Descriptive Metadata for Scrapbooks: Interdepartmental Collaboration to Support Efficiency and Discovery

David Gwynn, Anna Craft, and Kathelene McCarty Smith;
The University of North Carolina at Greensboro; Greensboro, North Carolina, USA

Abstract

In the fall of 2011, The University Libraries of The University of North Carolina at Greensboro embarked on the digitization of a collection of approximately 250 historical scrapbooks that documented university history and spanned most of the twentieth century. This collaborative project brought together personnel from Digital Projects, Special Collections and University Archives, and Cataloging. Personnel faced a variety of challenges involving both technical and physical aspects of the project. Among these challenges were the deteriorating physical condition of the scrapbooks and the need to construct project workflows that would maximize efficiency and effectiveness in regards to both staff time and end product creation. In meeting and overcoming these challenges, project personnel came away with lessons learned in the areas of collaboration, communication, and appropriate workflows for a project of this type and scale. These lessons can be taken forward for application in other collaborative digital initiatives.

Introduction

There are a few things one can be sure of when embarking on a scrapbook digitization project: No two scrapbooks are exactly alike, they can include a wide variety of distinctive and unique materials, and working with them is likely to produce both surprises and challenges. The University Libraries of The University of North Carolina at Greensboro (UNCG) faced these challenges with the recent University Archives scrapbook digitization project, which created online access and discovery for 243 scrapbooks dating between 1906 and 2002. The physical condition of the scrapbooks and the challenges associated with their contents necessitated a unique and organic collaborative approach involving archivists, catalogers, and information technology (IT) staff. Ultimately, the goal was to make the scrapbooks available online in a format that mimicked their physical characteristics as closely as possible, while facilitating discovery through descriptive metadata.

About the collection

The University Archives Scrapbook Collection is an important archival resource which documents the academic, social, and cultural history of UNCG through the twentieth century. Founded in 1891 as the North Carolina State Normal and Industrial School, the college's main purpose was to train young women to teach at a time when the state's literacy level was at a low point. In addition to the pedagogy curriculum, domestic science and commercial, or business, classes were offered. The scrapbooks illustrate the student experience as the college progressed from its early years as a women's teaching college, through its integration in 1956, and finally to its establishment as a co-educational state university in 1963. They form the collective memories of student organizations, dormitory life, campus events, and academic and athletic accomplishments. Items discovered within the scrapbooks include formal and candid photographs; programs and posters from productions and cultural events; personal correspondence, dance cards, and invitations; newspaper and publication clippings; and colorful ephemera. They also offer a more complete source of class songs and poems than any other archival record. Perhaps the most unique item of ephemera found in the collection is a peanut from a 1917 scrapbook which is dressed in an "Alice Blue" gown. These scrapbooks have been particularly helpful in researching the history of student life on campus. In several instances, they have provided the most comprehensive information available about unique campus events and traditions such as "Morning Watch" and "Tree Night." The University Archives proposed the digitization of the Scrapbook Collection because it was an invaluable, but fragile, resource whose digital preservation would ensure current and future access to students, faculty, alumni, and researchers.

Planning and digitization challenges

It was generally agreed that the Scrapbook Collection was of significant historical importance to the university, but it was necessary for a structured, goal-driven plan to be submitted to the library's Digital Projects Priorities Team for approval. Once the project was approved by this advisory committee, an interdepartmental project team was assembled consisting of the digital projects coordinator (Electronic Resources and Information Technology Department, ERIT) as chair, the photograph and special projects archivists (Special Collections and University Archives, SCUA), the metadata cataloger (Cataloging Department) and other staff members from each department. This collaboration among the three different library areas was a recurring theme in the project. Personnel in all three involved units collaborated on scanning the objects as well as on metadata creation. None of these units share office space (or even the same floor within the building, for that matter). The distance of the physical spaces, combined with different levels of metadata experience, made communication and documentation critically important to the success of the project.

Particularly critical was the ability to track the progress of individual scrapbooks through the project workflow. Early in the project, tracking was an online process, but as the workflow developed, it gained a physical component. Because numerous staff members from three different departments were handling the material, it was initially difficult to track which department had an individual scrapbook at any given time and what stage of the digitization process had been completed. In response, the archivists instituted a system where a paper form would be kept

with each scrapbook through its journey and would document basic information such as the size, condition, title, item number, and date of transfer between departments.

The physical aspect of the scrapbooks, which were of varying size and condition, was the next major challenge. This previously unprocessed collection, which had been stored in various areas of the University Archives, ranged in size from the smallest, 6" x 4 ¾", to the largest, measuring 25" x 19 ¾". The majority of the scrapbooks had originally been commercially produced, and had to be disassembled for scanning. In many cases the pages were discolored and crumbling, with items that were originally glued to the paper now torn and detached. Merely opening the scrapbooks and turning the pages sometimes caused further damage to the pages and content. Organic materials within the scrapbooks and past storage methods created further problems, including discoloration and decomposition. Additionally, the physical challenge of scanning fragile and cumbersome bound books on flatbed scanners was considerable; many of the scrapbooks were literally crumbling when touched and damage was a significant possibility. Even more daunting was that fact that any given page of a scrapbook might have several items attached that were multipage or folded documents (e.g. programs, greeting cards, letters, etc.).

Name	Date modified	Type
Page_017.tif	8/22/2011 2:19 PM	TIF File
Page_018.tif	8/22/2011 2:22 PM	TIF File
Page_018_Foldout	8/22/2011 5:03 PM	File folder
Page_019.tif	8/22/2011 2:51 PM	TIF File
Page_020.tif	8/22/2011 2:54 PM	TIF File
Page_020_Foldout	8/25/2011 1:40 PM	File folder
Page_021.tif	8/22/2011 2:57 PM	TIF File
Page_022.tif	8/22/2011 3:01 PM	TIF File
Page_023.tif	8/22/2011 3:04 PM	TIF File
Page_024.tif	8/22/2011 3:07 PM	TIF File

***Figure 1**. Example of filenames and folder structure.*

The structuring of the digital counterparts to the physical items also presented concerns. Chief among them was how to provide the most helpful browsing experience for users. File naming conventions and folder structure were set up to address this, and became essential to the organization and tracking of the project, as well as becoming significant components of the project metadata. A hierarchical system was employed wherein each "master page" was scanned as a base, to show how the page might appear to a user leafing through the scrapbook. These files were named "Page_001.tif" and so on. A sub-level was added to show "internal" pages of any attached documents; each of these individual items became a discrete folder when the scans were stored (fig. 1). The folders were give unique names reflecting both the master page and the item's content—for example, "Page_002_Program" contained scans of a program that was attached to page two of the scrapbook—so that the file and folder names would constitute a reasonably descriptive title for each scan. This allowed for automated uploads into CONTENTdm, which duplicated the file and folder hierarchy and displayed the file names in the DC.Title field for each image (fig. 2).

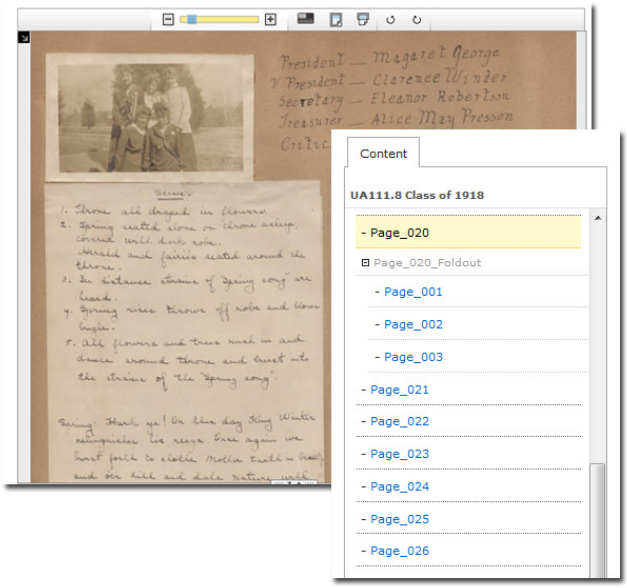

***Figure 2**. Example of CONTENTdm display with item hierarchy.*

Interestingly, the physical digitization process revealed another unexpected interdepartmental challenge—differing terminology among the departments. IT librarians often have very different visions from archivists when faced with terms like "file" and "folder." These differences started to became apparent in some of the earliest discussions involving project planning. When discussing the physical material, the archivists often used the two terms interchangeably, where an IT staff member saw them as separate and distinct. Catalogers brought their own local terminology to the discussions as well, with "record" meaning one thing for them, another for the archivists. Then there were questions such as what constitutes the actual title of a scrapbook, particularly when there is a series of annual scrapbooks from the same organization, some of which may have also added unique names to individual scrapbooks. Should the standardized title (e.g. *University Chorale Scrapbook, 1981*) or the one-off custom title (e.g. *Here We Go Again!*) prevail? The decision on that count was to use two title fields - a transcribed title for information taken directly from the resource, a standardized title to provide users with a clearer picture of the actual item. As will further be discussed in the metadata section, the project participants strove to create data that would satisfy the research needs of the patrons and the internal needs of the University Archives, all while meeting appropriate national standards for description and access.

Metadata creation

Setting project metadata standards

At the outset of the project, the participants agreed that the Scrapbook Collection needed fuller descriptive metadata than the base level used for many local digital collections. The Dublin Core metadata scheme was chosen and enhanced with some local fields and a local controlled vocabulary. Because the collection included a wide variety of material forms—from visual to textual to tactile—the collection was not an appropriate candidate for

scanning with Optical Character Recognition (OCR). With no OCR transcripts to provide keyword results in online searches, the project team would need to manually document the content of each scrapbook and create metadata to describe the notable items. But these tasks had to be completed while balancing both efficiency and the opportunity for discovery: the team would need to create useful descriptive data for researchers while working under the reality of limits on time and personnel resources.

Of particular importance to the digital projects coordinator was the need to allow for discovery both through the library's internal platforms (OCLC's CONTENTdm and WorldShare Management Services) and through external applications such as OAI-PMH harvesting and Google Search. As it was assumed that the scrapbooks would eventually be aggregated with other library collections to create custom websites and digital exhibits, it was essential that the metadata be compatible with other CONTENTdm collections managed by the library. With this in mind, the bulk of the descriptive data would be entered in fields mapped to Dublin Core.

As a mixed-media collection that does not focus on a particular topic or group of topics, the scrapbooks presented challenges when it came to controlled vocabularies. Two main controlled vocabularies were used for the project. Library of Congress Subject Headings (LCSH) were used as a broad, overarching access point, with all scrapbooks assigned the same general blanket headings to allow for discovery and access in cross-repository searches. The project team also chose to create a local controlled vocabulary, referred to here as "Content Formats." This field was used to note specific types of items found in each scrapbook, and the vocabulary was constructed of terms drawn from Getty's Art and Architecture Thesaurus (AAT) and the Thesaurus for Graphic Materials (TGM), with local terms added as needed where neither existing corpus provided an appropriate term. This field does not describe every item appearing in a given scrapbook, instead it brings together material types that appear regularly throughout the collection and may be of particular interest to researchers. Example terms from AAT and/or TGM include advertisements, brochures, correspondence, and dance cards. Example local terms include Class Day programs, dried flowers, play programs, and song lyrics. The vocabulary was created by the metadata cataloger in consultation with personnel from SCUA. At the beginning of the project, team members discussed known types of materials found in the collection, as well as material types that were likely to be found and would be of particular interest to researchers. The metadata cataloger drew up a draft list of terms as the start of the Content Formats controlled vocabulary. As the project progressed there were some additions and changes to the vocabulary as additional notable item types were discovered in the scrapbooks.

Metadata workflow

The decision at the start of the project was that archivists would provide most of the descriptive metadata because of their familiarity not only with the collection materials and general university history, but also with the research interests and needs of the faculty, students, and other university patrons. While archivists would provide knowledge and context of the content, the catalogers would determine descriptive standards, provide metadata training, apply subject headings and controlled vocabularies, and provide quality control and standardization for the descriptive work of the archivists. Digital Projects personnel also had a role in the metadata creation. As they scanned the scrapbooks, they began the metadata workflow by noting the measurements of each scrapbook and other basic information such as title and date.

As with the physical tracking, an online approach was originally planned for the metadata collection, with all participants entering information into a shared Google spreadsheet. But as the project progressed, the team realized that while the catalogers were well-versed in the creation of field-based metadata and the use of controlled vocabularies, the SCUA staff members were less experienced in these areas. Therefore, initial work on the project progressed slowly and produced some amount of frustration for all parties. After discussion between project personnel, the metadata process was revamped to split apart some of the pieces: creation, data entry, and standardization. This splitting brought the Cataloging Department into metadata creation at an earlier point in the process, and allowed each department to further play to its strengths within the overall workflow. Creation of the majority of descriptive data stayed within SCUA, but moved from electronic to paper-based, with each scrapbook now having an accompanying paper metadata form.

Digital Projects continued to initiate the metadata process, entering the basic descriptive data on the paper form at the time that the items were scanned. After scanning, the scrapbooks went to SCUA, where they were further described in free-form notes that provided information on item types and people, events, and other notable things that appeared in the books. SCUA staff members were encouraged but not required to use terms from the local Content Formats controlled vocabulary in their notes, and to suggest additional terms to better describe unusual items and ephemera as they were discovered. Once the paper forms were complete, the forms were sent to Cataloging, where catalogers and student workers entered data into the unified metadata spreadsheet. This involved standardizing terminology, applying terms from controlled vocabularies, and constructing descriptive narratives from the free-form SCUA notes. After the data was entered, the metadata cataloger quality-checked the data, checked entries in the Content Formats field to make sure all terms conformed to the controlled vocabulary, edited and added to that controlled vocabulary as needed, and consulted with SCUA on any questions that arose. Personnel from SCUA read over the descriptive metadata records once they were complete, and further changes were made as needed. The paper forms were stored in a binder after they were entered online, and this binder became an important reference point for project personnel. Once the new system was implemented, metadata work began to move at a swifter pace.

Lessons learned

While the University Archives Scrapbooks project presented many challenges, it also presented opportunities for learning to carry forward to future endeavors. The project team identified three areas of particular importance, relating to collaboration, communication, and project workflow/process.

Collaboration

As is true in so many collaborative endeavors, this project presented learning opportunities for all members of the project team. In order to develop workflows that would shepherd materials through the three participating departments, team members had to increase their familiarity with terms and processes common in the other areas. This included SCUA staff learning more about field-based metadata as well as the construction and use of controlled vocabularies, and Digital Projects staff learning more about care and preservation issues relating to historical scrapbooks. Another significant area of learning involved university history. As they created controlled vocabularies and descriptive metadata records based on notes from SCUA staff members, catalogers faced local terminology they had never seen before. In order to write descriptions of events such as "Morning Watch," "Tree Night," and other local traditions, catalogers frequently consulted their SCUA colleagues, learning about university history along the way. This experience and increased awareness has better prepared Cataloging Department staff to work with university history collections in the future. Staff members in SCUA were the indisputable content experts on this project, but even they encountered new and previously undocumented information about university history and the day-to-day lives of the students in earlier eras. Through the digitization of these scrapbooks, this information can now be shared with the greater university and larger world communities.

This project also provided a reminder to staff about managing expectations and not making assumptions about the work experience of personnel in other areas. Connected to this is another lesson learned - the importance of focusing on the strengths of each project partner. In designing a metadata workflow at the project's inception, the decision was made that SCUA staff members would create and enter descriptive metadata for the scrapbooks, with training from staff from Cataloging. The reasoning here was that staff in SCUA have the most in-depth content knowledge of the collection, and therefore would create the most accurate descriptive data for the items. But this decision did not take into account the full need for training in field-based metadata and controlled vocabularies that SCUA staff would need. As discussed above, the restructuring of the metadata workflow allowed each department to contribute in the most effective and efficient manner.

Communication

Another recurring theme throughout the project was the importance of communication in making the project a success and in building stronger working relationships between the involved departments. The departments were physically spread across the library, meaning staff members did not regularly see each other in their day-to-day work. With nearly 250 scrapbooks to move between SCUA and Digital Projects, and metadata sheets to move between all three departments, it was necessary for team members to regularly communicate on the status of the project through a variety of means. Most often these were face-to-face meetings and emails. Both avenues provided opportunities for status updates and needed clarifications on issues as they came up. And regular communication built and strengthened the working relationships between the departments, helping put in place systems that can be carried forward to future projects.

Workflow

As referenced above, the project team learned that existing workflows and structures cannot always be imposed on a new project - sometimes the process must be built organically to function successfully and showcase the strengths of its component partners. The initial plan to institute a paperless system for metadata and tracking did not meet the needs of the project nor of the participating staff so a combination paper and electronic system was devised. The continued success of a hybrid system means that it may be considered for future use in appropriate projects.

Conclusion

Scrapbook digitization will always be a challenging endeavor, fraught with both anticipated and unanticipated obstacles. The end product, however, can become a significant resource for researchers and casual observers alike. Scrapbooks can provide amazing insight into popular culture and everyday life and can also be a critical source of institutional history. By providing a digital version of the University Archives Scrapbook Collection, the UNCG University Libraries will dramatically increase opportunities for access to this resource while simultaneously reducing or even eliminating stress on the physical materials. By stressing an organic, collaborative approach based on open communication and emphasizing the strengths of each participating department, the Libraries have established a framework for similar digitization projects in the future. Specifically, this approach has recently been applied to a new $200,000 LSTA-funded grant project, a large component of which involves digitizing additional scrapbook collections held by UNCG and several sister institutions in the Greensboro area. It is hoped that the best practices for interdepartmental collaboration developed through the team's initial experience with scrapbooks will prove scalable to a project involving multiple institutional partners as well. At a minimum, though, the lessons learned through the digitization of University Archives Scrapbook Collection will provide a useful starting point.

Author Biography

David Gwynn is digital projects coordinator for the University Libraries of The University of North Carolina at Greensboro (UNCG). He received his MLIS from UNCG in 2009. Anna Craft, metadata cataloger at UNCG, earned her MSLS from The University of North Carolina at Chapel Hill in 2008. Kathelene McCarty Smith, artifacts, textiles, and photographs archivist for the Martha Blakeney Hodges Special Collections and University Archives at UNCG, was awarded her MLIS from UNCG in 2009 and her MA in Art History from Louisiana State University in 1985.