2006

South Carolina Goes Digital: The Creation and Development of USC's Digital Activities Department

Kate F. Boyd
University of South Carolina - Columbia

Douglas King
University of South Carolina - Columbia

September 22, 2005
By Kate Foster Boyd and Douglas King
South Carolina Goes Digital:
The Creation and Development of USC's Digital Activities Department

INTRODUCTION
Within just a few years, the University of South Carolina's Libraries have developed a progressive and active Digital Activities Department that is engaged in offering a wide variety of online digital collections which follow international imaging and metadata standards. The department's rapid growth is attributable to many factors, including the administration's commitment to online access to digital resources, strong yet flexible leadership within the department, the expertise readily offered from library staff members with a wide range of specialized skills and interests, and an overarching collaborative spirit.

HISTORY
The strong interest in digitization, however, is recent at the university libraries. A few people in different areas had some ideas and projects. The Music Library had begun work on its Digital Sheet Music Project <http://sheetmusic.library.sc.edu/Default.asp>, with its own funding, and the Map Library had scanned some maps onto CD-ROMs. In addition to those projects, Rare Books and Special Collections had a number of exhibits on the web, and a few of the other libraries on campus had some images and finding aids online. However, there was no concerted or programmatic effort across collections for the usual reasons: lack of funding and lack of staff. For instance, the Systems Department consists of only three people to support eight libraries. Therefore, there was no shared equipment, database, or quality control. Nor did the libraries share their technical expertise with each other, or communicate about possible online collaboration. This continued until 2002, when a loose coalition of librarians began a "Digital Initiatives Group," to determine interest in digitization within the various areas and to discuss possibilities for a digital collections program. Group members were interested in making rare materials, books, manuscripts, maps, images, and maybe even some video and audio resources from the library's collections available online. Fortunately, because the university was not one of the early creators of digital libraries, some schools having begun as early as 1985, the group members were able to take advantage of a tremendous amount of knowledge readily available and to focus on creating a solid program based on current international standards.

The Digital Initiatives Group began a communication network among the interested librarians. They surveyed the equipment and projects belonging to the various departments and created an internal website to share information. They became familiar with the literature, including *Moving Theory into Practice (*Kenney and Reiger, 2000), *The Digital Library: A Biography* (Greenstein and Thorin, 2002), and "A Framework of Guidance for Building Good Digital Collections" <http://www.niso.org/framework/Framework2.html>, created by IMLS (International Museum and Library Services) in 2001 and now available from NISO (National Information Standards Organization). One of the librarians visited the University of

Virginia's Digital Library Production Services, and two librarians attended the Digitization for Cultural Heritage Professionals Institute at the University of North Carolina at Chapel Hill.

Within a year, the library administration, with the vision and support of the new dean of libraries, saw the need for a more defined Digital Activities Team. The team consisted of a librarian or specialist from each of the rare materials and special libraries on campus, two systems staff, a special materials cataloger, two reference librarians, the preservation librarian, and eventually a faculty member from the School of Library and Information Science. The special libraries represented included the Rare Books and Special Collections, the South Caroliniana Library (the University's library of South Carolina culture), the Map Library, the Music Library, the Newsfilm Library, and the Government Documents Library - all different libraries and all managed in different ways in different parts of campus.

It was determined that the new Digital Activities Team would follow international standards for all of its projects. Reading "A Framework of Guidance for Building Good Digital Collections" confirmed the need for adherence to international standards widely regarded as authoritative within the digital library community. The next step was choosing a best practices standard that would be most suitable for the libraries' needs. After much review, the team determined that the Western States standards for imaging and metadata, which were created using a grant from IMLS, would be appropriate. The team was confident that these documents provided obtainable and appropriate standards for high quality image scanning and equally high quality Dublin Core metadata, facilitating optimal resource discovery and object retrieval for its online patrons. Thus, Western States Digital Imaging Best Practices <www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf > was chosen as the Library's preferred standard for scanning its images; likewise, Western States Dublin Core Metadata Best Practices <cdpheritage.org/resource/.../documents/WSDCMBP_v1-2_2003-01-20.pdf> became its preferred standard for descriptive metadata. Consistency of standards for metadata and imaging is needed to provide the broadest possible access to the collections and to facilitate their durability. Working together, the team created an internal handbook, defining these standards and processes for creating digital library collections at the university.

Team members soon discovered that it was difficult to find the time to fit digitization into their already full schedules. Who was going to decide on the first projects? Who was going to scan and create the metadata? Who was going to manage the database? Luckily, the libraries' administration realized the need for a full-time librarian to manage the projects for the different library departments. In April 2004, a reference librarian was named the new digital activities librarian, creating the Digital Activities Department. Initially, the new digital activities librarian spent a great deal of time learning about similar projects at other schools. She read books and articles about digital libraries, and called a number of practicing digital librarians to learn how they ran their departments and managed digital projects. Along with other colleagues, she also visited two schools, the University of Georgia and the University of Tennessee at Knoxville, to see those

programs. She discussed ideas on a regular basis with the Digital Activities Team and began building the department.

CHOOSING A DIGITAL REPOSITORY
An early but crucial decision was choosing the type of online database from which the digital items would be made accessible. After looking at a number of digital repositories, including Greenstone, Fedora, DLXS, DSpace, and CONTENTdm, the library decided to purchase OCLC's CONTENTdm JPEG2000 Version 3.8 database. The staff was strongly interested in open source software and the idea of designing its own database and site was very appealing. However, the need for a full-time programmer to help setup, design, and maintain the open source database was a huge drawback. CONTENTdm, once installed, could be managed with training by someone without programming skills. In the department's estimation, the searchable online database also displayed the cultural heritage materials in an easy-to-use manner. Other aspects that were appealing were the OAI (Open Archives Initiative) compatibility and the JPEG2000 file format. The OAI compatibility would allow larger repositories to harvest the records of the collections and the JPEG2000 format would add zooming capabilities to the images, an important feature for large images such as maps. Again, without the programming support it would have been difficult to set up a JPEG2000 or an OAI –PMH (Open Archives Initiatives-Protocol for Metadata Harvesting) server.

One of the features of CONTENTdm to which the team was attracted was its use of Dublin Core as its default metadata scheme. It was clear from the beginning that because non-catalogers would be creating and editing descriptive metadata records for the collections, it was important that the team utilize a metadata standard that would not be difficult to apply. The special materials cataloger, who had prior experience applying Dublin Core to digital resources, began thoroughly researching Dublin Core to learn how it could best be applied to digitized print resources such as photographs, maps, and manuscripts. A task force of catalogers from the team met twice to adapt the Western States Metadata Best Practices to the Libraries' local needs. Then, the special materials cataloger and digital activities librarian compiled these local standards into a document for use with all future digital collections
<http://www.sc.edu/library/digital/dacmetadata.html>. Early attempts at applying Dublin Core to digitized print resources stuck closely to the sixteen core elements, but after receiving input from involved librarians and other specialists as the collections grew, it became clear that refinements, or qualifications, would be necessary for optimal application. Therefore, the department began taking full advantage of the flexibility afforded by Dublin Core while striving to maintain as much consistency across collections as possible. Creating the best possible metadata records is clearly not a job for one person.
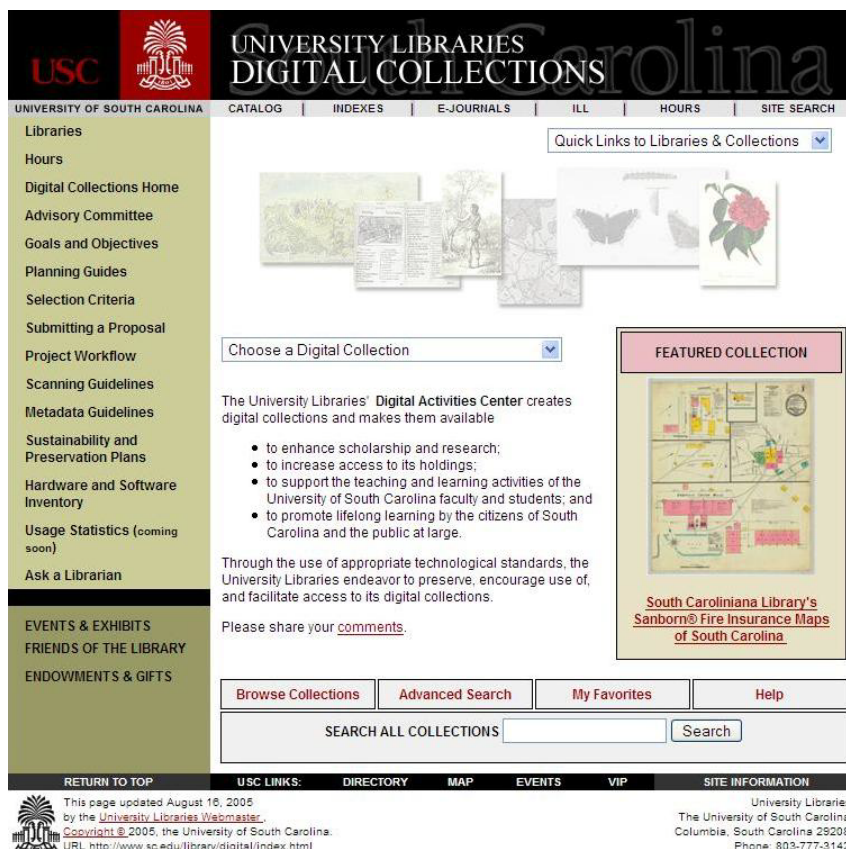
PUTTING THE PIECES TOGETHER

By fall of 2004 the new Digital Activities Department was coming together. With the continued help and support of the team, the digital activities librarian created a website for the department <http://www.sc.edu/library/digital>. She and the team drew up goals, objectives, selection criteria, and policies and procedures for creating the collections, and posted them on the site. Most of these were drawn from other school's programs, such as Columbia University's selection criteria and UT Knoxville's policies and procedures. Two part-time students were hired to start work in the upcoming year. Both library science students, one would do scanning twenty hours a week, and the other would create the home pages for the collections and other website maintenance ten hours a week. A successful evaluation of CONTENTdm had been conducted that fall and a full day of training by SOLINET (SOutheastern LIbrary NETwork) in December familiarized the team with the new product. Furthermore, an Oversight Group, consisting of the dean and three other administrators, was created to assist with choosing collections and deciding on the overall direction for the department.

To begin, the team focused on a simple setup of one flatbed scanner - a Umax PowerLook 2100xl flatbed scanner with transparency adapter that could handle up to 11x17 size documents. This scanner could handle a majority of the materials that the libraries initially wanted to scan, such as manuscripts, photographs, and transparencies.

4

In addition, the department was allowed to temporarily share a wide format scanner with another department on campus, which would prove useful for two map collections. The goal was to lease or outsource other scanner devices for particular projects when needed. To complete the scanning room, the library also purchased two Dell workstations with two gigabytes of RAM and 330 gigabytes of storage, and a server with a 400 gigabyte storage capacity for CONTENTdm. One of the workstations was connected to the scanner and the other was intended for the creation of web pages. The Systems Department set up a shared workspace on the CONTENTdm server for all those working on the projects to be able to access and review materials. The CONTENTdm Acquisition Station was made available on two computers in the department for uploading to the server. The archival master images were to be stored on computer hard drives, some external hard drives, and DVDs until a proper long-term storage system could be set up.

In January 2005, the new department began work on building digital collections. CONTENTdm was up and running, the students had started work, and the first collection was being scanned. As a result of all the preparation by the team and staff the previous year plus the help of the two students, the department was able to get off the ground quickly. The first collection of 100 images from the Rare Books and Special Collections, the *Otto F. Ege Collection of Medieval Manuscript Leaves* <http://www.sc.edu/library/digital/collections/ege.html>, was finished and available on the web site by March.   After a two-year process of setting up the department, it took only three months to begin creating digital collections.

WORKFLOW
To create the collections in the simplest and most efficient manner, the department relies on a small, dispersed group of colleagues and staff: the digital activities librarian as the project manager, the specialist librarians, cataloger, scanner, web designer, and preservationist. In addition, a faculty member from an academic department in the university is included when possible to add expert knowledge to the metadata or further information to the "About the Collection" pages and to ultimately bring the collection to his classroom. The digital activities librarian and the specialists work with the cataloger on the metadata records and with the student on the scanning specifications. The digital activities librarian is also responsible for making sure the project comes together, moving the project forward, eventually proofing and loading the collection to the database, and ensuring that the images and data are archived and properly maintained.

The current project workflow, written to fit the library's needs and the CONTENTdm Version 3.8 capabilities are as follows:

Proposal: A Selection Criteria Form and Proposal Form, available online are filled out and submitted to the Digital Activities listserv by a specialist. The project is reviewed and approved by the Digital Activities Advisory Team and the Oversight Group.

Administrative and Systems Work: Systems staff creates a folder on the CONTENTdm server for the new collection. The digital activities librarian registers the new collection in CONTENTdm and creates the fields.

Project Team: The Advisory Team designates the pertinent people to be involved, including the digital activities librarian, scanner technician, a specialist or collection administrator, the cataloger, and other staff as needed. For each collection there will also be a designated faculty liaison from an academic department who will advise as needed.

Images and Metadata Reviewed: The Project Team reviews the collection items for preservation and scanning-related issues, e.g. resolution, color, and server space. They also decide on file-naming and the metadata schema. A metadata template is created at this point (a basic Word document with a list of the fields and examples). Metadata decisions should be firm and an example created and approved before scanning begins or at least during the early stages of scanning.

Final Product Reviewed: The Project Team also decides on the "look" of the final product. This includes main web page; surrounding texts, e.g. transcripts and acknowledgements; image access design; browsing and searching options; watermarks and any other desired unique items that are to be accessed from the project's main web page.

Administrative and Systems Work: The digital activities librarian inputs collection administrative metadata, such as size of collection, names of project contributors, file-naming scheme, and date the project was started, into a Microsoft Access Projects Log database. She also composes any needed scanning or uploading instructions for the Project Team.

Images: The collection is brought to the Digital Activities Department and test images are scanned and uploaded to the CONTENTdm server shared workspace for review by the Project Team.

Metadata and Scanning: Once the Project Team has approved the metadata, the scanner creates an Excel file with file-names and any other simple fields that she can fill in. She then sends the file with the correct fields and some metadata to the metadata cataloger who finishes the records. Besides knowing the file-naming scheme, the cataloger is able to view the images as they are loaded onto the shared workspace.

Images: Once the images are approved by the team and the Excel file has been sent off, the scanning begins. TIFFs are created for master files, preservation and technical metadata are added to the TIFF headers, and JPEGs are created for the database. When the scanning is completed, the JPEG images are moved to the shared workspace for the Project Team to review.

Final Product: During the scanning, the digital activities librarian and a student begin to design the main web page and surrounding texts for the project in collaboration with the Project Team. The digital activities librarian writes an "Acknowledgements" and "Building the Digital Collection" for the site. The collection administrator or specialist writes an "About the Collection" and suggests links for further information.

CONTENTdm Upload: The digital activities librarian reviews the records in the CONTENTdm Administrative Section, adds the reviewed items to the collection, and builds the searchable text index on the CONTENTdm server, thereby making the images publicly available on the web. However, the new collection will not be advertised until it is reviewed by the Project Team and Digital Activities Advisory Team.

Items Reviewed: The digital activities librarian and collection administrator or specialist reviews the collection online for any final changes or corrections of the metadata records and images. The items in the collection are now available for public viewing from the CONTENTdm server.

Final Product: The final main web page and surrounding pages are completed and sent to the library's web administrator, who adds them to the library's web server, where the main Digital Collections web page will link to the specific collection's main web page.

Final Product Reviewed: The new collection site is reviewed by the Digital Advisory Committee and the Project Team before being publicly released.

Collection Released: The collection is publicly released by being advertised on listservs and local press releases. See http://www.sc.edu/library/digital/index.html.

SCANNING AND METADATA

All items are scanned following the *Western States Digital Imaging Best Practices* < http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf>, and are scanned no lower than 300ppi. Each collection is evaluated and tests are conducted to ensure that the proper ppi and bit-depth are achieved for maximum access and preservation quality of the image and minimum size for storing. Sometimes this is not straightforward and some images in a collection must be scanned at a higher resolution than the others. The scanning student and the digital activities librarian work closely together, frequently discussing issues that arise during the process of scanning. The goal is to avoid rescanning any of the original materials. Therefore, the images are scanned as TIFFs, and JPEGs are created from the TIFFs. The TIFFs are considered the master files and are burned to DVD and stored on a SAN server at Computer Services. The JPEGs are loaded into the database and are burned onto DVDs for backup. While creating the JPEGs, the scanning student reviews the files in Adobe Photoshop and adds preservation metadata to the TIFF header.

One of the most interesting aspects of the department's development, as well as one of its most important features, is the evolution of the Dublin Core metadata records. As stated earlier, the inherent flexibility afforded by Dublin Core was one of the reasons it was the preferred standard of the Digital Initiatives Group. Initial attempts at accurate, thorough, and "correct" Dublin Core records did not significantly stray from the sixteen core elements in strict Dublin Core. As the department added an increasing number of digital collections, it became obvious that qualifications, or refinements, were necessary to provide optimal information to online patrons. The department began utilizing an increasing number of non-Dublin Core fields, as deemed necessary for particular collections, but it was decided that it should create a list of required fields for records, regardless of collection, for the sake of consistency. Eventually, the special materials cataloger and digital activities librarian chose a set of ten core elements for local Dublin Core usage. All descriptive metadata records would include these fields, although their display order would possibly change, depending on the preferences of the specialist librarian and/or digital activities librarian. These local core elements include:

- Title
- Creator
- Description
- Date Digital
- Date Original
- Type
- Format
- Digital Specifications
- Language
- Publisher
- Contributing Institution
- Relation (used to denote the title of the digital collection)
- Web site (a hyperlink to the collection's main page)

Not every one of the above required fields is among the fifteen core Dublin Core elements, but they seem to be pertinent to all local metadata records.

In addition to descriptive metadata for the public's benefit, it is important to create administrative metadata for local, in-house usage. The digital activities librarian creates and maintains this information in a separate Microsoft Access database. Administrative metadata includes such information as names of the collection administrator and faculty liaison who contributed expertise, amount of storage needed for the digital images, local identification numbers, and the file-naming scheme.

Many people play vital roles in the creation of descriptive metadata. The digital activities librarian, who leads the overall direction of each collection and remains involved in every step of its development, begins shaping the metadata. This is done by discussing metadata issues with the person who proposed the digital collection. Those individuals, whether they are librarians or academic faculty, will have varied degrees of interest in, and knowledge of, metadata, so there is no single established role to be played by contributors. The digital activities librarian must determine how much involvement the department and others will have in metadata creation. The special materials cataloger works with the digital activities librarian and outside personnel to establish a template, which includes mandatory and optional fields for the records. This, of course, can only be done after viewing and studying a sample of the scanned images. The special materials cataloger and digital activities librarian create a sample metadata record and review it with the collection's expert, after which the digital activities librarian coordinates the actual creation of descriptive metadata records. These records are input into an Excel spreadsheet and uploaded into CONTENTdm. The scanning student begins the process by creating the Excel spreadsheet, adding filenames and titles as she is scanning, thus guaranteeing a direct link between the metadata and the images. The collection's expert is responsible for filling in the rest of the fields of the spreadsheet, and the digital activities librarian acts as editor. The special materials cataloger plays the role of advisor throughout the metadata creation process, and is available to answer questions and offer advice.

A large part of the project management is balancing the scanning with the creation of metadata. It usually takes longer to decide on the metadata details and create it in the database than it does to scan the materials, so it is easy for the scanner to get out ahead of the projects being loaded. (For example, the Paul Hamilton Papers were scanned in two days, but it was a month before the metadata was completed.) For this reason, the scanner is also assigned such duties as loading records, burning CDs, or fixing files.

While the metadata issues are being worked out and the scanning is begun, the student working on the web pages begins designing the home page for the collection. A formula has been created, so that all of the collection pages have some consistent components, including:

- the University of South Carolina website header and footer
- an image or images from the collection
- a title
- a search box
- at least one browsing option
- a link to the digital collections home page
- a link to the home page for the special library
- a link to the university libraries main page
- a link to an "About the Collection" page, which includes the background for the items, an acknowledgements, and how the digital collection was put together.

In this way the pages are somewhat similar, but each one has a few interesting, unique features that should maintain people's interest. To date, the student working on these pages has been using Adobe Photoshop to create images from the scanned items. He also works with the digital activities librarian and the Project Team on the browsing options for each collection. Sometimes there are a number of fields that the team wants to make available and other times it is a small enough collection that "View All Items" suits. If there are browsing options, the student uses the CONTENTdm custom query pages and some Java script to create these options.

From the outset, the library's approach to digital projects has been to focus on both high quality imaging and high quality metadata. The two work in concert to present optimally useful digital resources. Descriptive metadata records for each individually scanned item or group of interrelated images, such as a two-sided postcard or a book page, provides pertinent information that can assist the online patron in understanding what the image consists of, why it is of value, and how it is related to other items. Accurate and thorough metadata records clearly optimize the educational value of the images to which they apply. Thus, it is crucial that high quality metadata records be matched with high quality images.
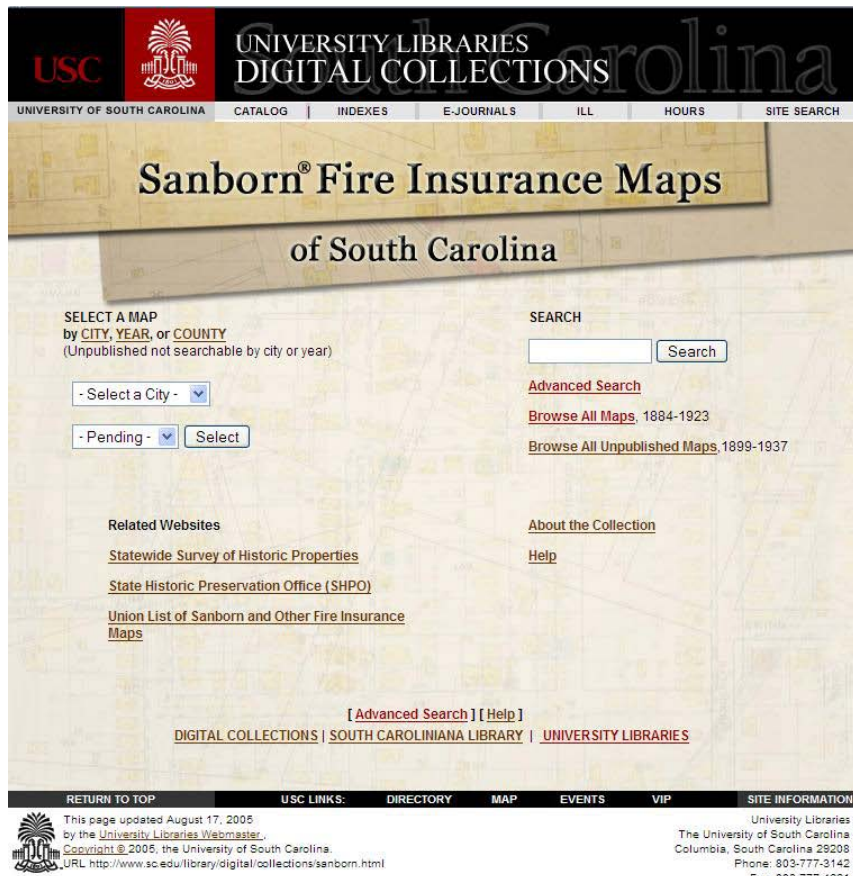
EARLY COLLECTIONS
As of summer 2005, the Digital Activities Department has finished six collections and is in the process of working on three more. Through working on these first six collections, the staff has found that each collection presents its own set of unique issues and

challenges. They have already learned from trial and error what can happen when a project is not completely planned out at the beginning, and they continue to learn from this and other mistakes.

The Otto F. Ege Collection was the first digital project undertaken by the Digital Activities Department, and it most closely resembles the workflow document. The Project Team consisted of a cataloger, rare book librarian, a faculty member from the English Department, and the digital activities librarian. The student scanner created acceptable samples and finished scanning the 100 images in a week and a half. The images were loaded on a shared folder for the team to review. The metadata was put together by the rare books librarian and approved by the cataloger. The front and back of the images were scanned, totaling 100 images, or 50 metadata records. The fact that MARC records for these items were already in OCLC helped with metadata creation. The team merely had to decide which fields it wanted to keep in the Dublin Core record. Finally, the images and metadata were loaded. The records were split up among five different people, in an attempt to save time, but the CONTENTdm database was not set up correctly on some of the staff's computers, so about thirty records out of fifty had to be reloaded. Then, of course, some of the metadata was changed as other collections were created. Luckily, it is possible to make global changes in the CONTENTdm database.

Whereas the Ege Collection was fairly basic with a few mistakes that were fixable, the next collections presented other issues that can arise from these types of projects. The South Carolina Sanborn Fire Insurance Maps were scanned in 2004; however, the library ran out of server space to store them, so the TIFFs ended up on two external hard drives and a computer. It was the new department's job to create JPEGs and metadata for all 2500 images, and then load the images and records into the new database. This took quite a long time.

As each new project began, new issues arose that had to be tackled. The USC Buildings and Grounds collection was the department's first transparency collection as well as the first to use a watermark. For the Development of the Printed Page, the department scanned both sides and was then told by the specialist librarian that only one side was necessary. The department began scanning the Historical Soil Survey Maps and then found out that over half of them were still at the Conservation Lab being repaired. The scanning student, who had experience with preservation, was able to help finish the conservation work. The Paul Hamilton Papers were scanned to the scanner's best ability, but because of the bleed-through and Japanese paper protecting them, they are still difficult to read online and really need accompanying transcriptions, even with the thorough descriptions. Unfortunately, this may not happen for a while because of limited staffing. All of these examples demonstrate the importance of clear communication, but also the challenges to persevere and complete the projects.

Throughout each project, a tremendous amount of collaboration goes into ensuring that the participants understand all the variables, and that nothing is left out. The digital activities librarian and two students closely working with these collections understand best how the images are going to be presented. However, the librarians and staff members from the various libraries and departments are more knowledgeable about the actual items. To create successful digital collections, this gap between the technology and materials must be bridged. Furthermore, as each collection is created, new information

comes to light, and each side needs to be aware of new developments. In addition, the staff is learning new short cuts and ways of better completing the projects in the most efficient and accurate manner possible. Moreover, everyone but the digital activities librarian and her students are taking on these projects on top of their already full workloads. To finish projects, it is important that the Digital Activities Department remain flexible and willing to work with others in any way possible. Every collection comes with its own unique issues that must be resolved. Therefore the department adapts to the needs of each project, yet still maintains the constants that are necessary for a high quality of work.

CONCLUSION
Now that the Digital Activities Department is up and running, it is important to maintain momentum and enthusiasm and to evolve as a department. In addition to adding new collections, there are other goals to achieve, such as improving the preservation of the digital assets, publicizing the collections, collaborating in consortia projects, and participating in repository harvesting.

The key to beginning a digital collections program is to start small, to have the motivation to do it, and to be willing to tackle any issue that arises. There are only a few components that are absolutely necessary to begin, such as a database, scanner and workstation. From there a few part-time staff with a little know-how can get the collections up and running. Most of the information needed to begin is mentioned in this article and is freely available online. Ultimately, it is the staff's hard work and their interest in completing the projects that will make the collections a success.

Resources

Bishoff, L. et al. (2001), *A Framework of Guidance for Building Good Digital Collections*, Institute of Museum and Library Services, Washington, DC.
http://www.niso.org/framework/Framework2.html

Bishoff, L. et al. (2003), *Western States Digital Imaging Best Practices Version 1.0*, Colorado Digitization Program, Denver, CO.
http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf

Greenstein, D. and Thorin, S. (2002), *The Digital Library, A Biography*, Council on Library and Information Resources, Washington, DC.
http://www.clir.org/pubs/reports/pub109/pub109.pdf

Kenney, A. and Reiger, O. (2000), *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Research Library Group, Mountain View, CA.
http://www.library.cornell.edu/preservation/tutorial/

Meagher, E. et al. (2005), *Western States Dublin Core Metadata Best Practices, Version 2.0*, Colorado Digitization Program, Denver, CO.
http://www.cdpheritage.org/resource/metadata/wsdcmbp/

OCLC's CONTENTdm, http://www.oclc.org/contentdm/default.htm

Open Archives Institute (OAI), http://www.openarchives.org/index.html