Digitization in the Real World
Lessons Learned from Small and Medium-Sized Digitization Projects

Edited by
Kwong Bor Ng & Jason Kucsma

Metropolitan New York Library Council
Building the ALBA Digital Library

Jill Anititto (Archivist)

Abstract
This chapter serves as a case study of how a professional digital library can be successfully built with a small staff and budget. It discusses the planning and experiments with beta versions of the Abraham Lincoln Brigade Archives (ALBA) Digital Library, the final version of which is available on ALBA’s website, through Metropolitan New York Library Council’s (METRO) Digital Metro New York program, and OCLC WorldCat. The sensitive issues of digitizing another institution’s collection while maintaining ownership of the final product are also explored.

Keywords: Copyright, CONTENTdm, Database, Digital archive, Electronic classroom, Educational resources, Funding, Indexing, Planning, Ownership issues, Small budgets

Introduction
In 1979, recognizing the vital importance of their radical history, and the need to collect writings, letters, photographs, oral histories and artifacts that would preserve their story, the Veterans of the Abraham Lincoln Brigade, the American volunteers who fought with Republican forces against Generalissimo Francisco Franco during the Spanish Civil War (1936-39), formed the Abraham Lincoln Brigade Archives (ALBA). Today ALBA lends its name to a major archive at New York University's (NYU) Tamiment Library and independently supports cultural and educational activities related to the war. The ALBA
collections are the most requested at the Tamiment Library. ALBA’s relationship with NYU is non-traditional, which makes the ALBA Digital Library unique. NYU owns and maintains the Archives and hosts many of ALBA’s programs, yet the ALBA name gives a different impression.

This chapter explores the challenges of digitizing a collection that is owned by another institution and how to overcome working with a limited budget, dated technology, and minimal staff to produce a professional digital resource. The ALBA Digital Library (Abraham Lincoln Brigade Archives, 2008) evolved from an ineffective form-based website tool to a fully indexed resource using CONTENTdm in a matter of months.

### ALBA Goes Digital

Until a full-time Executive Director was hired in 2007, ALBA was initially run by a group of volunteers followed by a string of part-time administrators. The ALBA Board of Governors hired me as the Assistant Director in May 2008 just as they began reconsidering their outdated website, a major step for the small organization.

By June 2008 the Executive Director had resigned, leaving me as ALBA’s only employee. It provided me with a great opportunity to help redesign the website alongside a subcommittee of the Board of Governors. The Web Committee, consisting of four history professors living across the United States, set a timeline of four months for the site’s overhaul, with an anticipated launch date of early October. The Board wanted to stake a claim as the premier electronic resource for information on the Spanish Civil War.

### The Impetus for a Digital Library

In July 2008 ALBA launched its first annual ALBA Teachers Summer Institute at NYU. The Institute hosts teachers from New York City public schools and exposes them to the history, art, and politics of the Spanish Civil War. Part of this immersion includes a trip to NYU’s Tamiment Library to view the Brigade’s Archives where teachers
receive an overview by the collection’s archivists as well as professors from various departments at NYU.

After this initial success, ALBA decided to sponsor two more Summer Institutes (as well as year-round professional development seminars) in Tampa and San Francisco in 2009. The new settings precluded a site visit to the Archives so we needed to find a way to bring the Archives to the teachers.

The solution to this dilemma was the creation of a digital library or archive to be launched in conjunction with the new website. The website redesign was going to make ample use of ALBA’s collection at NYU’s Tamiment Library anyway, slowly integrating digital collections of letters and photos, and eventually including video clips and oral histories. These primary resources would strengthen existing lesson plans and other educational modules once the site was completed and allow for a more dynamic classroom experience. Since the site was already being designed, we thought it would be a great opportunity to have a collections database created for ALBA by our website designers.

**Defining the Digital Library**

In my experience archives or library staff working in conjunction with a representative from the information technology department does most digital library planning. At ALBA, I was the only staff member and every decision required consensus of the Web Committee, busy professionals with jobs and other responsibilities. Every time an issue arose it could take nearly a week to be resolved. This is a problem that small museums with very active boards will be familiar with, particularly when board members have little time, experience or interest in the digital project at hand.

We quickly ran into a problem of defining the digital library; the Web Committee believed everything created by ALBA was archival. I was in favor of a more traditional, primary source, collections-based digital library with a thesaurus and cataloging guidelines. My idea for ALBA’s digital library was to create a system that would recall only primary sources from the war itself. The digital library would be its
own entity containing items that are separate from all other files posted to the website.

After several weeks of negotiations via email and conference calls, the Web Committee decided the best option would be to include all online documents in the digital library. The Web Committee charged the website designers with creating a web-based, simple form-based recall system as part of their contract.

The Beta Versions

**Albita**

Named *Albita* (or ‘little ALBA’) this resource was first launched in November 2008 along with the debut of the new website. It included the organization’s quarterly newsletter, book reviews, and transcribed letters, not primary archival sources. Uploaded items were listed in random order rather than alphabetically or by date. In the end the design did not conform to the standards of information professionals; *Albita* had become a “junk drawer” of every single file uploaded to the website.

**Document Library**

Even renaming it and reconfiguring the display, the *Document Library* was not robust enough. To recall specific items, the user had to consciously use certain keyword strings, e.g. “George Watt Prize Winner,” when retrieving items. It did not include a thesaurus and the plain-text description field did not allow for paragraph breaks. These descriptions were displayed as a solid block of text.

**Media Library**

Simultaneously added to the site was the *Media Library*. This database appears three pages deep within our Resources tab. It is the only way to access images (from archival photographs to logos of partner organizations) without using the ALBA Digital Library or slogging through the results from a general site search. Unfortunately because it is not linked to any other page, it is not often accessed. The
plain-text description field is displayed as a block of text and does not allow indexing.

Once the Web Committee saw Albita, the Document Library, and the Media Library in action the problems became apparent. We considered going back to the drawing board with our website designers, to create a thesaurus or a more sophisticated database system. Unfortunately, our original contract did not cover building a true digital library; the cost would have been prohibitively expensive.

I explained the possibilities of a professional digital library: adding our records to a consortium’s collection, making them widely available through WorldCat, including them in an NYU catalog at the item level. It was difficult to convince the Web Committee that we still had affordable, professional options that would blend seamlessly into the new website.

Building a Better Mousetrap

As the only person on staff who could design and implement the digital library, I had to balance time spent on digitization with my other duties, including grant writing, public programming, and administrative issues.

The first problem for ALBA to overcome was the lack of any type of digitization equipment. The nature of ALBA’s work and budget did not necessitate having a full flatbed scanner or laptop on hand.

Another major concern for ALBA was the high cost of software. A rough estimate for a full software package ran to the thousands of dollars. We didn’t intend to scan the entire collection, just highlights; for our purposes what we needed was something that would allow us to scan part of the collection but offer maximum exposure.

In January 2009 I became aware of the Metropolitan New York Library Council’s (METRO) Digital New York program. NYU’s status as an institutional member of METRO allowed ALBA to benefit from the digitization program, a partnership between METRO and OCLC and CONTENTdm. As a separate organization, ALBA's much smaller annual budget resulted in a nominal $200 annual fee, paid to METRO
for access to the CONTENTdm desktop module and the upload of 500 discrete items. With software issues out of the way, we could focus on permissions and access to the collection.

Since ALBA’s intention was to work with a collection that is owned by another large institution, we had to be very thorough and diplomatic about the project. We presented a sample record that would name NYU and Tamiment Library as the copyright holders, as well as instructions on obtaining permission to use the images. We assured NYU that the digital library pages on ALBA’s website would include the same information. A distinction is made on both the website and in each record that the ALBA Digital Library is published by ALBA while the copyright is held by NYU. After several weeks of negotiation, we were free to move ahead with the project.

With software and permissions settled we were left with the issue of hardware. ALBA Board Member and NYU Professor of Spanish and Portuguese James Fernandez offered his laptop and Epson flatbed scanner for the duration of the project. After several weeks of further negotiation in order to gain access to the collections, Michael Nash, ALBA Board Member and Director of Tamiment Library, allowed us to scan the items ourselves, free of charge.

**Selection, Policies, and Standards**

In general there is a lack of digitized archival resources available on the Spanish Civil War. The closest to any kind of digital information on the subject is through Spartacus Educational (Simkin, 1997), an online British encyclopedia dedicated to educating students on history, with a significant amount of information on progressive history. But even this site is lacking in primary source materials.

Besides the Abraham Lincoln Brigade Archives, there are other Spanish Civil War archives in the United States: the Southworth Collection at the University of California at San Diego (see University of California, San Diego, n.d.) and the Spanish Civil War Collection at the University of Illinois – Urbana Champaign (see University of Illinois at Urbana-Champaign, n.d.) Still, the Abraham Lincoln
Brigade Archives at NYU is the largest collection of American volunteers' archives in the US.

Our goal was to put forth a curated collection of the most compelling and historically significant items in the Archives; the original plan was to digitize 500 items, but the realities of time and staff restraints reduced this number to an initial 150 items. These items included postcards, letters, newspaper articles, a multi-lingual newsletter, and a telegram from Ernest Hemingway.

Dr. Fernandez performed the selection of four collections based on the following criteria: size (only one manuscript box each), condition (stable enough for handling and scanning), and variety (each collection included letters from a variety of people). These same collections had been used in the Teachers Summer Institute and they proved to be popular with the teachers.

When it came to setting the scanning standards I turned to the University of Wisconsin-Milwaukee Libraries’ Digital Collections (see University of Wisconsin-Milwaukee, 2006) pages, namely the *Transportation Around the World, 1911-1993* collection that I worked on as a graduate student there in 2003. The collection was built using CONTENTdm, then in its infancy. I followed UWM’s digitization standards and policies as well as their formula for long-term archival storage as a basis for the ALBA Digital Library.

**Digitization Standards**

All items were scanned in full-color at a resolution of 600 dpi using an Epson flatbed scanner. The items were saved as the highest quality TIFF files and stored on an external hard drive. We refer to these TIFF files as the Archival Images as they are used only to create access images (PDFs or JPEGs) and are otherwise not accessed.

Access Images were created for web delivery, in the form of thumbnails on the ALBA site and for display within the CONTENTdm records, as well as for everyday use. The letters in the collection do not have OCR capabilities nor are the PDFs searchable. These were saved on both DVDs and the external hard drive. Because the METRO contract only covered 500 discrete items, we used Photoshop to stitch
the JPEGs together to create a single PDF file for multiple-page items. Not only did this save room in CONTENTdm, it was also useful to have multiple-page documents combined for reproduction purposes and to better keep track of collections.

All of this information is posted on the Digitization page of the ALBA Digital Library section. Sharing this information shows researchers (and potential donors) that the digital library adheres to archival standards while also helping other librarians and archivists plan their projects.

To access the images it is important to label them properly. Some scanners and scanning programs assign numbers that may or may not be useful. Auto-numbering systems can cause problems if the items are not easily identified by sight (letters, manuscripts). I named the files according to the manuscript collection followed by a number that corresponded to the letter’s order in the folder, followed by a decimal that corresponded to the page number. For example, Lardner.1.4 would be the fourth page of the first letter of the Lardner Collection. The stitched PDF file would read Lardner.1.

**Indexing**

The documents were indexed using Dublin Core metadata (Dublin Core Metadata Initiative, 2010) which are standard in CONTENTdm, including: Creator, Date, Identifier, Type, Source, Description, Format, Coverage, References, Relation, Language, Publisher, Rights, and ALBA Reference Number.

Included in CONTENTdm is a default thesaurus (Thesaurus for Graphic Materials) for the Identifier field. Similarly, a default thesaurus provided the terms for geographic location in the Coverage field. Those fields for which the thesaurus was unique to the Abraham Lincoln Brigade Archives, I built a new thesaurus. These fields were: Type (e.g. text, image), Format (e.g. paper + size in centimeters), Identifier (e.g. envelope, autograph letter signed, typed letter), Publisher, Rights, and Creator. I added new Creators as they came up in each collection; some collections had as many creators as there were letters. In addition, for collections that had an online finding aid at NYU I included a link to that page.
In all, the ALBA Digital Library took four months for 150 items to be scanned, stitched, catalogued, and uploaded. I generally spent about fifteen minutes cataloguing each letter, sometimes longer depending on the length of the document. I worked on the digital library an average of eight hours per week with some weeks going by without any work at all. As collections were completed I added information to the ALBA Digital Library page and continued to redesign the website as necessary.

**Final Product and Reception**

The final product can be found at http://www.alba-valb.org/resources/digital-library. The Teachers Institute Alumni find the resource easy to use and helpful in the classroom, and ALBA even used the scanned collections to create facsimile copies of the Archives for the Tampa and San Francisco seminars.

As of publication, the digital library is available as a discrete collection on Digital Metro New York (http://cdm128401.cdmhost.com/cdm4/search.php), the digital program of the Metropolitan New York Library Council, the New York Heritage Digital Collections (New York Heritage, 2008) website, as well as at the item level on OCLC WorldCat (http://www.worldcat.org).

**Seeing a Digital Project from Beginning to End**

*Planning is Key*

Work backwards and set aside a day or two to really think about what you want to see in the final product. The small team (just myself most of the time) helped keep bureaucracy to a minimum. Remember: it is cheaper to do it right the first time.

*Unofficial market research*

I asked librarian and archivist friends what they liked to see in digital libraries and to give me advice on moving forward. I also reflected on what struck me in online collections, both the positive and the negative.
**Tread Lightly**

Occasionally it was necessary to receive approvals and it was important to remember that when dealing with large institutions there will be politics. This exists everywhere and I learned not to take it personally. Many of these issues were deep-seated and existed long before I joined the organization. I also learned that having a board member installed in a specific department or company does not guarantee easy access or donated materials.

Within my own organization, some board members did not understand the potential of planned, professional digital libraries, or that one person could undertake such a project. Price was also an issue; it was only after Albita and the Document Library failed was CONTENTdm considered and accepted as an amazing deal. In the end, waiting for the Web Committee members to come around in their own time was the best plan of action for this project.

**Push the PR**

As I completed each collection I sent information about the digital library everywhere: from Facebook and Archivists’ Roundtable of Metropolitan New York to ALBA’s listserv, e-news, quarterly newsletter, and fundraising appeals. While this publicity was mainly sent to people within the ALBA network, it also garnered the attention of local archivists and library students interested in doing small digital projects on limited resources.

**References**


