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
Collaborative Digitization Goes Local

Ken Middleton, Mayo Taylor
(Middle Tennessee State University)

Abstract
This chapter describes our efforts to build a digitization program in spite of limited funding. By obtaining small grants and forming partnerships with other campus units and institutions in the Middle Tennessee region, we have developed modest yet engaging collections and planted the seeds for additional partnerships. The article will also highlight the use of graduate students, efforts to streamline the digitization process, and lessons learned.

Keywords: Collaboration, CONTENTdm, Collaborative digitization; Oral histories; Work flow, Community projects.

Introduction
In an era when many libraries are struggling to stay relevant to today's Facebook generation, digital projects have injected new purpose into Middle Tennessee State University's Walker Library. Staff members are excited about digitizing oral histories of pioneering women on campus, documenting the agricultural history of a neighboring county, or preserving seminal research on local rare plants. The digital collections under development by the Library have required collaboration with other campus entities and external institutions. Few of the digitized objects have come from the Library's own collections. In this chapter, we will describe the benefits of these partnerships and how they can be nurtured. We will also explore
many of the practical issues that new digitization programs face, such as securing staff, streamlining the digitization process, handling diverse file formats, and customizing collections.

Efforts to collaborate with area institutions began as a natural extension of Walker Library’s involvement in Volunteer Voices (http://volunteervoices.org), Tennessee’s statewide collaborative digital project. MTSU librarian Ken Middleton served as the co-principal investigator of a three-year “Building Digital Resources” grant from the Institute of Museum and Library Services (IMLS) that coordinated the gathering of historical materials from organizations and institutions statewide. That project, based at University of Tennessee-Knoxville, focused on Tennessee history. Staffs were placed in each area of the state (west, middle and east) with a charge to collect unique photos, documents and other materials from cultural institutions. Selected items featured both major trends and figures in Tennessee history (e.g. early settlement, Ida B. Wells), and significant local items (e.g. Bradley County Courthouse). An important component of the project was to relate the digitized objects to the curriculum followed by schools in Tennessee. Walker Library was a satellite digitization center for the middle Tennessee region (Conner, Middleton, Carter, & Feltner-Reichert, 2009), with several librarians and staff participating.

As the IMLS grant project ended, Walker Library began exploring ways that it could build its own digital skills while continuing to assist area archives, libraries, and museums to digitize unique collections. We would quickly lose skills and momentum if we did not build on our Volunteer Voices experience by securing the institutional support to initiate our own projects. Noting that a key mission of the University is to develop “mutually beneficial partnerships” (Middle Tennessee State University, 2002) we examined appropriate models among more experienced institutions.

In 2002, The University of Pittsburgh’s Digital Research Library began a two-year project to digitize a collection of 7,000 historic Pittsburgh images in partnership with the Historical Society of Western Pennsylvania, the Carnegie Museum of Art, and the
University's Archives Service Center (Galloway, 2005). As evidence of the continued success of this approach, they have added three more partner institutions over the last five years and now have more than 14,000 images in the collection (University of Pittsburgh Digital Research Library, 2010). The Digital Research Library has also collaborated with University of Pittsburgh faculty to create such projects as *American Left Ephemera and Visuals for Foreign Language Instruction* (Galloway, 2009). This program offers home pages for each of its 70 digital collections; each collection home page offers search and browse options, and a link to the partner institution’s home page.

IUPUI Library’s Program of Digital Scholarship has collaborated with an impressive array of campus and local institutions. The collection websites and promotional material give every indication that the Program of Digital Scholarship treats these institutions as equal partners: the “Partners” page provides background information about each institution; logos of relevant partners appear on each collection home page; and a link for contact information appears in each record. Its list of partners continues to grow. Building on its initial collaboration with Conner Prairie Museum that features textiles from the museum’s collection, the Library recently announced two Conner Prairie collections based on 3-D imaging technology.

After touting the benefits of such collaboration (e.g., cost-effectiveness), Bishoff (2004) offers further guidance for librarians who want to forge lasting partnerships with other cultural heritage institutions: involve all partners from the start; respect each institution’s mission and culture; and ensure that each partner benefits from the end product.

Building on these models we outlined a vision for our own digital project design and received the support of our Dean to purchase a server, large flatbed scanner and a license to CONTENTdm software. We then leveraged that support to secure a $5000 grant from the University Special Projects Committee to fund graduate students for our first project, MTSU Memory. More on that collection will follow.
Institutional Context

Walker Library serves 25,000 students on a rapidly-growing campus. Its funding and staff levels are significantly lower than most comparable institutions within Tennessee and nationally – approximately 60% of the average of our designated peer institutions on a per-student basis in 2005-2006 (W. Black, unpublished data). Digital work has been accomplished by reassigning small portions of time for several librarians and one staff member, making extensive use of graduate students and purchasing equipment and software on an as-needed basis, usually at the end of the fiscal year when other needs have been satisfied. Digital projects management has been added to the load of our small staff. We have no programming support for customizing CONTENTdm or working with open source programs. This has hindered our ability to implement advanced features such as embedded audio and video players.

Staffing

The situation for digital projects did improve recently when the Library administration designated our working group as “the Digital Initiatives Team,” thus formalizing our role in the Library, and clearly recognizing the roles of the team members: Ken Middleton, Reference/Microtext Librarian (Team leader); Mayo Taylor, Reference/Access Services Librarian; Fagdeba Bakoyema, Reference/Web Librarian; Lynda Duke, Microtext manager; and Jo Williams, Cataloging Librarian.

Because these team members can only devote limited, irregular chunks of time to digital work and because they are drawn from different departments of the library, communication is the key to moving forward. We work hard to avoid problems by holding weekly meetings that also include our student workers, maintaining a wiki for the management of meeting minutes and other information, and by sending monthly reports of major activities to the Library Dean and Coordinators of both Library departments.

We are able to balance the restrictions on our professional staff time by liberal use of graduate students from the Public History
program on campus. They bring to the work an awareness of the historical and cultural significance of the collections, a desire to gain skills that are highly relevant to their discipline and a facility with the technology involved. Because many of these students will enter careers in history museums or archives they also make ideal liaisons to local historical agencies. Our graduate students have been supported primarily by small grants received in support of specific collections, as discussed below. We have supplemented this labor force with an undergraduate scholarship student from the Electronic Media Department, at no cost to the Library, and by utilizing students employed at the Microtext service desk for projects that can be done when they are not busy with patrons.

Software /Hardware

Staffing issues also heavily influenced our decision to purchase the CONTENTdm digital content management software. Although its shortcomings will be duly noted, CONTENTdm offers institutions the ability to quickly create digital collections with minimal systems support. In addition at least six institutions in Tennessee are using the system, creating opportunities for collaboration.

The Library has purchased a good mix of digitization equipment. An Epson Expression 10000 flatbed scanner is large enough to handle most photographs and documents that we have selected for scanning. The Library recently purchased a Bookeye 3 planetary scanner with OPUS software. Intended primarily for the Interlibrary Loan department, it is available for digital projects outside standard business hours for fragile books and large format materials. We also have two portable scanners and a laptop computer to assist area institutions in digitizing their materials.

Collections

Campus Partnership: MTSU Memory

In anticipation of MTSU’s centennial year in 2011, MTSU Memory (http://library.mtsu.edu/digitalprojects/mtsumemory.php)
will include photographs, documents, books, and oral histories that document our institution’s growth from a small Normal School to a doctoral university with Tennessee’s largest undergraduate enrollment. The Library’s Special Collections hold many items of interest, including the institutional copies of yearbooks, faculty directories and the student newspaper. However, a larger repository of essential items is at the Albert Gore Research Center, which serves as a de facto campus archive. An early emphasis was to consult with the Director of the Gore Center to establish a working relationship and procedures for pulling and scanning materials.

**Collection Focus**

The digital projects team wanted to maximize usage of the collection by focusing on topical areas: the Great Depression, World War II and aftermath (e.g., GI Bill), the Vietnam War and student activism, and African Americans and women on campus. We expected the majority of items in the collection to be photographs, but wanted to include other types of primary sources (artifacts, newspaper articles, and oral histories) to present a multifaceted portrait of the University’s history. The selection process for this collection has in fact been rather unusual because we initiated it without knowing what we would find, adjusting our expectations as new materials have been revealed.

**Early Implementation**

Walker Library launched its installation of CONTENTdm in late summer of 2007. Installation of the dedicated server and implementation of the software was handled by our Systems Librarian. Initial configuration of the CONTENTdm software was done by Ken Middleton when our first collection, MTSU Memory, was established. The remainder of the Digital Initiatives team members had been recruited and we had received guidance from our Dean on the direction that he would like us to take. Everyone on the team had some experience with digitization projects but none had worked with CONTENTdm. In an ideal world we would have enjoyed a beta period of testing the system and training the graduate students. In the real world, the clock was already ticking on a grant project and the
graduate students needed work to do. They worked on the project more hours in most weeks than did their supervisors, leaving the professional team members hard pressed to hone their own skills to keep up with the graduate students. The students began selecting materials and scanning materials before any of us were proficient in using CONTENTdm and before we completed the configuration of the MTSU Memory field structure. One example of the consequences: as each record is created it is put in an approval queue to be reviewed and approved by one of the librarians. If the librarian isn’t able to keep pace with the student, simple errors are repeated many times before they are corrected, which exacts a toll in terms of time wasted on corrections.

To slow down the rate at which items were loaded into CONTENTdm we gave the students guidance on the types of items we were looking for and sent them over to the Albert Gore Research Center. They concentrated on selecting items, but without the benefit of prior experience in completing metatadata they didn’t know exactly what information they needed to capture to assure complete documentation of the items. The result was that many objects were loaded with incomplete data and improper form. Predictably, we are still returning to some of the earlier records to fill gaps and fix inconsistencies.

By summer of 2009 we again posted a job notice with several graduate programs on campus and again selected two Masters candidates from the Public History program. This time we started each of them with an introduction to the collections they would work with and the metadata guidelines available on our wiki followed by CONTENTDM tutorials and documentation. Finally, they began by importing and creating metadata for objects selected by our previous workers. Only after these initial projects were complete were the students free to find new materials for the collections.

**Expanding Opportunities**

Our original vision for MTSU Memory was a collection composed primarily of photographs and documents. However, we adjusted that
concept when given the opportunity to include excerpts of audio and video interviews and oral histories.

**Audio materials:** As our graduate students investigated materials at the Gore Research Center they discovered an excellent set of oral histories that had been conducted with alumni from past decades. From that extensive collection, twenty-five were selected because of their relationship to the lead themes of our collection, such as World War II, Civil Rights, etc. Our initial plan was to digitize and display the typed transcript of each interview and then offer selected audio excerpts as a supplement. When a new Director joined the Gore Center he decided that the transcripts needed additional work to be ready for online publication. We adapted by finding a work around to offer only the selected audio excerpts. It consisted of loading MP3 files to our server, but not within CONTENTdm, importing a photo of the subject as the digital object for the CONTENTdm record, and then creating a metadata field to include links to the MP3 files. The work of selecting, reviewing and excerpting the oral histories was done by one of our graduate student workers as an internship for course credit. She selected and downloaded the freeware program, Audacity, to manage the sound files and participated in designing the customized process for creating the records. We hope to include the entire transcript of each record at some point. At present our CONTENTdm records direct users to the items available at our partner across campus.

**Video materials:** Another treasure trove at the Gore Center was a collection of interviews that were broadcast on the campus television station in the 1980s and 1990s. Of particular interest was a series of four lengthy interviews with Miss Mary Hall, an Education professor important to MTSU and the related University demonstration school from 1929 to 1963. We again turned to one of our graduate students who had recently completed a course in documentary filmmaking. She was able to use her personal computer (a MacBook) and software (iMovie) to create a series of video excerpts that encapsulate important phases of Miss Hall's life.

**Book-length publications:** As we researched our own campus for the MTSU Memory project, we frequently consulted *The First Fifty*
Years, a campus history written by alumnus Homer Pittard, that covered the period 1911-1961. Because MTSU held the copyright for the book we decided to digitize the entire volume. This initiated an intensive learning experience about the pros and cons of compound objects and ultimately the process of creating complex PDF files. The process of scanning the pages and doing OCR through the Abbyy Fine software was quite straightforward. A definite learning curve was encountered in figuring out how to structure a compound object from the hundreds of page images. However, once the book was loaded into the MTSU Memory collection, we were dissatisfied with the results, largely because queries could return dozens of individual pages intermingled with other materials. In addition, when we figured out the compound object expense in terms of licensed objects used, we rethought our original approach. In brief, a 200 page book exacts a cost of 201 objects utilized, 200 for the images and one for the combined record. By comparison, a book imported as a PDF made from the same original image files costs only one licensed object. While this may seem an arcane concern, with a licensing level of only 10,000 items we could envision our capacity diminishing quickly if we continued to do book-length items. So, The First Fifty Years was reloaded as a PDF as was our second book, The Raider Forties, and a thesis. Smaller multi-page items continue to be treated as compound objects.

It is worth mentioning that the experience of the user is quite different for audio, video and large PDF text files in our collections when compared with objects loaded as single JPEG files, compound objects or small PDFs. Whereas those files provide a display image immediately upon opening the record, the other file formats all depend upon software programs external to CONTENTdm (e.g. Adobe Acrobat or Quicktime) to open the file. In the case of large media files or PDFs the process can be relatively slow. The record display is also marked by the rather unimaginative default message of “Access this Item,” which is not far removed from the archaic “click here” message of early web pages. We hope to find a partial solution to this problem through embedded media players.
Customizing Our Flagship Collection in CONTENTdm

Although CONTENTdm is an excellent tool for creating digital collections quickly, it does not provide a particularly user-friendly public interface. All searches retrieve results from all collections, which can be problematic when you have collections as varied as university history and ecology. We have created a separate entry page for each collection with a brief description of the collection, a search box, a link to an advanced search, and browsing options.

The lack of default browse options for subjects, time periods and formats has been particularly frustrating. CONTENTdm requires us to create the browse lists manually by running “custom queries” for every subject, for instance, that appears in a list. This process requires a significant amount of time. We have recently discovered Elias Tzoc's web-based tools for creating CONTENTdm browse lists (Tzoc, n.d.). Creating the list of 200 subject headings in MTSU Memory would have taken approximately 10 hours if we had used custom queries. After learning a few fairly simple steps, we created the list in just a few minutes using Tzoc's tool.

One extremely useful feature of a recent CONTENTdm upgrade is that search results now include facets that help users narrow their searches. For example, a simple search retrieves a list of records that can be narrowed by time period, subject, and genre. However, making changes to the default facets required too much time. Rather than simply check the desired facets in an administrative module, one has to find the relevant line of code in a PHP file, use correct PHP syntax, and determine the correct code for each facet. Many libraries may lack the technical staff to make such changes.

We are making efforts to attract users through additional access methods. Our metadata librarian has set up a profile in OCLC’s digital collection gateway to streamline the creation of WorldCat records for our digital items. We are working at the state level to use the CONTENTdm Multi-site Server, which would create a central site for searching across CONTENTdm collections in the state.
Community Partnership

During 2009 the Digital Initiatives team received a $1000 grant from the MTSU Public Service Committee to assist several community associations with digital projects. Selected organizations were invited to propose projects; the first application received was from the Arts Center of Cannon County (ACCC). They had been given a collection of nearly 1000 photographs that were taken between 1944 and 1977 by William L. Clements, District Conservationist for the Soil Conservation Service (SCS). He served Cannon County, a small agricultural community just east of Murfreesboro and MTSU. The photos documented the farming methods that were disseminated by the SCS in that period. Significantly, each photo had descriptive information typed onto the image back, which made the task of describing each image quite easy.

The Clement Collection was much larger than our initial idea of what would be handled for each project with our small grant, with nearly 1000 photos available, but it was so compelling that we worked to find a way to accept it. The answer came through a supplemental $300 grant from our local Gannett newspaper and through use of a student worker who needed to fulfill work hours for his scholarship. The Cannon County, Tennessee Collection (http://library.mtsu.edu/digitalprojects/cannon.php) includes approximately 400 photographs selected from the total available.

We accepted the Clement photos because they beautifully illustrate and describe the details of the County’s predominant occupation, farming, over a period of three decades. In addition, the collection was an ideal test case for working with area institutions to create digital collections:

- The Arts Center is an essential institution in a community where farming is still economically and culturally significant. Many of the families pictured are still on their farms today.
- The Arts Center staff had already worked on a promotional plan for the collection and have developed plans for creating an exhibit based on the photos.
• Additional rich contextual material is available. Three hours of recent oral history interviews with the photographer will be included along with an interpretive essay from Evan Hatch, folklorist with the Arts Center.

• There were no copyright issues because the photos are the product of a federal employee.

• The photos have been relatively easy to scan and process because the photographs are all the same size. Moreover, detailed metadata is already recorded on the back of each photograph.

The homogeneity of the collection also provided an opportunity to develop streamlined processes, for example by loading nine images at a time onto the large format scanner, then utilizing Photoshop tools to deskew and adjust each image before saving it individually. We also were able to make optimum use of the template option within the CONTENTdm Project Client to automate the entry of identical information into selected metadata fields as each object is imported. In this manner large groups of records could be created in the Project Client and then uploaded in a batch operation for approval. The approval process itself was efficient because only the relatively few unique fields in each record (e.g. title, LC subject heading) needed to be carefully checked by the approver.

We have also used Elias Tzoc’s web-based tool to create a tag cloud based on subject headings in the collection. Tag clouds offer a useful alternative to traditional subject heading lists because they provide users with a quick sense of the major subjects in a collection.

The Digital Initiatives team has made every effort to involve the Arts Center of Cannon County in developing the collection. Initial meetings with Evan Hatch, the Center’s folklorist, covered everything from selection issues to scanning standards. The Center’s logo appears on the collection web site, and links to the Center’s web site. Treating the Center as an equal partner improved the quality of this initial digital collection, and it has also increased the likelihood of future partnerships. Early success of the project has encouraged discussions of additional materials that may be forthcoming to document the history and deep crafts tradition of Cannon County.
New and Forthcoming Projects

**Buchanan Collection** - will document the history of a local family that has included a Nobel Laureate MTSU alumnus, a governor, and an early settler of Nashville. The collection is being compiled in cooperation with family members.

**Cedar Glades Studies Collection** - a partnership with the MTSU Center for Cedar Glade Studies will digitize research on the rare cedar glade ecosystems of Middle Tennessee. Most notably the 1948 doctoral thesis of Dr. Elsie Quarterman was digitized with support from a small grant passed through the Biology Department. Her thesis was handled by two practicum students from the UT-Knoxville School of Information Science.

**Shades of Blue and Gray: Reflections of Life in Civil War Tennessee** - will include digital images of 250 artifacts from small archives, libraries, and museums across the state. In partnership with Celia Walker, director of Vanderbilt University’s Peabody Library, we will create a metadata template in CONTENTdm, enter metadata, and customize the CONTENTdm database for items collected by other entities.

**Stones River Battlefield Historic Landscapes** - is digitizing deeds and other materials to document the largely forgotten African American community of Cemetery, which developed after the Civil War on land that would become the Stones River National Battlefield. Created by MTSU history professor Rebecca Conard and her graduate students, it is hosted by the Library and Digital Initiatives staff have played an advisory role.

**Lessons Learned**

The Digital Initiatives team’s collaborative approach has made it possible to take on more projects. In turn, our list of lessons learned is comparatively long and keeps growing as we face challenges specific to new projects. However, as the list below illustrates, many of the lessons apply either to every digital project.
• Libraries that lack a digital projects department (and the corresponding budget line) should consider applying for small grants. In our case, the three grants have been enough to fund students doing the scanning and metadata entry.

• Recruit student workers / interns with relevant skills, and offer them a rewarding work experience. Recognizing that our public history interns are indispensable to the success of our digital projects, we have treated them as equal partners on the Digital Initiatives team. We have encouraged them to attend our weekly meetings, welcomed their suggestions for the improvement of digital projects, and tailored their internships to their interests and skills as much as possible.

• Be flexible. One of our best decisions has been to include interviews in audio and video formats in MTSU Memory, in spite of production delays. In addition, we adjusted our very modest approach to working with area institutions when we saw the potential of working with the Arts Center of Cannon County.

• An effective communication structure is essential, particularly for new digitization programs. By having weekly meetings, we have addressed metadata and technical problems before they become so large that they are difficult to correct.

• Communication is important when working with collaborative partners; tact, patience, and ability to explain both the process and benefits of digitization can not only help put partners at ease, but also pave the way for additional joint projects.

• Invest time in strategic planning. By focusing so much on the details of specific projects, we have not devoted sufficient time to plan for a robust approach for digital preservation.

**Conclusion**

Perhaps the most important lesson learned mirrors our experience with the statewide Volunteer Voices project. Small libraries, archives and museums have valuable collections, and often want assistance in digitizing these collections. By sharing our successes and growth pains
in working with an academic archive, a small arts center/museum, and a growing list of other partners, we hope to encourage other institutions to proactively seek out collaborative digitization partnerships at the local level.

References


