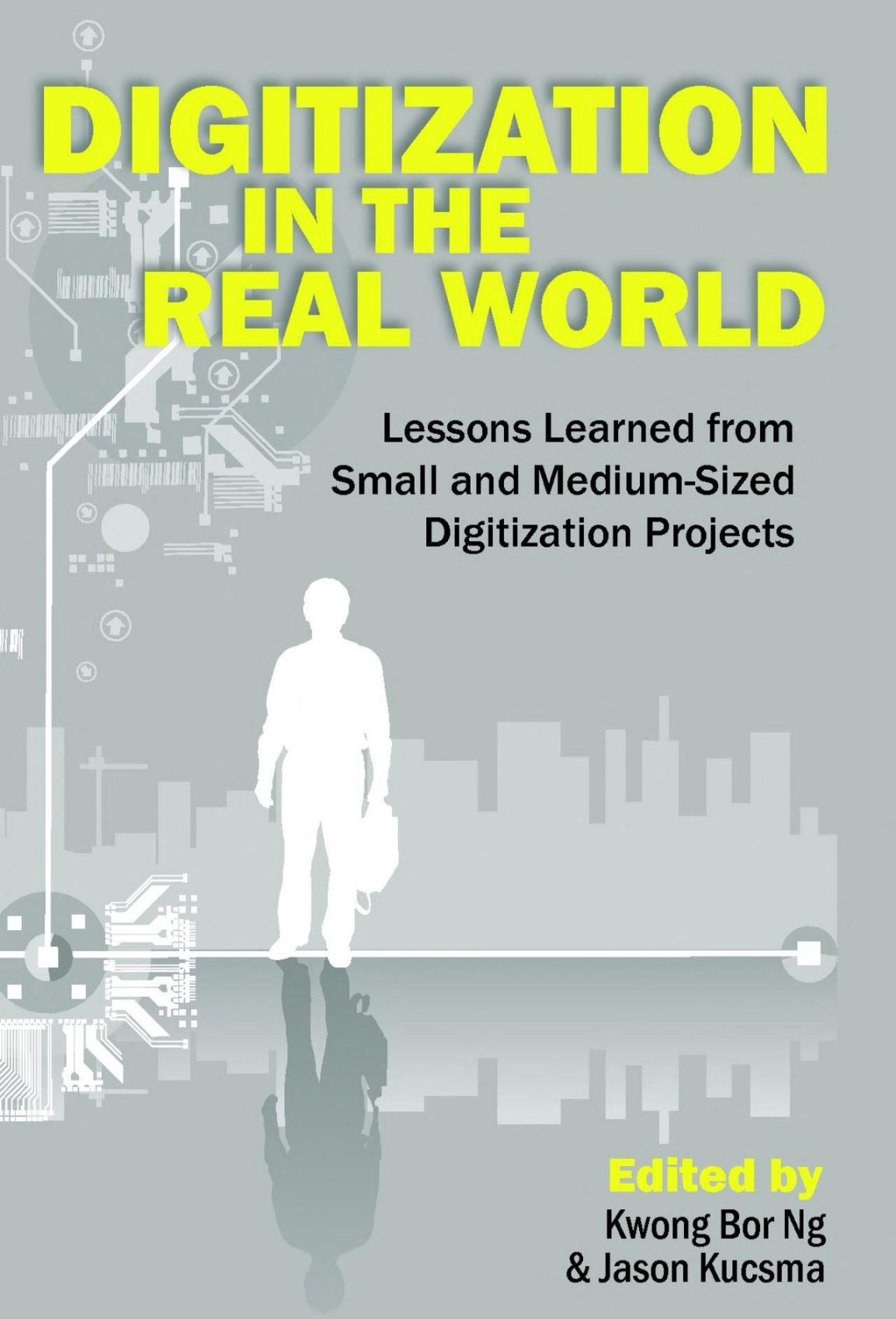
DIGITIZATION IN THE REAL WORLD

Lessons Learned from Small and Medium-Sized Digitization Projects



Edited by
Kwong Bor Ng
& Jason Kucsma

Digitization in the Real World

Lessons Learned from Small and Medium-Sized Digitization Projects

Edited by
Kwong Bor Ng & Jason Kucsma

Metropolitan New York Library Council

Published in the United States of America by
Metropolitan New York Library Council
57 East 11th Street, 4th floor
New York, NY 10003-4605
p: (212) 228-2320 f: (212) 228-2598
Web site: http://www.metro.org

ISBN: 978-0-615-37998-2

Cover Design: Jason Kucsma *(illustration by Smartone Design, licensed via iStockphoto.com)*

Reviewers Committee: Mark F. Anderson, Jill Annitto, Anna Craft, Jody DeRidder, Renate Evers, Wei Fang, Maureen M. Knapp, Sue Kunda, Mandy Mastrovita, Ken Middleton, Emily Pfotenhauer, Mark Phillipson, Alice Platt, Mary Z. Rose, Stacy Schiff, Jennifer Weintraub, Andrew Weiss.

The views expressed in this book are those of the authors, but not necessarily those of the publisher.

Table of Contents

Foreword

Dottie Hiebing (METRO Executive Director)

For more than 45 years, METRO has worked to provide opportunities for libraries to share best practice strategies to address many critical needs. In these efforts, we have often seen that the best learning comes through examples of libraries that have addressed important challenges successfully – and sometimes not so successfully. This has been especially true in efforts to support large and small digitization projects.

For more than a decade – and continuing today – digitization has been established as an essential focus for many libraries as well as for research centers, museums, and cultural and arts organizations. METRO has worked to support our members in these efforts with a range of grants, training programs and instructional materials.

Digitization in the Real World represents a significant new milestone in our commitment to providing library professionals with the hands-on experience and guidance they need to plan, execute, and manage digitization projects over the long term. In many ways, the examples presented in this volume show library professionals how to maximize the value and impact of digitization efforts for their libraries and their users.

This book also represents the first self-published text METRO has ever sponsored. As we continually look for new ways to help libraries stay ahead of the curve in digitization, technology and other areas, this strategy has the clear potential to be a major focus of our work in the years ahead. We will welcome your feedback and look forward to seeing how self-published materials such as this can support our mission and your needs moving forward.

On behalf of METRO, I would like to congratulate and thank editors Kwong Bor Ng and Jason Kucsma and all of the members of the library community who supported this project and who contributed of their time and insight in the development of these outstanding digitization case studies. They have created a vital new resource to help libraries continue to advance important digitization projects, and their efforts will have a profound and lasting impact on the future of these efforts in the years ahead.

Preface

Kwong Bor Ng & Jason Kucsma

For more than a decade, digitization has been both a critical need and a formidable challenge for libraries, archives, and museums around the world. To support these important projects, the Metropolitan New York Library Council (METRO) has been awarding annual grants to support digitization projects in New York City and Westchester County since 2005. Thus far, METRO has provided support for approximately 40 digitization projects at 25 different institutions. In those five years, we have learned a great deal about managing digitization projects effectively. In these efforts, METRO members have also shared best practice strategies in digitization through project showcase events and through the work of the METRO-sponsored Digitization Special Interest Group.

All digitization projects begin with some critical questions. How do we start a digitization project? What standards should we use for digital conversion and metadata? What are the best practices for workflow? What equipment or software should we use? Should we digitize in-house or outsource? What organizational or technological obstacles should we anticipate, and how should we negotiate them? Where can we turn for help in the middle of a project?

Naturally, the response to these questions will differ for different institutions. Even discrete projects within an institution will have many unique characteristics and challenges. But shared stories of successes (and failures) can be immensely helpful in supporting future digitization projects. To that end, Professor Ng came to METRO in the summer of 2009 with a great suggestion. Why not collect some of the most compelling examples of recent digitization projects? Many of us are familiar with the large-scale mass digitization projects of recent years. But Ng suggested — and we agreed — that there was a great

opportunity to share insights from lesser-known examples from the "real world." That's not to say that large-scale projects don't pose their own unique issues and learning opportunities for librarians, archivists, and technologists. But many libraries are more likely to proceed with smaller-scale digitization projects made possible by a special need or unique opportunity, a first-time grant, or the special dedication of a team of library professionals. Collectively, these efforts can provide many invaluable perspectives and procedural models.

This book was initially conceived as an opportunity to highlight digitization efforts in the New York metropolitan area. Our research quickly showed that there were many other project examples worth sharing. The response to our initial call for proposals was overwhelming; we received hundreds of chapter proposals from all over the world in just the first few months. Contacts from many of the world's leading knowledge-based organizations, cultural institutions and university libraries presented examples of projects representing a wide range of topics, perspectives, approaches, concerns, and lessons-learned.

The effort to choose from among these examples the examples that would be presented in the book was a daunting task. We were unable to include many great case studies. Each of the chapters presented was reviewed in a double-blind peer-review process to assess quality, accuracy and relevance. The 34 papers presented in this book represent our best effort to present a diverse and comprehensive overview of key issues in the management and realization of digitization projects.

We have divided the case studies into four primary groups. The first section focuses on small projects. They are digitization endeavors that moved forward with limited resources and staffing. The second group showcases digitization projects from diverse cultural institutions including public libraries, museums, research institutes, and cultural organizations. The third group consists of digitization projects based on medium-sized collections at universities and their libraries. The last group features projects that brought together

multiple institutions to work in collaboration on a project of mutual interest.

This book would not have been possible without the participation and hard work of all of the authors and reviewers involved, including those who submitted chapters that we were not able to accommodate. We're also greatly indebted to Dottie Hiebing, Executive Director of METRO, for recognizing the need for this important resource and for supporting this effort from inception. This is the first of what we hope will become a series of instructional self-publishing projects supported by METRO in the years ahead.

This is, above all, a book written by practitioners for practitioners who together recognize the critical needs and goals in digitization in our industry. Our hope is that it will be useful to students who are preparing for a career in library or research science and to practitioners who will shape the future of digitization for the library community. We know reading these stories has been enlightening for both of us, and we hope it will be for you as well. Thank you for reading.

Part I – Small is Beautiful: Planning and Implementing Digitization Projects with Limited Resources

DIY Digitization: Creating a Small-scale Digital Zine Exhibit

Melissa L. Jones (College Summit)

Abstract

The Barnard Library Zine Collection is an innovative special collection of dynamic popular culture artifacts. The zines in the collection provide a democratic and vibrant glimpse into the movements and trends in recent feminist thought through the personal work of artists, writers, and activists. The author finds that in order to improve access to and generate interest in such niche collections, institutions have a responsibility to overcome barriers to digitization and begin sharing their collections online. This chapter discusses the development of Barnard's first zine digitization project: *the Elections and Protests: Zines from the Barnard Library Collection Online Exhibit,* launched in the summer of 2008. The successful project demonstrates that it is possible to build effective and engaging small-scale digital collections using simple and inexpensive technologies.

Keywords: Barnard College Library, Copyleft, Copyright, Education, Elections, Lesson plans, Online exhibit, Political zines, Primary sources, Protest, Special collections, Zines.

Introduction

The Barnard College Library began collecting zines in 2003 in an effort to document third wave feminism and riot grrrl culture. Zines are self-published, usually inexpensively produced works by writers who subscribe to a Do It Yourself (DIY) philosophy. Generally, zines

are created out of an interest to communicate or express ideas that might not otherwise find acceptance in the mainstream media. Although zines as we know them today were born from the punk movement of the early 1970s (Duncombe, 1997, p. 21), they are part of a long history of small-run and "amateur" publication. Whether calling colonialists to arms in the days of the American Revolution or subverting censorship and challenges to free speech in Soviet Russia (Wright, 1997), alternative publications are a natural and important tool for preserving free speech.

Although zines are low rent ephemera, several public and academic libraries across the country have begun to recognize their value. At the forefront of the field, Barnard's collection has nearly 2,500 holdings providing unmediated access to the voices of young women on such subjects as race, gender, sexuality, childbirth, motherhood and politics. Zine Librarian Jenna Freedman's outreach and advocacy work helps to legitimize zines, not as radical historical footnotes but as valid literary and historic works worthy of collection, preservation and study.

As the Zine Intern in summer 2008, my role was to help Freedman to increase access to and interest in the Zine Collection. The result of my work was Barnard Library's first digital collection, an online exhibit entitled, *Elections and Protests: Zines from the Barnard Library Collection*. This project employed a DIY approach to digitization, making use of materials and resources at hand to solve problems and overcome challenges rather than relying on mainstream or out-of-the-box technologies. This project demonstrates that small-scale digitization projects can be topical, useful and impactful for a variety of stakeholders.

Literature Review and Needs Assessment

The literature surrounding zines reveals that, as unique primary source documents, they can serve as valuable research tools. Alternative press advocates such as librarians Chris Dodge and Jim Danky argue that self-published ephemera like zines, handbills, and military newspapers can provide a glimpse into a part of history that

includes the voices of marginalized individuals and groups which would otherwise be lost were they not collected (Dodge, 2008).

Dempsey (2006) notes that to collect the ephemeral and radical "long tail" is not enough; institutions have a responsibility to provide users with access points and contextual materials in order to maximize use. Liu (2007) notes that in order to better serve users, "academic library Web sites should ... switch the focus from presenting information arranged according to library functions and resources to providing targeted and customizable tools and services to library users ... and give users opportunities to express, share, and learn." In addition to their value as historical documents, zines also serve as powerful teaching tools for media literacy (Wan 1999; Congdon, 2003; Daly, 2005), but scholars and teachers need both access to zines and support for teaching with these unique documents in order to capitalize on this potential.

Lesk (2007) acknowledges the legal and philosophical issues that are inherent in digitization work, but advocates strongly for institutions and copyright holders to work together to overcome challenges due to the potential value of digital materials for research. In order to support online research, some public and academic institutions have begun digitizing their special collections. Unfortunately at the time of this project, no public or academic institution had moved to digitize their zine collections.

The lack of high-quality materials for studying and teaching zines online makes interacting with the genre impossible for anyone without physical access to a collection. Most public and academic institutions allow access to their zine collections mainly through catalog search. Some institutions occasionally mount online exhibits that include scans of zine covers only.

This has been due, in part, to the same barriers that hinder other digitization projects such as prohibitive cost, lack of time, and technological limitations. Additionally, zine librarians and scholars identify the intrinsically physical nature of the genre as another reason not to prioritize zine digitization. Migrating zine content to a digital form is seen by many in academia to undermine the very heart

of the genre, which is to be rooted in physical interaction between zinester, zine, and reader. Duke University's Zine Librarian argues that, "...zines are created by hand, crafted with paper, scissors, tape, glue, staples. They were meant to be handed from person to person, physically shared. The experience of handling zines in person, turning each page to reveal intimate secrets, funny comics, and poetry, can't be duplicated on-line. You would get the content, but miss out on the physical experience (Wooten, 2009)." Any academic digital Zine Collection would need to be very conscious of its treatment of digital surrogates.

Concerns about copyright, permission and privacy create another barrier to digitization. Copyright is a sticky issue when it comes to zines as a genre, which, by definition are created to be shared. Thus, many zines contain a copyleft statement, or some other notation of whether the owner has given permission for its contents to be reproduced. "Copyleft" is a term coined by open-source software pioneers to describe a "flipping" of traditional copyright laws that allows content owners to grant broader permission for their work to be shared. This "General Public License" can be applied in any situation where copyright might apply, including software, books, images and music (Söderberg, 2002). Generally, copyleft permission or GPL is considered to be conditional; zinesters who select copyleft status for their work, or those who claim no legal protection at all, still expect to be credited, or at least respected, for their work. It is poor zine etiquette to steal, borrow, or sell someone else's zine for personal gain.

Private zine online library and archive groups, run by zinesters and fans, have developed to fill the void of zines on the web. The sites digitize a large number of zines and serve as valuable repositories of content for experts in the field. Because they have grown organically from the zine community, these sites maximize their relationships to avoid and address concerns about copyright.

For Barnard Library, the the benefits of digitization digitization provided an incentive to overcome potential barriers, challenges and costs. An *Access and Use Survey* of known users administered in

2008 revealed that, while the Zine Collection has a strong contingent of feminist and zinester stakeholders, Barnard Library could be doing more to attract users outside the immediate scholarly and cultural community (see Figure ZINE-1.).

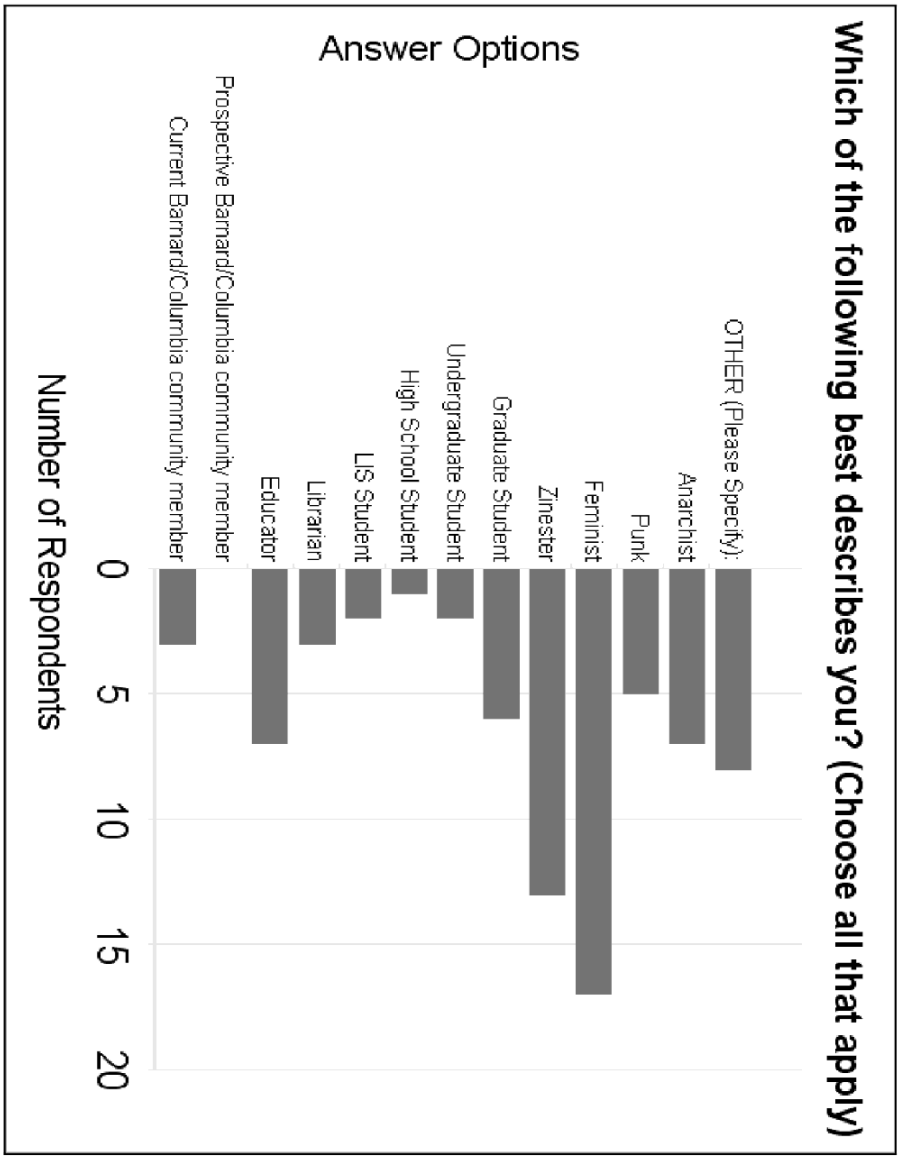


Figure ZINE-1: Results of the Barnard Library Zine Collection Access and Use Survey, administered to 25 known users in July 2008.

The survey confirmed that a small-scale digitization project would be a valuable addition to Barnard Library's existing services. 81% of known users reported that they would use curated, online exhibits about zines and zine history. Additionally, 81% of users said that they would use digital scans of selected zines.

The findings of the survey reflect the expectation by users that a library's website provide more than just access to information. By digitizing the popular and well-respected zine collection, Barnard Library could capitalize on the strength of its special collection to meet the needs of existing users, attract new users, and fill a need in the existing digital zine landscape. Additionally, a digital Zine Collection could help spread the word about the value of zines as historical documents and teaching tools to a new generation of potential stakeholders.

Project Planning

After making the decision to create a small digital zine collection, I created a project plan that included setting clear goals for the project.

Goal Setting

Digitizing even a small portion of Barnard's Zine Collection would have many benefits for the institution, its users, and the historical record. These included:

Improving access: Currently, membership in the Barnard/Columbia learning community is required in order to secure free access to the zine collection. Digitization would allow zines to be downloaded and shared easily, improving the ability of people from across the globe to access and learn from the collection.

Raising awareness about zines as legitimate historical objects: Freedman's work as an advocate for zine and other radical special collections would be complemented by a well-selected digitization project that is supported by descriptive and educational materials.

Highlighting Barnard's women's studies collection and drawing researchers to the institution: Barnard's Zine

Collection sets its women's studies research collection apart from other academic institutions. An online exhibit of materials from the Zine Collection could emphasize its uniqueness and eventually bring more researchers to the collection.

Preservation of the collection: Because most zines are produced cheaply using poor quality paper and inks, long-term conservation can be an issue. Digitizing zines makes their content available to future generations of researchers, students, and other stakeholders while preserving their physical form.

Project Scope

The scale of this project was by necessity very small. No fund was designated for the project. The site would need to be built and function within the existing Library website's structure; no money was available for purchasing a Content Management System or developing a complex metadata or image database. All work would need to be done using hardware and software already in Barnard Library's possession, or available open-source on the Web. As the Zine Intern, I would be the sole staff member available to work on the project. Freedman would supervise and approve my work. The project would need to be completed over the course of my summer internship, lasting only 100 hours over the course of 10 weeks.

Content Selection

Digitizing the collection as a whole proved to be too time consuming and technologically complex given these limitations. Selecting a small group of zines in a given theme or subject area to digitize first seemed a good model to begin with. Prioritizing digitization by demonstrated user need is a model that has been successful for other institutions. The University of Warwick in the UK, for example, developed an innovative research project in which students created digital surrogates of the 18th century French plays they used in their coursework (Astbury, 2006). Following a similar needs-based model would ensure that Barnard's first digital collection would be used by its most immediate stakeholders.

In 2008, the country was gearing up for an historic presidential election. Earlier that year, the previous Zine Intern, Julie Turley had created an exhibit of "Election and protest themed zines" to connect the institution's holdings with current events. The physical exhibit, which lived in the library's lobby, featured copies of selected zines and photocopied extracts of pages. From the Republican National Convention to the presidential election, from deciding to take your child to a political rally to challenging politicians to be responsible to their electorate, the featured zines addressed participation in -- or protest against -- the American political process.

The exhibit was a natural fit for this digitization project. The presidential election was only months away, we knew conversations about the political process would be a hot topic on campus. The selected zines offered a little-seen counterpoint to mainstream political coverage, rejecting voting as the sole means to make change in this nation. Moreover, educators across the nation would be looking for ways to talk about elections and the political process in their classrooms. It would be an excellent opportunity to demonstrate that zines can be relevant political and educational tools. Since zines are political in nature and often overtly political in topic, our digital collection would be reflective of the genre as a whole, even though we could only digitize a small number of zines. Finally, because the zines in this subset were already on display in the lobby, we knew that none were in need of conservation work or otherwise in danger of being damaged by the process of digitization.

Project Implementation

To maximize the benefits of digitization while addressing the barriers faced by the institution, I undertook a multi-step process for digitizing and presenting Barnard zines online. The process, like the zines themselves, was low-rent, low-tech, and outside the mainstream. The DIY approach was limiting in many ways, but also served as an excellent opportunity for learning and innovation.

DIY Digitization Project Timeline – May 5th through July 24th, 2008

Task	Timeline
Needs assessment and literature review	Prior to project start
Goal setting and scope definition	Prior to project start
Content selection	May 5th
Generating metadata	May 12th
Competitive landscape analysis	May 12th through May 20th
Creating site maps and wireframes	May 12th through May 13th
Copyright requests to publish sent to zinesters	May 19th
Scanning and digitization	May 19th through June 17th
Designing an intuitive user interface	May 20th through July 21st
Writing original content	June 1st through July 18th
Usability testing	July 19th through July 21st
Site launch and publicizing	July 23rd through July 24th
Evaluation and reporting	Ongoing

Figure ZINE-2: Project Timeline – May 5th through July 24th, 2008

Copyright Status and Securing Permissions

After selecting the zines to be digitized, securing permission to present their content on the web was the next step. Educational use, such as the creation of an exhibit, would likely fall within any zinester's definition of copyleft. Only one zine of the ten selected, "Radical Cheerbook," contained an explicit copyleft statement. We felt confident that we could use its content in the exhibition.

Because the other nine zines selected for this exhibit contained some kind of copyright statement or did not contain an explicit copyleft statement, an effort was made to contact and secure permissions from the original author. This effort was difficult, however, since many zines were published using pseudonyms or contain contact information that is out of date. To track down the zinesters, I used a combination of Google searches, MySpace, and a pre-catalog Microsoft Access database that Freedman maintains to identify current email addresses. For one zinester, I was only able to identify a mailing address, so I sent a letter and awaited a response.

By the time the site was ready to go live in mid-July, I received written permission to publish from six zinesters, with most expressing excitement about the project. One zinester requested that I send scans of the specific pages I'd hoped to include before giving permission. At the bottom of each zine's page on the site, I made a note that the copyright holder had given permission for Barnard to use scans from the zines.

In three cases, I was not able to secure permission before the launch of the website. In these cases, I added a note to each zine's page that we had made a diligent effort to contact the copyright holder and would remove the images used in the event that there was an objection. I also made the decision to include only minimal excerpts from these zines as compared to the more extensive scans used from the zines for which we had permission.

Site Design & Comparative Landscape Analysis

Once permissions requests were sent, I focused my work on designing the site's architecture and layout. Close analysis of the features of similar sites can be a good way to begin planning. In order to understand how zines and zine-like publications can be presented online, I analyzed five sites with similar collections to Barnard. Because there were at the time no academic institutions with large-scale digital zine projects, I reviewed three sites run by private groups. I also reviewed two academic digital collections that feature radical or obscure publications.

My analysis revealed several qualities that most online exhibits of zine-like material share.

Asset management: (1) All five sites included full-color image scans with legible text and graphics; (2) All but one site included an option to download the asset in PDF form; (3) Four out of five sites included descriptive metadata about subject, author, and publication date to aid in discovery and to give context to the asset

Navigation: (1) Every site evaluated had a descriptive homepage and a consistent look and feel; (2) All five sites utilized global navigation on each page to keep the user oriented.

Search and discovery: (1) Four out of five sites allowed users to browse for a zine by title; (2) Four out of five sites offered a keyword search function; (3) None of the sites offered a search by author or issue number function; (4) Four out of five sites made searching or browsing for a known-item simple and pleasurable.

Tools and customization: Every site evaluated offered a "printer-friendly" version of their assets

Aesthetics and usability: (1) Every site took care to ensure that all links and functions worked as they were expected to; (2) Four out of five sites used some type of backend content-management system to organize assets; (3) For the qualities adopted by all sites evaluated, I attempted to include them.

Site Name and URL	Launch Date	Assets
Zine Library.net http://www.zinelibrary.info/	None given	"hundreds" of zines
The Queer Zine Archive Project http://www.qzap.org	Nov 2003	154 issues
Punk Zine Archive http://www.operationphoenixrecords.com/archivespage.html	2004	120 issues
Ling Long Woman's Magazine @ Columbia University http://www.columbia.edu/cu/lweb/digital/collections/linglong/index.html	2005	241 issues
Anarchism Pamphlets in the Labadie Collection @ The University of Michigan http://www.lib.umich.edu/spec-coll/labadie/	1999	478 pamphlets

Figure ZINE-3: Sites Evaluated for Competitive Landscape Analysis

There were other qualities present in some sites but not in others. These included RSS feeds, customizable user accounts and high-tech page turners. Because these qualities appeared in only some sites, I considered them to be optional for my site.

Interestingly, none of the sites offered any curriculum or supporting finding aids that would add necessary context to the materials. I planned to include lesson plans and a bibliography to accompany my zine scans.

Creating a sitemap and wireframes.

I first sketched wireframes for my site using paper and pencil, then translated those sketches into digital files. The wireframes turned out to be ambitious, and due to time and skill constraints, I was forced to scale down my original vision, but the creation of the sitemap and wireframes helped me synthesize all my ideas for the site into one visual presentation.

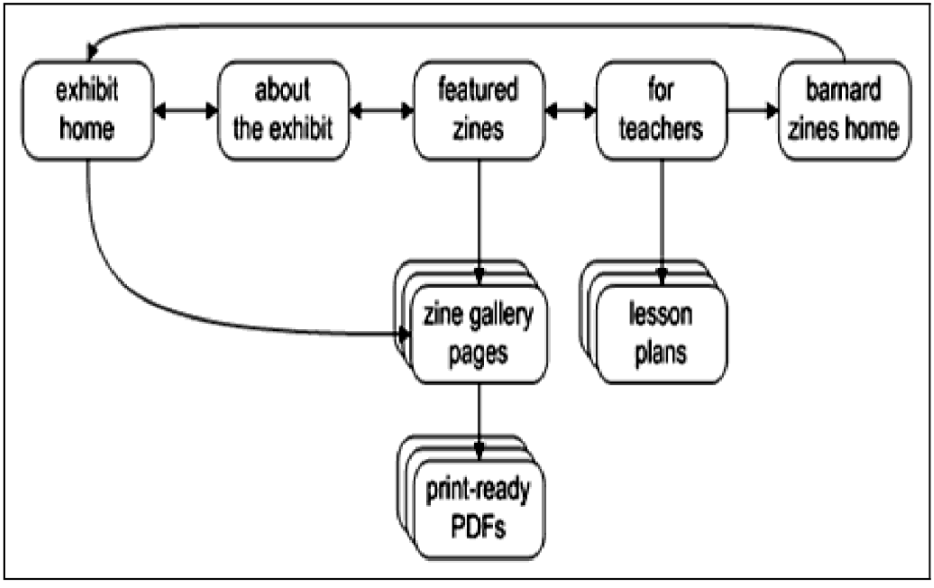


Figure ZINE-4. Final site map for the exhibit.

Generating Content

Scanning and digitization.

Ideally, for this project, preservation-quality scans should be made of each zine, along with its entire contents at the highest

possible resolution in addition to any presentation and/or thumbnail versions. However, I was limited in my ability to create high-resolution scans due to several factors. The first issue was the constraints of the hardware and software at my disposal. I worked on scanners and computers that Barnard undergraduates have access to in the library's computer lab. More sophisticated equipment might have resulted in better scans.

The condition of the zines themselves also contributed to poor resolution. Because most zines are produced using cheap materials, many of the oldest were beginning to deteriorate as the paper yellowed and the ink faded. Additionally, poor photocopying resolution in the original zine made some of the digitized page images appear grainy and pixilated.

After much experimentation, I balanced preservation and presentation needs with time and resource scarcity by scanning each zine once at 600 DPI or higher. Next, I saved two JPG versions of each zine: a presentation copy at a maximum height of 600px and a thumbnail copy at a maximum height of 90px. I preserved the aspect ratio of each scan each time I resized the image. When necessary, I used image editing software -- either the open-source GIMP (http://www.gimp.org) or Adobe Photoshop, depending on what was installed on the computer I was working on that day – to tweak the contrast levels of the scan and improve legibility. For each zine's cover, I created a slightly larger thumbnail which is presented on the Home page and the "Featured Zines" page. The last step was to convert all of the JPGs to PDF and create a print version of each zine for users to download.

These digitization decisions allowed me to produce legible copies of each zine while maximizing disk space. Because I didn't have access to an image database or a content management system, I simply organized all of the files in a series of folders on my desktop, giving each file a descriptive name following a clear convention. When the exhibit was complete, these folders were uploaded, along with the HTML and CSS files to the Barnard server.

Generating metadata.

I was able to take the metadata for the exhibit from the existing OPAC records. Each zine is currently assigned cataloged in a MARC record as part of the Barnard/Columbia joint OPAC, CLIO. Current metadata includes: title, an author or creator (when applicable), physical description, a publisher and date, Library of Congress subject headings and a summary or abstract. For each zine's gallery page, I used only the author, title, summary and call number fields. Additionally, I included a link to CLIO so users could locate the zine, check on its availability, and order it through interlibrary loan. I was able to add additional metadata about individual zinesters who responded to my copyright requests, including links to each zinester's current projects or personal websites.

Writing Original Content

A major component of the site was the contextual material that would add value and meaning to the zine scans for users. To meet this need, I wrote an "About" page describing the exhibit, as well as a "For Teachers" page that included:

- a brief explanation of why zines make good teaching tools
- three essential questions related to zines which could be used to frame curriculum planning
- A list of suggested resources for educators

The most time consuming content pieces to develop were the three lesson plans designed to help educators teach with the zines in the exhibit. Drawing on my background as a teacher, I designed these lesson plans around essential questions related to media literacy and social science content areas, then aligned them to three different learning levels: intermediate, secondary, and post-secondary. The lesson plans are student-centered and challenge students to interact with the zines in the exhibit through discussion and evaluation.

Building and Testing the Site

Ideally, the user interface for any site should be intuitive and promote discovery. For this project, I was limited to very basic web

design software and programming languages that have a low-barrier to mastery, but I was determined to make the site as usable as possible given the constraints.

To build the site, I again used hardware and software in the Barnard Library computer lab. I first attempted to build the site using Microsoft FrontPage, and then switched to an open-source HTML editor called Mozilla Kompozer (http://www.kompozer.net). Both FrontPage and Kompozer have "WYSIWYG" interfaces – an acronym for "what you see is what you get" – in that they allow users to create web pages using an interface that mirrors how the final product will appear (Myers, 1998). I needed to have a firm grasp on what was happening in the actual code behind my pages as I made changes. Additionally, I wanted to be able to customize my site, and the templates available in FrontPage and Kompozer felt limiting.

After a few days of struggle, I settled on developing the code for my site by hand using the simple text editing software available on most computer operating systems. The flexibility of being able to work on my files from any location made this project much easier to complete in only 10 weeks. I created and edited my files in Microsoft Notepad at Barnard Library, and could easily open them at home on my Apple laptop using either TextEdit or a free trial version of the excellent application, Coda (http://www.panic.com/coda/) which combines a WYSIWYG interface with an easy-to-use text editor. During the initial coding phase, I began by building a framework for each site using a common layout, menu bar, and footer using simple HTML tags such as those for images, links and tables. I also selected a patriotic red-white-and-blue color scheme and created an image banner to run along the top of the page that included the site's title in a typewriter-style font and some randomly-placed stars to mimic a zine-like feel. Later, the color scheme was changed to a more punk-inspired pink and black, to better reflect the lack of overt patriotism expressed in the zines themselves.

Once a basic page template was complete, I created a Cascading Style Sheet (CSS) file which governed the look and feel of each page and made coding the rest of the site easier. Using a CSS file is a simple

way to add style (such as fonts, colors and spacing) to every page in a website without having to edit each page individually (Bos, 2010). In order for the CSS file to work, I added a line of code in each page's HTML file that directed a user's web browser to "link" to my style sheet file, called "text.css". This file dictated the background and font colors of each part of each page, the margins for the different dividers and tables, and even the spacing of thumbnail images in my photo gallery. Before beginning this project, I had never worked in CSS. However, I found the language simple to learn and easy to use once I understood the fundamentals. As I worked, I referenced countless tutorials and open-source code available on the web to hack my way through the rest of the coding process.

The biggest coding challenge was creating a workable photo display gallery to present my zine scans. During the site mapping and wire-framing (see Figure 8), I had determined that I wanted users to interact with thumbnail versions of a zine's pages and select which pages they'd like to see in a larger, presentation-quality view. At first, I tried copying an open-source photo gallery HTML file that I found online. This allowed me to build the bare bones of my page layout, and indeed included thumbnail images and presentation-quality views. However, the photo gallery had one weakness: every time I moved my mouse off of the thumbnail I wanted to view, the presentation-quality view disappeared! This would have made my site very difficult to use, as it was impossible to scroll, zoom, or even save the presentation-quality view while still keeping your mouse over the thumbnail view. Finally, a programmer friend-of-the-collection helped out by producing a small piece of JavaScript code that allowed me to keep the presentation-quality view open without having to keep my mouse perfectly still. This code allowed me to create a version of the site that I was excited to test with users.

Usability Testing

Before going live, I tested the site's usability with three different test subjects, each representing a different group of Barnard Library stakeholders. Each test subject was given the same set of tasks to

complete. Figure ZINE-5 lists the tasks administered and whether or not the subject was able to complete the task without guidance.

Task	A	B	C
Choose and read through all the pages of a zine you're interested in.	✓	✓	✓
Find the author of the zine, "Don't blame me..."	✓	✓	✓
Find out if Barnard has permission to publish, "Subliminal Criminal"	✓	✓	×
Find out who first put together the zines in this exhibition and when.	✓	✓	✓
Locate information about how to use these zines in your classroom.	✓	✓	✓
Send an email to the zines collection	✓	✓	✓
Download a copy of page 3 of "Mama Sez no war"	✓	✓	✓
Find out why zines are used for protest	✓	✓	✓
Go to the homepage of Barnard Library	✓	✓	✓
Go to the homepage of the zine collection	✓	✓	✓
Download a lesson plan for use in your classroom	✓	✓	✓
Use CLIO to check the availability of the zine, "Dear Mr. Bush"	✓	✓	×

Summary of Test Subjects:

Subject A: Female, Age 20, Barnard Undergraduate Student
Subject B: Female, Age 28, Librarian
Subject C: Male, Age 31, Columbia Alumnus

Key:

✓ Task completed
× Task not completed

Figure ZINE-5: "Exhibits and Protest" Site Usability Tasks and Results.

In addition to identifying tasks that would be difficult for users to complete, I also made anecdotal records of the test subjects' comments and feedback. Based on the results of the testing, I made the following improvements to the site:

- Increased the size of all fonts used by 1px
- Added a mouse-over function to each zine cover image on the homepage that listed its title to aid in identification
- Added a "download all" link to a PDF containing all image files associated with each zine
- Made the copyright documentation on each zine more prominent

Launching and Publicizing

Upon completion of testing, the final version of the online exhibit (http://www.barnard.edu/library/zines/exhibits/online/elections/index.html) was launched on July 23rd, 2008. A link to the exhibit was posted on the Barnard College Library homepage, and a blog post about it was added to the institution's Livejournal (http://barnardzines.livejournal.com/). Emails were sent to a zine librarians' listserve and to other contacts and friends of the Zine Collection. I also sent a link to the exhibit to colleagues and friends in K-12 schools across the country. Since the initial launch, Freedman has continued to publicize the online exhibit in her talks and outreach activities for the collection.

Results and Next Steps

Informal evaluation of the project demonstrates that it has begun to meet its goals. User feedback on the site has been overwhelmingly positive. By making Barnard's zines accessible on the web to millions of people across the globe, the exhibit has indeed improved access to the collection. In an age when discovery on the web is primarily done through Google or other search (Belden, 2008), it is encouraging that search terms such as "zine lesson plans," "zines and elections," and "teaching with zines" consistently return the site in the first page of

search results. This is an indication of how many sites continue to link to the exhibit since its launch.

It is obvious from the *Access and Use Survey* that such an effort is both desired and respected by stakeholders. The exhibit only scanned selections of the zines featured, not entire issues, it may not be seen to contribute to the long-term preservation of the individual artifacts. However, creating a home on the web for zines around a contemporary issue can be seen to be contributing to the long-term preservation of the genre by making zines relevant in the digital age.

It remains to be seen whether this site will indeed drive users to the Zine Collection's other resources. Further evaluation should be done in order to determine whether or not this exhibit is directly contributing to increased access or use of the collection by Barnard/Columbia community members, outside researchers, K-12 educators, and other stakeholders. Repeating the Access and Use Survey annually may be a step in that direction.

Long-term sustainability and continued effectiveness of the exhibit are an issue. Freedman or future interns will need to take on the responsibility for maintaining and updating the exhibit as necessary over time. As Barnard Library further develops its web presence, the exhibit's look, feel, and even its content could become outdated. At this time, however, there is no reason why the exhibit cannot stay live for the foreseeable future without financial cost or significant staff time commitment. In order to maximize the exhibit's effectiveness, Barnard could consider:

- Continuing to promote and publicize the current exhibit, focusing on alternative outlets such as Wikipedia, educator websites, and media literacy blogs
- Developing an evaluation plan to determine the impact of the online exhibit on the stated project goals.
- Securing site analytics data on page usage and download stats to measure usage of the exhibit, as well as effectiveness of marketing techniques

Due to the support for this project from users, Barnard could consider digitizing more of the collection. Next steps could include:

- Creating more online exhibits around themes or subjects of interest to stakeholders if this first exhibit proves valuable
- Creating more subject guides, lesson plans and bibliographies about zines and zine history, and making them available online
- Collaborating with Columbia's New Media Teaching and Learning group in order to ensure that the user interface promotes teaching and learning with zines as primary sources, art objects and media literacy teach tools.

An open question is whether or not Barnard should move forward with digitizing the entire collection. Although this project was able to overcome many barriers to digitization of the genre (Wooten, 2009), a larger-scale project might open the door to more difficulty with copyright, permissions, privacy, and preserving the user experience of interacting with a zine's physical form.

Conclusion

This project showed that it is possible to create innovative web resources for a variety of stakeholders with a minimum level of technological and know-how. It provided a great opportunity for Barnard to continue to lead in the field of zine librarianship. This online exhibit supports teaching, learning, and research with quality and findable digital assets that highlight Barnard Library's strengths. With the current low barrier to web authorship, it is not enough for academic special collections to simply have a web page. Instead, more libraries and institutions can take advantages of the resources available to them – whether it is an eager intern, an exciting collection, or a timely theme –to create a resource that will meet the needs of users and make their holdings accessible to and available for generations to come.

References

Astbury, K. (2006). *French theatre of the first empire: Enhancing research-based learning*. Warwick Interactions Journal 28. Retrieved from http://www2.warwick.ac.uk/services/cap/resources/pubs/interactions/archive/issue28/abastbury/astbury

Belden, D. (2008). Harnessing social networks to connect with audiences: If you build it, will they come 2.0? *Internet Reference Services Quarterly, 13*(1), 99-111.

Bos, B. (2010). *Cascading style sheets*. W3C Consortium. Retrieved from: http://www.w3.org/Style/CSS/

Congdon, K. G., & Blandy, D. (2003). Zinesters in the classroom: Using zines to teach about postmodernism and the communication of ideas. *Art Education. 56*(3), 44-52.

Daly, B. O. (2005). Taking whiteness personally: Learning to teach testimonial reading and writing in the college literature classroom. *Pedagogy, 5*(2), 213-246.

Dempsey, L. (2006). Libraries and the long tail: Some thoughts about libraries in a network age. *D-Lib Magazine, 12*(4). doi:10.1045/april2006-dempsey

Dodge, C. (2008). Collecting the wretched refuse: Lifting a lamp to zines, military newspapers and Wisconsonalia. *Library Trends, 56*(3), 667-677.

Duncombe, S. (1997) *Notes from underground: Zines and the politics of alternative culture*. New York : Verso.

Jones, M. (2008). *Elections and protest: Zines from the Barnard Library collection*. Barnard Library Website. Retrieved from http://www.barnard.edu/library/zines/exhibits/online/elections/index.html

Lesk, M. (2007). *From data to wisdom: Humanities research and online content. Academic Commons*. Retrieved from http://www.academiccommons.org/commons/essay/michael-lesk

Liu, S. (2008). Engaging users: The future of academic library web sites. College & Research Libraries, 69(1), 6-27

Myers, B. (1998). *A brief history of human computer interaction technology*. ACM interactions, 5(2), 44-54. Retrieved from: http://www.cs.cmu.edu/~amulet/papers/uihistory.tr.HTML

Söderberg, J. (2002). *Copyleft vs. copyright: A Marxist critique. First Monday, 7*(3). Retrieved from: http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/938/860

Wan, A. J. (1999). Not just for kids anymore: Using zines in the classroom. *Radical Teacher, 55*, 15-19.

Wooten, K. (2009). *Why we're not digitizing zines*. Duke University Libraries Digital Connections Blog. Retrieved from http://library.duke.edu/blogs/digital-collections/2009/09/21/why-were-not-digitizing-zines/

Wright, F. (1997) *The history and characteristics of zines. The Zine & E-Zine Resource Guide*. Retrieved from http://www.zinebook.com/resource/wright1.HTML

Digitizing Civil Rights: An Omeka-based Pilot Digital Presence for the Queens College Civil Rights Archive

Valery Chen, Jing Si Feng, Kevin Schlottmann
(Queens College, CUNY)

Abstract

The Queens College Civil Rights Archive of the Department of Special Collections partnered with the Queens College Graduate School of Library and Information Studies to create a pilot web presentation using the open-source Omeka platform. Phase I of the project, conducted during the Spring 2010 semester, outlined a method for institutions of limited means to enter into the world of digitization using existing resources while highlighting the difficulties involved with metadata and IT support, and the advantages of involving graduate students.

Keywords: Civil rights, Digitization project, Omeka, Plug-ins.

Introduction

In the spring of 2010, the Queens College Civil Rights Archive of the Department of Special Collections partnered with the Queens College Graduate School of Library and Information Studies to create a pilot web presentation using the open-source Omeka platform. The synergy created between the technological skills found in the library school faculty and student body and the desire of the Civil Rights Archive to begin digitization of key holdings allowed the rapid

creation of a powerful web presentation platform. The process also outlined a method for institutions of limited means to enter into the world of digitization.

Queens College Civil Rights Archive

The Civil Rights Archive of the Queens College Department of Special Collections and Archives collects published and unpublished works relating to civil rights activities such as personal papers, community materials, organizational records, non-print materials, and artifacts. It also conducts oral histories to supplement its collections. The archive is particularly strong in materials documenting civil rights work by Queens College students during the early 1960s. The Archive seeks to provide evidences of the under-documented Northern involvement in the civil rights movement.

The Archive was founded in late 2008 around an estimable collection of personal papers donated by alumnus Mark Levy. Since then almost a dozen other personal collections relating to civil rights work in the 1960s have been donated by College alumni, and the Archive continues to actively collect in this area.

Queens College Graduate School of Library and Information Studies

The Queens College Graduate School of Library and Information Studies prepares library/information service professionals to meet the information and literacy needs of the New York metropolitan region and beyond. It is the only American Library Association accredited program for library and information studies within the City University of New York. The school prepares graduates to serve a broad segment of the metropolitan area's multicultural, multiethnic and multi-lingual population in a variety of institutional and informational settings. Through research, publication and other forms of scholarly activity, the school contributes and transmits new knowledge to society and the profession. The faculty provides opportunities for students to attain the competencies needed to participate in the evolving

electronic age by providing a technologically rich teaching/learning environment.

Project Origin

The Queens College Department of Special Collections, wherein the Civil Rights Archive is located, was acutely aware of the need for its collections to have a digital presence. It will soon be true that archival materials that are not electronically accessible in some way, whether via an OPAC or on website, will be no better served than in a dark archive. Given the limited resources of an urban public university, the Department had been unable to secure sufficient financial and technological support for an independent digitization project. Head of Special Collections Dr. Ben Alexander is also teaching in the Graduate School of Library and Information Studies, and he approached Dr. Kwong Bor Ng to discuss a mutually beneficial way to begin the process of building a digital presence. Drs. Ng and Alexander decided to expand the Special Collections Fellowship program, which provides archival graduate students at Queens College with a broad range of professional archival experience, to include a technology component. Dr. Ng selected two graduate students with extensive coding skills to do the actual work of creating an Omeka presentation website. The hope was to create a mutually beneficial arrangement: under Dr. Ng's supervision, the two graduate students were able to gain real-world experience in building an Omeka platform, while the Civil Rights Archive was able to lay sufficient groundwork to seek grant funding in support of a larger digitization project.

Staffing/Workflow

Drs. Alexander and Ng served as project coordinators. The semantic team, which was co-extensive with the Department of Special Collections staff, consisted of Dr. Alexander, Archives Adjunct Katie Hughes and Archives Assistant Kevin Schlottmann. The technical team was headed up by Dr. Ng, who supervised two of his technology graduate students, Valery Chen and Jing Si Feng. They were given academic credit as independent study students and Special Collections

Fellows to build the Omeka website. The semantic team was responsible for selection, digitization, and metadata creation. The technological team was responsible for the installation and development of the Omeka presentation. Both teams were involved in the development of the Dublin Core metadata schema, and they also collaborated in creating the user experience of the website.

The project began with a meeting in February 2010. This meeting was initially a brainstorming session, but the teams were able to agree to the basic semester goal of a pilot website, as well as a rough timeline. Once the semantic team completed selection and the metadata schema was ready, a few items from the Civil Rights Archive were digitized per week and forwarded to the technological team. Meanwhile, the technological team was preparing the Omeka website for import of digital items. This continued as an iterative process for the entire Spring 2010 semester. At biweekly meetings, the website and the digital items were discussed, and both were constantly improved.

Implementation – Semantic Team

The semantic team, consisting of the Department of Special Collections staff (Department Head Dr. Benjamin Alexander, Archives Adjunct Katie Hughes and Archives Assistant Kevin Schlottmann) was responsible for selection, digitization, and metadata creation.

Selection

When the semantic team began discussing what to digitize, it considered materials from the Civil Rights Archive, the College Archive, the Performing Arts collection, and the Rare Book, Zine, and Artists Book collections. It quickly became clear that the civil rights materials were best suited for this pilot project, for reasons such as processing status, donor relations, fitting into the Archive's specific mission of engaging with the broader community, copyright status of the materials, and the attraction of having students continue to work with the material.

First and foremost, the Civil Rights Archive had the best-cataloged collections in the Department. The majority of the Archive's holdings are fully processed, have finding aids, and are under archival control. This level of processing also allowed for existing contextual information to be used by the teams.

The civil rights materials are among the most prized and high-profile holdings in Special Collections. The recent founding of the Archive garnered attention from campus and activist communities, and received local press coverage as well. From a donor relations viewpoint, digitizing these collections would generate goodwill within the alumni activist community that contains future donors and supporters. The Archive's living donors are still very interested in their materials and legacies, and have repeatedly expressed strong interest having their materials digitized.

The semantic team also conducted a review of existing civil rights digital archives, which showed many worthy efforts already underway. The University of Southern Mississippi, which holds one of the largest archival collections about civil rights work in Mississippi, created the Civil Rights in Mississippi Digital Archive, an "Internet-accessible, fully searchable database of digitized versions of rare and unique library and archival resources on race relations in Mississippi" (University of Southern Mississippi Special Collections, 2006a). The Civil Rights Digital Library, hosted by the University of Georgia, is "a partnership among librarians, technologists, archivists, educators, scholars, academic publishers, and public broadcasters" that provides federated searching of digital civil rights materials from almost 100 different institutions (Digital Library of Georgia, 2009).

These two excellent examples, among dozens of others, illustrated two major reasons why the semantic team chose to digitize materials from the Civil Rights Archive, rather than from other special collections. First, it was found that there exists a vigorous online community that any institution holding archival civil rights materials must join to remain relevant and accessible. Second, the team found a paucity of material that relates specifically to Northern contributions to the civil rights movement, and thus digitization of the Queens

College Civil Rights Archive would add new perspectives to the online community, in keeping with the Archive's stated mission of engaging the broader archival civil rights community.

The copyright status of the many of the materials in the Civil Rights Archive was clear, because photographs and personal papers of known provenance could be easily cleared by the creators with whom the Department has a relationship. Many of the other areas in special collections have a murkier copyright status, which is a major potential impediment to digitization.

Finally, the civil rights materials were primarily processed by archival students from the Queens College Graduate School of Library and Information Studies as part of the Special Collections Fellowship program. It seemed natural to continue the collaborative effort between the library school and Special Collections by having the student-processed materials brought into the digital realm by Fellows as well.

Digitization

The actual scanning of the items was not the focus of this pilot project. A proper scanning procedure that will create archival images in the TIFF format, such as that developed by the University of Southern Mississippi (University of Southern Mississippi Special Collections, 2006b), will be developed during Phase II in conjunction with implementation of a digital asset management system. The items digitized for this project were scanned with an Epson 10000XL scanner using Adobe Photoshop CS3. After cropping, deskewing, and adjustments to contrast and level, the images were saved as 300-dpi Web-optimized jpegs and provided to the technical team.

Metadata

The development of a robust metadata schema was a primary goal of this project. The semantic team examined a range of available schemas, such as METS, MODS, and PREMIS, but it very quickly became clear that Dublin Core (DC) was ideal for a variety of reasons. Unqualified DC is compatible with basic Omeka; it is a simple and easily-understood schema; it can be extended by using qualified DC; it

is well-established; and similar projects are using it. The latter was a particular influence, both because the Civil Rights Archive hopes to digitally collaborate with other institutions as well as because there are a wealth of relevant resources available. The two key sources used by the semantic team were published DC schemas from the University of Southern Mississippi and the North Carolina Exploring Cultural Heritage Online project (Graham & Ross, 2003; NC ECHO, 2007).

The semantic team created a detailed qualified Dublin Core schema, but after much discussion the teams decided to work with an unqualified schema, because that was the Omeka default. The semantic team continued to create qualified DC, so that in the future the project will be able to implement this more detailed schema.

The project utilized many controlled vocabularies. Library of Congress Name and Subject Authorities were used for person and subject terms; DCMI and IAMA for type and digital format, respectively; and ISO 8601 for dates. For the analog Medium field, the Getty's Art and Architecture Thesaurus proved most useful. Geographic data were placed in the Coverage field. LC subjects were used for the general geographic area, such as the town or state, while latitude and longitude data were taken from Google Maps by manually entering a known address and harvesting the geospatial data provided. The technology team was able to use the latitude and longitude data to create a Google Map reflecting the geographic location of the digitized items.

The teams both felt it important to offer maximum searchability, and the semantic team thus also provided the full text of digitized items, using OCR software. This proved to be a time-consuming additional step, in particular the proofreading of the computer-generated text. The print quality of some materials was quite poor, and the many drawings and photographs were also difficult for the software to interpret. Better software and more experience creating OCR text should make this easier as the project moves forward.

Omeka also has the ability for users to add tags. The teams decided to take selected controlled-subject values from the Dublin Core metadata and use them as tags as well, to allow testing of

features such as the tag cloud. The teams engaged in an interesting discussion about how tags would be used in this project – some wanted to keep the vocabulary controlled, while others wanted to encourage users to add tags as they saw fit. The teams decided to allow tags to be used as user-generated metadata, in keeping with the Web 2.0 spirit of Omeka.

The metadata was created manually by the semantic team in an MS Word table, and transferred into Omeka by the technology team. A future goal is an automated process for metadata transfer.

Implementation – Technology Team

The technology team for the project consisted of two independent study graduate students, Valery Chen and Jing Si Feng, as part of the Special Collections Fellows program, and their instructor and project coordinator Dr. Ng. These Fellows began working on the project after the first meeting with the semantic team on February 16, 2010. Over the next ten weeks, the Fellows downloaded, installed, and modified Omeka, an open-source web-publishing system, and completed the first phase of the project on May 13, 2010.

Why Omeka?

The purpose of this project was to create a web presentation to showcase the unique and valuable holdings of the Queens College Civil Rights Archive, and at the same time provide the Fellows an opportunity to learn how to build a digital archive using a web-publishing system. In any project, it is important to consider the use of proprietary system versus nonproprietary/open-source. Omeka is relatively a new software package that describes itself as a web-publishing platform on its website:

Omeka is a free, flexible, and open source web-publishing platform for the display of library, museum, archives, and scholarly collections and exhibitions. Its "five-minute setup" makes launching an online exhibition as easy as launching a blog. Omeka is designed with non-IT specialists in mind, allowing users to focus on content and interpretation rather than programming. It brings Web 2.0

technologies and approaches to academic and cultural websites to foster user interaction and participation. It makes top-shelf design easy with a simple and flexible templating system. Its robust open-source developer and user communities underwrite Omeka's stability and sustainability. (Omeka, 2010, Project section.)

Omeka is an open-source web-publishing system developed by the Center for History and New Media at George Mason University. According to the Omeka web site (Center for History and New Media, George Mason University, 2010) Omeka is easy to install, allows great flexibility for customized web interface, and supports multiple plugins. All these features were appealing for this project.

Another compelling reason to choose Omeka was the potential inherent in the exhibit feature. Omeka has the Web 2.0 ability of allowing users to create their own exhibits from the digital collections. A primary goal of the Civil Rights Archive is to engage the educational community and encourage use of its materials. By providing digital surrogates and contextual information, this website would allow a teacher or professor to tailor their use of the materials in an exhibit, and also make them accessible to other educators seeking similar uses. This type of educational contextualization was a key reason to digitize the collection.

Description of Phase I (Feb 16, 2010 – May 13, 2010)

This section discusses the installation of Omeka, the addition of various plugins, the details of the most heavily manipulated pages, and examines particular technical problems and solutions encountered during Phase I of this project.

Installation

Initially, separate Omeka instances were created for each Fellow on the technology team to experiment independently. After both instances were adequately developed, the best features were selected from each and transferred over to a new Omeka installation, running on version 1.2.

Omeka 1.1, the latest version available at the time of the first installation process, was downloaded by the technology team and

installed. The Omeka system was in a zip file, and the technology team had to unzip the file to extract all the necessary files for installation. Each Fellow downloaded the zip file and unzipped the files successfully. The next step was to connect to the server remotely and upload the files for installation. Omeka consists of thousands of files, and Adobe Dreamweaver could not handle such a massive upload. Using an FTP client, such as FireFTP, was found to be the best practice. FireFTP supports large uploads and does not terminate in the middle of an upload. If termination does occur, FireFTP automatically reconnects to the server to continue with the upload.

The Omeka installation folder in the directory was removed by bash shell script for security. Administrator and Super accounts were created, and the system was up and running.

Creating Items, Tags, and Collections

A collection can be created in the Omeka Admin page by filling in the name and description of a collection. An item can be created by filling in the Dublin Core fields in the Omeka Admin page and adding the item to a collection. Tags can be added to each item to create more access points.

All image values in the Omeka General Setting should be defined before importing any image files. The values Fullsize Image Size, Thumbnail Size, and Square Thumbnail Size are crucial for Omeka to generate image output. Omeka automatically generates full size image and thumbnails during item creation. The technology team decided to change the full size image output and thumbnail size in the middle of the project, thus resulting in two different image sizes throughout the site. This meant that all the files needed to be uploaded again at the conclusion of the project, to ensure uniform image sizes.

Selecting a Theme

The public interface of Omeka was controlled by the files inside the "themes" folder. The Super can log on to the Omeka Admin page to choose a desired theme. More themes can be downloaded from the Omeka website and uploaded to the server. Both Fellows selected a different theme for their individual pilot Omeka sites. One chose "santa-fe" while the other chose "spring."

Plugins

Many plugins are available for download from the Omeka website. The technology team installed Geolocation, Simple Pages, Dublin Core Extended, Dropbox, ExhibitBuilder, and Lightbox for the pilot site. Several outside interactive effects were also installed, including animated collapsible panels, text truncation, and a slideshow. Most of the plugins were easy to use and install without any hassles; the plugins specifically named above are discussed in more detail below. One plugin issue was that some of the plugins were written in plain JavaScript while others were written using jQuery. This often created clashes in the code as the dollar sign symbol ($) was used for different purposes in both JavaScript and jQuery. In JavaScript, the $ indicates a variable, while in jQuery the $ represents the start of a command. Since jQuery is technically a JavaScript library, the double meaning of the $ was problematic.

Geolocation

Geolocation uses features from Google Maps and allows users to view items in various geographic locations by clicking on the balloon pointers. The technology team decided to install this plugin because it gives users a visual, spatial representation of the materials in the archive. This visual aid can be useful for quickly identifying locations of particular interest, and it allows users an additional non-textual method of accessing digitized items.

In order to use the Google Map functionality, the technology team first had to obtain a Google Maps API (application programming interface) key by creating an account with Google. Once the key was obtained, the team downloaded the packaged Google Map plugin from Omeka and uploaded it to the server. The key was then supplied through the administrator interface to activate the plugin. From the administrator interface, it is possible to customize the Google Map API. For example, the administrator can set the default location, which is controlled by latitude and longitude, and the zoom level ranging from 1 to 20, where 1 displays a view of the Earth and 20 displays a street level map.

The geographic location was added manually through the administrator interface. The latitude and longitude were entered for each digital item's metadata. The technology team hopes to create a plugin in the future that will automatically locate the spatial location information from the Dublin Core metadata and supply it accordingly so that as items are uploaded they will automatically appear on the map.

All items with geographical coordinates are represented by the default red pinWhen users click on the marker, the corresponding item will pop up in a balloon with limited metadata identifying the item. Users can then click on the balloon, which brings them to the particular item's display page. The team plans to enhance the map by creating a color-coding scheme that will correspond to an item's categorization. For example, items relating to a protest could have a red marker, while Freedom School items could have a blue marker. This functionality will add additional visual meaning to the map.

Simple Pages

Simple Pages provides the ability to handle html codes input by the Super at the admin page. The team used the Simple Page plugin to create the "About Us" section of the website.

Dublin Core Extended

Omeka defaults to Unqualified Dublin Core. A Dublin Core Extended Plugin can be installed to create more descriptive fields. A primary consideration when installing Dublin Core Extended is whether all the additional fields are necessary for a given project. Too many fields slow down the data entry process and scrolling time, and since all the Omeka data entries are web-based, a wrong click by the mouse can result in loss of data. The teams hope to take advantage of qualified Dublin Core in Phase II.

Dropbox

Omeka uses a web-based data entry system; each field needs to be typed in separately. A Dropbox plugin can be installed to facilitate bulk file uploads, allowing multiple files to be uploaded at the same time into a Dropbox Folder. When adding a digital item to the

collection, the user can then select that item from the Dropbox. Dropbox also supports the creation of multiple items, and allows bulk creation of Tags, Collections, and Titles. Other Dublin Core fields still need to be entered separately.

ExhibitBuilder

ExhibitBuilder was a more complex installation, as the architecture of the exhibit pages had to first be determined. The structure of the exhibit was defined as a Main Title Page with descriptions, a Sub-Section with descriptions, and individual pages with items in the layout of choice. The interface of an exhibit can be changed by accessing the screen.css file under the ExhibitBuilder directory.

Lightbox

In order to enhance the user experience, a Lightbox was installed for viewing full size images. The Lightbox feature creates an overlay around an image and enhances the clarity of the image in the center. The latest version of Lightbox was downloaded from http://www.huddletogether.com/projects/lightbox2/, and installation process was closely followed from the Omeka Documentation page (Ebellempire, 2009).

Heavily Manipulated files

This section describes two particular areas of the Omeka platform that were heavily edited by the technology team.

Item Display Page

The default display page for each item is a non-stylized, exhaustive list of metadata that requires extensive vertical scrolling. Moreover, some of the DC and Omeka metadata are duplicative. In order to create a more user-friendly display of information, multiple files were heavily manipulated and outside plugins were incorporated to make the display page more interactive.

The technology team found it challenging to determine which files control the item display page, because there are many files with the same name in different directories. However, once the structure of

how Omeka arranges its files was ascertained, the team was able to manipulate the code and customize the display.

The first step after locating the correct file was to understand the predefined functions created by Omeka. This was achieved with the help from the documentation on the Omeka website. This stage was time consuming, as the team used a trial-and-error approach to learn how to use the different available functions. Once it was determined which function handled the selection of individual metadata, the team was able to control where each piece of metadata would be displayed.

The display page was designed so that the most important information was provided in a clear, easy-to-read format. To minimize the scrolling, certain technical metadata was hidden from the user. The technology and semantic teams jointly examined each DC and Omeka data field, and decided on a short of list of fields to display. These included the collection, subject, tag, and rights fields. Creator, date, location and citation were included under the item's thumbnail image. Hidden fields included type, source, publisher, upload date, and extent. These fields were merely hidden – users may click on the "show" link to reveal the hidden metadata if desired. The show-and-hide functionality was made possible by installing jQuery, a lightweight JavaScript library, and the appropriate toggle-display code. The teams agreed that while an average user would not need to see this information on first viewing a page, more sophisticated users might want the option. The semantic team in particular found in its review of similar web projects that many sites did not include technical metadata, to the disappointment of archivists and librarians seeking to learn from how other institutions handle metadata.

If a particular item had an accompanying full text, it was made available to the user. However, since a typical full text would cover an extensive area of the display page, the technology team used a truncation function to display a snippet of the full text along with a link to show more. This was also accomplished through jQuery.

While most of the display page reorganization was controlled by the php files, CSS files controlled the final style touches, such as spacing, font sizes, and background color.

Homepage

The default homepage also underwent a major redesign. Instead of the default static thumbnails representing the featured item, the team added a dynamic slideshow with captions that rotate through the featured items. The initial plan was to implement the slideshow using Adobe Flash. However, that idea was discarded as inefficient because new .swf files would have to be created for each new imaged added and each image taken down. Therefore, the best solution with the least amount of maintenance was to find a method to automatically pull images for display. The team achieved this by installing a jQuery slideshow. These images are thus easily maintained by a few lines of code.

The technology team also added a right sidebar that allowed users to browse by various categories, including tag, date, location, subject, and type. The sidebar was initially occupied by a list of recently accessed items, but in conversation with the semantic team it was decided that creating multiple access points would be more useful to the user for navigation purposes. Currently, the list of categories is manually created through simple HTML code. The technology team hopes to automate the generation of categories in this sidebar, to simplify site maintenance.

Goals and Discussion

Phase II Goals

The teams have many goals for Phase II. The stabilization and scalability of the project are a top priority, and this includes development of a scanning process and a streamlined metadata creation process, as well as finding stable server space and implementing a digital asset management system behind the Omeka website.

In terms of the Omeka, the technology team hopes to add more sophisticated features to the map, such as item categorization. On the metadata side, it will implement qualified Dublin Core, and may

attempt to install the OAI-PMH Harvester. Finally, a long-term goal is the creation of a plugin that would allow use of TEI.

Discussion

This pilot project identified issues that might be similar in other institutions with limited resources. The three key lessons from Phase I of this project relate to technical support, metadata, and the involvement of graduate students.

Technical Support

Proper server space is absolutely vital for any digitization project. The extremely limited IT resources available at Queens College mean that this pilot project is limited in its ability to grow beyond Phase I. This also impacts the Department of Special Collection's ability to implement a digital asset management system. If continued efforts to obtain College server space are unsuccessful, the project will examine other options such as approaching another CUNY school for a partnership, or perhaps even renting commercial server space.

Metadata

The semantic team spent the majority of its time and effort on metadata, both in developing the schema and creating the actual records.

Creating metadata from scratch is extremely time-consuming. For each individual image, the team created titles and descriptions, applied a variety of controlled vocabularies, harvested geographic data, and proofread OCR text. With the current human resources, this process would be impossible to scale up. Thus, in Phase II the teams will attempt to streamline the process by automating some aspects of metadata creation. For example, technical specifications could be automatically imported and drop-down menus created for oft-repeated controlled vocabulary terms. Once a robust and well-described process is in place, Special Collections Fellows working on civil rights materials could be recruited to enter much of the metadata, which would provide the administrative team time to engage in other tasks, such as selection, quality control, and project planning.

Another issue the teams discussed was the extent to which existing archival description is helpful. As archival description by definition is not item-level, existing finding aids were only of limited use this pilot project, primarily providing contextual information. In the future however, incorporation of EAD finding aids would certainly allow additional types of categorization and searching.

Bulk Metadata Creation in Omeka

In order to facilitate the slow process of metadata entry for each item in Omeka, the technology team is experimenting with CSV Import in Phase II of the project. CSV Import, a plugin that handles bulk metadata creation, allows each Dublin Core field to be imported via a comma-separated file. However, more technical support is needed to define the PHP-CLO path setting with the server administrator.

Graduate Students

The collaboration of the Department of Special Collections with the Graduate School of Library and Information Studies was beneficial to both parties. Special Collections was able to lay the groundwork for future digitization and also to create a prototype Omeka website to show potential donors and partners, while the graduate Fellows gained valuable skills in actual implementation of a sophisticated software package. It should be noted that such a collaboration requires continuous interaction between and dedication from both the library school faculty and Department of Special Collections, to maintain standards and momentum as different classes of Fellows participate in the semantic and technology sides of the project.

References

Center for History and New Media, George Mason University. (2010). *Omeka: about.* Retrieved from http://omeka.org/about/

Ebellempire. (2009). *Adding Lightbox to Omeka.* Retrieved from http://omeka.org/codex/Adding_LightBox_to_Omeka

Digital Library of Georgia. (2009). *Welcome to the Civil Rights Digital Library.* Retrieved from http://crdl.usg.edu/?Welcome

Graham, S. R. & Ross, D. D. (2003). Metadata and authority control in the Civil Rights in Mississippi Digital Archive. *Journal of Internet Cataloging 6*(10), 33-42.

NC ECHO. (2007). *North Carolina Dublin Core implementation guidelines*. Retrieved from http://www.ncecho.org/dig/ncdc2007.shtml

University of Southern Mississippi Special Collections. (2006). *About the Civil Rights in Mississippi Digital Archive*. Retrieved from http://www.lib.usm.edu/~spcol/crda/about.htm

University of Southern Mississippi Special Collections. (2006). *Guidelines for digitization*. Retrieved from http://www.lib.usm.edu/~spcol/crda/guidelines/index.html

Digitization on a Dime: How a Small Library and a Big Team of Volunteers Digitized 15,000 Obituaries in Just Over a Year

Elizabeth Goldman (Kingston Frontenac Public Library)

Abstract

In 2006 and 2007, Chelsea District Library, a small public library in Chelsea, Michigan, digitized a collection of 15,000 obituaries on a small budget by using a staff of nearly 50 volunteers and open source software. The author describes the research and planning that led up to the project; unique aspects of the staffing and technology for the project; and the resulting database, which contributed to the library being named "Best Small Library in America" for 2008 by *Library Journal*. The database continues to be updated, expanded, and improved, and the use of volunteers as the primary workforce has had long term rewards for the library. The chapter demonstrates the accessibility of digitization projects to libraries, even those without pre-existing expertise, large staffs, or big budgets.

Keywords: Database, Digitization, Genealogy, Obituaries, Open source, Volunteers.

Introduction and background

Especially in small towns, the public library may serve multiple roles as library, museum, archives, and community center. In many cases, libraries accept donations of rare or unique historical material from

families in the area with little thought for long-term maintenance. Technological advances in the last few decades have given libraries new options for preserving local history collections and making them more accessible through digitization. While even the smallest libraries have staff educated in reference and circulation procedures, however, relevant training in archival methods and technology is rare, making the prospect of a digitization project overwhelming. At the same time, small public libraries often have little or no funding for such projects.

Chelsea District Library is a single-branch public library serving 14,000 people in southeast Michigan, about 50 miles west of Detroit. The library was established in 1932 by the local Women's Club and, in 1999, became a district library system serving both the town of Chelsea (population 5,000) and surrounding townships. The earliest settlements in the Chelsea area date back to the 1830s and many local families can trace their roots back to the town's founders, resulting in a rich history. From its earliest days, the library served as a repository for local historical and genealogical material, housing a local history room on the cramped third floor of the McKune House, its location from 1959-2000. In 2006, after extensive renovations and additions to the McKune House, the library moved from temporary quarters back to its historic home on Main Street in Chelsea, leading to renewed interest in the local history collection.

A collection of about 50,000 index cards known as the Family History Index made up a major component of the local history material. A retired lawyer and amateur genealogist named Harold Jones started the collection as a hobby, clipping obituaries from the local Chelsea Standard newspaper and other sources and pasting them onto 4 x 6 index cards, along with cross references that allow women to be located by maiden name. Upon his death in 1987, Jones' family donated the collection to the library, where it received extensive use by local and visiting genealogists. Library staff and volunteers completed a project from 2000-2002 to clean up and update the collection, since then volunteers have continued adding new clippings.

As part of a planning process leading up to an election to fund an expansion of the building and the staff, the library identified local

history as a priority for Chelsea area taxpayers and the Family History Index as a prime candidate for digitization work. At the time, the professional librarian staff consisted of the director, three department heads (adult services, youth services, technology services), and a part-time librarian, leaving few resources to focus on a project of this scope. In the spring of 2005, the library replaced the departing part-time librarian with a full-time librarian (the author), adding additional duties of managing the library's website and digitizing the Family History Index.

This paper will describe how a librarian and a team of four dozen volunteers completed the digitization of 15,000 records from the Family History Index in just over a year, resulting in a highly usable database that helped Chelsea District Library earn its distinction of "Best Small Library in America" for 2008 from Library Journal and the Bill and Melinda Gates Foundation. The first section will describe preparation for the project, including research in archival standards and digitization techniques, as well as the recruitment and training of the volunteer workforce. Section two will describe the decision-making process that went into the choice of open source software in order to create a user-friendly, free database of the records on a limited budget, as well as the work of developing and testing the database itself. Section three will offer an overview of the workflow for staff and volunteers as they did data entry, scanning, and proofreading of the records. Finally, the paper will describe the resulting database, current upkeep and expansion, and how the project served as a model for additional digitization work.

Project preparation

Research

Initial research focused on archival and digitization standards. No one on the staff had a specific background in either archives or genealogy that would fit the requirements of the project, so research started from a very basic level. Research was conducted largely online and in books, as well as by speaking with archivists and libraries that had done similar projects. Internal research played a role, too:

understanding how the Family History Index had historically been used and its importance to the community; exploring the scope, size, and fragility of the collection; and agreeing on project goals.

At the time, in mid-2005, quite a few libraries had launched efforts to put obituary indexes online for use by genealogists and historical researchers. The vast majority of these projects resulted in static webpages or simple databases that provided access to citations but not complete text. This early wave of digitization projects represented an important first step on the path toward full electronic access by giving researchers more complete information about the contents of a library's collection. However, those who found an obituary citation online still had to contact the library to obtain more detailed information or the complete text of the obituary.

Chelsea District Library was lucky enough to have a collection that included full-text clippings of obituaries on a significant portion of the cards in the Family History Index. Because of this and advances in technology, one of the library's primary goals was to make complete text available freely online, meaning genealogists and researchers had at least the possibility of meeting their research needs without leaving their computer or contacting the library at all. To further this goal, the library sought to develop a database that was free, simple to use, and contained relatively small image files accessible even by those with dial-up internet connections.

Important sources, although they sadly have not been updated in recent years, were the book and accompanying website, "Moving Theory into Practice," and the Making of Modern Michigan project. Moving Theory into Practice: Digital Imaging for Libraries and Archives by Anne R. Kenney and Oya Y. Rieger (Mountain View, CA: Research Libraries Group, 2000) provided a good grounding in what to consider when planning a digitization project, as well as minimum standards for use and long-term preservation. An online tutorial of housed at Cornell University Library (2010) offered step-by-step instructions for taking a project successfully through to completion. The Making of Modern Michigan was an IMLS-funded joint effort by the Michigan State University Library, the Library of Michigan, the

Michigan Library Consortium, and others (Michigan State University Libraries, 2005). Although its heyday had already passed by the time Chelsea embarked project, the website provided helpful background information on the structure of digitization projects and, especially, scanning equipment that had been vetted. The scanner models in the Making of Modern Michigan are no longer produced and libraries may have moved to higher minimum DPIs for scanning, but the advice offered at both sites remains sound.

Two more recent resources, available freely online, are BCR's CDP Digital Imaging Best Practices, from the Bibliographical Center for Research in Colorado (Collaborative Digitization Program, 2008) and NISO's IMLS-funded A Framework of Guidance for Building Good Digital Collections (National Information Standards Organization, 2007), which within the framework includes links to many other resources on more specific topics.

Project scope

Armed with this information, Chelsea District Library took a look at the Family History Index. While the cards had been stored away from the light in filing drawers, they also had been used heavily by patrons over the years, resulting in some wear and tear. In addition, neither the index cards nor the glue were acid free, and newsprint is one of the most acidic papers. The oldest cards and the oldest clippings dated back to the 1950s, resulting in some that were in very delicate condition. Along with the unique and irreplaceable nature of the collection, this delicate condition led the library director to decide that the cards should be scanned manually in-house rather than sent to a contractor who would likely feed them into a scanner, possibly resulting in damage.

This decision meant significant labor for library staff members and volunteers. In May 2005, the library applied for a grant from the State of Michigan to fund the digitization, which would have allowed for the hiring of contractors, but the application was turned down. Luckily, the library already had a well-established volunteer services program. Lacking any funding beyond $5,000 committed by the library, the project manager developed a volunteer-driven plan and

made the choice to seek an open source solution for the database, resulting in savings on the equipment side. An RFP was opened to competitive bidders in the fall of 2005, with selection of a contractor and design and testing of the database completed by early 2006.

During this time, the library also made decisions about the scope of the project. While the collection itself was estimated to consist of more than 50,000 cards, closer to 25,000 obituaries were represented, due to Jones' system of cross-referencing women by maiden name. A database eliminated the need for this. The remaining set of obituaries came from a number of sources:

- gravestone transcriptions from Chelsea's three cemeteries and several others in neighboring areas
- notes culled from early histories of the area, as well as scrapbooks and other material in the local history collection
- notes from death notices published in the local newspaper, The Chelsea Standard, taken from microfilmed versions of the paper, dating from about 1887-1950
- complete obituaries from The Chelsea Standard, clipped and pasted on cards, dating from about 1950 to the present
- complete obituaries from the newspapers in two nearby cities, Ann Arbor and Jackson, clipped and pasted on cards, dating in the 1970s and 80s.

Of this material, the first three could be included in the project without further consideration of copyright, as the donation of the Family History Index to the library included rights to copying the material. The Chelsea Standard, a weekly publication owned by Heritage Newspapers, supported the project from the beginning, granting copyright release for material originally printed in its pages.

Unfortunately, the publisher of both the Ann Arbor News and Jackson Citizen-Patriot declined to grant copyright permission. The library considered this a minor setback, as obituaries from those two newspapers covered a span of only about 20 years. The impetus for collecting from newspapers in the neighboring cities was that some Chelsea area residents chose to publish obituaries only in these publications. The number of citizens who fit this scenario, however,

was greatly outweighed by the number of obituaries of people who had no connection to the library's primary service area at all. The library considered entering data from these obituaries but refraining from scanning them, which would not have violated copyright law, but chose instead to leave the cards for a potential future project and focus instead on truly local residents.

Staffing

At this project's initiation, Chelsea District Library had about twenty employees. The project could not be completed by paid staff, and the small budget precluded the hiring of contract labor.

The library had a well-established and strong volunteer program, including a tradition of volunteers working with the Family History Index. Started as a volunteer effort in 1932, the library had always had strong support from volunteers, and genealogy in general is a topic that draws volunteers. Nevertheless, bringing in volunteers double the size of the library's own staff would not have been possible without a coordinator, who had developed procedures, documentation, and processes for intake, training, and evaluation.

Building on this strong foundation, the library advertised through its newsletter, the local newspaper, word-of-mouth, and presentations to organizations such as the county genealogical society. The Family History Index was a well-used collection and one of the best sources for obituaries for the area, so the library was able to to draw in volunteers who did not live in the Chelsea district itself. Volunteers filed standard application forms, which covered basic contact information, times available for work, and special skills. The initial group of about three dozen volunteers received training at one of two sessions set up in early February 2006, after which additional volunteers received training one-on-one or in small groups as they signed on. Later, existing volunteers would train new recruits. Over the course of the project, nearly 50 volunteers contributed to various aspects of the project.

Database development

In considering technology options the library's predominant constraints related to both funding and expertise. In 2005, Chelsea District Library contracted most of its technology services to the library cooperative of which it was part, including website hosting. The library itself at the time had only six public computers and about a dozen non-networked staff computers, with the single on-site server running the public computer time management system. No one on staff had knowledge of server administration, leaving staff nervous about hosting a server for the digitization project but also open to any of a number of configurations.

Open source software has seen increasing adoption by public libraries in recent years. While often referred to as "free," open source software is monetarily free only in the sense that to obtain a copy of the code requires no exchange of funds. In the truer sense, "free" refers to the user's freedom to view and adapt the software, generally with an agreement to then share improvements with the larger community. Chelsea District Library initially considered both proprietary and open source options for this project, as well as both in-house and contract solutions. While the librarian hired as project manager had extensive experience with Microsoft Access, the licensing costs for the accompanying Microsoft SQL server allowing multiple simultaneous users made that option prohibitive. An open source database based on PHP and MySQL appeared to be a more realistic option, with the drawback that no one on staff had the requisite familiarity with these programming languages. The library decided to solicit requests for proposals to get a better sense of its options. Replies to the RFP highlighted the range of options: from a $40,000 proposal that involved proprietary software and taking cards offsite for more efficient scanning to a $1,500 proposal based on open source software and leaving data entry and scanning purely up to the library. The library selected the able services of a programmer who worked at a nearby library and who recommended the purchase of a server and quickly designed and built a PhP/MySQL database meeting specifications. While open source was not the initial goal, that such

software ended up forming the basis for the database contributed greatly to the library's ability to produce a high-functioning, easy-to-use database on a limited budget.

Once the library selected the underlying software, the real work of database design began. Priorities were a simple interface on both the administrative and public ends; completely web-based access for both data entry and retrieval; and the ability to attach multiple images to each record. The library pictured a database that would be simple and fast for access by users all over the world and one that could be expanded to meet larger goals for the local history collection. Starting with attached obituary images, the library envisioned eventually allowing researchers to submit their own photos, family trees, marriage licenses, or other material that could supplement the library's own collection to tell the broader story of each person represented in the database. Flexibility for growth of both the size and scope of the collection was important.

These ambitious plans remained in the future. In the near term, the library had to balance providing extensive access to the obituaries with completing the work within a relatively brief time frame. The poor quality of the newsprint and cards meant doing optical character recognition (OCR) was not realistic. Thus, while researchers would be able to view an image of the complete obituary, searching would be limited to data entered by volunteers. This meant maximizing the number of access points was ideal; at the same time, too many access points could slow work to a crawl. In the end, after consultation with genealogical researchers, the library chose the following fields:

- first, middle, last, and maiden names of subject in separate fields
- first and last names in one field for: mother, father, spouse(s), children
- date and place (city, state) of birth and death
- cemetery and funeral home
- metadata covering obituary source and date and source of digital record

The database allowed for multiple spouses and children. It also included a notes field which, at the early stages, was left blank but proved to be invaluable for later expansion.

The database would be accessed via the library's website by users through either a basic or advanced search. Upon visiting the Family History Index Online, users see a basic search screen for the name field, which searches all name-related fields. This is often enough to get users to the obituary they need.

Users also have access to an advanced search screen, which searches first, last, and maiden name as separate fields. It also allows access via date and place of birth or death, cemetery, and funeral home, making it useful to those who may not be searching for a specific person but for more general historical information. Data typed into any of the fields on the basic or advanced search screens will also search the notes field, which may contain additional data from various sources.

The designer brought another feature to the database that would set the *Family History Index Online* apart from similar projects at other libraries, a hyperlinking feature that makes Chelsea's project unique in allowing researchers to jump from record to record, following the obituaries of family members represented in the database. If the parent, child, or spouse of an obituary subject is represented in the database, his or her name will appear as an active link. Clicking on this link takes the user to that person's obituary record. In this way, researchers may discover family connections they did not know existed and be better able to visualize how families relate to each other. Rather than noting the names of relatives, backing out to the initial search screen, and starting a new search, database users can simply hop from one relevant record to the next.

Once the database design was complete, both staff and an initial group of volunteers participated in testing. Volunteers tested for ease of use from the administrative and public perspectives as well as for how the database would meet the needs of genealogy researchers. The library was lucky to have a number of experienced genealogical researchers among its volunteer corps. They provided invaluable

feedback throughout the project. The database designer and project manager worked together to tweak the database in late 2005 and early 2006, leading up to training of volunteers and beginning of data entry work.

Timeline

- April 2005 – Chelsea District Library creates a librarian position with duties including digitization of the Family History Index
- October 2005 – Database development begins
- January 2006 – Database development and testing completed
- February 2006 – Volunteer training and data entry work begins
- May 2006 – Scanning of obituary cards begins
- October 2006 – Data entry of 15,000 records completed/ library closes to move to new building
- January 2007 – Scanning resumes; proofreading and database updates continue
- June 2007 – Final image attached to database
- October 2007 – Family History Index Online released to public; timed to coincide with Family History Month in Michigan

Digitization process

The library had two old PCs available for use by volunteers in a back office, a setup that proved beneficial to the project, as volunteers found they could often focus better on the mundane task of data entry while working in pairs. Because the database was entirely web-based, no software installation was required. Volunteers also used each other as resources to answer questions such as the interpretation of unclear wording or how to enter data in a particular field. Each volunteer had committed to working two hours per week for a period of at least three months, in an attempt to minimize the amount of retraining that would need to be done. Most volunteers stayed much longer, seeing the project through to completion, and many also worked multiple shifts each week. Through the volunteer program, these workers were able to sign up for shifts during all hours the library was open, including evenings and weekends, maximizing the number of people who could be involved.

As the project progressed, it became clear both that some volunteers wanted to participate but lacked interest or ability in data entry and also that volunteer tasks existed beyond what the library had originally envisioned. This led to some refocusing of efforts before a final workflow developed. In particular, a pair of volunteers took on the task of taking cards from the filing cabinet and sorting out those that need digitization. They stored the sorted cards in a box and transported them to the office where volunteers were engaged in data entry. Cards not in use remained in the files, which prominently displayed signs explaining the project. Volunteers doing data entry took the sorted cards and entered them into the database. Cards that brought up questions went into a separate pile for review by the librarian. The rest went into a "completed" file and moved on to step 2, proofreading. Two volunteers with especially good eyes for detail, as well as genealogical research experience, were recruited as proofreaders, tasked with checking every fifth card. While it would have been ideal to have a second set of eyes on every single database record, this was not realistic, and conferral with professional archivists confirmed that a 20 percent rate was more than sufficient.

After proofreading, cards moved on to scanning. Scanning represented another challenge, in part because the library had only one scanner which was also used for other purposes. In addition, while the scanning software that accompanied the purchase HP Scanjet 5500c was relatively simple, not all volunteers felt they had the requisite level of computer skills. As a result, scanning did not begin until about three months after data entry and was handled by a subset of about 10 volunteers. Volunteers scanned cards at 300 dpi and saved them in the archival standard TIF format. Because they were on black-and-white newsprint and newsprint already has a very low resolution, a higher resolution would not provide any benefits. These archival copies of the cards have been retained in separate, backed up files so that the original cards should never require rescanning. The entire database, including these archival images, was set to copy to a tape drive, with the tape changed daily. After one week, tapes were reused for new backup copies, leaving the library with multiple recent backups for added security.

At this point, the cards were refiled by a volunteer, returning access to the public who still relied on the paper file for research. The digital images then underwent additional processing. The TIF images were converted, using Adobe Photoshop Elements into compressed JPEG images with small file sizes. In general, the image of each card posted to the database was no more than 100 Kb. It would download quickly even with dial-up internet connections. Volunteers manually attached these images to each database record, completing the cycle.

Data entry work progressed remarkably quickly, with volunteers putting in close to 2,000 hours between February and October 2006. In October 2006, the project went on hiatus while the library packed up its temporary quarters and moved into a new facility. Volunteers completed data entry for the final card just before the move commenced, adding more than 15,000 records into the database in eight months. Between half and two-thirds of the scanning had also been completed at this point. The move to the new facility caused significant delays in the project due to other priorities for the library's technology staff. Scanning resumed in February 2007, however, and the final image was attached to the database in June. The library used the next few months to continue proofreading and to test the robustness of the database, releasing it to the public in conjunction with Michigan's Family History Month in October 2007.

Results

Chelsea District Library's Family History Index digitization project was a major success. Through a combination of creativity, open source software, and volunteer contributions, the library produced a highly usable online database providing full access to more than 15,000 obituaries for less than $5,000, completing work in about 15 months. It demonstrates that even at an institution with little staff and little specific expertise, a project can be developed to meet both the community's needs and the standards set by the library and archives world. The community involvement and grassroots nature of this project made it truly special for those who participated. It drew attention to the library, increased the volunteer corps, and gave

community members a sense of ownership for a key piece of the library's collection.

The flexibility with which the database was designed has proven vital to its success. Once done with the initial work of populating the database, the library turned its attention to improvements. One goal was to provide full-text access to all obituaries, even those from newspapers old enough that the only existing copies were on microfilm. In early 2008, the Friends of the Library applied for and received grant funding to help purchase a digital microfilm machine, which volunteers are currently using to scan obituaries from the Chelsea Standard going as far back as copies survive, into the late 1800s. These digital images are then either added to existing database records or used to supplement the database.

There also remained the problem of providing access to obituaries of local residents that were printed in neighboring newspapers. To this end, the library looked to its partnerships with local businesses. Chelsea is represented by two local funeral homes, and directors of both proved willing to provide the library with access to their files. When the funeral homes submit obituaries to any area newspaper, they also send an electronic copy, including photo where available, to the library. These partnerships have allowed the library to enrich the database with full-text access and color photos for more recent obituaries. The text that is sent electronically is pasted into the notes field, and images are higher quality than those ultimately printed in the newspaper. Along with this material, if the final obituary is printed in the *Chelsea Standard*, the scanned newspaper clipping is attached. The funeral homes have also indicated that they have computerized files going back a number of years, and the library is investigating the possibility of further enhancing the database with this material.

As another extension, the library returned to its initial contractor in 2008 to develop a database on the same platform to house the library's local history collection, which had been brought out of storage and organized only in 2007. Much of this material was even more valuable and unique than the obituaries, leaving the library with a strong desire to have it made accessible to the public primarily, if

not exclusively, in digital format. Scanning of material and database development have continued in 2009. While working on this and other local history initiatives, including a series of oral history projects, the library has found additional material to enhance the obituary database records of members of Chelsea's founding and prominent families.

Aside from the primary lesson that ambitious digitization projects are not beyond the reach of even very small libraries, the Family History Index digitization project provided additional lessons that apply to similar projects and beyond:

- The existence of a well-setup volunteer program allows a library to think much bigger than would otherwise be practical. People are out there in all communities who have the time, expertise, and interest to contribute.
- Planning is good but flexibility is essential. Projects may stray somewhat from their original vision or carefully thought out procedures, but that isn't necessarily bad. Being open to new opportunities and listening to workers and users can ultimately make a project much richer.
- Open source doesn't have to be terrifying. In fact, open source solutions are generally very stable, as the Family History Index Online has proven to be. Aside from minor software upgrades, the server has provided consistent access to the database for more than two years with virtually no staff intervention required.
- There's nothing wrong with thinking big, but it doesn't hurt to be realistic. Ambition leads to projects being even more successful than initially imagined. That said, a realistic assessment of aspects such as which tasks could be handled in-house (project management) and which should be contracted out (database design) prevented later stumbles.
- Partnerships enhance any library activity. In this case, a good working relationship with the local newspaper eased the process of gaining copyright access, and new partnerships with funeral homes have enhanced both the database and the library's reputation in the larger community. In other situations,

partnerships could be used to gain access to services or material the library cannot pay for or obtain on its own.

Small libraries often represent their communities' best hope for preserving local history. Preserving this history, even with little or no budget, is within reach if libraries combine their expertise in information management with technology decisions geared toward simplicity and a lack of hesitation in taking advantage of the knowledge and goodwill in their communities. Digitization projects not only preserve the past but provide an opportunity for greater community involvement, partnerships, and identification of the library as a key to the community's overall health.

References

Chelsea District Library. (2009). *Chelsea District Library Family History Index Online*. Retrieved on Jun 30, 2010 from http://fh.chelsea.lib.mi.us/

Collaborative Digitization Program. (2008). *Western States digital imaging best practices*, version 2.0. Retrieved on March 31, 2010 from http://www.bcr.org/dps/cdp/best/wsdibp_v1.pdf

Michigan State University Libraries. (2005). *The making of modern Michigan: Digitizing Michigan's hidden past*. Retrieved March 15, 2010, from http://mmm.lib.msu.edu/.

Cornell University Library. (2010). *Moving theory into practice digital imaging tutorial*. Retrieved March 15, 2010, from http://www.library.cornell.edu/preservation/tutorial

National Information Standards Organization. (2007). *A framework of guidance for building good digital collections*. Retrieved March 15, 2010, http://framework.niso.org/

Building the ALBA Digital Library

Jill Annitto (Archivist)

Abstract

This chapter serves as a case study of how a professional digital library can be successfully built with a small staff and budget. It discusses the planning and experiments with beta versions of the Abraham Lincoln Brigade Archives (ALBA) Digital Library, the final version of which is available on ALBA's website, through Metropolitan New York Library Council's (METRO) Digital Metro New York program, and OCLC WorldCat. The sensitive issues of digitizing another institution's collection while maintaining ownership of the final product are also explored.

Keywords: Copyright, CONTENTdm, Database, Digital archive, Electronic classroom, Educational resources, Funding, Indexing, Planning, Ownership issues, Small budgets

Introduction

In 1979, recognizing the vital importance of their radical history, and the need to collect writings, letters, photographs, oral histories and artifacts that would preserve their story, the Veterans of the Abraham Lincoln Brigade, the American volunteers who fought with Republican forces against Generalissimo Francisco Franco during the Spanish Civil War (1936-39), formed the Abraham Lincoln Brigade Archives (ALBA). Today ALBA lends its name to a major archive at New York University's (NYU) Tamiment Library and independently supports cultural and educational activities related to the war. The ALBA

collections are the most requested at the Tamiment Library. ALBA's relationship with NYU is non-traditional, which makes the ALBA Digital Library unique. NYU owns and maintains the Archives and hosts many of ALBA's programs, yet the ALBA name gives a different impression.

This chapter explores the challenges of digitizing a collection that is owned by another institution and how to overcome working with a limited budget, dated technology, and minimal staff to produce a professional digital resource. The ALBA Digital Library (Abraham Lincoln Brigade Archives, 2008) evolved from an ineffective form-based website tool to a fully indexed resource using CONTENTdm in a matter of months.

ALBA Goes Digital

Until a full-time Executive Director was hired in 2007, ALBA was initially run by a group of volunteers followed by a string of part-time administrators. The ALBA Board of Governors hired me as the Assistant Director in May 2008 just as they began reconsidering their outdated website, a major step for the small organization.

By June 2008 the Executive Director had resigned, leaving me as ALBA's only employee. It provided me with a great opportunity to help redesign the website alongside a subcommittee of the Board of Governors. The Web Committee, consisting of four history professors living across the United States, set a timeline of four months for the site's overhaul, with an anticipated launch date of early October. The Board wanted to stake a claim as the premier electronic resource for information on the Spanish Civil War.

The Impetus for a Digital Library

In July 2008 ALBA launched its first annual ALBA Teachers Summer Institute at NYU. The Institute hosts teachers from New York City public schools and exposes them to the history, art, and politics of the Spanish Civil War. Part of this immersion includes a trip to NYU's Tamiment Library to view the Brigade's Archives where teachers

receive an overview by the collection's archivists as well as professors from various departments at NYU.

After this initial success, ALBA decided to sponsor two more Summer Institutes (as well as year-round professional development seminars) in Tampa and San Francisco in 2009. The new settings precluded a site visit to the Archives so we needed to find a way to bring the Archives to the teachers.

The solution to this dilemma was the creation of a digital library or archive to be launched in conjunction with the new website. The website redesign was going to make ample use of ALBA's collection at NYU's Tamiment Library anyway, slowly integrating digital collections of letters and photos, and eventually including video clips and oral histories. These primary resources would strengthen existing lesson plans and other educational modules once the site was completed and allow for a more dynamic classroom experience. Since the site was already being designed, we thought it would be a great opportunity to have a collections database created for ALBA by our website designers.

Defining the Digital Library

In my experience archives or library staff working in conjunction with a representative from the information technology department does most digital library planning. At ALBA, I was the only staff member and every decision required consensus of the Web Committee, busy professionals with jobs and other responsibilities. Every time an issue arose it could take nearly a week to be resolved. This is a problem that small museums with very active boards will be familiar with, particularly when board members have little time, experience or interest in the digital project at hand.

We quickly ran into a problem of defining the digital library; the Web Committee believed everything created by ALBA was archival. I was in favor of a more traditional, primary source, collections-based digital library with a thesaurus and cataloging guidelines. My idea for ALBA's digital library was to create a system that would recall only primary sources from the war itself. The digital library would be its

own entity containing items that are separate from all other files posted to the website.

After several weeks of negotiations via email and conference calls, the Web Committee decided the best option would be to include all online documents in the digital library. The Web Committee charged the website designers with creating a web-based, simple form-based recall system as part of their contract.

The Beta Versions

Albita

Named *Albita* (or 'little ALBA') this resource was first launched in November 2008 along with the debut of the new website. It included the organization's quarterly newsletter, book reviews, and transcribed letters, not primary archival sources. Uploaded items were listed in random order rather than alphabetically or by date. In the end the design did not conform to the standards of information professionals; *Albita* had become a "junk drawer" of every single file uploaded to the website.

Document Library

Even renaming it and reconfiguring the display, the *Document Library* was not robust enough. To recall specific items, the user had to consciously use certain keyword strings, e.g. "George Watt Prize Winner," when retrieving items. It did not include a thesaurus and the plain-text description field did not allow for paragraph breaks. These descriptions were displayed as a solid block of text.

Media Library

Simultaneously added to the site was the *Media Library*. This database appears three pages deep within our Resources tab. It is the only way to access images (from archival photographs to logos of partner organizations) without using the ALBA Digital Library or slogging through the results from a general site search. Unfortunately because it is not linked to any other page, it is not often accessed. The

plain-text description field is displayed as a block of text and does not allow indexing.

Once the Web Committee saw *Albita*, the *Document Library*, and the *Media Library* in action the problems became apparent. We considered going back to the drawing board with our website designers, to create a thesaurus or a more sophisticated database system. Unfortunately, our original contract did not cover building a true digital library; the cost would have been prohibitively expensive.

I explained the possibilities of a professional digital library: adding our records to a consortium's collection, making them widely available through WorldCat, including them in an NYU catalog at the item level. It was difficult to convince the Web Committee that we still had affordable, professional options that would blend seamlessly into the new website.

Building a Better Mousetrap

As the only person on staff who could design and implement the digital library, I had to balance time spent on digitization with my other duties, including grant writing, public programming, and administrative issues.

The first problem for ALBA to overcome was the lack of any type of digitization equipment. The nature of ALBA's work and budget did not necessitate having a full flatbed scanner or laptop on hand.

Another major concern for ALBA was the high cost of software. A rough estimate for a full software package ran to the thousands of dollars. We didn't intend to scan the entire collection, just highlights; for our purposes what we needed was something that would allow us to scan part of the collection but offer maximum exposure.

In January 2009 I became aware of the Metropolitan New York Library Council's (METRO) Digital New York program. NYU's status as an institutional member of METRO allowed ALBA to benefit from the digitization program, a partnership between METRO and OCLC and CONTENTdm. As a separate organization, ALBA's much smaller annual budget resulted in a nominal $200 annual fee, paid to METRO

for access to the CONTENTdm desktop module and the upload of 500 discrete items. With software issues out of the way, we could focus on permissions and access to the collection.

Since ALBA's intention was to work with a collection that is owned by another large institution, we had to be very thorough and diplomatic about the project. We presented a sample record that would name NYU and Tamiment Library as the copyright holders, as well as instructions on obtaining permission to use the images. We assured NYU that the digital library pages on ALBA's website would include the same information. A distinction is made on both the website and in each record that the ALBA Digital Library is published by ALBA while the copyright is held by NYU. After several weeks of negotiation, we were free to move ahead with the project.

With software and permissions settled we were left with the issue of hardware. ALBA Board Member and NYU Professor of Spanish and Portuguese James Fernandez offered his laptop and Epson flatbed scanner for the duration of the project. After several weeks of further negotiation in order to gain access to the collections, Michael Nash, ALBA Board Member and Director of Tamiment Library, allowed us to scan the items ourselves, free of charge.

Selection, Policies, and Standards

In general there is a lack of digitized archival resources available on the Spanish Civil War. The closest to any kind of digital information on the subject is through Spartacus Educational (Simkin, 1997), an online British encyclopedia dedicated to educating students on history, with a significant amount of information on progressive history. But even this site is lacking in primary source materials.

Besides the Abraham Lincoln Brigade Archives, there are other Spanish Civil War archives in the United States: the Southworth Collection at the University of California at San Diego (see University of California, San Diego, n.d.) and the Spanish Civil War Collection at the University of Illinois – Urbana Champaign (see University of Illinois at Urbana-Champaign, n.d.) Still, the Abraham Lincoln

Brigade Archives at NYU is the largest collection of American volunteers' archives in the US.

Our goal was to put forth a curated collection of the most compelling and historically significant items in the Archives; the original plan was to digitize 500 items, but the realities of time and staff restraints reduced this number to an initial 150 items. These items included postcards, letters, newspaper articles, a multi-lingual newsletter, and a telegram from Ernest Hemingway.

Dr. Fernandez performed the selection of four collections based on the following criteria: size (only one manuscript box each), condition (stable enough for handling and scanning), and variety (each collection included letters from a variety of people). These same collections had been used in the Teachers Summer Institute and they proved to be popular with the teachers.

When it came to setting the scanning standards I turned to the University of Wisconsin-Milwaukee Libraries' Digital Collections (see University of Wisconsin-Milwaukee, 2006) pages, namely the *Transportation Around the World, 1911-1993* collection that I worked on as a graduate student there in 2003. The collection was built using CONTENTdm, then in its infancy. I followed UWM's digitization standards and policies as well as their formula for long-term archival storage as a basis for the ALBA Digital Library.

Digitization Standards

All items were scanned in full-color at a resolution of 600 dpi using an Epson flatbed scanner. The items were saved as the highest quality TIFF files and stored on an external hard drive. We refer to these TIFF files as the Archival Images as they are used only to create access images (PDFs or JPEGs) and are otherwise not accessed.

Access Images were created for web delivery, in the form of thumbnails on the ALBA site and for display within the CONTENTdm records, as well as for everyday use. The letters in the collection do not have OCR capabilities nor are the PDFs searchable. These were saved on both DVDs and the external hard drive. Because the METRO contract only covered 500 discrete items, we used Photoshop to stitch

the JPEGs together to create a single PDF file for multiple-page items. Not only did this save room in CONTENTdm, it was also useful to have multiple-page documents combined for reproduction purposes and to better keep track of collections.

All of this information is posted on the Digitization page of the ALBA Digital Library section. Sharing this information shows researchers (and potential donors) that the digital library adheres to archival standards while also helping other librarians and archivists plan their projects.

To access the images it is important to label them properly. Some scanners and scanning programs assign numbers that may or may not be useful. Auto-numbering systems can cause problems if the items are not easily identified by sight (letters, manuscripts). I named the files according to the manuscript collection followed by a number that corresponded to the letter's order in the folder, followed by a decimal that corresponded to the page number. For example, Lardner.1.4 would be the fourth page of the first letter of the Lardner Collection. The stitched PDF file would read Lardner.1.

Indexing

The documents were indexed using Dublin Core metadata (Dublin Core Metadata Initiative, 2010) which are standard in CONTENTdm, including: Creator, Date, Identifier, Type, Source, Description, Format, Coverage, References, Relation, Language, Publisher, Rights, and ALBA Reference Number.

Included in CONTENTdm is a default thesaurus (Thesaurus for Graphic Materials) for the Identifier field. Similarly, a default thesaurus provided the terms for geographic location in the Coverage field. Those fields for which the thesaurus was unique to the Abraham Lincoln Brigade Archives, I built a new thesaurus. These fields were: Type (e.g. text, image), Format (e.g. paper + size in centimeters), Identifier (e.g. envelope, autograph letter signed, typed letter), Publisher, Rights, and Creator. I added new Creators as they came up in each collection; some collections had as many creators as there were letters. In addition, for collections that had an online finding aid at NYU I included a link to that page.

In all, the ALBA Digital Library took four months for 150 items to be scanned, stitched, catalogued, and uploaded. I generally spent about fifteen minutes cataloguing each letter, sometimes longer depending on the length of the document. I worked on the digital library an average of eight hours per week with some weeks going by without any work at all. As collections were completed I added information to the ALBA Digital Library page and continued to redesign the website as necessary.

Final Product and Reception

The final product can be found at http://www.alba-valb.org/resources/digital-library. The Teachers Institute Alumni find the resource easy to use and helpful in the classroom, and ALBA even used the scanned collections to create facsimile copies of the Archives for the Tampa and San Francisco seminars.

As of publication, the digital library is available as a discrete collection on Digital Metro New York (http://cdm128401.cdmhost.com/cdm4/search.php), the digital program of the Metropolitan New York Library Council, the New York Heritage Digital Collections (New York Heritage, 2008) website, as well as at the item level on OCLC WorldCat (http://www.worldcat.org).

Seeing a Digital Project from Beginning to End

Planning is Key

Work backwards and set aside a day or two to really think about what you want to see in the final product. The small team (just myself most of the time) helped keep bureaucracy to a minimum. Remember: it is cheaper to do it right the first time.

Unofficial market research

I asked librarian and archivist friends what they liked to see in digital libraries and to give me advice on moving forward. I also reflected on what struck me in online collections, both the positive and the negative.

Tread Lightly

Occasionally it was necessary to receive approvals and it was important to remember that when dealing with large institutions there will be politics. This exists everywhere and I learned not to take it personally. Many of these issues were deep-seated and existed long before I joined the organization. I also learned that having a board member installed in a specific department or company does not guarantee easy access or donated materials.

Within my own organization, some board members did not understand the potential of planned, professional digital libraries, or that one person could undertake such a project. Price was also an issue; it was only after *Albita* and the *Document Library* failed was CONTENTdm considered and accepted as an amazing deal. In the end, waiting for the Web Committee members to come around in their own time was the best plan of action for this project.

Push the PR

As I completed each collection I sent information about the digital library everywhere: from Facebook and Archivists' Roundtable of Metropolitan New York to ALBA's listserv, e-news, quarterly newsletter, and fundraising appeals. While this publicity was mainly sent to people within the ALBA network, it also garnered the attention of local archivists and library students interested in doing small digital projects on limited resources.

References

Abraham Lincoln Brigade Archives. (2008). ALBA Digital Library. Retrieved March 30, 2010, from http://www.alba-valb.org/resources/digital-library

Dublin Core metadata initiative. (2010). Retrieved March 30, 2010, from http://dublincore.org

New York Heritage. (2008). Retrieved March 30, 2010, from http://www.newyorkheritage.org

Simkin, J. (1997). *Spartacus educational*. Retrieved March 30, 2010, from http://www.spartacus.schoolnet.co.uk/Spanish-Civil-War.htm

University of California, San Diego. (n.d.). *Southworth Spanish Civil War Collection*. Retrieved March 30, 2010, from UC San Diego Libraries website, http://libraries.ucsd.edu/locations/mscl/collections/southworth-spanish-civil-war-collection.html

University of Illinois at Urbana-Champaign. (n.d.). *Spanish Civil War Collection*. Retrieved March 30, 2010, from University of Illinois at Urbana-Champaign Rare Book and Manuscript Library website, http://www.library.illinois.edu/rbx/SCWPeople.htm

University of Wisconsin-Milwaukee. (2006). *Transportation around the World, 1911-1993*. Retrieved March 30, 2010, from Digital Collections -Transportation around the World, 1911-1993 website, http://www4.uwm.edu/libraries/digilib/transport/index.cfm

Digitization and Access of Louisiana Oral Histories: One Oral History Center's Experience in the Digital Realm

Gina R. Costello (Louisiana State University Libraries)

Abstract

The Louisiana State University (LSU) Libraries Center for Oral History began an effort to digitize at risk and high demand collections in 2007. The Center acquired digitization equipment, server space, and collaborated with the Libraries Special Collections Digital Services librarian to offer digitized oral histories online via the statewide Louisiana Digital Library (LDL). This paper details the history of the ongoing development of a digitization program for oral history materials using two staff members and limited resources. Decisions about what materials to digitize and how, equipment and software, and issues with access and preservation will be discussed.

Keywords: Audio digitization standards, CONTENTdm, Digitizing audio, Digitization equipment, Digital library, Digitization workflow, Oral history, Oral history interviews.

Introduction

The Louisiana State University (LSU) Libraries T. Harry Williams Center for Oral History began to digitize at risk and high demand collections in late 2007. Planning for the systematic digitization of the primarily analog collection began a year prior to any digitization efforts. The Center sought advice from an expert in the field, acquired

digitization equipment and server space, hired a full time employee to manage digitization, and collaborated with the Libraries Special Collections Digital Services Librarian to offer digitized oral histories online via the statewide Louisiana Digital Library (LDL).

The Center staff and the Digital Services Librarian have prioritized collections for digitization based on fragility or patron demand, made decisions about organization and access of the audio materials for the public, and addressed copyright issues. Only a small number of oral history collections have been added to the LDL, although over 700 hours of tape have been digitized so far.

This paper details the history of the ongoing development of a digitization program for oral history materials with one full time staff person and partial effort from another staff member. Decisions about what materials to digitize and how, equipment and software, and issues with access and preservation will be discussed. Results of the digitization and online access efforts have been mixed, but may serve as an example for oral history programs wishing to develop a more programmatic approach to digitization.

Center History and Description

The T. Harry Williams Center for Oral History at LSU Libraries Special Collections documents the social, political, and cultural history of LSU and the state of Louisiana by conducting, collecting, preserving, and making available to the public oral history interviews of folk artists, war veterans, governors, congressmen, state and local officials, civil rights activists, and other historically prominent figures in Louisiana. The Center maintains over 4,000 hours of tape-recorded interviews. The three person staff and a number of student workers transcribe, index, and deposit oral history interviews for archival storage at LSU Libraries Special Collections.

The Center, opened in 1991, is named after a man who helped legitimize the field of oral history. Dr. T. Harry Williams, a popular and acclaimed southern history professor at LSU spent more than ten years researching the biography, Huey Long. Published in 1969, this Pulitzer Prize and National Book Award winning book drew upon

Williams' tape-recorded interviews with nearly 300 individuals. Williams used a 30 pound Webster Electric Ekotape reel-to-reel tape recorder to capture the interviews.

The primary mission of the Center is to document the history of LSU. Since the history of the state and university are closely intertwined, many broader Louisiana subjects are documented as well. Public outreach through training workshops, consultations, and collaborations with individual researchers, community groups, classes, and institutions, enhance oral history collections throughout the state. Often, the collections are donated to LSU Libraries for preservation and public access. In many cases copies are provided to libraries, schools, museums, providing access for members of the communities in which the oral histories were collected.

The Center differs from some oral history centers in its commitment to providing fully edited transcriptions of all recorded interviews. Barring any restrictions placed on interviews by the interviewee or interviewer, the audio and a full transcription are made available to scholars and the general public. Because of the large volume of interviews that are collected each year, the Center maintains a backlog of interviews that are not fully processed (i.e., digitized if applicable, transcribed, audited, and cataloged). Interviews are organized into more than 40 different series, including Civil Rights, Military History, and Political History.

The Center Director has taken a more programmatic rather than project-based approach to the digitization of the collected oral histories. To ensure that preservation issues are addressed and collection access is a top priority, the Director employs a full time sound technician/webmaster at the Center. Center staff also works with the Special Collections Digital Services Librarian to mount oral history collections to the Louisiana Digital Library (LDL) (http://www.louisianadigitallibrary.org).

The Center makes available materials that are not restricted by the interviewee or interviewer. Interviews are digitized on demand for patrons, for preservation purposes, and for public access on the LDL. Prior to the acquisition of digitization equipment, patron requested

copies were recorded from cassette tape to cassette tape. Now materials are delivered to patrons via CD unless a cassette tape is requested. Copies are provided for a fee to patrons, although a small number of oral histories maintained by the Center are available for listening free online in the LDL. Center staff generally digitize fewer than five interviews per month for patron requests.

The funding for the Center is a mix of Libraries monies and endowment funding. The Libraries pays the salaries of the Director and two full time employees. Student workers' pay, a portion of travel money, and some supplies are also paid for by the university. The Libraries purchases and provides support for computers for the Center staff and student workers. Endowment funds cover most travel expenses, the majority of the equipment (specifically the field recorders, digitization station, software, fax machine, scanner), any Graduate Assistantships, additional student workers, and the majority of the transient workers' (e.g., professional interviewers, transcribers, editors) wages

Early Forays in Digital Access

One of the earliest digital projects the Center was involved with was a pilot project to digitize oral histories that are part of the University History Series sub-series, Integration and the African American Experience at LSU. The sub-series contains interviews with black students, faculty, and administrators at LSU during integration (1950-1970), plus interviews with lawyers and their clients who were involved in key lawsuits, as well as politicians and others who were vocal opponents or supporters of integration. The resulting digital collection, named "Integration and the Black Experience at LSU" (2003) contains audio files and transcriptions of three individuals interviewed between 1985 and 1998.

This legacy digital collection is scheduled to be revamped soon. The ".rm" or ".ram" audio files are available for listening only in RealPlayer and must be downloaded to the listener's computer before playing. The digital files were created more than eight years ago, so the sound quality could be improved and the information about

equipment and digitization method has been lost. The analog tapes will be re-digitized and optimized using current technologies.

Between 2001 and 2005, the Center utilized the skills of their part time webmaster and other staff members to create several online exhibitions and presentations (*T. Harry Williams Center for Oral History Exhibits and Presentations*, 2009) using readily available software and tools: simple HTML, PowerPoint, and Windows Movie Maker. Notable among these is the digital exhibition, "Baton Rouge Bus Boycott of 1953. A Recaptured Past" (2009) which includes a background and chronology of the event complete with photographs and audio excerpts. "Leaving Vietnam" is a nine minute presentation of audio clips from the Americans in Vietnam collection, featuring stories of escape from three Vietnamese refugees who immigrated to Louisiana around 1975 while fleeing Communist takeover. The presentation debuted at the 2005 Oral History Association annual conference and is currently available on YouTube, where it has been viewed over 6,000 times. Two other presentations were also mounted on YouTube to provide ease of access.

Center staff also began digitizing oral history transcriptions that were only available in paper format in 2004. They had some success using a HP Scanjet 5590 document feed scanner and an early version of Readiris optical character recognition (OCR) software. The software was lost, and the Libraries Systems department replaced it with Readiris Pro 11. Subsequent digitization efforts have been stymied by problems getting good readable OCR text, so the project has been put on hold. Student workers often are tasked with re-keying transcriptions.

In 2007 the Center Director, with the help of the LSU Libraries Special Collections Exhibitions Coordinator, curated a physical exhibition called "Have you Heard?: The Past in First Person from the T. Harry Williams Center for Oral History". The extensive exhibition contained ephemera and narrative relating to more than a dozen oral history collections. The Libraries provided two "listening stations", computers loaded with web-based presentations in the exhibit hall. In addition, exhibit-goers could check out MP3 players with pre-

recorded narration of the exhibition contents and snippets of oral history interviews. These digital offerings were made available with little cost using spare computers and a staff member as the voice of the narrator. No previous Libraries exhibition had employed technology in these ways. The Center Director counts the exhibition a success, as it led to a few collection development opportunities and awareness of the Center and its mission.

Digitization Station

After attending a digitization workshop at the Oral History Association annual conference in 2006, the Center Director decided that the systematic digitization of at risk and high demand analog collections should become a central focus for the Center. With the idea of "going digital" but with little research in hand they initially purchased two standalone analog to digital Lucid AD9624converters, which are designed to work in a recording studio setting. They realized belatedly that the converter units themselves were not useful without a digitization station, which would cost several thousand dollars. The Center made the all too common mistake of purchasing equipment without a clear plan how the individual hardware or software will interface with existing equipment. Fortunately they were able to later purchase a digitization system that uses one of the Lucid converters.

In order to ensure that in the future the Center made sound investments in technology and established a digitization workflow appropriate to their needs, the Director sought advice from oral history expert Doug Boyd at the University of Kentucky. Dr. Boyd visited LSU in March 2007 to evaluate the Center and conduct an introductory digital audio workshop for the Libraries staff. He generated a seven-page report with recommendations for equipment, collection development, and staffing.

Recommended Analog to digital work station equipment and software

1. Lucid AD9624 A/D Converter
2. RME Hammerfall DSP 9632 PCI Audio interface
3. 2 Yamaha HS50M 5" Active Monitor

4. 1 Tascam 202MKIII Dual Recorder Cassette Deck
5. 4 BP20 20' TRS - TRS Cable
6. 8 DKQR10 10' Dual RCA - TS Cable
7. 1 Furman PL8II 15 Amp Power Conductor w/Light
8. 1 DT770pro Closed Studio Mon Headphone
9. 1 Presonus Cent. Station Audio Control Center
10. 1 Plextor PX-716UF External CD-R/DVD+-RW
11. Sony Sound Forge 8.0 Audio Editing Software
12. Sony Noise Reduction 2.0 Noise Reduction Plug-In

The equipment recommended in the report was purchased with endowment funds nearly a year after Boyd's initial visit. Boyd returned to the Center to help set up the equipment and train a newly hired staff member.

Although not all institutions have the funds to hire a consultant, this less than $2,000 expenditure has proved money well spent for the Center. Without the vetting of the digitization program, the listed recommendations for equipment, and Boyd's encouragement to pursue positioning the Center as a leader in digitization efforts in the state and the profession, the Libraries administration might not have acted so quickly to support the endeavor. The administration approved reallocating funds to hire a full time staff member for the digitization and in less than two years, the Center has been able to digitize over 700 hours of interviews with their single digitization station.

With the addition of a dedicated digitization station and full time staff member to manage the process, the Center was ready to begin digitizing in earnest. It was immediately apparent, though, that server space and file redundancy would be an issue. The average file size of one hour of digitized uncompressed audio from analog tape is around 1.5 Gigabytes (GB). The Center only had access to a relatively small 74GB drive when digitization began.

Working with the Digital Services Librarian and the Libraries Systems Administrator, the Center temporarily located all digital audio files to a 5TB networked server that primarily serves as storage for TIFF images. In late 2009, a regional corporation donated used

storage equipment to the Libraries. The Libraries' Systems Administrator was able to configure four 2TB Raid 5 storage arrays, totaling approximately 8TB, for the Center's long term storage. This unexpected gift enabled the Center to continue digitization efforts, although they will still have to be selective.

The Center exists not just to archive, but to conduct research-based oral history interviews and to educate the community about conducting interviews. To fulfill this mission, the Center keeps a stock of digital audio field recorders to loan for oral history projects. As noted earlier, this equipment is purchased with endowment funds. The Center currently has four Edirol R-09 recorders, two Marantz CDR 310 recorders, and five Zoom H2 Handy recorders for loan. Center staff uses a Marantz PMD 661 for interviews.

The Edirol R-09 and Marantz CDR 310 are portable CD recorders and the Zoom H2 Handy records employ flash memory. Individuals borrowing the equipment are trained and instructed on its use. Digitally recorded interviews are brought to the Center either on CD or on secure digital (SD) flash memory cards. Interviews are saved to the Center's server and eventually processed.

Digitization Workflow

The digitization process is handled by one staff member, although he has recently trained a student worker to help run the digitization station. The staff member samples the audio to determine the optimal hardware and software settings and reformats the analog tape to a lossless uncompressed digital master WAV file. This master file is captured at a bit depth of 24 and a sample rate of 96 kHz in stereo.

The master WAV file is stored on a networked server, which is routinely backed up to a tape drive. This "master file" is not altered after the initial digitization process. Whenever possible, barring any time or funding constraints, a copy of every collection is also stored on an external hard drive as well as burned onto a gold archival CD.

The staff member then creates an optimized file from the master WAV file. Using Sound Forge software, he improves the signal strength and removes distortion from the audio. The optimized file is

saved as a WAV file to a different location on the server. He then generates a compressed MP3 file from the optimized file. This MP3 file is the use copy, and it is also saved to the server.

Unprocessed collections are digitized prior to processing to facilitate time stamping of the transcriptions. The Center uses Express Scribe Transcription Playback Software (http://www.nch.com.au/scribe/) and adds time stamps to the transcriptions based on the actual run time. Old transcriptions will be re-audited and time stamps added because the tape time stamps are arbitrary, often reset every time the tape player is used.

Metadata for the entire collection is kept in a Microsoft Access Database. All oral histories entering the center are processed based on a 13 page processing checklist. The processing checklist steps include 1) Accession 2) Transcribe 3) Audit 4) Send to Interviewee 5) Edit. This process is time-tested and thorough. The majority of the oral history collection is cataloged according to AACR2 standards in MARC format in the LSU Libraries online catalog (i.e., OPAC). The Dublin Core metadata in the digital collections is often copied directly from these catalog records.

Implementation and Access

The Center does not currently have a formal collection development policy to determine which oral histories are digitized The interviews that have been digitized thus far were identified as "high risk" on unstable medium or they were considered to be of particular interest to researchers and the public. Materials are also digitized "on demand" for patrons for a fee.

Tapes that were created prior to the Center opening in 1991 and later donated were assessed for deterioration and digitized as a means of preservation. For example, the 60 interviews in the Americans in Vietnam series, recorded between 1974-1977, were identified as at risk and were prioritized for digitization. Because of the content of the interviews, however, the digitized audio will not be offered via the LDL. In this situation, preservation of the materials outweighed the need to provide access.

Particular interviews and/or series of interviews, such as the Hurricane Betsy Series or the McKinley High School Series, were digitized because of their potential value to researchers and the general public. These collections will be uploaded to the LDL as soon as they are fully processed. Patron requested interviews that were digitized on demand for a fee are also candidates for the LDL.

During the past two years the Center staff and Digital Services Librarian have discussed workflows for uploading audio to the LDL. They consulted collections mounted by the University of Louisville (http://digital.library.louisville.edu/), Ball State University (http://libx.bsu.edu/), University of Nevada, Las Vegas (http://digital.library.unlv.edu/), and the University of California, San Diego(http://ceo.ucsd.edu/index.html) to facilitate decision making about the organization and display of online oral history materials.

The LSU Libraries serves all digital library materials via the Louisiana Digital Library, which was developed at the start of this decade by LSU Libraries and the LOUIS Library Consortium. LOUIS staff maintains the LDL for the nineteen participating institutions, including historical societies, libraries and museums. Individual institutions add content to the LDL and all materials are available for public use. The digital library is powered by CONTENTdm software and hosted by OCLC. LOUIS staff assists LDL institutions with customization of the software. LSU Libraries Special Collections maintains over 35 collections in the LDL.

Adding audio collections to the LDL has been a slow process that seems to move in fits and starts. Center staff and the Digital Services Librarian have held many meetings and exchanged numerous emails about serving digitized oral histories online. Debate about the topic centered around how the interviews would be organized and displayed. Many interviews, especially the life narratives, are topically related even though they are in different series. For example, university history overlaps with civil rights history in several interviews. Organizing the interviews both topically and by series can be achieved by using CONTENTdm custom queries to unite items from different digital collections, although this method does require

staff to re-create the collection custom queries and topics or series are added.

The CONTENTdm software seems more suited for its original purpose to serve digital images, and the default treatment of audio files is rather clunky. Audio does not play automatically, but instead the text "Access this item" appears at the top of the screen and metadata for the item below it. This presentation of the audio is somewhat confusing, because it is not even immediately clear that it is an audio file. Some institutions using CONTENTdm have devised workarounds that make serving audio in the software more usable.

In order to better group interviews together with the transcriptions and other related content, the Digital Services Librarian began uploading files as "compound objects" or multi-part files in CONTENTdm. Figure 2 illustrates this with the different files, abstract, transcription, and audio, hyperlinked in the left column. This display is not ideal since the metadata for the interview is on a separate screen and the "Access this item" text is still present. In addition to the cumbersome nature of the audio display, patrons wishing to listen to it are forced to download the often very large file to their computer. The Director felt strongly that other options not requiring the patron to download the audio be explored. Copyright would be difficult to manage if the audio was copied to different computers.

After reading about Ball State University development of a user-friendly embedded Windows Media Player above the PDF file within CONTENTdm (Hurford & Read, 2008), the Digital Services Librarian contacted LOUIS about implementing this method. LOUIS staff worked with the LSU Information Technology Services (ITS) department to obtain access to a streaming server from which the audio could be served. MP3 files are uploaded to the server via FTP software and the file path is linked to the item in CONTENTdm in the metadata field "Stream File".

The embedded player facilitates ease of use by providing the searchable PDF transcription to the patron as they listen to the audio. It does not require listeners to download the audio, thus it better

protects the copyright of the files. Information about copyright is included in the metadata for each item and future transcriptions may be watermarked with a copyright statement.

To organize the oral history collections in the LDL, the Digital Services Librarian used the "collection of collections" model that CONTENTdm employs to organize user collections on their website (http://www.oclc.org/contentdm/collections/default.htm). The individual series or collections are cataloged as a whole in the overall Center LDL collection. The series are represented by an image and selecting that image displays metadata taking the patron to the interviews. CONTENTdm software allows the creation of custom queries that will link the different collections and enable patrons to search across them. The individual series can be added to and the interviews and other materials in the collections will remain together, searchable alphabetically by title.

Problems and Some Solutions

Every digitization endeavor has its problems, but it is the individual institution's staffing, resources, and prior experiences that dictate the solutions. The Center, although small, is supported by a large university library. Digitization is a luxury that can be afforded because the Center has endowment money to purchase equipment and to provide staff with continuing education in the field. The time it takes to digitize resources is not a major factor in the continuation of digitization either because digitization is accepted as a part of the overall processing workflow. Digitization at the Center will be funded indefinitely and a full time employee will be dedicated to the effort if at all possible.

The Center is now two years into their programmatic digitization effort. At this point the digitization workflow has been well established and interviews from a few collections have been uploaded to the LDL. This section of the paper details problems encountered, such as legacy digitized collections, prioritizing digitization efforts, storage solutions, staffing, and digital access and display via CONTENTdm software, and how the Center staff and the Digital Services Librarian resolved or

did not solve them. Many problems could have been mitigated with more long-term planning, but the degree to which digitization efforts are currently supported and the ramifications of beginning a digitization program were not known at the start of these efforts.

The Center holds some legacy digital collections that do not meet the current standards for digitization. Prior to acquiring the digitization station and hiring an audio technician, Center staff did some preliminary digitization of analog tape using an external cassette tape deck connected to a computer. The sound was collected using a low end sound card to ram (Real Media Player) format in a process like the one that Washington State University Libraries used for their African-American Oral History collections (Bond, 2004). These early recordings were deemed important enough to place in the queue to be re-digitized according to the Center's current standards. For practical purposes, an institution may choose to keep legacy digitized items even if they do not meet current standards because the cost to re-digitize is high. For the Center the lessons learned with early experiments in digitization were important in shaping the future decisions to allocate more funds and staff to the digitization efforts in order to produce better quality sound.

The Center's at risk materials were digitized first, however, some of these materials are not good candidates for online access. The files will need to be stored long term, but because of restrictions they will be largely inaccessible. This falls within the mission of the Center, which includes collecting in addition to providing access to oral histories. Some audio files do not have completed transcriptions, rendered them unacceptable for immediate uploading to the LDL. The interview editing process is very time consuming and there is little immediate results (Bond and Walpole, 2006). Digitization priorities may differ depending on the institutional mission. If the mission is to provide access and preservation is secondary, then more popular or relevant collections should be digitized first. Institutions not supported by a parent institution, such as the Center is by LSU Libraries Special Collections, may not have the luxury to digitize collections just to archive them.

Another ongoing issue is long term storage solutions for the digitized files. The Center hoped to have files saved in at least three different places, a dedicated server in the main library, CD, and offsite storage. Some files are saved to an external hard drive in addition to the networked server, and born digital audio is saved to Gold CD. Ideally a copy of each master WAV file would be stored in offsite storage in a similar set up to the University of Kentucky (Weig, Terry & Lybarger, 2007), but this has not been implemented. The Libraries' server on which all audio files are saved is backed up incrementally to magnetic tape every night. Full backups take 40-120 hours because of the amount of data contained on the servers, so they are conducted once monthly. It is a secure system, but there is always a chance for failure. Future plans call for the Center to assess file storage and redundancy options.

The document "Sound Directions: Best practices for digital audio preservation" provides recommendations for long term preservation storage (Casey & Gordon, 2007), however many recommendations may not be feasible for small centers. The authors emphasize that file redundancy which is neither labor-intensive nor costly in media (e.g., CD or flash memory), should always be implemented. The majority of institutions will likely not have multiple terabyte servers and staff to keep them running, but files can at the very least be backed up to a more affordable storage medium such as portable hard drive or CD. Any storage medium can fail, however, so careful attention to this matter is imperative if an institution is interested in long term storage of files.

An issue that may require further review and assessment is the current standard of capturing audio at the higher sample rate of 24 bit 96 kHz. As server space fills and the Center and Libraries' budgets decrease, however, this standard may be reduced. Capturing audio at 16 bit 44.1 kHz reduces the file size by nearly half, and according to some experts it does not substantially decrease the quality of the WAV file (Weig, Terry & Lybarger, 2007). If the server is filled the Center may elect to save the derivative optimized WAV file to CD rather than the server. File optimization is time consuming, often taking the

length of the recording to complete, so deleting these files is not an option.

Before embarking on a digitization project, an institution should estimate the number of files that will be created and storage space needed. An institution may choose to capture audio at a lower and still acceptable rate to expedite the digitization process and conserve storage space. The institution should conduct an assessment of whether file optimization and multiple WAV files are needed before creating additional files that must be saved over the long term. Any derivative files can be recreated, so they should always be deleted or copied to more affordable storage media if server space is at a premium.

Another issue related to the audio capture standards is the Center's lack of written standards and best practices. Workflow principles and digitization methods are generally adhered to, but there is no guide or manual, just institutional knowledge. The workflow is based on recommendations by oral history expert Doug Boyd, who served as an advisor to the Center and also wrote the tutorials and information found on the Oral History Association website (http://www.oralhistory.org/technology/). The Center should apply the same level of detail and documentation to digitization workflow as they have for the processing workflow.

There are only two staff members who work with the Center's digital files, which could pose potential problems if either leave and has often caused bottlenecks in the workflow. At the Center all digitization is handled by one staff member with some student support. Other Center staff members do not have time to perform these duties, so little cross-training has been done. This is a risk because if the staff person leaves it will be difficult to continue digitization efforts. In the same vein, only the Digital Services Librarian currently uploads items to the LDL. This duty is usually shared by graduate assistants, but financial constraints have prevented hiring any additional help. Digitized files often do not get uploaded quickly because they are placed in a queue with all Special Collections digital materials. Cross-training between the digital

technician and Librarian is an option that should perhaps be explored. At the very least the two individuals, who are separated geographically across campus should establish better communication and more effective workflows. Information about which collections are ready to be uploaded to the LDL is sent ad hoc via email and there is no current mechanism for tracking the LDL files via the Accession database. Institutions should establish a clear workflow and assign responsibility for different aspects of the digitization process early on in a project. This will alleviate any potential miscommunication or turf war situation.

Before purchasing equipment and hiring staff to digitize audio, an institution should assess the environment where they will be located. At the Center the digitization station is equipped with the right hardware and software, but its location is less than ideal. The Center is located in an 80 year old house that is poorly insulated. The room in which digitization takes place is in the center of the house next to the building air handlers. The sound technician must use headphones while optimizing audio. If the Center is relocated much thought will be put into the location of the digitization station. In addition, Dr. Boyd recommended the Center purchase two digitization stations. When funding is available, the Center will explore this option.

A very important aspect of digitization efforts is providing access. The Center works with the Digital Services Librarian to upload items to the LDL, which uses CONTENTdm software. The software is less than perfect in its treatment of audio files, and efforts to retrofit the software to better serve audio are time consuming dependent on LOUIS staff expertise. LOUIS controls server access so software customization must go through them. The Center benefits from being a part of this consortium environment where an infrastructure is in place and support is offered at all times, but there are some constraints that this relationship brings. Small or not well-funded institutions interested in mounting collections online may be better served entering into a partnership with a larger institution or consortium.

An issue specific to the retrofitting of the software potentially affects patron access and sustainability. The embedded audio player that LOUIS retrofit for audio display does not display a time stamp so patrons cannot skip to a specific section of the interview. The audio player works well in the most current version of CONTENTdm, but the software is scheduled to be upgraded soon. Changes may affect the workflow and change the player functionality. The Center will rely on LOUIS consortium staff to recreate the embedded player in the upgraded software. Some institutions may not be able to expend a great deal of staff time continually addressing the interface when the software is upgraded, so this should be considered when addressing the sustainable access points.

In many ways the process for adding audio collections to the LDL has just begun. In 2008 all processed oral history collections which had been on a cataloging backlog were cataloged in the Libraries OPAC and WorldCat, which facilitates the metadata creation of records in the LDL. Changes in the CONTENTdm software in the past few years have made it more customizable. In 2009 the Center staff began producing audio and video podcasts with images and sound from the collections. The podcasts and information about hem are available on the Center's blog (http://oralhistory.blogs.lib.lsu.edu/). In order to maximize the amount of digitized materials that are available online, key players should outline a digital access plan wherein all materials that are currently ready for public display are listed and other materials are prioritized.

Conclusion

The T. Harry Williams Center for Oral History began a digitization program a little more than two years ago. Since then the Center has acquired digitization hardware and software, hired a full time staff member to perform digitization duties, and mounted several collections to the Louisiana Digital Library. By all accounts, the Center's efforts have been successful, although they hope to develop more sound workflows for digital access to enable them to add additional interviews to the online collections in the future.

Institutions wishing to emulate the Center should consult experts in person or through the literature, follow industry standards set forth by the Oral History Association (http://www.oralhistory.org), and, formulate plans based on best practices such as the CDP Digital Audio Working Group Digital Audio Best Practices (http://www.bcr.org/dps/cdp/best/digital-audio-bp.pdf). It is essential to plan ahead for storage space needs, keeping in mind that what one thinks you'll need is probably less than the reality.

References

Bond, T. J. & Walpole, M. (2006). Streaming audio with synchronized transcripts utilizing SMI., *Library Hi Tech 24*, 452-462.

Bond, T. J. (2004). Streaming audio from African-American oral history collections. *OCLC Systems & Services, 20*, 15-23.

Casey, M. & Gordon, B. (2007). *Sound directions: best practices for audio preservation*. Retrieved from: http://www.dlib.indiana.edu/projects/sounddirections/papersPresent/sd_bp_07.pdf

Hurford, A. A. & Read, M. L. (2008). Bringing the voices of communities together: the Middletown digital oral history project. *Indiana Libraries. 27*, 26-29.

Integration and the black experience. (2003). Retrieved December 14, 2009 from http://www.louisianadigitallibrary.org/cdm4/browse.php? CISOROOT=/IBE

T. Harry Williams Center for Oral History Exhibits and Presentations. (2009). Retrieved December 14, 2009 from http://www.lib.lsu.edu/special/williams/ep.html

The Baton Rouge Bus Boycott of 1953. A recaptured past (2004). Retrieved December 14, 2009 from http://www.lib.lsu.edu/special/exhibits/boycott/index.html

Weig, E., Terry, K. & Lybarger, K (2007). *Large scale digitization of oral history: A case study. D-Lib Magazine 13*. Retrieved from: http://www.dlib.org/dlib/may07/weig/05weig.html

Digitizing a Newspaper Clippings Collection: a Case Study and Framework for Small-Scale Digital Projects

Maureen M. Knapp (John P. Isché Library, New Orleans)

Abstract

How does a small specialty library establish, develop and maintain in-house digital collections? What are the considerations, challenges, and benefits they experience? This chapter describes one library's experience in turning an aging and inaccessible collection of newspaper clippings into a preserved and searchable online collection, which in turn laid a basis for other digital projects. This chapter also discusses considerations, challenges and opportunities observed during their first foray into creating a digital collection.

Keywords: Clippings, Digital libraries, Digital preservation, Digital projects, Digitization, Electronic preservation, Newspaper clippings file, Newspaper clippings, Press clippings.

Background

The John P Isché library is a mid-sized, urban, academic health sciences library serving six schools of health professions at the LSU Health Sciences Center (LSUHSC) in New Orleans, Louisiana. Established in 1931, the library has collected newspaper clippings related to the history and accomplishments of the health sciences institution since its inception, and even today monitors the local papers for pertinent news items. The "newspaper clippings file," as it

came to be called, is an astounding 70 year snapshot of the development of the health sciences in Louisiana. Over 6,000 clippings trace development of LSUHSC through the twentieth century, including such topics as: the people, places and events associated with the LSU School of Medicine, the growth of health infrastructure in Southeast Louisiana and New Orleans, and the development of 20th century health sciences education in Louisiana.

Digital Collection Origins

In 2002, access and preservation concerns with some of the earliest newspaper clippings encouraged the library to investigate digitization as a possible solution. Access points to the collection were limited. The only online access consisted of a locally-created subject database containing basic citations to newspaper articles from 1985 to present. Users had to search the local database by faculty name or department, and then locate the physical newspaper clippings in filing cabinets by call number. The remaining fifty-odd (1933-1984) years of the collection was indexed in a card catalog, stored in the library's back offices and only accessible to library staff.

Numerous problems plagued the physical collection. The newspaper clippings had been stored in filing cabinets as they were collected, which allowed the typing paper to curl heavily over the course of many years. The newsprint itself showed signs of age: rust marks appeared where staples and paperclips had once connected pages, and gaps in the collection were apparent.

A lack of funding and staffing was another concern. Any efforts towards creating a digital collection would have to be inexpensive and make use of staff and resources the library already possessed.

However, to truly understand to physical condition of the newspaper clippings file, and the challenges that would arise once digitization began, one must understand the collection process of gathering the original newspaper clippings. While no documentation exists, the library postulates that even back to the 1930s, a library member would skim the daily local papers from around Southeast Louisiana for any mention of LSU School of Medicine, and its faculty,

staff or students. Once an article was discovered, it was cut out of the paper, dated, and the name of the paper was noted. The articles were glued to standard 8 ½ by 11 inch typing paper, usually several to a page, somewhat in order by date, and the paper was assigned a numerical call number in the order they were received. Later someone would read the articles, underline named entities pertaining to LSU, and assign a subject heading, which was recorded in a small local card catalog. Finally, the pages of clippings were organized into manila folders by year and placed into filing cabinets until further needed. This entire process continued for 50 years.

So basically, the library had a unique local news collection, spanning the majority of the 20th century, collected and stored under questionable archival methods, with limited access to documents before 1985. In order to increase availability and use of the clippings, the library wrote a grant proposal for a small-scale digitization project to scan the newspaper clippings from 1933-1953, streamline cataloguing, and offer public access to the resource online. The grant proposed using Greenstone digital library software, an open source "suite of software for building and distributing digital library collections" (Greenstone digital library software, 2007), to provide access to the digitized newspaper clippings.

Stops and Starts

Though the grant proposal was rejected, the grant writing process did provide a catalyst for action within the library. The small grant requested $3,000 to purchase a flat-bed scanner, computer and optical character recognition software. Library administration was impressed enough with the grant's digitization plan that they provided funding for a scanner, software and travel to a continuing education class on digital projects in 2003. A library staff member began scanning the clippings. However, the library quickly ran into problems. The Greenstone software would not work properly on their secure intranet, and the library lacked a staff member with enough computer programming experience to install and troubleshoot the software properly. In addition, the image quality of the scanned

newspaper clippings was poor, which was attributed to a faulty scanner that did not produce dark enough images. Finally, copyright concerns made library administration hesitant to post the collection online to the general public.

By the time Hurricane Katrina struck New Orleans in August 2005, access, software and image quality issues had put the library's newspaper clippings digitization project on hold. The library's collection was undamaged from this natural disaster. However, it was moved to remote storage for over half a year and the entire library staff was displaced.

During the ensuing hiatus, library staff took several continuing education classes on digitization. "Digitization Fundamentals," a course offered by the Illinois Digitization Institute at the University of Illinois Urbana-Champaign (University of Illinois Library, 2009), was exceptionally useful, as it provided training in digital projects management, standards and organization, as well as an introduction to Photoshop software.

In 2007, an opportunity opened for the library to join the Louisiana Digital library, the state digital library consortium provided through LOUIS: The Louisiana library network (LOUIS: The Louisiana library network, 2009). The library was able to obtain access to OCLC's CONTENTdm platform, which was previously too expensive, as well as the technical infrastructure and support needed to store and access digital assets.

Consortial membership for digital library services addressed many of the problems faced by the library developing an in-house digital collection. The documentation on the technical and operational requirements for participation in the LOUISiana Digital library proved critical. The consortium's style manual for scanning and cataloguing provided guidelines for selecting collections to digitize, scanning practices, post-scanning image manipulation, project workflows, metadata standards, and quality control. Another practical advantage to consortial membership was LOUIS staff support, which provided advice on imaging standards, basic training on the

CONTENTdm software, and a shoulder to cry on when things went awry.

The library began their second try at developing a digital version of the newspaper clipping file in January 2008. As of December 2009, the library has not only met their original goal of digitizing and indexing over 1600 items in the collection from 1933-1953 (LSUHSC New Orleans library, 2009), but also created several other collections.

Work Flow, Image Manipulation and Standards

The format and organization of the newspaper clippings collection created a challenge in regards to digital manipulation and workflow. In order to achieve indexing of items on an individual level, some information that was included only once on a sheet of several newspaper clippings (for example, the name of the newspaper, the date, and most commonly, the clipping's call number) would have to be added to each individual item. Thus, several steps beyond simple scanning and image processing were included in the workflow.

Here are the workflow and standards for creating digital versions of the Newspaper Clipping File:

1. Following consortium standards for creating digital images for the Louisiana Digital library, the full-page newspaper clipping is scanned on an HP Scanjet 8390 flatbed scanner to create an archival black and white image at 300 dpi, 8-bit grayscale and saved as an uncompressed TIFF file on the library server. This creates an archival master version of the original digital image.
2. Using Photoshop, a copy of the archival master version is opened and saved according to file naming conventions for the digital library set forth by the consortium. This creates a duplicate of the archival master that can be manipulated to isolate an individual clipping. This file is the image that will eventually be loaded into the digital collection.
3. The duplicate is cropped to isolate a single newspaper clipping. Pages that have only one clipping on them are also manipulated and cropped to minimize file size.

4. If not visible, the call number, date and newspaper name from the original scan are copied, cut and pasted to the now isolated clipping.
5. Post capture processing is applied. The item is processed for alignment and an unsharp mask filter is applied to correct blurring that might have occurred during the scan process. In addition, the image's histogram is viewed to adjust color intensity.
6. The individual, processed image of the individual newspaper clipping is saved to the server.
7. For pages with more than one newspaper clipping, this process is repeated until all clippings have been isolated.
8. After digital manipulation, the TIFF of the clipping is loaded into the CONTENTdm Project Client. Cursory metadata is entered by a library staff member. The file name, size and location are recorded in a Scanning Log to track progress.
9. The librarian performs Optical Character Recognition (OCR) on the clipping to create an excerpted text field and assigns subject headings. OCR produces an abstract of the first 50 words of the article, which is keyword searchable in the digital library. This takes a bit of time, but it is a good way to review the article and assign the proper subject heading. After a final quality check, the item is approved and uploaded to the digital library. Upon upload, CONTENTdm converts the full resolution TIFF file to JPEG, which is what end-users access when viewing the collection online.
10. CONTENTdm also offers an Archival File Manager, which automatically archives collections in a location specified on our library server as they are uploaded to the online collection. Once a volume is full, it is burnt to an archival quality CD recordable disc, as well as saved on the server.

Cataloging and Metadata

The LOUIS consortium requires collections in the Louisiana Digital library to use the Dublin Core 15 metadata element set (Dublin Core Metadata Initiative, 2008), in addition to non-Dublin core structural

and administrative metadata. CONTENTdm allows up to 125 fields per collection. The library decided to add 3 more metadata fields to the newspaper clippings collection: Call number (to locate the item in the physical files), Full Text (for excerpted text) and Contact Information (so users can contact the library). The following lists the metadata fields used in the newspaper clipping collection.

Field Name (in CONTENTdm)	Type of metadata	Metadata Content	Added by
Title	DC	Title of newspaper clipping	LS
Contact Information	A	Contact information for library	T
Creator	DC	Author of clipping	L
Contributors	DC	Contributor to clipping (rarely used)	L
Subject	DC	Institutional controlled vocabulary, MeSH	L
Call Number	D	Call number for the original clipping	LS
Description	DC	"Newspaper clipping"	T
Notes	D	More descriptive information about content of original clipping, if needed	L
Publisher	DC	Newspaper title	L
Date	DC	Date of publication	L
Type	DC	"Text"	T
Format	DC	"TIFF"	T
Identifier	DC	Mandatory field directs users to identifier URL	T
Source	DC	Library name and homepage URL	T
Language	DC	"En."	T
Relation	DC	URL to homepage of Newspaper Clippings Collection	T
Coverage – Spatial	DC	"New Orleans (La.)"	T

Field Name (in CONTENTdm)	Type of metadata	Metadata Content	Added by
Coverage – Temporal	DC	Year of publication	L
Rights	DC	Copyright information	T
Cataloger	D	Initials of librarian	L
Cataloged Date	D	Date of cataloging	L
Object File Name	D	File name of item	LS
Image Resolution (Archival)	A	Dots-per-inch of scanned TIFF i.e.: "300dpi"	T
Image Bit-Depth (Archival)	A	"8-bit"	T
Color Mode (Archival)	A	Grayscale	T
Extent (Archival)	A	Pixel dimensions of image (WWWW:HHHH)	LS
Image Manipulation (Archival)	A	"Crop, alignment, unsharp mask, histogram"	T
File Size (Archival)	A	Size of TIFF image in KB	LS
Hardware / Software (Archival)	A	"HP Scanjet 8390, Photoshop, ABBYY FineReader"	T
Digitized By	A	Initials of library staff member	LS
Digitized Date	A	Date of digitization	LS
Full Text	D	Abstracted content from OCR	L

List of metadata elements used in cataloging items. Meaning of symbols: A is administrative; D is descriptive; DC is Dublin Core 15; LS is added by Library Staff, L is added by Librarian, and T is added by Template.

Many of these fields are inserted automatically via a template in CONTENTdm. The remaining fields are divided among project members. The most tedious data entry was entering the Extent and File Size fields for each item. Each clipping's dimension and size is

different, so library staff tends to write these down on a notepad as they scan images for entry, then record them in CONTENTdm and the scanning log later.

Another feature of Content DM is the ability to build a customized controlled vocabulary for the Subject field. This worked to the library's advantage, as the newspaper clipping file possessed a card catalog of subjects. The library uses the newspaper clippings card catalog as a basis to build an institutional controlled vocabulary in the digital library. The card catalog also serves as a reference point to verify names and spellings of affiliated persons. This institutional controlled vocabulary can be shared across digital collections, which is an advantage for future projects related to our institution.

The library soon recognized that other subjects would be necessary to adequately describe the digitized newspaper clippings. Original cataloging varied so much over the years that clippings might only include the name of the person or entity mentioned in the article. The library wanted to add more descriptors, so that articles describing conferences, publications, research grants or other common topics were easier to locate. When applicable, the library consults the National Library of Medicine's list of Medical subject headings (MeSH)(U.S. National Library of Medicine, 2009) for appropriate descriptors in the Subject field. For example, the MeSH term "Congresses as Topic" is used when a clipping discusses conferences, or the MeSH term "Publications" when a clipping mentions a new book or journal article published by one of the institution's faculty. Sometimes, MeSH is not useful, especially when discussing local events such as campus expansion or departmental news. In these cases, a subject heading is created and assigned by the librarian. Clippings in the digital collection can be browsed by year, subject, creator or title. Browsing by date is an interesting way to view the development of institutional history. To further open the collection, keyword searching is enabled in the excerpted text field.

Project Considerations

Storage, standards, documentation, training and staffing were all considerations for this project.

Storage was a huge concern. The deteriorating condition of older newspaper clippings made it evident that storing the physical newspaper clippings in filing cabinets was not conducive to preservation. To address the curling paper, books were used to weight down the paper for several weeks. This did not entirely fix the issue of curling paper, but it did help a little in preparing the clippings for a move to flat storage. After flattening, the files were transferred to acid-free archival folders and placed in clamshell archival storage boxes. Finally, the clamshell boxes of physical files were relocated to the library's humidity controlled Rare Books Room, in order to protect them from humidity and sunlight.

Likewise, the library was heedful of digital storage and the "digital mortgage": how will the library address transfer of archival TIFF files to new formats as software and hardware change? Though the library has yet to encounter a change in image format standards, they did attempt to prepare for this inevitability by storing the collection of archival images in multiple locations, as well as on multiple formats. Having multiple copies also addresses the possibility that some files might eventually become corrupted. TIFF versions of the images are burnt to an archive quality, professional grade CD recordable discs, as well as copied to a location on the library server, which is maintained by our institution and backed up daily to tape at a remote location. This is in addition to the processed JPG file that is available to the public on the Louisiana Digital library. A TIFF of the raw scan of the original newspaper clipping is also retained on the library server.

With multiple storage locations and a complicated workflow, documentation and staff training are also important concerns. The library's consortial membership provided a style manual for scanning, cataloging/metadata standards, and basic workflow suggestions. The library used this as a basis for creating a local workflow policy, which includes detailed directions on image scanning and manipulation as well as step by step directions on how to process the item in

CONTENTdm. A scanning log is used to track size and progress of a collection. The scanning log is simply an Excel file which records the file name, file size, and date of digitization, as well as locations to which the file has been saved.

Regarding training, the library realized it was critical that everyone involved with the project learn Photoshop. The LOUIS consortium takes a 'train the trainer' approach to CONTENTdm, so the librarian was responsible for training local staff on the software after initial training.

This project is staffed with one librarian and two library staff members, who devote about 10 hours a week to this project. Library staff is requested to scan and process 60 clippings per week. Scheduling issues quickly became apparent for the librarian project manager, who has bibliographic instruction and reference desk duties in addition to overseeing digital projects. A supervisor suggested setting aside one day a week to solely devote to digital projects. Friday has since become "Digitization Day" and has worked well in keeping the load of items to be processed and approved by the librarian to a reasonable amount.

Benefits and Challenges

One of the first challenges was software sustainability. The free Greenstone digital library software did not work within the institutional intranet and required higher level technical skills than the library possessed. In addition, problems with the original project scanner resulted in poor quality images that had to be redone.

Support from your institution from inception is critical. Administration has to be on board to provide funding and act as a liaison to other resources, for example, consulting with your institution's legal department about copyright questions. Support from information technology (IT) is also important. Getting our IT department to provide support for open source library software was a challenge that soon put the library's original plans to use Greenstone digital library software on hiatus. One of the benefits of membership in a state digital library consortium is that technical support is

provided in an automated timely manner. In addition, the consortium has direct contacts with the software developers at CONTENTdm, so software concerns are quickly addressed.

The newspaper clippings collection is unique in that it collects clippings from many regional news sources. All materials were published after 1923. Therefore, the work may be protected by copyright until 2018. Violation of copyright was a large concern, so the library decided to restrict access to the images within the newspaper clippings collection to the institutional IP address. In order to share the collection with a larger audience, the collection's metadata is searchable and viewable to anyone. This way, any user can find items in the newspaper clippings collection, and if they are not from the institution, the library works with them to get the information or clippings they need.

Funding is a final challenge. Consortial membership to the digital library is about $2000 a year, while hardware and software ran about $1500 in startup costs. In addition, the library director donated a 21" screen won at a library conference raffle for use with the digital projects computer. Digital imaging is much easier with a larger screen. Grants and scholarships are another source of funding. A scholarship from a regional medical library group helped fund attendance at the first continuing education class on digital imaging and metadata for the librarian project manager. An recent Institute of Museum and Library Services "Connecting to Collections" Bookshelf grant (Institute of Museum and Library Services, 2009) allowed the library to obtain a set of conservation resources and books, which was previously non-existent.

The library now has over 10 years of institutional history available online in a searchable database. Visibility and access to this collection has increased. Indexing though OCLC allows results to appear in Google. As a result, the library has received several inquiries about subjects indexed in the newspaper clipping file from the United States and Italy. The clippings file has also acted as a catalyst for change, inspiring library staff to organize the rare books room, research archival storage methods, and apply for grants. One of the benefits the

library is proudest of is the mentoring opportunity this created. A library staff member who helped start this project recently completed their library degree and went on to become a Digital Initiatives librarian at another local library.

The library has established a workflow and gained experience in digital imaging and management for future projects. Because of the success in creating the newspaper clippings collection, the LSUHSC School of Dentistry started a digital collection of historic photographs. In addition, the library worked with the LSUSHC Registrar's Office to digitize graduation program records, which are now available in a public, searchable collection. Finally, the library is in the planning stages of creating a digitized version of early volumes of the medical school student newspaper. The library also continues to add items to the newspaper clippings collection.

As one can surmise, it has been a long 4 years to produce this digital collection, but once the library established workflow and standards it was much easier to begin other projects. Support from the state library consortium certainly expedited and streamlined the process, and the library recommends state or regional consortium membership to any smaller institution considering developing a digital project. For all the tedious data entry and malfunctioning software, the creation of an enduring, searchable and accessible source of institutional history made the entire project worthwhile.

References

DCMI Usage Board. (2008). *DCMI type vocabulary*. Retrieved December 9, 2009, from http://dublincore.org/documents/dcmi-type-vocabulary/

Dublin Core Metadata Initiative. (2008). *Dublin core metadata element set, version 1.1*. Retrieved December 9, 2009, from http://dublincore.org/documents/dces/

Greenstone digital library software. (2007). Retrieved December 9, 2009, from http://www.greenstone.org/

Institute of Museum and Library Services. (2009). *Connecting to collections: A call to action*. Retrieved December 9, 2009, from http://www.imls.gov/Collections/

LOUIS: The Louisiana library network. (2009). Retrieved December 9, 2009, from http://appl003.lsu.edu/ocsweb/louishome.nsf/

LSUHSC New Orleans Library. (2009). *LSUHSC New Orleans newspaper clippings collection homepage*. Retrieved December 10, 2009, from http://www.louisianadigitallibrary.org/cdm4/index_LSUHSC_NCC.php?CISOROOT=/LSUHSC_NCC

U.S. National Library of Medicine. (2009). *Medical subject headings - home page*. Retrieved December 9, 2009, from http://www.nlm.nih.gov/mesh/meshhome.html

University of Illinois Library. (2009). *Digital services and development -- training*. Retrieved December 9, 2009, from http://images.library.uiuc.edu/projects/newproj.htm

METRO Grant Success Story: Waterways of New York Project

Claudia A. Perry and Thomas T. Surprenant
(Queens College, CUNY.)

Abstract

The concept of experiential learning is particularly useful when students are required to create database entries as part of an ongoing, real-life, online experience. A METRO grant resulted in an opportunity to use students to create a CONTENTdm database which, with the continued software support from METRO, has continued and evolved until the present. This chapter describes the experience of both faculty and students. Sections include the background, technical issues and implications for teaching, project procedures and workflow, successes and lessons learned, challenges and next steps. Of particular interest is the use of out of copyright postcards and the metadata that has resulted from intensive student study and evaluation of the data contained on these cards. Those contemplating a digitization project of their own will be able to learn much about best practices, project planning, management and the advantages/disadvantages of the CONTENTdm software.

Keyword: Best Practices, Canals, Case Studies, Cooperative Learning, Digitization, Digital Collection Management Software, Digital Collections, Digital Imaging, Experiential Learning, Library Education, Metadata, Postcards, Project Based Learning, Project Management, Project Planning, Quality Control, Standards, Student Developed Materials, Student Participation, Student Projects, Waterways.

Introduction

For many of us, hands-on learning is the best way to integrate an understanding of principles and best practices with a practical grasp of the actual challenges and learning opportunities of a project. This is particularly true for library school graduate students seeking to expand their theoretical, technical and management skills. As digitization is increasingly seen as a worthy endeavor for even the smallest institutions, it is worth considering the range of approaches available for gaining needed expertise, especially at the novice level. Examining the long-term development of an integrated, semester-long, course-based approach to digitization may be of value for those seeking an inexpensive approach for the creation of small to medium-sized digital collections.

A course entitled "Introduction to Digital Imaging" was first taught at the Queens College Graduate School of Library & Information Studies (GSLIS), City University of New York (CUNY), in the Fall of 2003. In the Spring of 2005, a year-long METRO-funded grant facilitated a co-operative project between the Rosenthal Library and the GSLIS to support student digitization of a portion of the Queens College Rosenthal Library Archives (e.g. see GSLIS, 2005-2009, Digitization projects). The project included a variety of forms and formats. The evaluation of this valuable learning experience identified a strong need to find a single standard format that was information rich and moderate in scope, but which lent itself to more uniform metadata standards and digital specifications. The evolving project, "Waterways of New York", an online digital collection of historical postcards, was created in 2006, and partially supported by METRO through continued access to CONTENTdm. It continues to be extended by GSLIS students each semester the course is taught.

Scope and Format

The most important feedback provided to our team by METRO digitization experts regarding our "Rosenthal Library Archives" initiative was the value of working with a limited number of manageable formats and a relatively focused subject area and time

frame. During the implementation of the grant a serious problem was the complexity resulting from too many different types of media, the overly wide range of subject matter, and the challenges these characteristics presented to the creation of consistent metadata.

One of the GSLIS professors, Thomas Surprenant, has an ever-expanding collection of Erie Canal and related New York State waterways antique postcards, which addressed many of the problems noted in the METRO feedback. In particular, by selecting a single, simple, information rich format—postcards published before 1923—copyright concerns were eliminated and only a single set of digitization specifications needed to be developed. METRO's willingness to host the collection on their CONTENTdm server simplified selection of Dublin Core as the metadata standard, and use of a subset of the Library of Congress Thesaurus of Graphic Materials (TGM) for standardized metadata terminology (Library of Congress, 2007). This greatly aided our ability to develop a manageable set of project-specific guidelines that could be adequately addressed by the evolving documentation.

The choice of postcards as the source medium turned out to be far more interesting to the students than was expected. An initial option to describe the backs--as well as the front images of cards—was enthusiastically embraced by virtually all of the students and became the norm for subsequent classes. Hand-written messages, address conventions, postmarks, trademarks and other attributes of the cards were at times as much or even more rewarding to analyze than the front images themselves. Further, student interest in the varied aspects of architecture and activities of daily living portrayed in the postcards led to an expansion of emphasis far beyond the initial focus of the project on locks, canal boats, shipping, waterways and transportation.

Background, Technical issues and Implications for Teaching

Any planning for digitization requires a detailed analysis of one's institution, and an assessment of where the proposed project fits into

its mission and priorities. Further, consideration of the potential audience(s), project goals and objectives, resources and limitations, oversight and long-term maintenance are among the many issues to be addressed (e.g. see JISC Digital Media, 2008: Project management; North Carolina Echo Project, 2007). These considerations inevitably will shape the nature of the evolving project. It is important that an honest appraisal be conducted, committed to writing, and approved by the appropriate governing bodies. However, the nature of digital projects ensures that adjustments inevitably will be required over time. Changing standards, software and hardware upgrades, technical glitches, and shifts in the growth of a project are just a few of the issues which must be dealt with, often on very short notice. Planning and documentation therefore should be viewed as an iterative process, where ongoing evaluation is used to address and correct for changing circumstances.

Creating a list of stakeholders and intimately involving them in this planning process is critical to success. In our own case, student feedback on procedures and emphasis has been an invaluable aspect of the evolving project. Each incoming class section serves as a de facto Advisory/ Editorial Board that contributes to the decision-making process. These contributions include identification of additional TGM terms for our thesaurus, the development of standardized Trademark descriptions, fine-tuning of documentation and lab handouts, and increasingly higher expectations for the quality of the metadata. Within a more traditional library environment, all members of the digitization team, as well as users and other staff members, undoubtedly will have many valuable insights to contribute.

Among the key elements shaping the evolution of a project-based digitization course at Queens College were the following:

Institutional characteristics

- When the initial course was developed it was necessary to have the course proposal cleared with the departmental Curriculum Committee after consultation with the Chair. This required the development of course goals and objectives, specific readings, and course assignments and activities.

- After three semesters teaching the course it was submitted to the GSLIS and College Graduate Curriculum Committees, Faculty Senate and, ultimately, the CUNY Board of Trustees for approval as a permanent course.
- An understanding of the possible pitfalls of the process at every step was important to ensure that all potential hurdles were considered and cleared.
- Even outside of explicitly academic environments, proper attention to obtaining documented approvals and support from key stakeholders--at all levels up to the governing board—will prove invaluable in avoiding challenges and ensuring continued buy-in by the institution and other funding agencies.

Lab facilities (capabilities and challenges)

- For our project we were able to use a 16 workstation Mac lab with direct connections to the Internet. The lab had been expressly designed by GSLIS faculty for digitization-related activities and hands-on learning, in close collaboration with the Queens College Office of Converging Technologies (OCT) and college architectural staff, in conjunction with the development of the course proposal. Appropriate institutional commitment to fund, support and regularly upgrade such a lab was, and is, essential to the continuing success of the project.
- Specifications included an instructor workstation (in addition to student Macs), ceiling mounted projector and wall screen for demonstrations, two (eventually three) flatbed scanners, SilverFast AI scanning software, Photoshop, and the Microsoft Office Suite, particularly Excel.
- A major continuing challenge concerns computer and software upgrades. The OCT staff do not always consider the rhythms of the academic year in making changes to the lab, which regularly causes problems, even after many years of teaching the course. For example, in the Fall 2009 semester alone, new computers were installed during the first week of the semester. This resulted in equipment and software glitches, and a delay in the availability

of the lab, as well as the need to test software functionality and then revise/upgrade lab handouts with minimal advance notice.

- While we were grateful for the new equipment (a regular replacement cycle is essential for ongoing functionality), timing issues resulted in a rough start to the semester.
- A major equipment problem for us was solved when Apple changed to an Intel CPU. The new Mac computers are now dual boot (Apple and PC Operating Systems), meaning that they can now run CONTENTdm (CDM) using the Project Client software interface. Previously, lab sessions had to be specially scheduled in a nearby PC lab (CDM Project Client software is not available for the Mac OS). However, dual boot capabilities have created additional problems of compatibility, accessibility and ongoing troubleshooting.

Software

- As noted above, new equipment means software installation and the attendant complications. The specialized nature of our lab, and lack of teaching assistants, necessitates that course faculty test all functionalities and work with OCT staff to address problems. Oftentimes this has meant repeated testing and troubleshooting, frequently a day or two prior to a scheduled class. Such technical malfunctions can wreak havoc on the best-planned teaching schedule.
- While CDM has been sufficient for our needs, and we are extremely grateful to METRO for their continuing support, there are still some issues that cause concern. The biggest is that students cannot directly upload their input into the database due to the administrative rights structure. This situation requires another level of review by the course instructors serving as database administrators/quality control experts, adding substantially to time demands near the end of the semester. In addition, after students submit their data entries for approval, editing ability on their side is extremely constrained, by both time and software limitations.

- More recent upgrades appear to have adjusted this limitation, permitting downloading of materials from the live database for additional editing if errors are detected. However, this creates additional levels of oversight and complexity, and assumes that the instructors will be able to approve the uploads in time for the students to review and make changes. This is simply not readily accommodated within a 15 week course schedule.
- Further, although recent versions have been more stable, in the past CONTENTdm has crashed frequently, causing much frustration on the part of both students and faculty.
- These points emphasize the steep, and ongoing, learning curve of digital project-based courses for faculty, support staff and students.

Support

- Adequate and timely support for equipment and software is essential to any technology-based project. The GSLIS has a number of student computer assistants and a campus-wide Help Desk, but as noted, the specialized nature of our lab sometimes puts it outside of the realm of their expertise.
- It is good practice to fully document and save ALL help desk requests and related support communication. These include emails, screen shots and help desk tickets. These records of ongoing and recurring problems have proved to be invaluable in our efforts to ensure follow-through, and to support our case when requests have not been fully resolved to our satisfaction or when problems repeat themselves.
- CONTENTdm Help seems to work best when we go through METRO. That means that an additional layer of contact needs to be activated anytime is a problem. That said, all relevant staff at METRO, over the years of this project, have been incredibly knowledgeable, supportive and responsive to our needs.

Staffing/oversight

- Experience suggests that having a subject expert for image content is a critical factor. The faculty have, or have developed over time, sufficient expertise to assist students in their metadata and description activities.
- Given the need to protect the postcards and the equipment the lab has to be under supervision whenever anyone is working. This greatly adds to the time burden of both faculty and staff.
- Postcards are stored in archival quality sleeves and students use white gloves when handling the postcards while scanning.

Class size and student characteristics

- The class size is dictated by two elements: the number of work stations and the work volume. Experience suggests that all students need to have access to their own workstations, and two workstations are dedicated to scanning use (a third must be shared between functions). The initial classes scanned, created metadata and submitted for approval six cards (front and back), but many of these initial canal cards were fairly simple rural scenes. Given the amount of detail that has emerged in postcards in later semesters, we have gradually reduced input to three cards (front and back), because the quality and quantity of the metadata has increased substantially. The time spent in quality control by instructors has increased commeasurably, despite having students doing quality control on their partners' work.
- Those involved in the digitization process are best served with at least an intermediate level of computer and software expertise. We have constructed our teaching labs in a step-by-step fashion, and utilize in-class time extensively. This allows the faculty to introduce and demonstrate skills and to detail the various steps of the project.
- In our experience, students who are highly competent in computers and/or relevant software or metadata creation have been more than eager to assist their classmates. This leads to a

highly supportive class environment in which all learn from one another, modeling (one hopes) the ideal workplace environment.

- However, a fair number of students have no previous familiarity with the Mac OS, Photoshop, scanning and related software, which complicates the pace at which the class can proceed. The nature of our curriculum and scheduling constraints make it difficult to require pre-requisites beyond the required core courses. Consequently, a teaching assistant to help in quality control, and in the provision of additional technical support in lab sessions, would be extremely desirable for all.

Evolving nature of the target collection

- Initially, the Waterways Post Card Collection consisted principally of cards of the Erie Canal, with collections of additional New York State canals such as the Oswego, Seneca, and Champlain Canals. As demand for the course has remained steady, indeed increased, the diminishing availability of canal-related cards posed a potential problem. (Most cards are currently obtained on eBay, and changing availability in geographic scope is an interesting topic for another paper.) On the other hand, many of those canal-related cards that have become available are increasingly distinctive.
- With the Quadricentennial Celebration of the discovery of the Hudson River approaching in Fall 2009, it appeared to be a logical extension of our collecting scope to extend to another key New York State waterway, the Hudson River. We included in our selection criteria cards depicting New York Harbor and the East River. The first such Hudson River cards were digitized in Fall 2008.
- The expansion in scope of the cards created fascinating but unanticipated challenges. Publishers, trademarks, and the increasing complexity of the images depicted required a substantial expansion of the TGM thesaurus, as well as the development of descriptions of an increasingly diverse set of trademarks, logos, stamp boxes and postmarks.

Consistency and accuracy

- Digitizing any collection over a period of time by a changing group of participants creates somewhat greater consistency and accuracy problems than might apply in a short-term. In spite of the iterative editing of documentation, inconsistencies and errors are regularly emerging in our project. To a large degree, this is due to the pressures of a fast-paced curriculum, a constantly changing panoply of operating systems, software versions, additions to our thesaurus and the ongoing, changing nature of the cards within our purview..
- Our experience has shown us that the students themselves are the best editors in catching errors and inconsistencies. It is obvious that the road to "perfect" metadata, documentation, labs and handouts is continuous, difficult, and perhaps, ultimately elusive. Such is the nature of a work in progress.
- It helps to have students who have a keen eye for detail as well. The project was significantly enhanced when then-student Susan Savage completed an Independent Study project in Spring 2007, that corrected many of our past mistakes, and developed the scaffolding of our current metadata documentation. It is now obvious that outside help in editing is an important part of the process (although not easily achieved).

Key readings and course activities and relevance for the project

- Clearly, carefully selected readings are critical to the success of a digitization project. We provide access to a range of resources in an effort to meet the needs of those at varying levels of familiarity with digitization and related issues. Alumni feedback has suggested the importance to many of providing continued access to the resources once our graduates are working in the field. Once involved in a real world project, many become even more aware of the importance of items that may not have seemed salient at the time of the course.
- Students complete a "Tech Review Exercise" in week seven, to document their understanding of key technical concepts. This

- That said, it continues to surprise the faculty that: 1) the quality of most metadata submissions is so impressive, and that 2) there remain previously unidentified errors in what seems to be a fairly strict process. In this regard the quality control process is working as envisioned.

Below is a transcribed example of the metadata for "1609 • HUDSON-FULTON CELEBRATION • 1909 [front caption] (1front) [h0189ac1]" after it has been uploaded to CDM.

Title: 1609 HUDSON-FULTON CELEBRATION 1909 [front caption] (1front) [h0189ac1)

Creator: Copyright 1909 J. Koehler, N.Y. [indicated on front only]

Subject—Front: Cliffs, Clouds, Flags, Grasses, Portraits, Rocks, Ropes, Schedules (Time plans), Ship equipment & rigging, Shrubs, Smoke, Smokestacks, Steam engines, Trees, Men, Passengers, People, Color postcards, Sailing ships, Side wheelers, Aerial views , Rivers.

Description–Front: A commemorative postcard celebrating the 300th anniversary of the discovery of the Hudson River, with portraits of Henry Hudson and Robert Fulton superimposed over a daytime aerial view of the Hudson River. Prominently featured are a sailing ship (circa 1609) and steamship (circa 1809) [presumably the Claremont] which together serve to commemorate the passage of time from discovery to the modern day. Soaring cliffs line the far bank and along the near bank; at right, there is a gathering of people (perhaps Native Americans). An information box titled 1609 HUDSON-FULTON CELEBRATION 1909, lists the following 15 events: Sept. 25 Commencement Day N.Y., Sept. 26 Religious Observance Day N.Y., Sept. 27 Reception Day N.Y., Sept. 28 Historical Parade N.Y., Sept. 29 Commemoration Day N.Y., Sept. 30 Military Parade Day N.Y., Oct 1 Naval Parade N.Y., Oct 2 Naval Carnival Parade N.Y., Oct 3. Religious Day Upper Hudson, Oct. 4 Dutchess Co. Day, Oct. 5, Ulster Co. Day, Oct. 6 Green Co. Day, Oct. 7 Columbia Co. Day, Oct. 8 Albany Co. Day, Oct. 9 Rensselaer Co. Day. COPYRIGHT 1909 BY J. KOEHLER, N.Y. [indicated on front only].

Coverage – Geographic: Hudson River, New York and New Jersey

Date Original: 1909?

Publisher: Graduate School of Library and Information Studies – Queens College (CUNY), New York, New York

Language: eng

Source Height: 3.5"

Source Width: 5.5"

Source: Waterways Post Card Collection of Thomas T. Surprenant: Hudson River

Type: Text; Image