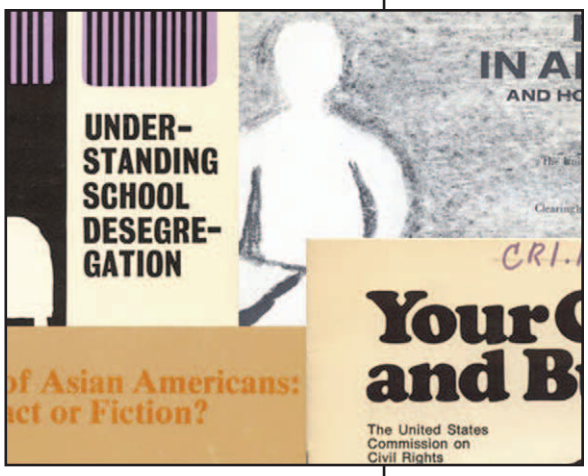

Graphic courtesy of Thurgood Marshall Law Library

Building a Digital Collection

The Making of *Historical Publications of the United States Commission on Civil Rights*

by Bill Sleeman

The Career Development Task Force, which coordinates the Desktop Learning Opportunity Series, encourages members to explore professional development offerings and opportunities at the local, regional, and national levels. We welcome your comments and article suggestions. Please contact Phyllis Marion at 619/525-1429 or pmarion@cwsl.edu.

AALL PROFESSIONAL DEVELOPMENT

To keep up in today's world, law libraries must identify fresh approaches to deliver client-centered services. One library reached out by developing electronic versions of a select collection.

In 2000, the Thurgood Marshall Law Library of the University of Maryland School of Law decided to create digital access to a small group of government documents produced by the United States Commission on Civil Rights (USCCR). After evaluating the collection and reviewing the strategic plan of its parent institution—commitment to teaching civil rights law—the library staff decided it was a good fit and a challenge worth tackling.

The staff found that hard copies of commission publications were underutilized—particularly the microfiche items. With improved access to the documents, students would better understand the history of federal and state efforts regarding civil rights initiatives. Additionally, the Thurgood Marshall Law Library has a rich collection of material, providing a range of unique resources from which to choose. Finally, being federal documents, most of the items were free of copyright issues.

Following is a personal account of what it took to reach this goal, from inception to completion.

Getting Started In-House

As the librarian in charge of getting the project underway, I had to become familiar with the state of the scanning community. Two resources quickly became my bible and hymnal: the *NEDCC Handbook for Digital Projects: A Management Tool for Preservation and Access,* edited by Maxine K. Sitts, and the January/February 2001 issue of *Library Technology Reports,* which is devoted to digital imaging technology. Although now somewhat dated, *Library Technology Reports* proved especially valuable in getting up to speed on then-current standards and tools.

The *NEDCC Handbook* offers a very good outline of how a digital project should commence, and some of the steps that were incorporated into our planning are included below.

1. Define clear boundaries for a digital conversion project.
2. Brainstorm in non-technical terms the desired outcomes for the source materials and functional requirements of the scanned versions.
3. Justify why digital, rather than analog, reproduction is necessary.
4. Project a lifespan for the digital reproductions.
5. Prepare a plan and budget (time and money) for the project.
6. Write documentation for the project.

And most importantly:

7. Learn from your mistakes.

AALL Professional Development Program Competencies of Law Librarianship

1 Core Competencies

Specialized Competencies
2 Library Management
3 **Reference, Research, and Client Services**
4 **Information Technology**
5 **Collection Care and Management**
6 Teaching

This article addresses the AALL Specialized Competencies on Reference, Research, and Client Services; Information Technology; and Collection Care and Management. The complete AALL statement of Competencies of Law Librarianship is online at www.aallnet.org/prodev/competencies.asp.

Step One: Form a Project Team

To start off, we formed a project team that consisted of a cross section of library staff; user services, technical services, and the Technology Assisted Learning Center staff all took part in brainstorming and boundary setting. We met several times and eventually decided that we would scan only documents in our collection and initially would not make an effort to acquire other materials to scan.

Since we would generally work with recent publications, our project would focus on improved access, not preservation (although any digitization project offers a component of collection protection in reduced handling of the originals). The final digital versions would replicate the originals as closely as possible.

Each converted document would be searchable by keyword and generally accessible by date, title, and SuDoc number. We also decided that each document scanned would have a MARC 856 tag added to the bibliographic record in our catalog,

and each bibliographic record would, if necessary, receive upgraded cataloging.

Step Two: Tackle Technology
Our next step was to work out our conversion standards. After reviewing the literature, we selected 400 dots per inch (dpi) as the resolution rate for our digital conversion. This put our scanning in the middle of the various ranges and norms discussed in the literature.

JSTOR uses 600 dpi, which is often quoted as an ideal rate for archival purposes, while several articles suggested scanning text-only documents at 75 to 100 dpi. The literature also suggested that text characters tend to break up at higher resolutions, and the higher the dpi, the larger the files.

Recently the Law Library Microfiche Consortium (LLMC), a major producer of law-related microfiche, decided to convert its fiche products to an electronic form. After extensive tests discussed in a mailing to AALL members, LLMC elected to scan at 300 dpi, based on both file size and presentation needs. This decision by LLMC validated our decision to scan at 400 dpi and suggested that, in fact, we could have used a lower dpi and still produced a quality product.

Next we determined the output. We knew that we wanted a product that would be an accurate copy of the original. Research of other digital conversion projects led us to Adobe's PDF as the best means to meet our goal. Adobe PDF, although proprietary, remains the method of choice within the government documents community. Given that the federal government and many state governments continue to produce original electronic documents in PDF, as well as converting paper and fiche to this format, we remain comfortable that our decision to provide public access in PDF was appropriate. We have recently begun experimenting with both an HTML and an XML presentation.

Step Three: Start Scanning
We first attempted to scan the documents in-house with mixed results. Small documents, such as the 1972 publication *Your Child and Busing,* were easily scanned and manipulated, but larger documents like the 1971 *Voting Rights Act: Summary and Text* proved to be difficult and time consuming.

Some of the challenges we faced included an inability to effectively manipulate large files with multiple images. The conversion process took a great deal of time; the auto feeds on our scanners would pull inconsistently, and we had to do large documents by hand. De-skewing consistently across individual images in a file was a problem. Also, adjusting the image brightness and clarity across files so that each file or document would appear exactly the same was extremely labor intensive and nearly impossible with our in-house equipment.

From a practical perspective, staff found it difficult to take apart large books. It is generally no problem to tear a small document apart, but to do so—while keeping the margins even and pages intact—without a commercial bindery instrument (or a table saw, as was suggested by one staff person), was problematic. Since getting the book disassembled in such a way that it could be rebound at the conclusion of the scanning process was a priority for us, this was no small consideration

Going Outside for Help

As the *NEDCC Handbook* suggested, "learn from your mistakes." We quickly learned that we could not do the volume and quality of scanning we had envisioned on in-house equipment with our staff members who were still doing their regular jobs. So, I called a colleague at the University of North Texas who had extensive digital conversion experience.

Important tip: Know when to ask for help and don't be afraid to do so. Throughout this project, vendors and experienced colleagues willingly shared their knowledge and expertise with us.

Step Four: Review Project Plan
The next task in our project, now called *Historical Publications of the United States Commission on Civil Rights* (www.law.umaryland.edu/edocs/usccr/html%20files/usccrhp.asp), was to review our initial research, focusing on how institutions utilize commercial digitization firms. We knew that this stage would likely involve money, of which we had very little, so we wanted to get the best possible plan in place.

Our core consideration was deciding what to convert. After several meetings and discussions (including some communication directly with the Commission on Civil Rights to ascertain their priorities in converting older documents to electronic form), we decided to convert the fiche versions of the USCCR publications.

Microfiche, while still one of the best long-term storage mediums, is everyone's least favorite tool. We were fairly confident that converting this body of material could only improve access to the documents.

Points to Ponder When Creating a Digital Collection

1. **Be realistic about the skill sets available among project team members.** While not every member of the team can, or should, be expected to perform each task in a technology-driven project, it is important that everyone involved have a basic understanding of the process. It is the responsibility of the project manager to ensure that everyone on the team is comfortable with the technology.
2. **Successful project management requires that all participants are able to contribute freely to the project planning and implementation.** Everyone is responsible for being open to new ideas and accepting criticism. The project manager sets the tone for the team.
3. **A successful project requires flexibility.** As the *NEDCC Handbook* recommends, "learn from your mistakes," and be ready and able to shift gears as the project progresses.
4. **Manage time effectively.** A large-scale digitization project can take on a life of its own. The project manager should plan out ahead of time how to best employ the team when personal schedules and project schedules begin to conflict.

Step Five: Pick a Vendor
We decided to approach OCLC's Preservation Services for help. Working with OCLC and its digitization arm was especially beneficial. Its staff is extremely knowledgeable and patient, and they worked closely with us to develop our experiences scanning in-house to create technical benchmarks that we could employ both with OCLC and with other vendors later in the project.

Preservation Services also recommended that we reconsider our proposed scanning standards; in particular it suggested that we

AALL recognizes major support from BNA, Inc. for the Professional Development Program.

scan in black and white instead of grayscale. Normally grayscale is used with non-color scanning to assure that depth and distinction are retained in items with multiple shadings. Since the bulk of our proposed documents were straight text, pure black and white scanning was cheaper, just as clear, and resulted in smaller files. Since scanning at black and white, we have not observed any loss of data or degradation of the image when expanded out on the PDF viewer.

Working with OCLC also provided the project team with a lesson in defining terms. Our final result with OCLC, while significantly better than our in-house attempt, wasn't exactly what we expected. We had planned for one document with a searchable PDF behind it. OCLC understood us to have asked for two separate PDF files: a clean presentation copy and a dirty (or unedited) OCR'd file. The contract language was ambiguous enough to go either way. After a meeting of the project team and help from OCLC, we all agreed that we could make the two files work and would still have a sufficiently large and useful test bed.

A final lesson came as we went back and forth with OCLC and it scheduled our work around its other projects: this project was going to take a great deal longer than we had anticipated. With our law library's move to a new building looming on the horizon, the amount of staff time available for the project was soon to become an issue.

Step Six: Find Additional Scanning Help

After converting the microfiche, we still hoped to move some paper titles to electronic form. The development of a course at the law school on voting rights provided us with an opportunity to recommend that several early commission *Voting Rights Act* publications be converted. This suggestion was approved by the assistant dean for library services, and she provided us with additional seed money to carry out the plan.

On the suggestion of a colleague, we made contact with Northern Micrographics and arranged to have them convert several reports and hearings, based on the specifications developed with OCLC.

As we wanted to get as many items scanned as possible with our remaining money, the project team elected to have Northern Micrographics produce only the basic files, while library staff would work on enhancing the PDF files (bookmarks, thumbnails, subjects, etc.). It was an exciting moment in the project to see for the first time that what we had envisioned for both the presentation and access side of the conversion was possible. Unfortunately, we had completely exhausted the money that was provided for us to prove our concept.

We now needed to either find the funds to hire a commercial firm or find a commercial firm that wanted to do the work for free.

Resources for Digitizing

"The Circle of Life: Managing a Law Library Web Site Redesign Project." Bonnie Shucha. *Law Library Journal,* v. 95, no.1 (Winter, 2003), p.47-67.

D-Lib Magazine. www.dlib.org/.

Electronic text center. University of Virginia Library. http://etext.lib.virginia.edu/.

A Framework of Guidance for Building Good Digital Collections. Institute of Museum and Library Services. www.imls.gov/pubs/forumframework.htm.

Library Technology Reports (January/February, 2001).

NDLP Project Planning Checklist. http://memory.loc.gov/ammem/techdocs/prjplan.html.

NEDCC Handbook for Digital Projects: A Management Tool for Preservation and Access, Maxine K. Sitts, Editor. Andover, Ma.: Northeast Document Conservation Center, 2000.

"Why Digitize: Principles in Planning and Managing a Successful Digitization Project." Susan Jephcott. *The New Review of Academic Librarianship*, v.4 (1998), p.39-52.

"Why Web Projects Fail." Darlene Fichter. *Online*, v.27, no.4 (July/August, 2003), p.43-45.

Site Design

Providing access to the materials on the site is the key to its success. We had to identify the best structure for doing this at the same time that we were honing our digital conversion and project management skills.

Step Seven: Determine Site's Purpose

Using Dreamweaver, we designed the site to be a simple discovery tool rather than a true database with access via broad subjects, title, date, and SuDoc number. Two of the project staff with Web authoring skills and extensive public service experience collaborated to help refine the site.

We also wanted resources to be available outside of the site via the library catalog and Internet search engines. After considering server requirements, examining the literature, and communicating regularly with colleagues in the government documents community, we decided that the project called for two approaches.

First, library project team members worked to enhance our local bibliographic records for each title we scanned with a MARC 856 field and appropriate subfields. We also discovered that many titles needed re-cataloging as older titles were done under earlier cataloging standards.

For the second approach, we opted to explore the use of basic Dublin Core resource tags to create a set of metadata files for several of the converted documents. The tag content, some of which was pulled from the MARC fields, was based on the specifications available from the Library of Congress' Dublin Core site. Our goal was to use this metadata to aid potential users in other non-library and non-MARC environments to more readily find the site and use the resources.

Step Eight: Team Up with Vendor and Everybody Wins

At the American Library Association midwinter meeting in 2002, I met a representative of MarcLink, the library retrocon company. MarcLink had been using digital technology to scan catalog cards when converting libraries to automated catalogs and was now interested in applying that technology to move into the document conversion business. We discussed the Civil Rights Commission project, and MarcLink agreed that our project was the type of initiative that they were looking for.

Project staff met with representatives of MarcLink several times in the first few months of 2002, and together we agreed on the outline of a program whereby we would prepare a contract and standards. In turn, they would provide complete service for upwards of 20 titles in order to gain experience in ramping up a commercial scanning facility.

This proved to be an ideal situation. Access Imagery, as the company came to be called, had very helpful and knowledgeable staff on the digital side but had limited experience developing a product of this type for the library market. Working together we were able to develop detailed specifications for how the documents would be scanned, the PDF files and DC metadata prepared, and the material delivered.

But the process was not one-sided. We were able to advise Access Imagery on current standards for preservation metadata, we reviewed and recommended changes to how it presented its final product package, and we shared our research on image types employed by archives and libraries. We also advised the company to purchase a bindery-type guillotine for cutting spines. MARC Link and Access Imagery have recently merged into one organization now known as Backstage Library Works.

Future Plans

Access Imagery currently is converting another 30 documents for us—this time at cost with money provided by the library.

We have experimented with harvesting data from the official USCCR site. Many of the commission's electronic publications are not included as part of the federal depository program—they are fugitive documents—and their long-term access is not guaranteed.

Since July of 2002, project team members have visited, downloaded, and reformatted documents released by the commission for inclusion on our site. The commission tends to put the files up in a bare-bones state, so in order to make the material useful, bookmarks, summaries, thumbnails, and metadata all need to be added. Unfortunately this makes harvesting the publications from the USCCR site just as labor intensive as scanning in-house, so we may not continue to pursue this.

Step Nine: Put Systems in Place

Our next step is to systematize the PDF and site creation steps—we plan to write a manual—so that the project is not dependent on a few staffers to continue.

> **"We developed a model for digital projects that is now employed on new projects within the library and law school."**

Another goal is to work with our law school information technology group to create a reliable methodology for testing usage. Depending upon the results of this effort, we will make a decision about moving forward with efforts to fund the conversion of the remaining 1,000-plus paper titles in our collection.

During the past three years, staff at The Thurgood Marshall Law Library has worked hard to bring this site along. We developed a model for digital projects that is now employed on new projects within the library and law school. We have learned a great deal about project management and digital conversion and have demonstrated that it is possible for a small institution with minimal financial outlay to produce a research tool that benefits the legal and academic community.

Bill Sleeman *(wsleeman@umaryland.edu) is information services librarian at the University of Maryland at Baltimore Thurgood Marshall Law Library.*