Volunteer Voices: Tennessee's Collaborative Digitization Program

Tiffani Conner
Ken Middleton
Melanie Feltner-Reichert
University of Tennessee, Knoxville, mfeltner@utk.edu
Andy Carter

Follow this and additional works at: https://trace.tennessee.edu/utk_libpub

Part of the Library and Information Science Commons

Recommended Citation

This Article is brought to you for free and open access by the University Libraries at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in UT Libraries Faculty: Peer-Reviewed Publications by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.
Abstract

This article provides an overview of Volunteer Voices, Tennessee’s statewide digitization program. The authors focus on the three-year Institute of Museum and Library Services (IMLS) National Leadership Grant that provided the foundation for future growth of the digitization program. In addition to an overview of the content selection, metadata issues, software selection, digital preservation, and K-12 education emphasis of the grant project, the article includes a detailed description of the work done by the digitization and content specialists from across the state who selected and scanned items. The article concludes with a look at post-grant efforts to promote the sustainability of Volunteer Voices.

Keywords: Digitization; Digital Libraries; Project Management; History; Digital Humanities; Grants; National Leadership Grant; IMLS; Volunteer Voices

Editorial Note: Collaborative digitization projects continue to be an attractive and popular method of meeting the growing demand for electronic resources. This article presents a substantive account of the experience of Tennessee’s Volunteer Voices. Due to time and length constraints, not all significant questions related to this project could be addressed in this article. Thus, the authors welcome your inquiries; Collaborative Librarianship also welcomes readers’ response. Contact the General Editor, igaelz@regis.edu.

Introduction

Many states have developed collaborative digitization programs in an effort to save money, to allow smaller institutions to enter the digital library movement, and to promote digital collections to the K-12 community and lifelong learners. Currently, there are ongoing collaborative digitization programs in at least forty states. Programs in California, Colorado, Georgia, Kentucky, and other states have provided useful models for designing and implementing collaborative programs.

Tennessee’s statewide digitization program, Volunteer Voices (VV), began in 2003 under the leadership of Tenn-Share, an organization that promotes resource sharing among libraries. A 2004 Tennessee Digitization Survey found that although there was a relatively low level of digital collection development activity in the state, there was a high level of interest in learning how to create digital collections. Furthermore, large institutions (University of Tennessee Libraries, for example) housed most of the digital collections in the state that were being developed. This discrepancy fueled discussion of an approach tailored to assist smaller institutions in creating digital collections. This
led program leaders to seek grant funding for a collaborative digitization project.

In 2005 the University of Tennessee Libraries received a three-year National Leadership Grant from the Institute of Museum and Library Services (IMLS). With nine partner institutions, the project employed or otherwise engaged more than forty individuals from across the state to complete content selection, scanning, metadata creation, training, networking, conference and group presentations, computer programming, and preservation work. The goals of the project were to provide open access to ten thousand digital items from Tennessee cultural institutions; to provide access, use, and technical training to K-12 educators and students as well as cultural and information agency employees; and to initiate the development, growth and sustainability of a statewide digital library that would make Tennesseans proud.

This article explores our experiences implementing the IMLS-funded project. The authors offer perspectives of the Project Director, the grant Co-Principal Investigator, the metadata specialist, and the Digitization & Content Specialist (DCS) who selected and scanned items in the western part of the state. The issues addressed will interest collaborative digitization program planners and managers. These issues raised many questions: What collections are worth digitizing? How can the project make the best use of digitization centers, portable digitization equipment, and digitization personnel? Which metadata schema should be used? How do we encourage use of the digital collection in the classroom? Above all, what steps need to be taken to build a sustainable digitization program?

In this article we will present in detail what worked, what did not work and why, and what we would have done differently.

**Content Contributors and Content Selection**

Project staff cast a wide net when considering which institutions to target for contributing content. Although the collection
Figure 2. This map represents the distribution of content about specific Tennessee counties in the Volunteer Voices database. The stars represent the types of institutions that contributed content: libraries (solid blue stars), museums (open blue stars), and archives (red stars).

represents libraries of all sizes, a special effort was made to include content from other cultural heritage institutions. Museums contributed images of artifacts, offering end users a perspective that is often missing in even large primary source digital collections. Images of artifacts contributed by these cultural institutions include musical instruments from the Museum of Appalachia, a piece of a patterned flour or feed sack women used to make clothing from the Englewood Textile Museum, and a segregated drinking fountain sign from the Roy Bailey African-American Museum and History Center. In addition, twenty county archives contributed a wide range of county records, including land grants, Chancery Court records, apprentice indentures, and bills of sale for slaves.

The theme of the collection, “The Growth of Democracy in Tennessee,” was selected to highlight the many relevant historical topics that address national themes such as slavery, the Scopes trial, World War II, and civil rights. This thematic approach not only helps teachers to integrate the collection into the K-12 curriculum, but also attracts an audience beyond Tennessee. Targeting these themes also increased the likelihood that multiple institutions would contribute items about similar issues. For example, Scopes trial materials include sheet music from Middle Tennessee State University’s Center for Popular Music, as well as photographs and additional documents from Bryan College, the Tennessee State Library and Archives, and the University of Tennessee.

Figure 3. Worker at machine (Courtesy of Union University)

Work of Digitization and Content Specialists

The Digitization and Content Specialists who worked in the field travelling across the state to collect materials on behalf of the project were the points of contact for contributing institutions and they were responsible for the majority of the collection development. Each DCS assumed responsibility for one-third of the workload and covered a territory roughly equal to one-third of the state. The East Tennessee representative worked from the University of Tennessee in
Knoxville; the Middle Tennessee representative worked from the Tennessee State Library and Archives in Nashville; and the West Tennessee representative worked from the Central branch of the Memphis Public Library & Information Center in Memphis. From these regional bases, the DCSs would travel to surrounding counties to meet with interested librarians, archivists and museum curators.

The DCSs had several responsibilities including outreach, item selection, scanning, image editing, and metadata creation. These tasks proceeded in a loose order that remained similar from institution to institution. A DCS would contact an institution, or, as word-of-mouth spread, respond to an inquiry from an institution. After the initial contact, the DCS arranged an introductory visit to discuss the project, the needs of the institution, and the resources that Volunteer Voices could provide to meet those needs.

After the first visit—and assuming a working agreement—the DCS scheduled an extended on-site visit during which the final selection and scanning of materials would take place. The majority of the DCS’s time was spent scanning images and creating the appropriate metadata before sending the images and their associated files back to the University of Tennessee for final processing.

Having a DCS working on-site was one unique aspect of the grant project. This arrangement meant using portable digitization equipment and establishing a clear workflow that would accommodate their various responsibilities. It was not feasible, given both the nature of archival materials and the limited resources of the DCS, for the items to be moved to a centralized site prior to scanning. To facilitate this work, project personnel at the University of Tennessee created two web accessible tools: an administrative database (AdminDB) that was used to create filenames and track collections, and the Volunteer Voices Metadata Object Description Schema Workbook (MODS Workbook) that was used to create metadata records. Combining the portable hardware with online tools freed the DCSs to work in ways that best accommodated the diverse needs of the institutions contributing to the project. Typically, on-site work lasted from a couple of days to a full week and often required the DCSs to stay in town. Because transporting the equipment was necessary (an oversize flatbed scanner, a laptop computer, and an external hard drive), each DCS was given a flat-bed dolly; strapping; and padded, water-resistant cases to protect the equipment.

The DCS frequently worked at a spare table in a reading room or back office. Materials were scanned at 400ppi using SilverFast SE to manage the image quality and to make color corrections. Prior to saving each image as a TIFF file, all images were de-skewed and cropped in Photoshop. Each TIFF file was sequentially numbered and saved on the external hard drive. In order both to provide these files with identifiers and to begin the process of creating metadata for the XML records, the next step was to create records in the AdminDB.

The AdminDB tracked contributions to VV at the institution, collection, and item levels. The database automatically generated a unique filename for each image based on the institution, collection, and item level. Each item record in the AdminDB contained fields for descriptive and administrative metadata. At the beginning of the project, extensive descriptive information was entered into the AdminDB. However, the DCSs identified many constraints and quality control issues when attempting to work in the field with limited or no Internet connectivity. In order to address these issues, University of Tennessee programmer Christine Deane developed a metadata creation tool. The Volunteer Voices MODS workbook, using a simple graphic user interface, pulled information from the AdminDB including identifier, title, creator, and description, and posted it to the appropriate field in the XML record. Prompts embedded in the code required DCSs to complete obligatory fields, and a number of scripts and queues such as pre-loaded controlled vocabulary...
lists and Tennessee-specific Library of Congress subject headings (LCSH), helped to overcome difficulties in using LCSH and choosing other standardized, descriptive elements. After completing both the scanning and AdminDB data entry, the DCS used the MODS workbook to generate MODS XML compliant records for ingestion into the Digital Library eXtension Service (DLXS) database. Because creation of the MODS records was the most time consuming step in the process, this work was completed at the DCSs home institution.

**Metadata**

Potential descriptive formats for the collection included Dublin Core (DC) and Metadata Object Description Schema (MODS). While DC is widely used for collaborative digital projects, the metadata specialist chose MODS because of the functionality features and the depth of description desired by project partners. Using the 2005 draft release of the Digital Library Federation MODS Implementation Guidelines for Cultural Heritage Materials as a starting point for best practices, the Volunteer Voices (VV) wrote its MODS profile both to conform to the emerging standards for shareable MODS records and to meet the specific needs of the VV collection. Developed in close collaboration with project partners across the state, the VV MODS profile articulates the controlled values and vocabularies and the required, optional, and repeatable elements of MODS that are applied to VV metadata records. Extensive tailoring of the VV MODS profile enabled the project to provide adequate representation of cultural heritage materials. For instance, physical attributes of artworks, such as medium and technique, were included. The result was a flexible and rich schema providing a deep level of descriptive access to contributed materials. Additionally, the use of locally controlled vocabularies enabled DCSs to tie the content concretely to Tennessee history and curriculum. For example, every object in the collection is both connected to an era...
in Tennessee history and associated with subjects in the state social studies curriculum.

**Digital Content Management Software**

The University of Tennessee Libraries, the institution processing the project records, was using Digital Library eXtension Service (DLXS) for other digital collections. For this reason, DLXS was selected as the digital content management software for the project. The ability to handle the MODS metadata schema and the advanced full text search options in the DLXS interface were additional advantages. However, we discovered that DLXS is not a sustainable option for moving the project beyond the grant period. The reasons include a considerable time delay in processing records, requirement of a full-time programmer, a difficult end-user interface, and record display problems.

**Digital Preservation**

The University of Tennessee Libraries controls and maintains digital preservation for the project. Their system employs a Sun Storage L100 Tape Library for offline storage of all digital content, and a second copy of the digitized archival masters is stored on a dedicated (Dell 2850) server with a seventeen terabyte RAID array. Online archival masters and their metadata are monitored daily by MD5 checksum verification cron scripts and a series of fail-safe backup measures, including flat-file backups for checksums, two MySQL databases (one for checksums and another for script logging), and sequential cron scripts on three servers with email updates sent automatically to system administrators. The system is designed to prevent altered files from being written to tape and ensures that valid backup tapes still exist to replace any corrupted files. Incremental tape backup occurs nightly with a full backup completed each week. The current backup protocol uses three sets of tapes with off-site rotation occurring daily.

**Education and Outreach**

Another distinct aspect of the Volunteer Voices project was the inclusion of a licensed educator who travelled throughout the state training teachers in the K-12 system (both public and private) to use Volunteer Voices in their classroom. The Education Coordinator travelled extensively both organizing daylong training sessions and working directly with nearly forty-five teachers and media specialists who were identified as master teachers by their school districts or principals. In exchange for participating in the all day training sessions, these master teachers received a stipend, lunch, and reference materials for their classroom. They created lesson or activity plans to share via the project web site and gave a presentation to their own school or district about Volunteer Voices.

Overall, the education workshops were a rousing success and met our goal of spreading the training and news of Volunteer Voices across the state. The following comments by master teachers further reflect the positive effects of the workshops:

- “I am amazed at all the available sources on the Internet. I will be researching and using more of these sources.”
- “I am excited about tying Tennessee history into my regular curriculum.”
- “I love knowing of a simple website where I am able to get primary sources that are reliable.”
- “Volunteer Voices opened up a whole new source to help make my teaching more interactive.”

The education coordinator also reached administrators and teachers beyond the classroom by giving presentations at conferences as varied as the Tennessee Educational Technology Conference, Tennessee Association of School Librarians, and the Tennessee Geographic Alliance. These presentations included short overviews of the project and pre-conferences that provided hands-on experiences for attendees. As the project progressed, the demand for the Education
Coordinator’s presentations increased to nearly all counties in Tennessee, another indication of the importance of the project and of interest on behalf of teachers.

Digitization Training

Librarians and cultural agency employees were the primary target groups for digitization training. Many libraries and cultural institutions across the state were interested in creating digital collections, but few were creating content, and those who were involved in digitization identified a lack of knowledge and equipment as the primary barriers. The grant provided in-person training sessions across the state for cultural agency employees to gain a general understanding of the back-end work of creating a digital library collection. During the three-year period, four intensive digitization workshops were held in each region of the state: at the University of Tennessee, the Society of Tennessee Archivists Conference, Cleveland State Community College, and Jackson State Community College. The workshop attendees included library, museum, and archives employees. These free sessions allowed personnel to learn about the current state of digitization across the state, to receive basic technical instruction with specific software, to have discussions concerning selection of items for inclusion, copyright issues, and ways they could get involved. The project director led these training sessions and created a technical manual for the sessions. In total, more than eighty cultural heritage institution employees received training and information for creating their own digital collections as well as tips for getting involved with the statewide project, Volunteer Voices.

Summary of Challenges and Lessons Learned

The Volunteer Voices project illuminated several issues that arise when librarians create a multi-institution, multi-type digital collection. The individuals and institutions involved learned valuable lessons throughout the three-year grant—lessons that can assist others as they contemplate their own digital programs. While not all of the recommendations that follow are applicable to all institutions, the authors believe that they provide useful insight into the realities of executing a digital project.

The DCSs encountered multiple challenges conducting fieldwork. These challenges included time management, mobility, metadata creation, and transfer of files. According to the DCSs, metadata creation was the most challenging task. The records created for VV were extensive. Each record included several subject fields created for specific browse functions. For example, each record contained a county designation and as many as three fields that correspond to Tennessee’s state history curriculum. In most cases, little item level information was available, requiring extensive historical research to establish names and basic descriptions. There was a struggle between the desire to create rich records and the impending deadline by which the project had to have ten thousand images. Time management was difficult because each item required different levels of research and each location presented different challenges regarding workspace and Internet connectivity. To address the deadline and metadata issues, graduate students in information sciences worked at the three regional sites as volunteers, practicum students, or employees. These students worked with the DCSs and the Project Director to complete the metadata records prior to processing.

A second problem inherent to the division of labor was maintaining consistency in subject analysis and the application of Library of Congress (LC) subject headings. While each DCS held graduate degrees in both history and information science, making them ideal candidates for this project, identifying appropriate subject headings was complex. The Project Director and Co-Principal Investigators called on several individuals with cataloging expertise to assist the DCSs and the graduate students. Additionally, the Tennessee State Library & Archives staff provided the project with Tennessee specific
LC subject headings. These subject headings appeared as pop-up lists in the MODS workbook to help mediate the difficulty inherent in subject heading assignment. In hindsight, the DCSs and the project would have benefitted from advanced, formal training in cataloging and indexing prior to beginning fieldwork.

Communication between the DCSs in the field and the project staff was also challenging. The project plan provided office space, telephones, computers, and other necessary equipment to accomplish the tasks, but fieldwork presented logistical challenges. The DCSs found themselves in isolated, rural areas without a telephone or the Internet. Many times they were forced to use their own cell phones to communicate, or they would wait until they returned to their home office to pose field related questions. Because of the demands of travel and being required to stay on-site for weeks at a time, receiving answers to questions in a timely manner was nearly impossible. The DCSs resolved some of these challenges by engaging in chat sessions when possible, or collecting multiple questions and scheduling additional visits to sites to complete the work once issues were resolved. The advisory board discussed several solutions to this problem, including wireless network cards or project cell phones issued to each DCS, but no resolution was ever finalized. The authors recommend that groups considering extensive fieldwork and travel anticipate and plan alternative methods of communication, including accepting reversed phone charges or paying institutions directly for long-distance charges.

The project faced another issue in its choice of presentation software. Because of some unforeseen problems with the original software chosen, the lead institution was forced to find an alternative. The solution was the Digital Library eXtension Service (DLXS) platform, a trusted, robust program used for many digital projects at the University of Tennessee Libraries. DLXS and the back-end work required with that version created long delays between initial scanning and online presentation. These delays created strain between the project staff and anxious repository staff who wanted to share their digitized materials soon after contribution. In the end, the project advisory board determined that DLXS was too difficult to use for distributed input because contributors had extremely varied levels of technical expertise.

One of the most important lessons from this grant project was that on-site scanning proved to be a viable means of digitization. The ability for resource-poor archives to be included in the project was the primary advantage of this approach. A secondary benefit was that VV was able to avoid the liability and cost of moving archival materials. The trade-off for mobility of the DCS was a limit to the size and type of materials that could be scanned. Items had to fit within the dimensions of the flatbed scanners thus limiting the ability to scan some maps and posters, fragile objects, and many bound items. In the context of the goals of VV, these limitations were acceptable, and the inclusion of a wider range of institutions compensated for the narrower range of item type.

While the teacher workshops were successful, and everyone seemed to enjoy them, we ran into problems coordinating sites and gathering participants. In retrospect, we learned that we needed a better understanding of the work schedules, and, hence, availability of K-12 school teachers. Despite the fact that we paid teachers a personal stipend and compensated the school for a substitute teacher for the day, we ran into difficulty finding teachers who were able to attend for a variety of reasons. More often than not, the school simply could not let them go for a full day. Additionally, the cost of travel was problematic, and communication within the school systems was challenging. One recommendation from the authors would offer based on this experience is to plan the workshops ahead of schedule or perhaps offer them in conjunction with a professional conference for teachers.
Our initial view of empowerment changed during the course of the grant. We began this project with the goal of having the DCSs provide individualized training to the staff at the institutions they visited. Due to time constraints, the DCSs ended up scanning materials for institutions and were involved in relatively little training. Our empowerment plans shifted the focus from one-on-one training efforts provided by the DCSs in the field to a more formalized training of staff at libraries, museums, and archives. Through this experience, it is believed that more of these formal training sessions would have benefitted both the project and the institutions that were included.

Consistent communication is vital to the outcomes of a multi-partner project. Although the project established an advisory board to provide feedback loops to administrators and project staff, there was a consistent breakdown in lines of communication. Upon reflection, a communication model established on the front end and adhered to throughout the project would have improved the flow of information and maintained a uniform level of involvement.

Finally, in an ideal scenario, official confirmation from participating repositories and collections would occur prior to the beginning of the grant. While the DCSs followed up with identified repository staff, many were not prepared to participate until well into the second year of the grant. The project staff determined that an alternative model that required the DCSs to finalize the list during the first six months would have given institutions time to prepare and to select items on their own. Additionally, this undertaking would have decreased the load on the DCSs and increased the relevance of chosen items to the key themes of the collection.

Sustainability Issues

While some statewide digitization programs have been vibrant and sustainable, others have tended to stall after grant funding has ended. Recognizing this and not wanting to lose momentum, project staff began addressing sustainability issues during the second year of the grant period. We desired an approach that did not rely on large grants exclusively, retained the collaborative and “big tent” culture of the grant period, was scalable for all institution types and sizes, maintained a reasonable growth schedule, continued to employ state-of-the-art standards, and kept Volunteer Voices in the public eye.

Software

Recognizing that DLXS would not contribute to a sustainable digitization program, grant personnel turned to CONTENTdm’s Multi-site Server as an attractive alternative. There were several reasons this option was attractive:

- eight institutions in the state use CONTENTdm to build local digital collections,
- several other multi-institution projects use Multi-Site Server with relatively little technical demand,
- CONTENTdm’s transferable station software increases the ability of institutions to create their own digital objects and metadata records, and
- OCLC’s offer to make a hosted version of CONTENTdm a part of its FirstSearch base package should increase this number significantly.

Furthermore, the searchable collection would grow as objects are added to local installations of CONTENTdm, rather than become stagnant while waiting for the next big grant to fund building a large collection.

The reasons that DLXS was not a sustainable option for Tennessee include considerable delay time in processing records, the requirement of a full-time programmer, a difficult end-user interface, and record display problems. As mentioned above, because of the distributed nature of the project and an underestimation of programmer time needed, sometimes months passed before records were available online. Sustaining a
distributed digital program requires that records be posted in a relatively short time-frame in order to both maintain interest in and progress the collection. The version of DLXS that this project team used did not have an interface that allowed contributors to create records, thus the MODS Workbook was created. In addition, direct ingestion into DLXS was impossible for the project and contributed to an increase in demand for programmer time. From an end-user perspective, especially for first-time and K-12 users, DLXS is difficult to navigate. Two critical components for these target groups are the site’s low learning curve regarding internal navigation and the simple and elegant record display.

Management

Developing a sustainable management structure for Volunteer Voices has been equally challenging. Efforts to sustain the program still reside with members of the Volunteer Voices Committee of Tenn-Share, the original sponsor of the digitization program. The project management team created and proposed three models for sustainability. However, due to economic and infrastructure concerns, none of the models proposed have worked for institutions and discussions of sustainability continue.

One problem with multi-institution projects like Volunteer Voices is finding a permanent home once the grant project is completed. To circumvent this issue, administrators of partner and contributing institutions should be part of the sustainability conversation early in the project. Increased front-end involvement of decision-makers would allow necessary after-project partnerships and agreements to be planned and carried out.

Collaboration Continues

In spite of the challenges noted above, several collaborative projects have developed as offshoots of the IMLS grant. These projects, in addition to the growth in CONTENTdm sites in the state, offer promising approaches to collaborative digitization in Tennessee.

The Tennessee State Library and Archives is developing the Tennessee Regional Libraries Digitization Project. Each of the state’s twelve regional libraries will offer a scanner and digitization training to area cultural heritage institutions. These regional libraries, strategically positioned to serve rural areas across the state, have the potential to reach institutions that would otherwise not consider digitizing their collections.

Two other projects apply the collaborative model at the local level. Middle Tennessee State University’s Walker Library has received a small grant to support graduate assistants’ work with personnel at small institutions to digitize their materials. These items will be cataloged using Walker Library’s CONTENTdm software and added to the “Middle Tennessee Communities” collection. Staff at Cleveland State Community College continue to work with area cultural heritage institutions to build the Southeast Tennessee Digital Archive (SET-DA). If adopted by other libraries in the state, this local collaborative approach would be another method of keeping Volunteer Voices vibrant with minimal funding at the state level.

Conclusion

Few collaborative digitization projects evolve as originally planned in a grant proposal. The IMLS Volunteer Voices grant project was no exception. The program experienced personnel changes, technical hurdles, and communication issues. However, project staff at archives, libraries, and museums worked together and both overcame these obstacles and created a multifaceted collection of more than ten thousand primary source materials. This collection, unmatched in its coverage of Tennessee history in the online environment, offers an excellent opportunity for ongoing collaboration between libraries, museums, archives, and schools. By building on these relationships, Volunteer Voices can evolve from a grant-
based digital project to a dynamic, sustainable digitization program.

End Notes


2 “Volunteer Voices.” For a comprehensive account of the program, see: http://www.volunteervoices.org.


8 Examples of statewide digitization programs that are using the CONTENTdm multi-site server include: AlabamaMosaic (http://www.alabamamosaic.org/), Iowa Heritage Digital Collections (http://iowaheritage.org/), and New York Heritage (http://www.newyorkheritage.org/).
