Digitization in the Real World

Lessons Learned from Small and Medium-Sized Digitization Projects

Edited by
Kwong Bor Ng & Jason Kucsma

M Metropolitan New York Library Council

**The views expressed in this book are those of the authors, but not
necessarily those of the publisher.**

Digitizing the Rare Book Collection of the Leo Baeck Institute

Renate Evers (Leo Baeck Institute, New York)

Abstract

The article focuses on typical problems encountered during the establishment of a digitization program for rare books at a small special library – from finding the best solutions for rare book cradles to organizational and technical challenges and pitfalls.

Keywords: Digitization, Judaica Collection, Rare Book Collection, Special Library

The Library of the Leo Baeck Institute: A Medium-sized collection

The Library of the Leo Baeck Institute (LBI) in New York is internationally recognized as the most comprehensive repository for books documenting the history and culture of German-speaking Jewry. Over 80,000 volumes and 900 periodical titles provide important primary and secondary material. Most of the collection deals with central European Jewry during the nineteenth and early twentieth centuries. It also includes material dating back as far as the 16th century and is as current as the Jewish population in Germany today. The focus of our collection is on the diverse culture of German-speaking Jewry, especially in the arts, sciences, literature, philosophy, and religion. The Institute was established in New York in 1955, one of the largest centers of the exiled German-Jewish community. In 2000

the Leo Baeck Institute became part of the Center for Jewish History, an umbrella organization for 5 Jewish research institutions.

The Rare Book Collection of the Leo Baeck Institute

The Rare Book Collection of the Leo Baeck Institute (LBI) consists of ca. 2,000 volumes primarily in the field of German Judaica, dating from the earliest period of printing in the 15th century and continuing through the annihilation of European Jewry under the Third Reich. Rich in rarities ranging from early 16th century writings to Moses Mendelssohn and Heinrich Heine, first editions and dedication copies of works by more recent prominent writers, many of its volumes were salvaged from famous Jewish libraries that were confiscated and dispersed by the Nazis. In many cases these primary sources cannot be found anywhere else in the world. An important part of the rare book collection is a collection of illustrated and art books. Among them are many limited editions of twentieth century artists' portfolios and several illustrated eighteenth century books on Jewish customs. The Institute also has a significant collection of books dealing with "Science of Judaism" (Wissenschaft des Judentums). The emergence and establishment of modern Jewish scholarship in the 19th and 20th century in Germany and Europe became the cornerstone and predecessor of Jewish Studies.

Project: Digitizing the Rare Book Collection of the Leo Baeck Institute:

The goal is to digitize the rare book collection in order to make these exceptional primary sources easily available to researchers on and offsite, and to minimize use of the originals. To begin, the focus is on titles which are in highest demand.

We started a pilot project with a selection of 38 books in 2005 when the Leo Baeck Institute was awarded one of the first grants in the newly established Metro Digitization Grants Program. The experience gained in this project enabled us to continuously enlarge our digital rare book collection. Working closely with scholars in the

field and using feedback from our users in order to identify a meaningful selection, so far we have successfully digitized 167 rare books (about 22,000 pages) from our rare books collection. Among those books are 33 illustrated and art books which were funded by another Metro Grant awarded in 2008. By now we have worked out the technical details and workflows for our digitization program and are identifying new funding sources in order to continue the process. Newly digitized books are continually being added to our digital management system and made available on the Internet.

Future projects: in 2008 the Archives at the Leo Baeck Institute embarked on an interesting new project - "DigiBaeck," which is destined to put all LBI archival holdings onto the World Wide Web in close cooperation with the Internet Archive. In this framework we consider to digitize parts of our book collection as well. Besides, we are collaborating with other libraries and have applied for shared digitization projects.

Web interface and Examples

The digitized books are accessible through the web interface of our digital management system Digitool (*Ex Libris the Bridge to Knowledge, 2010*). The books can be retrieved via a search interface or can be browsed by various categories. Searchable are the bibliographical data and the full-text, most of the books underwent OCR (optical character recognition).

Additionally we reference the digital objects in our online library catalog records by adding the URL in MARC field 856 (see *856 - Electronic Location and Access*, 2003.)

Challenges

This chapter focuses on the challenges faced and solved mainly during the implementation of LBI's digitization program in 2005/2006 as well as during follow-up projects. The problems were typical for projects which engage in new technologies, ranging from technical to organizational hurdles. Some of the difficulties which we encountered were connected to the special medium rare books.

Challenge 1: Outsourcing a Digitizing Project

Choosing the right vendor, developing a set of criteria for finding the right vendor including testing of sample files, negotiating a contract, monitoring and testing the results proved to be a very challenging endeavor.

During the selection process for a digitization vendor, we developed the following set of criteria and asked vendors to submit a proposal with information and price quotes for the following areas:

- Master files (TIFF) 300, 400, 600 ppi in bi-tonal, grayscale and color
- Derivatives (jpeg, jpeg 2000) 300, 400, 600 ppi in bi-tonal, grayscale and color
- Scanning oversized materials, foldouts
- Technical and structural metadata
- File naming
- Cameras used
- Book cradle
- File storage and delivery
- OCR (Optical character recognition)
- Insurance
- Security (handling of books, security of lab area, book tracking system, etc.)
- Shipping and delivery
- Quality control
- Other costs
- References

We provided vendors with samples from our collection and asked them to provide us with scanned images and metadata. The scanned images underwent a quality check, the metadata was checked against standard recommendations. The best thing to do would have been to load the sample files into our digital management system, but at the time of our vendor negotiations our digital management system was not yet fully implemented.

Overall we compared the vendor proposals and the samples very thoroughly, especially the price quotes for different scanning scenarios, the setups of the scanning labs, the quality of the scanned images, and the vendors' ability to produce structural metadata for our digital management system. The determining factors were the book cradle solutions and security features (e.g.: handling of books, security of lab area), the quality of the scanned images, the ability to provide complex metadata as well as the price differences.

During the negotiation process we encountered variations of important technical details (e.g. file types, file naming, cameras used, technical and structural metadata, OCR) which helped us to make decisions which we then incorporated in the final contract. The contract spelled out our decisions in the listed areas. We learned that it is very advisable that the contract describes as many technical details as possible.

Challenge 2: Medium Rare Bound Books from Previous Centuries

One of the biggest challenge to digitize rare books from previous centuries is the fact that flatbed scanning is usually not an option since bindings cannot be removed for the scanning process. Rare books in our collection are usually treated as artifacts, the content is as important as the bindings and physical appearances. Older books are often very tightly bound and cannot be opened 180 degrees, it is very common that they can only be opened 90 to 120 degrees.

Many of the scanning services which we approached were only equipped for flatbed scanning and declined to take on projects with bound books. Commercial solutions explicitly for rare book cradles do not seem to exist. Those few vendors we encountered which also scan bound books usually built or customized their own solutions for rare book cradles. Some used commercial book cradles for modern books which can be opened 180 degrees and added features to make them suitable for books which cannot be opened 180 degrees. Others vendors constructed adjustable boards to hold the books in place without applying pressure during the scanning process.

The Gruss Lipper Digital Laboratory at the Center for Jewish History designed and uses two book cradles, one cradle has a fixed 90 degree angle, the other one a 120 degree angle. The cradles are made of acrylic and covered with black velvet. Velcro straps keep the book in place. The use of 2 "fixed" angle cradles turned out to be a simple, but successful solution.

The different book cradles which we encountered represent different solutions for accommodating books which can only be opened at an angle less than 180 degrees. It is hard to make a general recommendation for a best solution. Depending on the type of books and condition of their bindings some cradles are better or less suited for the task of supporting fragile bindings and spines and of exerting as little pressure on the books as possible. For some books it is possible to use glass plates, but glass plates put too much pressure on most of the books. It is a trade-off - if glass plates cannot be used it very often results in shadows which often cannot be removed through the adjustment of the lightning. Overall those solutions which had a support for both front and back bindings were better suited for our specific collection.

Rare books require additional manual handling time in securing the volume after each page is turned which makes the process more costly than scanning regular books.

Challenge 3: Technical and Structural Metadata for Compound Objects:

Not all scanning vendors are capable to provide the necessary technical and structural metadata for compound objects. During our search for a scanning service in 2005/2006 we realized that digitization vendors came from different backgrounds – photographers, microfilming companies, library services, bookbinders, etc. – and were thus often not aware of or capable of adjusting to emerging library best practices and standards. Many of them were able to deal with single digital objects and equipped to deliver the technical metadata, but not able to provide ready-to-go structural metadata in order to build complex digital objects.

Our digital management system DigiTool (Ex Libris) requires the use of METS (Metadata Encoding & Transmission Standard, 2010) for storing structural metadata along with descriptive, administrative, technical and other kinds of metadata in an XML "wrapper." During our vendor search in 2005/2006 we found out that at that time many vendors were not aware of or capable to provide this type of metadata with the scanned images. If the scanning process is outsourced it is very advisable that the vendor provides structural metadata at the same time, otherwise it becomes a very tedious and time-consuming endeavor to match the digital objects to a structure map afterwards. During the testing phase we asked the vendors to supply us with metadata, preferably METS. During our initial project we could only find one vendor who could provide us with METS files. During the first phase of our project we additionally discovered that we needed a translation code between the METS ALTO version that the vendor could create and the METS ALTO version that our digital management system understood.

Challenge 4: Organizational Parameters & Limitations of Digital Management Systems

We had the "problem" that we had to use a digital management system which was chosen by our umbrella organization and which was not fully implemented at the time when we started our digitization program. The challenge was to bridge shortcomings of the system and shortcomings of what digitization vendors could provide, a typical problem when a technological field has not yet developed mature and commonly accepted standardization guidelines. The first versions of our digital management system did not accept structural METS files in METS ALTO standard, so we had to find additional funding for developing a translation code between METS ALTO (what Vendors were able to produce) and the METS "dialect" that our system understood, which resulted in higher costs.

Organizational set-ups and the choice of the technical system have a great influence on the direction and the outcome of a project. These factors cannot be underestimated and they influence the timeline and the financial framework of a digitization project.

Challenge 5: Keeping up with Evolving Digital standards

One of the challenges we faced in follow-up projects was to keep up with evolving and maturing guidelines for digitization. One of the lessons learned was that it is necessary to spell out technical requirements in detail in a contract and to make sure that they are applied, and not to take them for granted. General formulations can lead to ambiguous interpretations.

Example 1: An area of dispute in one of our project was if it is "allowed" to enhance or sharpen master files during the scanning process.   Many labs follow the NARA guidelines (see National Archives, 2004) and do not apply ANY sharpening or other "enhancements" during capture or to master files.   If errors are detected during the capture, then the camera is calibrated. One of the vendors enhanced the master files instead of calibrating the camera. Our contract read: "Quality of image: Files will be inspected for sharpness, image contract, density and faithfulness to original and improved if necessary."   - that should have only be allowed for derivatives, and not for master files. The master files should be true to the original materials; colors and contrast of the digital files should represent accurately the original pages. Many of the master images had too much contrast (visible in bad histograms) which probably came from sharpening during the scanning.   In that respect our contract was not specific enough, we did not specify that only derivatives could be altered, but not master files.

Example 2: It proved to be necessary in our 2008 project to spell out in the contract to check regularly if the calibration of the camera results in an accurate image.   In our contract we had specified that the images produced should be "faithful to the original" materials. The colors and contrast of the digital files should represent the original pages accurately.

Our first vendor calibrated the camera only once at the beginning of the scanning process to the "default" NARA guidelines. The results were aggravating: the majority of the image files had extensive color bleeding, a severe image quality problem.   Ultimately we had to rescan our books at another digitization facility.

Our contract was sufficient in the sense that we could rightfully claim that the bad image results were a breach of contract. But we had to rescan all our books – something which is definitely not advisable for rare books. A line in the contract about periodic camera calibration could have prevented this dilemma. We should have also done quality checks periodically.

Lesson learned: The contract texts for projects in emerging technological fields are often vague and only give a narrative description of the wanted outcome since various schools of thought exist in terms of technical specifications.

In many cases a narrative formulation in the contract about the desired outcome is the only possible way to address an area of concern, but in some cases it is not sufficient. The more hardcore technical details can be specified in a not ambiguous way the better. It is also very advisable to control the project frequently, to do several test phases, and not to rely on past vendor performances.

Conclusion

The successful implementation of our digitization program was and is based on a combination of persistence, serious initial testing, vendor comparison, ongoing testing of results, questioning of vendors, collecting of technical information, balancing of outside factors, changing of vendors when the outcome did not meet our expectations. The most challenging aspect of our rare book digitization project has been the steep learning curve of getting into the technical intricacies of creating digital content, and balancing outside factors to get the project on its way and moving. The most rewarding aspect has been to make hitherto relatively unaccessible books accessible to a broader audience.

An important lesson learned was that is necessary to keep up with evolving and maturing guidelines for digitization – and to make sure that they are really applied, and not to take them for granted. Even established workflows have to be reevaluated from time to time, since the field changes all the time, and the technical details become more

detailed and sophisticated. Digitizing of rare books is a very manual, labor, and ultimately cost intense work.

References

856 - Electronic Location and Access. (2003). Retrieved May 9, 2010 from http://www.loc.gov/marc/bibliographic/bd856.html

Ex Libris the bridge to knowledge, Overview. (2010). Retrieved March 15, 2010, from http://www.exlibrisgroup.com/category/DigiToolOverview

Metadata Encoding & Transmission Standard. (2010). Retrieved May 9, 2010 from http://www.loc.gov/standards/mets/

National Archives. (2004). *Technical guidelines for digitizing archival materials for electronic access: Creation of production master Files - raster images.* Retrieved May 9, 2010 from http://www.archives.gov/preservation/technical/guidelines.html