

Sixteen Students, Fourteen Weeks: Building a Digital Library through Collaborative Learning

Since 2005, the Graduate School of Library and Information Science (GSLIS) at Simmons College has provided students with the opportunity to take part in the creation of a digital library highlighting alumni scrapbooks from Simmons College Archives. These projects have been incorporated into the classroom as part of a Digital Libraries course taught by Professor Candy Schwartz. In May 2012, the Paula Kantala McAndrews Scrapbooks were added to the Notable Women of Simmons College Digital Scrapbook Collection after a semester of intensive training and collaboration among the library science and archives students. Students, as team members responsible for the library's creation, learned as they developed workflows, honed project management skills, and managed a successful technology project.

During the digital library's 14-week development period, students acquired specialized knowledge on their committee's topics as well as gaining a general knowledge of other aspects of digital libraries. Professor Schwartz gave guidance when needed but encouraged students to make decisions independently on all aspects of the project, including deadlines, criteria, and technical specifications. These methods of constructivist and experiential learning left students with practical experience and a "t-shaped" skill set to aid them in future digital library projects. Roda et al. (2005) refer to a similar case study of "participatory design," in which the students at the American University of Paris participated in constructivist learning by becoming involved in the development of an academic digital library.

Digital library creation as a pedagogical tool familiarizes students with the technological aspects of digital librarianship as well as introducing concepts related to digital preservation, rights management, metadata standards, and interdisciplinary teamwork. This first-hand experience for novice digital library developers has set the stage for future digital project management endeavors and has opened up opportunities for a redefinition of what it means to be a librarian, archivist, and curator.

Content Committee

The Content Committee was responsible for placing the scrapbook in its historical and contemporary contexts. During the initial stages of the project, they researched Paula Kantala McAndrews's time at Simmons College and what her life as a student in Boston would have been like. This research provided the groundwork for the full content of the

website, including the educational materials for students in middle school, high school, college, and teaching English as a second language (ESL) or other language classes.

The bulk of the Content Committee's work had to be completed early in the semester, as many other committees' work was dependent upon it. The Marketing Committee relied upon deliverables from Content in order to populate Facebook and other media outlets, as well as print advertisements and promotional materials. The Systems Committee also required the full text of the website for the Content Committee. In order to keep the workflow on schedule, the entire class assisted in proofreading the website content prior to the digital library's launch.

Digitization Committee

The Digitization Committee focused on scanning and managing over 400 items including photographs, newspaper articles, brochures, and other memorabilia. Three students were responsible for researching digitization standards, coordinating object evaluation and dissemination, setup and scanning environment maintenance at the Simmons College Archive, and developing the digitization workflow and the training of classmates, most novices to digitization practices.

The committee's work started one week into the Spring 2012 semester and began researching standards and equipment to ensure the objects were properly handled, digitized, and stored. Various resources such as the Digitization Guidelines by the Federal Agencies Digitization Guidelines Initiative (FADGI, 2013), Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images by National Archives and Records Administration (2010), and DLF (Digital Library Federation) Benchmark for Faithful Digitization Reproductions of Monographs and Serials by the Digital Library Federation (2002) were consulted to develop guidelines and procedures for the project. Two of the three team members hold photography degrees and understood the technical needs and capabilities for digitization. Limited research went into equipment resources because the project was supplied with two Epson Expression large bed scanners and two Apple laptops loaded with the Epson scanning software and Adobe Photoshop CS5. The team discovered the scanning environment did not completely conform to FADGI guidelines, but ensured the scrapbooks were handled properly during scanning by providing dust spray cans and demonstrations on handling materials. Along with these guidelines, the committee also worked closely with the Preservation Committee to ensure preservation standards such as image file types and image pixel resolution.

While research was conducted, the team evaluated the scrapbook materials and devised a plan for coordinating

page-scanning responsibilities among their classmates. First, all scrapbook objects were given an image filename; image naming conversions were created by the committee to give contextual meaning and organization to each file name. Scrapbook pages were then categorized by level of difficulty, and given designations of simple, easy, or difficult. By the time the committee selected which classmates got which objects, the project's schedule was finalized, which determined the digitizing workload for each person. Photographs of each scrapbook page were taken and made web ready to ensure if any object fell off a page, it would be paired back with its proper placement.

Before students started their digitization responsibilities, the Digitization Committee had to train classmates how to handle, scan, storage, and edit the scrapbook objects. The team created training manual visualizing the step-by-step processes. The team also held an in-class training session for working with the objects and opened the floor for any concerns the students had. Many students had never used Epson scanners or Photoshop before this experience. First, the team instructed on how to handle objects. Next, students were shown specific technical requirements for the Epson scanners, such as navigating through the software and showing hardware features. The hardware was not brought to the demonstration; instead images showing the step-by-step process were presented. After students were shown how to make images preservation friendly and how to create image file names, they were taught how to use Photoshop. Majority of the class were novices to image editing and not clear on how to use this software. Because image editing takes specialty and skill, the team decided to use auto levels, tone, and colors for modifications. In an ideal situation the process would have abided by FADGI guidelines, but since this was an introduction to building a digital library and on very tight deadline, the team decided to make this compromise.

Guidance for saving files to network storage was also given to students. To keep everyone's work organized the team created individual folders for students to place their work. Once students were working independently, the Digitization team divided management responsibilities amongst themselves. Two members divided the class up and were the point persons for those individuals if they ran into technical issues or needed imaging assistance. These team members were also responsible for conducting image quality control to ensure students properly followed digitization instructions. The third member of the team acted as an administrator and liaison between other project committees, such as metadata and systems. This person also managed and coordinated the scanning calendar with the Archives and students since the department was only open Monday through Friday during normal workday hours.

After all the objects were digitized, the files were automatically resized in a batch process using Adobe Bridge and passed onto the Systems Committee for upload into Omeka.¹ Despite the amount of planning involved, the

Digitization Committee faced a number of unexpected challenges while the scanning and editing was executed. Early into the project, it was discovered that one of the flatbed scanners was failing, and causing slight color distortions in the digitized files. This necessitated the rescanning of a number of scrapbook pages and complicated the shortened time frame. Some work had to be delegated to students who had completed their assignments due to scheduling conflicts.

Creating a professional product was also a challenge with a novice workforce, many of the students involved were not familiar with image editing software. Despite the creation and testing of a manual, it was not always consulted. It was clear during the quality control process that not all steps were being followed consistently, leading to poor image corrections, excessive debris on images, and incorrect image sizing. The Digitization Committee also needed to make a number of compromises on quality control during the project. We were unable to change the archives space to minimize potential color distortions since walls could not be painted to a neutral grey, nor could the outside light be controlled.

The decision to use Photoshop's autotone and autocolor features directly conflicted with the recommendations of the *Federal Agencies Digitization Guidelines Initiative's Cultural Heritage Materials: Creation of Raster Image Master Files* (Osborne et al., 2010). Autocolor correction had the side effect of creating colorcasts on scanned pages, often leaving images with a distinct pink or blue tone. Additionally, all color corrections had to be performed away from the actual objects due to constraints of using space within the archive. Ideally, students would have performed all corrections with the objects in hand, and using finer controls on image correction.

Environmental Scan Committee

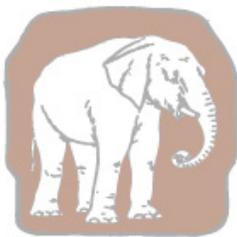
Before the construction of the digital library was underway, the Environmental Scanning Committee was instructed to discover and analyze scrapbook projects similar to our own. The committee provided for contextualized analyses of the digital scrapbook landscape, describing the current state while highlighting trends.

The committee's goal was to systematically collect information about digital scrapbooks created outside of Simmons and settled on evaluating the libraries based on their own class committee infrastructure. Researching digital scrapbooks similar to the Paula Kantala McAndrews project proved to be a bit of a challenge, but the students found three to comparatively analyze: the James Carroll Scrapbook (Catholic University of America: American Catholic History Research Center. n.d.); the Mildred Wirt Benson Memory Book (Iowa Digital Library, n.d.); and the Clarence D. McLain Scrapbook (Digital Collections Wolfgram Memorial, n.d.) After reviewing these scrapbooks, the Environmental Scan Committee created survey questions based specifically on the other committees' needs. For

example, questions about image quality target the needs of the Digitization Committee, and questions about layout, appearance, and content management system were tailored for the Systems and Web Committee. Ten questions for each scrapbook were entered into Survey Monkey and then distributed to all class members. Discussion of the survey results set expectations for the class's own project.

Marketing Committee

The Marketing Committee was responsible for drafting the project's vision, mission and user scenarios, as well as branding and advertising the scrapbook. Working alongside the Systems Committee, they helped to develop the elephant logo using ephemera found within the scrapbook. With the visual branding of the scrapbooks established, Marketing then set up a Facebook page to promote the digital library and our fundraisers to the community beyond our class, both on campus and off.



*Paula Kantala McAndrews
Scrapbooks*

The fundraisers themselves were also coordinated by the Marketing Committee.

The first was a late 1920s-themed dance hour on campus. The second was a two-day bake sale, also on campus. The proceeds covered the cost of printing promotional materials, as well as that of the digital library's unveiling party at the end of the semester.

Over the course of the semester, each of us logged the hours we worked each week in a Google spreadsheet. Using Microsoft Project, Marketing calculated the hours required to complete the creation of the library, coming up with a total of 1,457.25 for the semester. Based on current salaries for comparable work at the time, they estimated that salaries for the project would have total a cost of \$44,698.

Metadata Committee

Like many of the other committees in the class, the Metadata Committee's first task was to produce a set of detailed white papers. Two members of the committee researched the different metadata schema options for the project, and the other two committee members reviewed the range of controlled vocabularies. The metadata schemas that were considered included Dublin Core, MODS, METS, MARCXML, and VRA Core. The precedent from past classes was set using Qualified Dublin Core for the metadata, so after thoroughly analyzing the different alternatives and presenting to the class, the group collectively decided that DC would be the best option for the digital

library. Every decision that was made included a thoughtful consideration of our classmates' limited experience with cataloging. Dublin Core was relatively easy to learn, and perfect for a group of relative novices to use. The group arrived at a similar conclusion with the controlled vocabularies, choosing the relatively straightforward Library of Congress Thesaurus for Graphic Materials and the Notable Women of Simmons College Subject Access, which had been constructed specifically for the digital library projects by a student in a previous class. Through the research process, the metadata committee gained a strong understanding of schemas and controlled vocabularies on the whole, preparing the committee to present and impart this information to the rest of the class. It also gave the committee the background needed to move onto the development of instructions for metadata creation.

In the next step of the process, the metadata committee produced detailed guidelines for the entire class to create metadata. As part of the class, each student was required to submit records for objects in the digital library, ensuring that everyone in the class developed a basic understanding of Dublin Core and the process of metadata production. The metadata group developed the necessary expertise to guide the rest of the class through record creation. Each item on each page of the scrapbook was assigned to a class member, who was responsible for filling out the appropriate Dublin Core fields, using the instructions constructed by the committee. These instructions were presented to the class both orally and in written guidelines, offering not only formatting rules to ensure consistency but also some advice on how best to approach more difficult tasks, such as subject cataloging. The benefits of the process was that every individual in the class received some basic training in creating records and developed first-hand experience in cataloging digital objects. In turn, the metadata committee developed a deeper knowledge of material by instructing the rest of the class.

Following the creation of the records, the metadata committee rigorously reviewed all the submitted metadata for quality assurance. One of the biggest challenges that we faced as a group was ensuring consistent formatting and use of vocabulary across records. Subject classifications in particular proved difficult for some. When you have sixteen different people filling out records with relatively little cataloging experience, you are bound to encounter a wide range of different approaches. The metadata committee ended up thoroughly combing through each record to lend some added consistency among the records. This process, however, provided us some insight into the difficulties of producing good metadata and helped us more fully appreciate the considerations that go into creating records for digital records. We were also able to provide some feedback to the class, though the cramped schedule only allowed us to do this in a limited capacity.

Preservation] Committee

The Preservation Committee was responsible for ensuring the long term functionality of the Digital Library. The first task undertaken by the two-member committee was to recommend optimal file types for long term storage. The committee recommended using robust file formats - formats that were well documented, open source, and with minimal compression. These types are likely to continue to be supported in the future, and already are widely used. Ultimately, we chose to use TIF, PDF/A, and CSV formats. Furthermore, the committee examined the ideal scanning resolution for the scrapbook pages. Due to the limited amount of time and space available for the project, the committee chose to go with the minimal resolution for scanning recommended by the Federal Agencies Digitization Guidelines Initiative (Osborne et. al., 2010), 300 dpi, as it would allow all the relevant information to be recorded.

A large portion of the work performed by the committee involved assessing the current state of the Notable Women of Simmons College collection. Unfortunately, because of how the overall project was structured, there was very little ongoing maintenance of previous digital libraries. This lack of oversight had caused the loss of some of the earlier libraries during a server upgrade. Future planning for the library was complicated by its lack of a dedicated caretaker. Ultimately, a recommendation was made for the creation of a digital stewardship position so that one individual would be able to maintain the files over the long term.

Since future maintenance of the Paula Kantala McAndrews digital library was uncertain, our best defense was to ensure the library files were fully documented and stored in multiple locations. Preservation metadata was harvested from all image files utilizing Mac Terminal commands. All metadata was stored in a comma-separated-variable spreadsheet; using fields adapted from PREMIS (Preservation Metadata) standards. The information collected included details about the creating software, file size, file location, and image specific metadata, such as resolution and color space. Checksums were generated so that a digital steward could validate files in the future, potentially. In addition to the publicly accessible copy on the server, additional copies of the image files were saved to a dedicated external hard drive. A third copy was to be saved to cloud storage purchased through Amazon; however due to time restraints the backup was unable to be executed. The storage space was obtained and will be able to be used by future digital libraries classes.

The final effort undertaken by the Preservation Committee was to try to create an archived version of the website, which would live on the backup drive. On the recommendation of faculty, the committee utilized SiteSucker software for Mac OS (SiteSucker, n.d.) to create a local copy of the Paula Kantala McAndrews Digital Scrapbooks.

While the software created a functional local copy, it did not create a perfect backup. The software was not compatible with JavaScript, so the scrapbook navigation was not properly rendered on the backup.

Project Management Committee

The primary initial responsibility of the digital library's Project Manager was to determine the critical path of the project; in other words, to establish the specific activities required of each committee and to assign deadlines for each individual or group task. As a preset course requirement, we used Microsoft Project project management software, though other proprietary and open source solutions were researched for a white paper on project management. Microsoft Project gives the project manager the opportunity to view the project as a whole and to see the full scope of work. It has functions for managing time and budget, start and finish dates for specific tasks, and the relationships between those tasks. Another strength of this program is its capacity for generating reports and charts based on the data entered. Although it has a familiar Microsoft interface, it is not necessarily intuitive, and allowing some training time prior to beginning the project is recommended. In a class setting, the main downfall of using a proprietary program was the fact that the entire class did not get firsthand experience using the software, though some overview and instruction was presented throughout the semester.

In addition, the Project Manager was responsible for establishing communication methods for the course using a variety of soft tools, including a Wordpress blog for posting announcements and meeting agendas, Google calendars for sharing deadlines, and Google documents for collaborative work. When making decisions about communication methods, keep in mind that using too many disconnected tools can cause communication to become either overlapping and repetitive, or disjointed and confusing. Software created for project management can help streamline the information documented with lightweight tools. In our experience, combining these communication methods helped to keep the class at large up to date on changes to the workflow and topics to be covered at weekly committee meetings in a fairly consistent and streamlined manner. This was particularly critical due to the interdependence of committee work; no committee operated within a vacuum, and if one committee was late delivering work to another, the risk of scheduling bottlenecks at the end of the semester increased.

Some of the main challenges associated with digital library project management are inadequate time for planning, insufficient communication between stakeholders, and poorly defined acceptance criteria for deliverables. Making use of Microsoft Project and other deemed-appropriate tools in order to avoid these trappings was an ongoing project management responsibility. In our experience creating a digital library, these did prove to be apt concerns.

In time-sensitive projects, breaking larger deadlines down into step-by-step tasks is critical to avoiding workflow bottlenecks during the final stages of the project.

Systems/Database/Web Development/Usability Testing Committee

The systems team members worked with a Simmons College systems administrator to upload files into Omeka and design a web portal for the digitized scrapbooks. As there was little time for team members to develop new technical skills, the four committee members were chosen based on prior experience and subdivided into a front-end team focused on design, CSS coding and usability and a back-end team focused on scripting, uploading files to the database and tailoring Omeka plugins for the project's needs. Collaboration with the Simmons technology staff was necessary, as the students in the class lacked the necessary server access and permissions to install the software and plugins used for the library.

The group began by surveying existing systems for their suitability for the project, as an academic exercise only. The instructor had already determined that Omeka would be used for the library. The team reviewed available Omeka plugins and lists of plugins used by previous courses, selecting OAI-PMH Harvester and Repository, Exhibit Builder, Access Keys, COins, Browse Subjects, and Browse Types. When the digitization process was complete, the team uploaded the image and metadata files to the database, then corrected them after a proofreading and quality control process. The back-end developers of the team developed a page viewing function so the scrapbooks could be viewed page by page with the items on each page displayed in thumbnails beside the scanned pages.

The front end developers on the team presented logo and color scheme designs to the rest of the class for input and approval and liaised with the marketing team on the designs. A theme was modified from the Emiglio theme. The team also set up a card sort for the class to participate in to determine the site's information architecture. The systems team worked with the content committee, who developed supplementary historical and education materials that were converted into web pages by the systems team. The team developed and administered a usability test, which was administered to an undergraduate student volunteer during a Digital Libraries class meeting.

The developers had to link the metadata records to the items manually, which was time consuming, and getting PHP and JQuery to function with Omeka required hours of tweaks and testing. Ensuring that the functions and appearance of the digital library were consistent across different web browsers also took time and testing. Much of the systems group's work depended on the work of other groups being completed first, so by necessity was weighted toward the end of the semester and created an intense workload in the final weeks before launch.

Rights Management Committee

The Rights Management committee was responsible for researching copyright laws that apply to digital libraries, and obtaining permissions for objects in the scrapbooks, as well as creating use statements for the website. This included statements for items for which Simmons College owns the copyright, items for which the copyright issues were considered “low risk” and users should inquire to the College Archives regarding permissions, and items for which permissions should be acquired before we could comfortably include the item in the digital library. The only item that proved to be problematic was a Camel Cigarette ad; the committee was unable to obtain permission to use the ad, and instead used a placeholder image created by the Systems Committee. The placeholder informed users that the redacted item could be viewed in person at the Simmons College Archives, but could not be made available online. The decisions made by the Rights Management Committee were informed by the Fair Use Act (US Copyright Act 17, Section 108), which is commonly used by libraries when making materials available online for education purposes.

One option that we discussed after the launch of the digital library was that of Creative Commons licenses. This could be an interesting avenue for future classes to consider, particularly if they should choose to use Flickr as an additional social media platform for promoting the digital library.

Conclusion

Through collaborative teamwork and intensive peer training, the class was able to produce a complete digital library accessible to the general public (<http://alanis.simmons.edu/paula/>). More importantly, though, the students in the class emerged with critical skills in project management, digitization and metadata, and a variety of different specializations based on specific committee work. In their article “Digital image library development in academic environment: designing and testing usability”, Claudia Roda, Ann Murphy Borel, Eugeni Gentchev, and Julie Thomas refer to a similar case study of “participatory design”, in which students (the intended end-users) became involved in the development of an academic digital library. Roda et al. (2005) discuss the learning objectives, which included “developing the skills necessary to work in interdisciplinary teams; appreciating the skill necessary to participate in, and manage a large project; learn about the affordances of some of the essential hardware and software components of information systems;...learn about the essential issues in digital assets management... and how they apply to different types of use” (266). Out of the Digital Libraries course at Simmons, students obtained a comparable understanding of the communication, management, and technical skills needed to execute a digital project.

The practical experience of developing a digital library offered many challenges and a steep learning curve, but

ultimately resulted in a strong basis of knowledge for further work in digital libraries. Students learned to collaborate with different personality types and work together to come to a consensus as a project team. H. Frank Cervone (2005) cites the importance of influencing people and finding common ground as important techniques in the project management of digital library project teams. The class also encountered many of the frequent problems seen in the development of digital libraries, such as breaks in communication, limited time for extensive planning, and the need for risk management. Frederick Zarndt (2011) discusses some of these in the article "Project management 101". Through this experiential learning, however, the class learned the value of management and collaboration in the overall process of project development. With peer training and continual teamwork, the class as a whole was able to develop a wide variety of technical skills in specialized areas and overall understanding of what goes into digital library creation.

References

- Catholic University of America: American Catholic History Research Center. (n.d.). James Carroll Scrapbook. *Acua Photograph Collection at Catholic University of America*. Retrieved from <http://www.aladin0.wrlc.org/gsdll/cgi-bin/library?c=cuphoto&a=d&cl=CL6&d=HASH01f1524299823f8f06877d8e>
- Cervone, H. F. (2005). Influencing: a critical skill for managing digital library project teams. *OCLC Systems & Services* 21(2), 105-109.
- Digital Collections Wolfgram Memorial. (n.d.). Clarence D. McLain Scrapbook. *Widener University Archives*. Retrieved August 11, 2013, from <http://digitalwolfgram.widener.edu/cdm/landingpage/collection/p270801coll4>
- The Digital Library Federation Benchmark Working Group. (2002). *Benchmark for Faithful Digital Reproductions of Monographs and Serials*. Retrieved August 14, 2013 from <http://old.diglib.org/standards/bmarkfin.htm>
- Federal Agencies Digitization Guidelines Initiative. (2013). *Federal Agencies Digitization Guidelines Initiative: Digitization Guidelines*. Retrieved from <http://www.digitizationguidelines.gov/>
- Iowa Digital Library. (n.d.). Mildred Wirt Benson Memory Book, 1922-1928. *University of Iowa Libraries*. Retrieved August 11, 2013, from <http://digital.lib.uiowa.edu/cdm/search/collection/mwb>
- Osborne, R., Scott, C., Griggs, K., Rhodes, E., & Puglia, S. (2010). Cultural Heritage Materials: Creation of Raster Image Master Files. *Federal Agencies Digitization Guidelines Initiative*. Retrieved from <http://www.digitizationguidelines.gov/guidelines/digitize-technical.html>
- Roda, C., Borel, A. M., Gentchev, E., & Thomas, J. (2005). Digital image library development in academic environment: designing and testing usability. *OCLC Systems & Services* 21(4), 264-284.
- SiteSucker. (n.d.). SiteSucker.com. *Sitesucker*. Retrieved August 11, 2013, from <http://sitesucker.us/home.html>
- Thornhill, K., Powers, E., Toner, E., Gunn, C., & Tanguay, C. (2012, October). Sixteen Students, Fourteen Weeks: Building a Digital Library Through Collaborative Learning. Poster session presented at the annual meeting of Association of Information Science and Technology, Baltimore, MD.
- Zarndt, F. (2011). Project management 101: Plan well, communicate a lot, and don't forget acceptance criteria! *OCLC Systems & Services* 27(3), 170-174.

Notes

¹<http://omeka.org/>