YES	YES	YES
Accessible Format (text-based)	N/A	YES	YES	N/A	YES



OTHER ARTICLE

The evolution of electronic reference sources

Electronic
reference sources

287

Amy S. Van Epps
Purdue University, West Lafayette, Indiana, USA

Received 10 August 2004
Revised 4 October 2004
Accepted 14 November 2004

Abstract

Purpose – To provide a historical look at the development of web versions of reference materials and discuss what makes an easy-to-use and useful electronic handbook.

Design/methodology/approach – Electronic reference materials were limited to handbooks available on the web. Observations and assumptions about usability are tested with an information retrieval test for specific tasks in print and online editions of the same texts.

Findings – Recommended adoption of those elements which create a well-designed book in combination with robust search capabilities and online presentation result in an easy-to-use and useful electronic reference source.

Research limitations/implications – The small sample size that was used for testing limits the ability to draw conclusions, and is used only as an indication of the differences between models. A more thorough look at difference between electronic book aggregates, such as ENGnetBASE, Knovel[®] and Referex would highlight the best features for electronic reference materials.

Practical implications – Advantages to particular models for electronic reference publishing are discussed, raising awareness for product evaluation. Areas of development for electronic reference book publishers or providers are identified. Work in these areas would help ensure maximum efficiency through cross title searching via meta-searching and data manipulation.

Originality/value – The paper presents results from some human computer interaction studies about electronic books which have been implemented in a web interface, and the positive effects achieved.

Keywords Reference services, Information retrieval, Electronic media

Paper type Technical paper

Introduction

The ease with which electronic reference materials can be used is increasing, as development moves beyond a static representation of the print edition and begins exploiting the capabilities of the electronic environment. Presented here is an overview of the history and development of web-based handbooks and what may be driving some of the decisions behind the online options some publishers are offering. Technology has progressed to a point where electronic information is at less of a disadvantage and more materials are becoming available online (Webster, 2003). User preference seems to be for electronic information whenever they can get it (Lehto *et al.*, 1995; Marcum, 2003; Gray and Langley, 2002). Development of online materials during the last eight years, particularly reference materials, compelled the discussion of why people would use an electronic version that appears to take longer to access than the print, and may not be as easy to use. When a resource is available on the desktop it can save a trip to the library, and therefore be perceived as saving time. This article



examines why electronic may be preferable in some situations and not others. Discussion is not restricted to locating information online, but includes reading and using the data as well.

The literature about electronic resources includes articles discussing user interface and usability, (Tennant, 1999; Nielsen and Norman, 2000; Thong *et al.*, 2002) commentary or reviews of particular products, (Gibson, 2002; Kirkwood, 1996; DePetro, 2000; Arnold, 2004) and bibliographies with commentary about those reference materials which are available in electronic format (Juhl, 2001; Wilkinson, 2002; DiBianco and Chapman, 2003). Discussion of the best format for particular types of information, or why publishers would choose to pursue the electronic environment can be found in the Virtual Book and WEB book experiments (Landoni *et al.*, 2000; Wilson *et al.*, 2004). Materials currently available electronically are largely the result of transferring print materials to an electronic format, with the occasional inclusion of information that cannot be incorporated in print, such as music. Culp (2002) discusses the new life given to several multiple volume reference sources in chemistry, by a switch to a searchable electronic format. Current electronic reference materials show that unwieldy print sources can become usable and user-friendly in electronic format and thereby gain speed of use and functionality. The nature of electronic resources will continue to change and develop as the materials are natively produced in electronic format (Lynch, 2003).

Background

For the purposes of this article, electronic resources refer to electronic versions of reference sources, primarily handbooks, and specifically web-based versions. CD-ROM versions will not be considered for a number of reasons. CDs typically run on stand-alone, dedicated computers in the library, and thereby do not address the desktop delivery needs of many of today's users. In some instances a CD resource is added to a campus network, but there is still a need for the user to install software to ensure the native interface will run properly as a remote application. Additionally, CDs often use programs that have been written specifically for a given computer type and are designed to access specific data, which can allow for more programming flexibility than is available in a web interface. Comparisons between the two technologies are difficult to make.

Handbooks are often described as a single volume, written for practitioners in a field to be used as a quick resource for facts, figures and equations relevant to their discipline. Handbooks are most common in scientific and technical fields of research. Other types of handbooks exist, but a single volume or two-volume set is the primary content focus for this discussion. Handbooks are a useful tool for finding an item of information when the book is easily accessible (i.e. sitting within easy reach, for example on a shelf over the researchers' desk or in their lab). As pointed out by Culp (2002), "[handbooks] are used for consulting, not reading". This consultative use is referred to as "reading-to-do" use by Lehto *et al.* (1995) and is the reason Landoni *et al.* (2000) used scientific books for their experiments, which are typically used in this fashion as opposed to being read linearly. Handbooks become inconvenient to use when they are not easily accessible (i.e. down the hall or even several buildings away in the library). Without a convenient resource a user faces a decision: whether the time it would take to make a trip to the library to find the information is worth the effort, or if

there might be another way to get what they need. A nearby resource will seem more convenient, even if completely unsuited or inefficient for the task-at-hand. With ubiquitous Internet access in academic settings, the trend is for a researcher to go online to find the needed information. If handbooks are in electronic format, researchers can have access from their desktop computer, thus eliminating the proximity issue and increasing the ease of use. Evidently, since libraries are showing lower gate counts, (i.e. numbers that say just how many people are entering the library) people are no longer going to the library to take the book off the shelf. However, if the electronic interface is cumbersome, not only is it faster to visit the library to use a print copy of the handbook than work through the one online, it could forever turn a user off the resource (Nielsen and Norman, 2000). Thus, more convenient and user-friendly ways of delivering the library collection must be built.

Perceived ease-of-use or availability is only part of the usability equation. For some people, the desire to find the information online outweighs the need to visit the library (Webster, 2003; Subramanian, 1998; Gray and Langley, 2002; Marcum, 2003; Landoni *et al.*, 2000). Lehto *et al.* (1995) demonstrated a user preference for electronic text with online books using hypertext, even for tasks shown to be poorly suited to electronic presentation. For those people, any progress in delivering resources online will be accepted as a positive change.

Added ease-of-use and functionality of electronic books has been shown to generate a positive user response (Catenazzi and Sommaruga, 1994; Landoni *et al.*, 2000; Wilson *et al.*, 2003). Combine the added functionality with the advantages gained from one company producing and providing the interface for a collection of electronic handbooks and a model such as Knovel[®] appears. This model combines an ability to search any word on any page of a collection of handbooks, familiar page representation via Adobe Acrobat PDF and the ability to select and manipulate data from handbook tables. The model represented by Knovel[®] is a natural evolution of the positive elements of an e-book as shown in the WEB book experiments. These include user requested enhancements (Landoni *et al.*, 2000; Wilson *et al.*, 2003), which result in users being able to retrieve exactly what they need with a minimum of scanning to find their information. Many of these functions were available with custom programs years ago, (Catenazzi and Sommarugs, 1994; Lehto *et al.*, 1995) but only in the last few years have all the pieces been available for this type of service to be web accessible in a cross browser compatible fashion. Users can browse the pages of the text, navigate via the table of contents or index if desired, while choosing to search will return information presented anywhere on a page, including table captions and column headings. Specific advantages presented by the Knovel[®] model include interactivity, such as an equation accepting direct user input at the click of a mouse (Arnold, 2004). Similarly, active tables allow users to manipulate and extract data quickly and thereby save time in data analysis (Gibson, 2002).

As computers have gotten faster and web response times have sped up, loading large PDF files has become less time consuming. Given that PDF can be cumbersome to use, it may be surprising that it is used heavily in electronic resources. The main reason for this is that it “reproduces a book’s graphics, page layout, fonts and other elements with high fidelity”, which is exactly what Adobe developed the product to do (*Publishers Weekly*, 2000). Huttenlocher and Moll (2000) make the case that the graphic layout and design of the paper page is important in conveying the information and

meaning, so pages need to be reproduced accurately. Adobe's intent when developing the Acrobat software during the early, 1990s was to create a method by which people in a company could share a document, review it and mark changes in it and all see exactly the layout and font the original author intended, regardless of what computer platform they used to read the document (Carlton, 2003). Len Kawell, director of e-book development at Adobe Systems, Inc., believed the high fidelity reproduction of a page makes PDF the ideal format for reference sources, (*Publishers Weekly*, 1998) although others disagree. If the layout is truly essential to the presentation and understanding of the information, then PDF, TIFF or DigiPaper is well suited to the task. If information can be enhanced by being presented in tables where the user can manipulate the data, or combine information from different locations in one book, then layout takes a secondary role to usefulness (Webster, 2003).

Web capability is the lowest common denominator for widespread accessibility where there is likely to be a mix of Wintel, UNIX and Mac computers. This level of availability makes the web the preferred delivery mechanism for electronic content. Development of interactive, easily usable web versions of handbooks has taken longer than for other reference sources due to the pace of web development. Publishers appear to have been waiting for web technology to catch up with programming capabilities in multiple platform and browser friendly environments. Early in web development, client-side small application processes were able to run on Wintel machines using Java. These programs are called applets. Despite claims that Java ran on all machines, Java programming was not the same for Wintel and Mac machines. Therefore, using Java applets for an interactive reference source meant cutting out a part of the user population (generally those on Macs). Quirks still exist between browsers, as some follow the standards closer than others. If a program or interface has been written for a specific browser (e.g. Internet Explorer), it is possible all or part of it will not run when a user accesses the site with Mozilla, Opera, Netscape, Safari or any number of other browsers. This is a large hurdle developers face in creating usable online handbooks.

Use of the web developed before PDF became a widely accepted format, so the limitations of Hyper-text Markup Language (HTML) dictated some of what was available in electronic format and how it was presented. HTML focuses on textual content and its appearance. The HTML standard provides authors with the ability to include headings, text, tables, lists, and photos; retrieve online information via hyperlinks; design forms for gathering information or feedback; and including spreadsheets, video clips and sound clips directly in the document. (W3C recommendation, 1999) All of these pieces fit together and allow the author to control the general appearance of a webpage, but not the exact layout of pieces of information and how they may appear in relation to one another, as those variables are generally determined by the computer, browser and screen resolution being used to view the page. Dictionaries, encyclopedias and other materials where the content is primarily textual were converted to web-based formats before handbooks most likely because the information is more easily adapted to HTML, being less constrained to precise page layout. Linking sound and images with the text is relatively simple and provides an added value to the printed material. Thus, for relatively static information, like an encyclopedia article, dictionary entry, or other known-item search, HTML is suitable. However, for data-rich content, where research and discovery of information is the intended use, more sophisticated solutions are needed.

Response time and testing

During early development of online handbooks, it took less time to grab a paper copy of a resource and look up an item than to navigate the web version. Take, for example, an early online version of the *Statistical Abstract* of the USA (www.census.gov/prod/www/statistical-abstract-us.html) a useful source well suited to print format. The first online incarnation of *Statistical Abstracts*, in approximately 1996, consisted of scanned PDF versions of all the pages in the book, including the index. Portable Document Format (PDF) was an emerging format in the early 1990s, developed by Adobe Systems, the purpose of which was to ensure the same representation of a page of information on all machines, regardless of operating system, etc. To be able to read the files, the users' computer needs to have an additional piece of software, Adobe Acrobat Reader. No searching was available. This is still the case with the online version of this source. To look up information it is necessary for the user to load PDF pages of the index, scan these pages to find a table number, then return to the web page table of contents, load the PDF of the desired chapter of the book and manually scroll through each page in this chapter to find the needed table.

For purposes of rough comparison, a simple timing experiment was done by the author using print and electronic versions of *Statistical Abstracts*. The task was to look up one fact in the given resource and the amount of time elapsed to find the information was recorded. With the print resource, timing began with the book in hand. For the electronic version of the source, timing began with the browser already at the main page for the resource. Three people, who had varying levels of experience with the source, were timed for each search and the time and average are shown in Table I.

The task was to use the 1999 *Statistical Abstracts* to look for the percentage employed as engineers in 1998 who were women.

As seen in Table I, the information look up in print took an average of just under 2 minutes. The search was then repeated online. In the electronic version, with a variety of computers, the information locating process took an average of three-and-a-half minutes, as shown in the second column of Table I. On a slower computer, the online process includes several steps that take long enough to be outside the tolerable wait time for most web users (Nah, 2004). The last row of Table I shows the average numbers for the print and online look-ups, and it can be seen that with the two options equally available, it is faster to grab the print edition, if it is at hand.

Webster (2003) makes the point that for some resources, (e.g. dictionaries) the print is still the best and fastest options, and it seems that *Statistical Abstracts* is one of those sources. For those who don't have easy access to a print copy, this cumbersome online delivery method stills saves time over traveling to the library. Since unlinked PDF files are still the method the Government Printing Office (GPO) uses to deliver *Statistical Abstracts* online, it appears to meet a need for some users, and that usefulness can outweigh ease-of-use in some cases. Using straight PDF files is the least expensive

User	Time in seconds for print	Time in seconds for electronic
User 1	54	152
User 2	210	375
User 3	61	104
Average	108	210.3

Table I.
Search times in *Statistics Abstracts*

method to make this information available electronically. Other options, including a searchable, interactive web index linked to the proper pages and tables, or PDFs employing Acrobat's linking capabilities, would require more expense in programming and maintenance on the part of the GPO.

Machinery's Handbook, online via Knovel[®], uses a full-text search engine and leverages Acrobat's capabilities, thus providing a more sophisticated solution for locating information. The result of the interface differences can be seen in a comparison similar to the earlier one on amount of time needed to perform a task. Using the same rough method as earlier, the task was to find the tensile strength of red spruce in *Machinery's Handbook*. Table II shows the times for each of the three users for the print, and online versions and the averages.

The average look up time for the print edition was around 48 seconds. It seems an electronic book has to be well produced to be faster than the print, and the timing showed that the average time to locate the same information in the electronic format was about 42 seconds. By looking at the two average times in Table II, it can be seen how using a searchable index and the computers ability to scan the page for the desired information results in about the same amount of time spent on the task.

Issues and considerations

Usability includes more than a good user interface and quick response time, even though those two factors are important. The technology acceptance model (TAM), described by Thong *et al.* (2002) is a measure of perceived usefulness and perceived ease-of-use (usability) that helps determine a persons' likelihood to use a resource and can be used to test electronic resource acceptance. A cumbersome or unwieldy user interface is one reason users may continue to use the print option, or a different web option, when available (Nielsen and Norman, 2000). With some print reference sources, not only is the organization of information more familiar and therefore more comfortable, but in cases of a poor interface or slow response time online, faster to use. While the user interface and response time are vital in establishing regular use of an item, much more is involved in creating an efficient electronic handbook (Nielsen and Norman, 2000; Thong *et al.*, 2002; Nah, 2004). A resource must also be useful.

Creating an electronic resource should provide the ability to "transcend limitations inherent in the printed page" (Winter, 2000). In the case of the electronic version of resources such as Statistical Abstract this feat was not accomplished; they simply duplicated the print for online access. This model of static page images works well for electronic journals, when a user is looking for a specific article and will want the complete article. When the material being searched is a number or series of numbers in one table out of a chapter, a reader will not want to scan the entire chapter. Therefore, presenting these users with the whole chapter is a disservice.

User	Time in seconds for print	Time in seconds for electronic
User 1	56	35
User 2	55	40
User 3	35	52
Average	48.6	42.3

Table II.
Search times in
Machinery's Handbook

When publishers decide to produce a resource electronically, they must decide if they will create the electronic version, or let someone else handle it for them. The publisher or content creator relinquishing control of interface design to another vendor is a model that appears frequently with electronic journals. With journals it is common for the publisher to work with an organization that is willing to host the electronic publications for them, by providing the computer, the interface and the support. If a product that goes beyond flat PDF copies of pages is desired, programming time, design time and computer space are required. It is not surprising that a small book publisher would follow a hosting model similar to that for some e-journals and look to a company willing to develop and host an electronic version of their material. This results in electronic book vendors who can offer a wide range of products, maximize programming and computer resources, provide a uniform interface and offer cross title searching.

Online resources leverage current common technologies and software to create the desired level of interactivity. These programs can include Adobe Acrobat, mentioned above, as well as Java, JavaScript, and Flash. There are several advantages to this approach for the interface provider, including a reduced load on their computers. Any data manipulation and rendering capabilities that can be performed on the client machine reduces the need for the server to handle functions. Another reason to leverage the currently common programs is because developing another reader or software plug-in that users would need to download is seen as a barrier to ease-of-use by patrons.

With online sources there are other hurdles in finding the needed information, for instance the time necessary to consult multiple “books”, compile the information needed from each and then continue with the research at hand. In another parallel with the e-journals model, one interface provider loading many different materials and using the same search interface can provide an economy of scale that becomes attractive. Knovel[®] has done this for e-books from many publishers. ENGnetBASE (www.engnetbase.com/) and Referex (www.ei.org/eicorp/referexengineering.html) are electronic handbook providers with cross title searchability, but they are limited to materials from a single content provider, or publisher. Limiting a researcher in this way for books is very much like a researcher who uses only ScienceDirect for finding journals articles; the search has not been comprehensive in a given subject. Instead, it has been limited to one publisher. In a society inundated with information, the ability to maximize user efficiency is critical. The value of a researcher’s time is important, and equates directory to dollars for companies (Thilmany, 2003). It is important to maximize efficient use of that time, and delivering reference sources to the desktop saves the time a trip to the library takes (Gray and Langley, 2002). Therefore, the ability to search multiple sources in one location, even those that may be skipped by a user as out of scope, can save time and money.

Productivity is one of the driving forces behind the move toward electronic delivery of resources. In the “economy of human attention”, or the time it takes a researcher to sift through the multitude of data that a search can return, less attention is given to each source, which increases the likelihood of missing pertinent information. “In an information society, the scarce commodity is not information – we are choking on that – but the human attention required to make sense of it” (Lanham, 1997, p. 164). Therefore, any application which accurately pinpoints resources to be consulted and

specific data within those resources will increase the efficiency of a person's time and effort.

When searching a large pool of knowledge for information, accuracy and relevancy of results is critical to success in today's society. Success in this area is one of the reasons Google™ (www.google.com) has become such a popular search engine. Highly relevant results save the searcher time, as evidenced by users who "vote" the only way they can, by using one source over another. Perceived ease-of-use and perceived usefulness have coincided to create a site where users are comfortable and productive, so they return.

Solutions and recommendations

Complete electronic delivery of information has been a goal for some publishers, and many users, not only because of accessibility but also because it "allows for the development of reference content that wouldn't be as feasible in a printed work" (*Publishers Weekly*, 2000, p. 41). Some advantages presented by electronic access to books include rapid access to chunks of information via linked indexing, an ability to search the text and more frequent updating (Catenazzi and Sommaruga, 1994; Landoni *et al.*, 2000; Lynch, 2003; Winter, 2000). The goal of moving beyond the bounds of a printed volume means programming must be done to enhance usefulness. Generally, in the area of encyclopedias, content enhancement means including multimedia, such as a music clip of Beethoven's Ninth Symphony linked to biographical information about the composer. The enhancement of handbook content includes adding functionality which becomes available only when the information is in electronic form and able to be manipulated. Linking between related pieces of information, fully searchable text and an interactive table of contents are important, along with interactivity features such as being able to sort a table into the most relevant order at the moment, graphing the information presented in a table, or comparing several items at one time on the same graph by choosing them out of a list of materials and properties (Landoni *et al.*, 2000; Gibson, 2002; Wilson *et al.*, 2003; Arnold, 2004).

It takes both human and processing time to convert information in a static table (as displayed in a print resource) into a web table with the capability to manipulate fields. Many publishers are successful with the print versions of well-known sources and are not visibly dedicating resources to electronic development of their products, but in acknowledgement of the need for electronic access, have licensed their content to providers who develop and market the interface. In the model created by Knovel®, one company takes on the challenge to develop and enhance online access to books in particular subject areas, with the goal of bringing life and a new level of information usability to handbook data, and making it accessible at the desktop. Individual sources become easier to search for data in this online format, and by having one company put many resources online it becomes possible to search a collection of handbooks at once. A user does not have to make a "best guess" at what book might answer their question or progress through a number of sources until they find the right one. While a large portion of the over 600 titles included in Knovel® are not deep-searchable at this point (their name for titles including interactive tables, graphs and equations), cross-source searching is available, and the full capability of Adobe Acrobat's text search is put to use. Within an Acrobat version of a page or pages the user is taken directly to the first instance of the searched term and the word is highlighted.

Additional publishers are pursuing similar content searching and manipulation of data via the web, including CRC Handbook of Chemistry and Physics (www.hbcnpnetbase.com) and CINDAS Thermophysical Properties of Matter Data books (<https://cindasdata.com/Products/TPMD>). Both sources are quite useful, but as stand-alone resources do not take advantage of cross-source searching. A researcher is expected to identify the right source and then search it. It may not be enough that using electronic sources may speed up the searching time and give new life to some older volumes that were cumbersome in print. If the first title chosen by the researcher does not contain the needed information, the researcher is in the situation that is always present in print format. They must choose another title and try again. Cross-title searching takes out much of the guesswork inherent in this part of an information seeking process. A user may still need to look at several titles to find what they need, but a full-text search usually turns up some additional references to the compound or data for which they are searching, and provides a starting place for more thorough investigation. Thus, the aggregate electronic search process is immediately more productive than making a best guess on a source.

Future trends

The future of online resources is in working with materials that are native to the electronic environment (Lynch, 2003). Many handbooks have years of history in their fields as the primary trusted print resource, and that trust in a source must follow it to the electronic format (Webster, 2003). Choosing to deliver the information in a different fashion does not change the information being delivered. In an attempt to maintain a level of comfort for users and a level of credibility for sources, materials are not yet being envisioned and created in a completely electronic mode. Many of the print features, like pages, are still present. The shift is beginning, but it will likely be another generation or two before users are ready to fully abandon a print metaphor in online resources.

XML and other programming developments that will allow for metasearching may be the answer to cross-title searching without all titles having to be digitized by one company. Metasearch technologies allow simultaneous searching, from a single access point, across many different information resources. (Mischo and Schlembach, 2004; Fryer, 2004) To make metasearching work, all players in the field must agree to follow standards and work with a system that will benefit everyone, not just themselves (Webster, 2004). Metasearching is still in its infancy, and creating products that are capable of searching a variety of online sources from different publishers at one time has been a challenge. As an example, Knovel[®] has written and uses a propriety system architecture to create its interactivity, which would need to be reconciled with any standards developed. Similar to metasearching of electronic journal databases, the combined search would not provide power-searching capabilities. Those would continue to reside in the native interface for each resource. Standards development work has begun with NISO to look at authentication, standard search protocol, common descriptors and taxonomies, among other items (Fryer, 2004). And while the future seems to reside with far reaching metasearching capabilities, many product vendors are not enthusiastic about working toward this model (Webster, 2004).

Conclusion

Efficient use of time and money is critical for researchers, and the information glut that is part of everyday life slows people down. The time necessary to sift through all the responses to a general search, could be more effectively used by applying relevant information to finding a solution. When attention, searching and the location of information is streamlined, clearly focused and allows data manipulation into a format the researcher needs, an incredible edge is gained for searching.

Cross-title searchability and compatibility appears to be driving the comments from researchers in industry about the speed, accuracy and ultimate time savings involved in a product such as Knovel[®] (Thilmany, 2003). The advantages the deep-searchable titles have to offer, even over other online versions of the same titles, are seen as worth pursuing.

The first steps have been made in streamlining the usability of handbook data in an electronic format, but more could be done. A case study of Knovel[®], investigating their method for searching and presenting information online, their level of success in the field and the potential for expanding or licensing their technologies would be instructive. Investigation of the developing metasearching standards and discussion with Knovel[®] company representatives regarding the compliance of their e-book information could help raise awareness of the need for all electronic book companies to embrace the move toward cross-product searchability. A comparison study between several of the publisher specific online handbook collections (e.g. ENGnetBASE and Referex) with Knovel[®] could be used to further build on the WEB book experiments to define the best aspects of electronic reference material, particularly for a scientific community. The study could be used to determine more precisely where the line is between maintaining the familiarity of the book presentation and creating interactive data and when it is most logical to cross that line.

Full-text searching and results ranking algorithms could benefit from a close look and evaluation of the most effective methods for finding specific information. Most librarians cringe at the thought of searching the full-text of journal articles, so advances need to be made, or more widely adopted, which make full-text searching useable. Very accurate searching speeds up the research process and increases the usability, so it is worthwhile to determine how companies are determining accuracy of the results when searching entire books. Knowing the current process for a variety of companies will inform all of the possibilities and can only improve full-text searching.

There is much more to do, particularly to include the data from companies that ignore standards and continue to create their own online versions instead of contracting with another vendor. Additional research on the application of metasearching to electronic books and other content beyond library catalogs and online databases could ensure that future metasearching engines can handle the full scope of electronic materials available. Once electronic book materials are searchable through a single interface that offers comprehensive topic coverage, the productivity of researchers will increase, and access to information will be enhanced for everyone. Since, "... the library's business is orchestrating human attention structures", (Lanham, 1997, p. 165) it is the librarian's job to find sources which can best provide fast, accurate results for users.

References

- Arnold, S.E. (2004), "Interactive technical books: a Bloomberg for engineers", *Searcher*, Vol. 12 No. 3, pp. 45-8.
- Carlton, J. (2003), "Adobe Systems to unveil program to let computers ship complex documents", *Wall Street Journal (Eastern Edition)*, 14 June, p. B4.
- Catenazzi, N. and Sommaruga, S. (1994), "Hyper-book: a formal model for electronic books", *Journal of Documentation*, Vol. 50 No. 4, pp. 316-32.
- Culp, B. (2002), "The handbook: glorious past, parlous present, no future?", *Proceedings of the 2002 Chemical Information Conference, 20-23 October, Nimes, France*, pp. 49-53.
- DePetro, T.G. (2000), "ENGnetBase engineering database", *EContent*, Vol. 23 No. 1, pp. 25-7.
- DiBianco, P. and Chapman, L. (2003), "Ready reference 2A/7", *Information Searcher*, Vol. 14 No. 2, pp. 5-14.
- Fryer, D. (2004), "Federated search engines", *Online*, Vol. 28 No. 2, pp. 16-19.
- Gibson, C. (2002), "Knovel[®]: review", *Charleston Advisor*, Vol. 4 No. 2, pp. 29-37.
- Gray, E. and Langley, A. (2002), "Public services and electronic resources: perspectives from the science and engineering libraries at Duke University", *Issues in Science and Technology Librarianship*, available at: www.istl.org/02-summer/article2.html (accessed 27 August 2002).
- Huttenlocher, D. and Moll, A. (2000), "On DigiPaper and the dissemination of electronic documents", *D-Lib Magazine*, Vol. 6 No. 1, available at: www.dlib.org/dlib/january00/moll/01moll.html (accessed 8 October 2004).
- Juhl, B. (2001), "Desk set: ready reference on the web", *Choice*, Vol. 38 No. 8, pp. 1385-407.
- Kirkwood, H. (1996), "Reference resources from Uncle Sam", *Business & Finance Bulletin*, No. 101, pp. 25-7.
- Landoni, M., Wilson, R. and Gibb, F. (2000), "From the virtual book to the web book: the importance of design", *The Electronic Library*, Vol. 18 No. 6, pp. 407-19.
- Lanham, R.A. (1997), "A computer-based Harvard red book: general education in the digital age", in Dowler, L. (Ed.), *Gateways to Knowledge: The Role of Academic Libraries in Teaching, Learning and Research*, MIT Press, Cambridge, MA, pp. 151-67.
- Lehto, M.R., Zhu, W. and Carpenter, B. (1995), "The relative effectiveness of hypertext and text", *International Journal of Human-Computer Interaction*, Vol. 7 No. 4, pp. 293-313.
- Lynch, C. (2003), "Digital library opportunities", *The Journal of Academic Librarianship*, Vol. 29 No. 5, pp. 286-9.
- Marcum, D. (2003), "Requirements for the future digital library", *The Journal of Academic Librarianship*, Vol. 29 No. 5, pp. 276-9.
- Mischo, W. and Schlembach, M. (2004), "Metasearching technologies in reference work, OAI, and search navigation assistance", *Proceedings of the 2004 American Society for Engineering Education Annual Conference and Exposition, June 20-23, 2004*, American Society for Engineering Education, Washington, DC, Session No. 2441.
- Nah, F.F.-H. (2004), "A study on tolerable waiting time: how long are web users willing to wait?", *Behaviour & Information Technology*, Vol. 23 No. 3, pp. 153-63.
- Nielsen, J. and Norman, D.A. (2000), "Web-site usability: usability on the web isn't a luxury", *Information Week*, available at: www.informationweek.com/773/web.htm (accessed 8 October 2004).
- Publishers Weekly* (2000), "Mind your P's and E's", *Publishers Weekly*, 18 December, pp. 40-8.

- Subramanian, J.M. (1998), "Patron attitudes toward computerized and print resources: discussion and considerations for reference service", *Reference Librarian*, Vol. 60, Spring, pp. 127-38.
- Tennant, R. (1999), "User interface design: some guiding principles", *Library Journal*, Vol. 124 No. 17, p. 28.
- Thilmany, J. (2003), "Too much information", *Mechanical Engineering*, available at: www.memagazine.org/backissues/june03/features/toomuch/toomuch.html (accessed 4 June 2004).
- Thong, J.Y.L., Hong, W. and Tam, K-Y. (2002), "Understanding user acceptance of digital libraries: what are the roles of interface characteristics, organizational context, and individual differences?", *International Journal of Human-Computer Studies*, Vol. 57, pp. 215-42.
- W3C Recommendation (1999), "Introduction to HTML 4, Section 2.2", 24 December, available at: www.w3.org/TR/html4/intro/intro.html#h-2.2 (accessed 10 October 2004).
- Webster, P. (2003), "Implications of expanded library electronic reference collection", *Online*, Vol. 27 No. 5, pp. 24-7.
- Webster, P. (2004), "Metasearching in an academic environment", *Online*, Vol. 28 No. 2, pp. 20-3.
- Wilkinson, S.L. (2002), "A guide to literature", *C&EN*, Vol. 80 No. 1, pp. 30-3.
- Wilson, R., Landoni, M. and Gibb, F. (2003), "The web book experiments in electronic textbook design", *Journal of Documentation*, Vol. 59 No. 4, pp. 454-77.
- Winter, K. (2000), "From wood pulp to the web: the online evolution", *American Libraries*, Vol. 31 No. 5, pp. 70-4.