

[This document contains the author's accepted manuscript. For the publisher's version, see the link in the header of this document.]

Paper citation:

Russell, Beth M. "From the Ground Up! Lessons Learned from One Librarian's Experience Digitizing Special Collections." *College & Research Libraries News* 62 (2001): 603-606.

Text of paper:

From the ground up

Lessons learned from a librarian's experience with digitizing special collections

by Beth M. Russell

Digitizing library collections is becoming a common request, and libraries are responding to these requests by attempting to make images of rare or interesting materials available to a wider, "virtual" audience.

My involvement in digitization began quite by accident; the chance to work on a large-scale digitization project crossed my desk completely unsolicited. As a special collections cataloger with just enough computer expertise to know that imaging was a complicated and demanding process, my first reaction was far from enthusiastic.

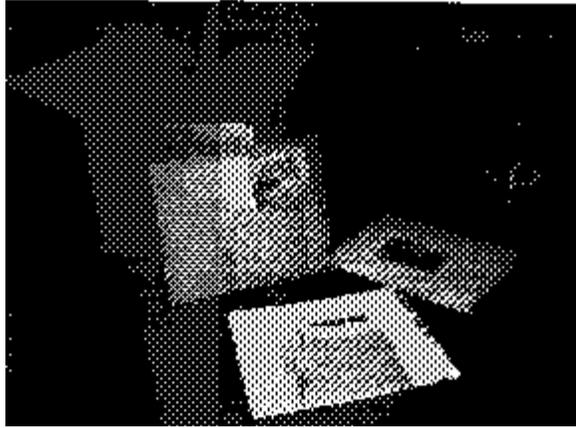
Nonetheless, more than two years after beginning the project. I can now point to a large database of searchable photographs that has already proven useful to researchers.³ My experiences are living proof that starting a digitization project from the ground up can be done.

This article will highlight how the digitization project was built from the ground up, stressing a few simple lessons learned that may prove useful for other librarians who find themselves in a similar situation. In short, I would advise librarians beginning such a project from scratch to keep in mind three simple maxims: 1) do your research, 2) nurture collaborations, and 3) be flexible.

Beginnings

A digitization initiative was in the works at Texas A&M University long before my arrival in 1996. I was unaware of this, though, when I was encouraged to pursue grant funding from TexShare to make accessible a collection of thousands of agricultural photographs. The program, I called a TexShare Access to Local Holdings grant at the time, and now named i TexTreasures, was designed to assist libraries to provide access to their special or unique local collection holdings and to make information about these holdings available to library users across the state."

This was clearly a case of having the right knowledge at the right time. I had been working with a collection of photographs from the university's Agricultural Communications Office for a short time, using several images to illustrate a library publication and researching the



The original boxes and housings of the material and the state in which it arrived at the archives, after many years of storage.

relationships between this material and historical archives of the Texas Agricultural Extension Service and other agencies held in the Gushing Memorial Library at Texas A&M.

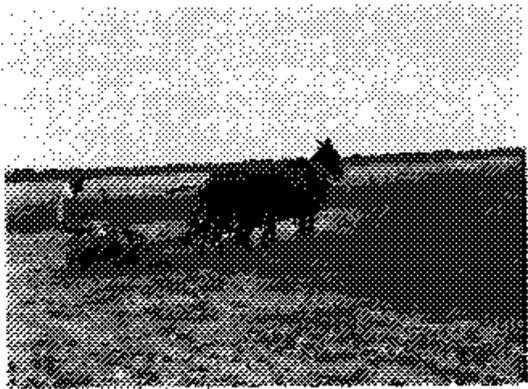
At first glance, the collection fit the goals of the grant program very well. The material was unique to Texas A&M, but featured images from all over the state. Historical interest in agricultural education and home demonstration was demonstrable, but since most of the photo-graphs also had descriptive captions identifying individuals and locales, they would have a much broader appeal. African Americans and women were also featured prominently in segments of the collection.

The photographs were fragile and deteriorating quickly: negatives were often paper-clipped and placed in acidic pockets, while the prints themselves were usually-glued to cardstock. If the images were to be preserved for future generations, something had to be done.

1. Do your research The opportunity to pursue a grant with a Texas theme was a good match with my knowledge of the collection. In order to promote the collection and to fully understand it, however, required much more in-depth research. I began researching the history of the collection itself. I was able to contact people who had worked with the photographs. Many aspects of the organization and content of the collection that weren't immediately apparent were made clear through these contacts.

This knowledge paid off in many ways. I was able to compose a suitable narrative about why the photographs were important: they documented extension activities for the whole state over a period of decades and highlighted a few important areas of the enterprise that scholars, genealogists, and the public might find interesting—specifically housekeeping instruction, agricultural education, publicity, and other activities. In addition, my newly gained knowledge allowed me to anticipate problems and those ever-present exceptions to the rule throughout the project design.

2. Nurture collaborations. I began working with Dilawar Grewal, then imaging manager for the Texas A&M Libraries. Grewal assured me that soon Texas A&M would have "multi-terabyte storage capabilities and commercial grade, parametric, full-text searching and retrieval capabilities" at its disposal. We discussed, in broad terms, the workflow and process by which



Accession Details	
Accession#	63
Reference Number	6300100-
Title	154C
County	Not Available
Photographer	Not Available
Date	1 SA Photo Glass
Title	Not Available
Caption	Not Available
Description	A man on a horse grazing
Available	Yes
Format	154C
Negative	Yes
Document	No

An example of a photograph from the collection and how the information appears in the database,

the collection could be digitized and indexed. I did not pretend to understand the details, but I used this information to design the project workflow in the grant proposal.

Project design

Once TexShare approved the broad outline I had created, I realized a more concrete project design would be necessary. I put a workflow in place immediately, since the deadline for completing the project was an ambitious 13 months away.

3 Be flexible. Again I consulted Grewal, who was in the process of establishing the Texas A&M University Digital Library. The software and hardware for this entity were not actually in place yet, but I had a large, fragile, and unprocessed, collection on my hands. I decided to spend some of the grant money on preservation supplies and hiring the first student worker. Instead of digitizing, she was put to work sleeving and foldering photographs, assigning unique item numbers to each based on the box number in which the photographs had arrived at the archives, initially. I thought this would get work underway for a few months, and that this preservation processing would eventually be carried out concurrently with the digitization.

To say that Grewal faced innumerable challenges in procuring hardware and software would be an understatement, and I am sure it comes as no surprise to anyone who has attempted to start up a similar facility. I will simply advise anyone planning on purchasing and setting up a state-of-the-art imaging center to remain flexible and plan ahead,

Faced with a deadline and no usable computers, I got creative. Preservation processing continued while a temporary Microsoft Access database was set up (with the help of the staff of the nascent Digital Library) for entry of descriptive records. The database was designed to allow input of all the information fields that appeared in the photographic files. The cardstock that the photographs were mounted often listed a negative number, photographer, county, date, or caption. Since some photographs did not have captions, or had captions that did not accurately describe the contents of the image, an additional field for a descriptive caption was added.

In retrospect, the design of the database was remarkably workable; although we encountered several problems that did not fit the "ideal" model, we were able to modify the data to record all the necessary information,

The database

Creating the database quickly revealed itself to be the most time-consuming part of the process. Of course, there was the limitation of typing speed, even though we had designed the data entry form to be easily navigable with minimal reliance on mousing. Students unfamiliar with the vocabulary of agriculture (or more specifically, the agriculture of mid-century America) composed some interesting descriptions of machinery, plants, and animals.

In fact, training and supervising the summary writing was a major concern because everyone had a different idea of how much information was enough and how to best compose that information. In fact, we were limited to a certain field length because of the setup of the database. This was only a problem a few times because some photos had pages and pages of typed "captions." For the most part, this limit forced the students to think of succinct ways to represent an image's content in text and will hopefully be helpful to users.

My goat, alter I had scheduled four students for the project, was to have one student working on data entry from 8 a.m. to 5 p.m., Monday through Friday (office hours of the Digital Library). Additional students worked on preserving the material, then the actual scanning, and finally, problem solving and cleanup.

The scanning phase

A number of unrelated factors kept the Texas A&M Digital Library from having its arsenal of scanning equipment up and running on the initial timetable. Instead, three Macintosh G3s were used, running first the software resident on the flatbed scanners and then on Adobe Photoshop, often all three at the same time.

Where negatives were present, these were scanned; otherwise the actual prints were scanned, cropped, and just barely cleaned up. Because we could not create digital watermarks, we decided to mount monitor-level images and urge people to contact the Gushing Library to obtain high-quality (digital or print) copies. Therefore, very little image enhancement was done. Despite these issues, the scanning phase was completed well ahead of schedule, and students began working on other projects while the database was being completed. Cleanup began, but it was complicated by



A student worker in the Digital Library manipulates an image from the Texas Agricultural Extension Service collection.

system differences between the (PC-based) database and the (Apple-based) stored images. All images were backed up on ZIP disks, and the unique numbers (which had been used as filenames) were sorted and eyeballed for any obvious errors.

Later, two students working together would call up a database record, check it for mistakes or for problems with the description, then consult the accompanying image; many problems were discovered this way. Some images had been very poorly scanned or over- or under-adjusted. Often there was a database record for an item but no scan or vice versa. These problems had to be resolved by pulling the files in question (which had been returned to their permanent home in the Gushing Library, a few buildings away), often re-scanning and re-keying data.

The Web phase

During cleanup, I began working with Digital Library staff to plan the Web interface for the database. Again, this was a trial-and-error process. While I had designed basic personal Web pages for years, I had no idea how to design a graphics-intensive site that would link to a database mounted on a server. I sometimes resorted to sketching out on paper how I wanted results displays to look. There was a high learning curve for the students, the Digital Library staff, and myself in the project. I would advise other librarians who anticipate a similar project to hire students familiar with digitization and photographic manipulation, if possible. The graduate assistant working for the Digital Library also had to learn database and Web site design from scratch, so there was sometimes difficulty in knowing just what we could do. Still, given the time constraints of the process, I was very fortunate to have worked with an enthusiastic and competent group of people. I certainly enjoyed the learning process, however, and the student assistants who went on to work on other projects with the Digital Library clearly learned their lesson.

Conclusion

If I can do this, anyone can. Certainly having a grant to hire students and a graduate assistant was a major factor in the success of this project, as well having access to the knowledge of others. However, I believe it was a firm understanding of the collection and the ability to think things through that really made this project work.

Regardless of the environment, it's likely that there is someone knowledgeable around to help. As the project manager for this grant, I relied heavily on others for technical expertise and troubleshooting. I would encourage anyone attempting such a project not to reinvent the wheel. Chances are good that there is someone around to help you.

I am very grateful to TexShare for their financial assistance, as well as to Dilawar Grewal and the staff of the Texas A&M University Digital Library for technical assistance throughout the project.

Note

1. The database can be accessed at <http://dl.tamu.edu/aggiana/collections/texshare/home.html>.