2002

Getting Digitization Projects Done in a Medium-Sized Academic Library: a Collaborative Effort Between Technical Services, Systems, Special Collections, and Collection Development

Michael Boock
Cleveland State University, michael.boock@oregonstate.edu

Bruce Jeppesen
Cleveland State University, b.jeppesen@csuohio.edu

William C. Barrow
Cleveland State University, w.barrow@csuohio.edu

Publisher's Statement

This is an Author's Accepted Manuscript of an article published in Technical Services Quarterly 2002, available online: http://www.tandfonline.com/10.1300/J124v20n03_02.

Getting Digitization Projects Done in a Medium-Sized Academic Library: a Collaborative Effort Between Technical Services, Systems, Special Collections, and Collection Development

By Michael Boock, Bruce Jeppesen and William Barrow

Abstract

Cleveland State University Library, a medium-sized academic library serving approximately 15,000 students, is engaged in large-scale efforts to digitize and make accessible online collections of unique Cleveland-related materials. The Cleveland State University Library Special Collections digitization and cataloging efforts use staff from several different library organizational units. The collaboration of staff with specific expertise in long-standing library functions -- special collections, cataloging, systems, archives, selection -- to create two Web databases is described. The collaborative effort has proven effective in getting resources processed, archived, digitized, described, promoted, and made accessible in a highly efficient and effective manner. The responsibilities of the different library units involved in the digitization project are described.

Collaboration, digitization, image processing, CONTENT

Michael Boock is Head of Technical Services at Cleveland State University Library, Cleveland, Ohio.
Bruce Jeppesen is Head of Systems at Cleveland State University Library, Cleveland, Ohio.
William Barrow is Special Collections Librarian at Cleveland State University Library, Cleveland, Ohio.

Michael Boock, MLS
Head of Technical Services
Cleveland State University Library
1860 E. 22nd St.
Cleveland, Ohio 44114
m.boock@csuohio.edu

Bruce Jeppesen, MLS
Head of Systems
Cleveland State University Library
1860 E. 22nd St.
Cleveland, Ohio 44114
b.jeppesen@csuohio.edu

William Barrow, MLS/MA
Special Collections Librarian
Cleveland State University Library
1860 E. 22nd St.
Cleveland, Ohio 44114
w.barrow@csuohio.edu

INTRODUCTION AND BACKGROUND

The editorial library of *The Cleveland Press*, a Cleveland, Ohio, daily afternoon newspaper published from 1878 to 1982, dominated the library's Special Collections before the hire of a full-time Special Collections Librarian in 1999. The Cleveland Press collection consists of approximately 500,000 black and white photographs and one million clippings. No formal collection development policies existed for the library's Special Collections, and little effort had been made to acquire new collections prior to 1999. Gifts were occasionally accepted, such as the construction records of The Cleveland Union Terminal, a private collection of rare bridge books from a notable local civil engineer, and the archives of two prominent composers from the Cleveland area.

A University Archivist was responsible for processing and supervising the Cleveland Press collection, and the Head of Collection and Database Management oversaw the various components that would become Special Collections, but until recently no permanent staff in Technical Services had any special-collections duties. Cleveland State University History Department graduate students were hired to work in Special Collections with the Cleveland Press collection. One of these students was later hired to process the Cleveland Union Terminal collection. This important collection includes the archives of the company that built Cleveland's Terminal Tower, a union passenger station, a complex of office buildings, post office, department store and an infrastructure of tracks, bridges, signals, electrical catenary structures, and yard facility buildings necessary to switch passenger coaches over from steam to electric power and bring them into the downtown Cleveland area.

One student remained with the University Library as Data/GIS Specialist while completing an MA in History and an MLS and became the library's first Special Collections Librarian in 1999. Thanks to his enthusiasm, the guidance of the Head of Collection Management, and the strong support of a new Library Director, who was hired in 1998, Special Collections began to grow rapidly.

The library acquired several large collections of materials in the following years, predominately relating to railroads, civil engineering, and music in Cleveland. Numerous books and collections of photographs were digitized and made accessible as browseable, static html pages.

The library's Cataloging Unit cataloged the new physical collections using collection-level catalog records. Links were provided from the catalog record to the online collections as the collections were digitized and made accessible from Web pages. The library's Systems Unit solved PC and network problems. But the Special Collections Librarian and his cadre of enthusiastic interns and student employees accomplished nearly all the work of Special Collections themselves, by acquiring, processing and digitizing collections, and creating Web pages.

The first Special Collections Web pages were mounted in 1996, when the Cleveland Union Terminal Collection site came online. Other collection home pages quickly followed. These sites initially consisted merely of descriptions, hours, location, and other basic information but quickly grew to include digitized images and texts pertaining to the collection and related topics.

Each site had its own list of relevant links to outside resources until that proved difficult to manage, at which time they were consolidated into one master list. Here the

links were divided into those that led directly to historical content and those that led to contemporary information such as institutional home pages, and the historical content links were arranged according to each of four facets: Subject, Date, Location and Format. This content site was named the Cleveland Digital Library and is available at http://web.ulib.csuohio.edu/SpecColl/cdl/.

As a result of these digital efforts, faculty, student and general interest in the collections increased to the point that two-thirds of all Web traffic coming into the library from off-campus locations visited the local history Web pages of Special Collections. That resulted in redoubled institutional support, and a new site was developed to access all the digitized Special Collections resources, called Cleveland Memory (www.clevelandmemory.org).

THE CHALLENGE

By 2000, these digitization efforts of Special Collections had received some renown, but clearly they could benefit from the expertise of Systems and Cataloging to better organize the materials and make them more easily accessible. The volume and sophistication of the work of Special Collections had increased without increasing staff. To increase the number of online digitized images and to preserve the physical materials from handling, large-scale scanning of the collections was needed. An online database that allowed items within digitized collections to be searchable and more easily retrievable was also deemed necessary.

Item-level cataloging of the collections under the direction of the cataloging unit was needed to enhance access to the items within large collections. Cataloging expertise was also needed to develop and implement standardized cataloging rules and schema. Also,

problems of file organization arose that could have been prevented by earlier involvement of Systems staff - e.g., several versions of digitized collections existed on multiple networked and hard-drive directories with no indication as to which was the master directory. Greater Web design expertise was also needed.

Library staff who had expertise in the particular new functions could accomplish the work. Catalogers could establish cataloging rules and schema, identify standards and controlled vocabularies, and perform original cataloging. Systems staff could be responsible for Web design and the maintenance of servers, hardware, and digital production software. Subject bibliographers could select digital images for specific digitization projects. Subject bibliographers could also catalog images utilizing their subject expertise. Also, subject bibliographers could help promote the resources to their department's faculty and students, and assist faculty in including the local history resources in their curricula. The challenge was how to get these disparate library employees to work together toward the same goal of providing enhanced online access to the increasingly digitized resources of Special Collections.

LITERATURE REVIEW

Ressel and Smith, in "A New Approach to Thesis Subject Analysis: A Collaborative Success," describe a successful collaborative approach between cataloging and collection development.[1] They recommend using the knowledge of subject bibliographers in applying subject analysis to specialized theses and dissertations for subject analysis to the resources in a timely fashion.

Bunker and Zick, in the article “Collaboration as a Key to Digital Library Development: High Performance Image Management at the University of Washington,”[2]

discuss the many collaborative challenges encountered in creating digital projects. Their Digital Libraries Initiative provides a context for effective collaboration among faculty, engineers, students, and librarians. The initiative serves as a focal point to produce a variety of electronic information services, resources, and systems.

In their March, 2001, ACRL Poster Session, "Digital Asset Management at the University of Washington Libraries: Teamwork and Technology,"[3] Zick, Ingram, Wilson, and Fluvog address the teamwork required to overcome the technological challenges of creating an image database system.

Withers, Presnell, and Schmidt, in their Academic Library Association of Ohio November 2001 conference presentation, "Creating Digital Projects with Cross-Departmental Teams,"[4] discuss the teamwork involved in first creating a database system for image collection organization, and then the teamwork required to plan and complete specific image projects.

Bond and Cornish, in their summary of an Online Northwest 2001 Conference program, "CONTENT: A Model of Collaborative Database Building," outline the "decentralized database building capability" of the CONTENT software.[5] The flexibility offered by the CONTENT software creates a system for efficient collaboration between various individuals and departments. The presentation also discusses digital imaging initiatives.

FIRST STEPS

In 1999, the library's Systems, Collection Management, and Cataloging departments successfully collaborated to create one of the first electronic journal databases. The electronic journal database is searchable online by title and arranged alphabetically and

by subject for browsing. A Microsoft Access database contains over 10,000 electronic journal entries available to Cleveland State students, faculty, and staff through local subscriptions, consortial subscriptions, and several aggregator databases. Technical Services maintains the database. Collection Management assists the cataloger in the assignment of subjects and fund codes to the electronic journals. Systems creates Web pages and the active scripting that updates the online database in real time.

The electronic journal database proved to be an excellent opportunity for library divisions to cooperatively create an online product. Together, Systems and Cataloging defined the purpose of the project, the functionality of the database, and the general appearance of the data. Systems found that this project gave it the opportunity to implement some of the database and html techniques that it had already been investigating and testing. In fact, the electronic journal database was the Systems Division's first online project using active server pages.

When the Special Collections Librarian approached Systems with the idea of transferring a database of Cleveland area postcards to the Web, the basic process of putting a database online was already in place. Special Collections had already scanned a large number of postcards, and placed some basic information into a Microsoft Access database. The techniques used with the electronic journal project suited this image project quite well. Unfortunately, the metadata within the database was minimal and required enhancement to provide more searchable content. Systems added links to the scanned images, designed a home page with links to related information, and added a pulldown menu to search the collection of postcard images. Executing a search yielded a list of related "hits" with image thumbnails and a brief title. Selecting one of these

entries led to a full-sized image with more complete textual information from the database.

This first Postcards of Cleveland Web site looked attractive and functioned efficiently. As the collection continued to grow, however, the limitations of this simple system became obvious: The system offered no keyword searching. Cataloging expertise was needed to add searchable metadata to the description of the images. Also, the system did not lend itself to multiple groups working on it simultaneously, increasingly necessary as more people became involved in the digitization efforts.

It was clear that an interface was required that could help create and manage a controlled vocabulary for the collection. This interface would also need the capacity to manage workflow from one processing group to another. The system also needed a good method of indexing the database and providing keyword searching. Systems, with limited staff, did not have the resources to add these features to the existing system.

These early collaborations among key players in the digitization process proved that good things could be accomplished if the right people were involved and they were asked to do what they knew best. It was clear that the involvement of staff with specific expertise from other library units would be required to select and implement a more sophisticated database. Recognizing this, a committee was created and charged with selecting and recommending for purchase a more sophisticated image database. Members of the committee included the supervisor of the new Digital Production Unit, the Head of Technical Services (cataloger), the Head of Collection and Database Management (subject bibliographer), the Head of Systems, and the Special Collections Librarian.

The committee recommended the purchase of CONTENTdm, a product developed at the University of Washington and now owned by DiMeMa, Inc. CONTENTdm uses the Dublin Core metadata standard and permits the use of controlled vocabularies, two very important criteria to the committee. The library didn't want to lose the information it had already develop during the creation of the postcard database, so the committee deemed that the ability to batch import media items from Microsoft Access was an important function. CONTENTdm satisfied this requirement.

The committee also wanted a system that could be housed on a local Microsoft NT server; could provide support for full resolution scanning, image storage and indexing; and allowed for the batch creation of thumbnails on import. The CONTENTdm system satisfied all of those criteria as well. The committee agreed that CONTENTdm would allow the library to digitize, catalog, and make images accessible online in a cost-effective and efficient way.

After the purchase of CONTENTdm, the committee began to consider processing issues, Web design issues, cataloging, prioritization, and file-storage issues. The committee expanded to include other key contributors to the digitization projects: the University Archivist, a Web-page designer, and an additional member of the library's Digital Production Unit. All members of the committee have specific digitization responsibilities and unique expertise that makes their participation necessary.

Some of the items discussed in committee meetings include collection-processing prioritization, assignment of responsibilities for each collection, and updates from appropriate staff on how work is proceeding. Issues relating to image description (what fields will be used for a particular collection, the controlled vocabularies planned, key

access points for a particular collection, how the collection will be searched) are discussed. The committee also establishes timelines and discusses systems issues such as scanning hardware/software and the status of problems that occur during the digitization process.

Informally, the committee serves to quell fears and define roles. Because digitization is so new to the library and many of the associated responsibilities are new as well, responsibilities related to the digitization process were at first undefined. The discussion and assignment of responsibilities in committee meetings helps clarify roles and gives individuals confidence in their responsibilities.

PROJECTS

Postcards of Cleveland

The *Postcards of Cleveland* database, accessible at http://www.ulib.csuohio.edu/postcards, includes thousands of digitized postcards from the postcard collection of Walter Leedy, Cleveland State University Professor of Medieval Art, Architecture, and Urbanism. Leedy's postcards are related to Cleveland history in some way; highlights include several panoramas and aerial views of the city. Others include major figures in Cleveland's popular history, such as Bob Feller and Satchel Paige, starting pitchers for the 1948 World Champion Cleveland Indians. Cleveland's industrial history and tragic events of Cleveland are well represented. The postcards permit viewers to observe Cleveland's changing cultural environment and architectural history. Searching the postcard database by Cleveland neighborhood, for example, one may visually recreate how a specific neighborhood's architecture changed throughout the twentieth century.

The *Postcards of Cleveland* database is now searchable by keyword and browseable by several predefined, broad subject terms. One may also perform advanced searches of the collection by searching within specific fields and using boolean operators. One may limit a search to The *Postcards of Cleveland* collection from the Cleveland Memory project home page, search across all Cleveland Memory project collections, or search the *Postcards of Cleveland* collection from the *Postcards of Cleveland* home page.

The Cleveland Press Shakespeare Photographs

The *Cleveland Press Shakespeare Photographs* database, accessible at http://www.ulib.csuohio.edu/shakespeare, includes approximately 400 images from publicity photographs. Shakespeare productions by regional theatre companies, New York stage productions, motion pictures, and televised productions are featured. Many notable actors and artists are highlighted in the database.

The database is browseable by play and genre and searchable by actor name, character name, or other names represented in credits fields-- producer, director, costume designer, etc. Photographs are arranged chronologically within each photo gallery. Each photograph is identified by production title; venue (i.e. stage, film, television, opera, or ballet); genre (i.e. comedy, tragedy, English history, Roman play, or romance); production date; production company, if known; place of production, if known; cast, photographer or studio; and credit information. Original captions from *The Cleveland Press* newspaper are presented with additional captions and dialogue provided as appropriate.

DIVISION OF FUNCTIONAL RESPONSIBILITIES – WHO DOES WHAT

Getting these two collections online required the coordinated efforts of individuals within several library units including staff from Collection Development, Technical Services, Systems and Special Collections.

Collection Development

At Cleveland State University Library, the Head of Collection Management oversees Technical Services, Special Collections, and subject bibliographers. Without the support of the individual in this position, the digitization efforts at Cleveland State University would not have proceeded efficiently. Collection development has several roles in the digitization efforts, including overall supervision of digitization projects, selection of images to be digitized, liaison work with faculty, cataloging of images, and promotion of the collections.

The Head of Collection Management encourages the library's subject bibliographers to identify collections for digitization that relate to their collection responsibilities. For example, the Humanities bibliographer, in discussions with the University's English Literature and Dramatic Arts department faculty, identified interest in having photos of Shakespeare plays in the Cleveland Press collection digitized, cataloged, and made accessible online for course use.

The library's subject bibliographers are responsible for identifying faculty who may be interested in specific projects. Subject bibliographers work closely with interested faculty in the design of a project's Web site. Subject bibliographers also ensure that the digitized collections will be useful to students and faculty.

Subject bibliographers are responsible for writing textual information that is to appear on the Web site for the project, including accompanying bibliographies,

acknowledgements, and introductory and explanatory pages. Because the subject bibliographer has a better understanding of the collection than anyone else does, the subject bibliographer is also asked to complete much of the cataloging. For the Shakespeare collection, the Humanities bibliographer completed the cataloging of the images based on rules and controlled vocabularies provided by the cataloger. The cataloger and digital production staff is then responsible for quality checking all cataloging work.

Technical Services

The role of Technical Services staff in digitization projects is large. Technical Services at Cleveland State University Library consists of eleven full time staff within four units: Cataloging, Monographic Acquisitions, Serials, and Digital Production (formerly Database Maintenance). Database Maintenance traditionally handled problems with the library catalog, processed library materials, inventoried the collection, and gathered statistics. The unit took on much of the management of the library's special collections as they burgeoned.

The unit now hires, supervises, and schedules Special Collections student assistants, staffs the Special Collections public service desk several hours a day, and processes and catalogs all the books in Special Collections. They catalog digital images and quality-check all digitization, cataloging, and preservation work. To perform these new responsibilities, Database Maintenance became Digital Production. Many Database Maintenance duties were assigned to other staff within Technical Services. In many cases, changing procedures slightly allowed some duties to be discontinued entirely.

Technical Services oversees all digital production work including the scanning, archiving, and preservation of the physical formats. Other digital production work includes the storage and preservation of digital images on compact disk. The cataloger develops rules for cataloging specific collections based on *LC Thesaurus for Graphic Materials* and *Graphic Materials: Rules for Describing Original Items and Historical Collections*. The cataloger, in close consultation with the subject bibliographer, determines the fields that will be used to describe images in particular collections.

The cataloger determines the properties of these fields. The assignment of field properties in the CONTENTdm database includes identification of the Dublin Core Metadata element that the field will be mapped to, the field data type, and the field size. Other field property decisions include determination of whether the field will be searchable, whether the contents of a field will conform to a controlled vocabulary, and whether the field will be hidden to the public. The cataloger is responsible for making these decisions after they are recommended, discussed, and agreed upon in committee meetings.

The cataloger also identifies, establishes, and edits appropriate controlled vocabularies for collection fields. Technical Services staff catalog images and ensure the quality of all cataloging that is done before the cataloging and images are uploaded to the CONTENTdm server and made accessible from the Cleveland Memory Web site.

When a collection is selected and developed by a subject bibliographer, the subject bibliographer has input regarding what fields are used to describe the collection. The subject bibliographer also assists in the creation of controlled vocabularies for the fields

that require them. The cataloger, with the input of the subject bibliographer, also chooses the field on which all search results will be sorted.

Systems

The role of Systems in digitization projects is quite diverse. At the processing end, Systems selects, installs, and supports a variety of scanning workstations. Currently, the library has four scanning workstations equipped with zip drives, CD-RW drives, three flatbed scanners (both legal and ledger size), one slide scanner, and a dye-sublimation photo printer. A fifth flatbed scanning station will be added soon. Systems installed the necessary scanning, image editing, and CONTENTdm acquisitions client software on these computers. Image editing and the CONTENTdm acquisitions software was also installed on about six other computers for additional processing power. After testing and becoming familiar with the CONTENTdm acquisitions client software, Systems provided general training to Technical Services, Special Collections, and Collection Development staff who took on digital processing and cataloging responsibilities.

Systems also selected and set up a Windows NT server to run the CONTENTdm server software. Using utilities included with the CONTENTdm system, the Systems staff imported the postcard database into the new system. The postcard database worked well in this new environment, but after running into a conflict with a new version of a popular virus shield, the library asked DiMeMa to send the Linux version of the CONTENTdm server. The NT server was completely reconfigured with Linux, and the postcard database was again imported. Systems also created the entire directory structure for all of the existing and proposed collections, set security permissions for all users, and helped troubleshoot some problems with controlled vocabulary files. When new

collections are desired, Systems staff set up all new directories and permissions on the CONTENTdm server.

Systems is also responsible for the design, navigation, and architecture of all library Web sites. Systems staff design main pages for each collection. These pages are integrated into a larger overarching Web site that contains all image collections and exhibits. Each collection introductory page includes custom search boxes that enable easier searching of the collection. These search boxes are built using custom scripts as well as query-building utilities included with the CONTENTdm server software.

Support for these digitization projects greatly stresses the already limited resources of the Systems Division. With the cooperative effort between divisions, however, a great deal of progress has been made in a relatively short period of time, and future projects should be completed more efficiently now that the initial "pains" of creating a new project are largely past.

Special Collections

Special Collections provides the historical materials to be digitized and sets the overall priority for projects. Some of the priorities are based on joint programs with other institutions, some are based on obligations incurred from grant funding, and some are based on anticipated demand from patrons. In the latter case, for instance, patrons order individual photographs from the Press Collection for their own use, and Digital Production digitizes the entire subject at the same time. In that way, Special Collections is gauging public interest on the basis of patron requests. The library is not attempting to digitize the 500,000 Cleveland Press photographs systematically, from A to Z, but rather

cherry-pick out the more interesting and requested shots in the normal course of doing business.

FUTURE PROJECTS

We will continue adding photographs from other collections, such as the Cleveland Union Terminal Construction Photographs we are scanning with support from the North American Railway Foundation. We are adding bridge pictures from the Wilbur J. & Sara Ruth Watson Bridge Book Collection and shipping photographs from the Steamship William G. Mather Museum Archives. Thousands of Cleveland Press images relating to Cleveland's ethnic communities, and photographs of prominent African-Americans are being digitized and cataloged.

Once the image database is fully operational, another project includes linking images to maps, using online Geographic Information Systems (GIS) technology in conjunction with the university's Urban College and their Sacred Landmarks Initiative. This will provide patrons with a geographically referenced finding aid to holdings on church landmarks in northeastern Ohio, thus providing a valuable data visualization tool.

CONCLUSION

In a medium-sized academic library such as Cleveland State, in which Special Collections staff consists of one librarian and several students, it is necessary to engage staff from other library units to assist in the work of Special Collections. In order to process, digitize, catalog and put thousands of selected images on the Web quickly, staff from Technical Services, Systems, Special Collections and Collection Development were called on to assist Special Collections. Staff performed the work for which they were otherwise responsible, working cooperatively toward the creation of the online Cleveland

Memory Project. The interdepartmental digitization efforts at Cleveland State University Library, which uses staff with specific expertise in traditional library functions, has proven effective in getting images processed, archived, digitized, described, promoted, and made accessible in a highly efficient manner.

NOTES

[1] Maggie Ressel and Vicki Toy Smith. "A new approach to thesis subject analysis: A collaborative success." *Cataloging & Classification Quarterly* 26 (3): 41-49.

[2] Geri Bunker and Greg Zick. "Collaboration as a key to digital library development: High performance image management at the University of Washington." *D-Lib Magazine* 5 (3), <http://www.dlib.org/dlib/march99/bunker/03bunker.html> (15 January 2002).

[3] Greg Zick, Geri Bunker Ingram, Betsy Wilson, and Jill Fluvog. *Digital Asset Management at the University of Washington Libraries: Teamwork and technology: ACRL Poster Session 2001*, <http://contentdm.com/docs/acrl-exec.html> (15 January 2002).

[4] Rob Withers, Jenny Presnell, and Bob Schmidt, "Creating Digital Projects with Cross-Departmental Teams" (paper presented at annual meeting of the Academic Library Association of Ohio, Columbus, Ohio, November 2001).

[5] Trevor Bond and Alan Cornish, *CONTENT: A model for collaborative database building: Executive summary of Online Northwest 2001 Conference*, <http://www.wsu.edu/~cornish/content/summary.htm> (15 January 2002).