2008

Digitizing criminals: Web delivery of a century on the cheap.

Ellen A. Sexton
John Jay College of Criminal Justice

Ellen Belcher Dr.
John Jay College

Digitizing Criminals:
Web Delivery of a Century on the Cheap

Ellen Belcher and Ellen Sexton
Lloyd Sealy Library, John Jay College of Criminal Justice, CUNY

Abstract:
Purpose – This article presents the process, challenges and lessons learned from carrying out a small digital project to create a web resource of unique historic materials related to crime in New York City. All aspects of digital project management are discussed including selection, infrastructure, budgeting, workflow and delivery.
Design/methodology/approach – Experiences from project administration, including management of a combination in-house and outsourced digitization and metadata are discussed. Formation and management of the resulting web resource is explained, which is the product of a creative amalgamation of commercial and open source software. Challenges encountered are presented with suggestions for practical solutions and considerations for future projects.
Findings - This grant funded pilot project presented foreseen and unforeseen problems. Lessons learned and solutions suggesting best approaches for a small-scale digitization project are presented here.
Practical Implications – In this article best practices and suggestions for managing a small digital project are presented, including financial, staffing and technical considerations.
Originality/Value – Unlike many other discussions that focus on management of large institutional projects, this study presents an incremental approach for small-scale digitization projects. Presented here practical uses of available applications for establishing project infrastructure to sustain and continue growth of digital content for small institutions.
Keywords - Digital Libraries, Criminal Justice, Photographs, New York, Trial Transcripts, Academic Libraries, Project Management
Paper type - Technical

The collections of the Lloyd Sealy Library of John Jay College, a small academic library, are internationally renowned for their specialized scholarly value, and are ripe for better access through digitization. But as part of a large under-funded public university, the library has limited monetary resources and overstretched staff resources. With little digital infrastructure, time, or money, the prospect of managing a large digital project was daunting. The solution was to obtain a grant for a small pilot project that would enable the incremental development of a digital library, while offering patrons unprecedented, sustainable and increasing web access to the unique materials.

For almost four decades, the library of John Jay College has concentrated its resources on gathering together remarkable criminal justice collections including unique holdings with a special focus on New York City (Egan, 2007). The library holds, for example, the papers and publications of prisoners and prison wardens; criminals and policemen; scholars of criminal behavior and supporters of crime victims. Together, these historically document the process of crime, its investigation and prosecution. The special collections contain traditional archival

formats as well as unique and unusual objects such as criminal trial transcripts, mug shots & rap sheets.

The catalyst for embarking on this project was the availability of funds for digitization projects and training by METRO, a local library consortium. This program has successfully stimulated smaller libraries in the New York City area to begin digitization efforts by providing grants for pilot projects and support in the form of training workshops. The library applied for and was awarded a "Digital New York" grant amounting to $8,750 in order to support a one-year digital library venture. The proposal was the creation of digital access to portions of three of the most popular special collections. The aim was to begin building a digital library incrementally, easily and cheaply, and to ascertain the feasibility of managing a digital project using existing resources, knowledge base and staff.

Before digitizing materials commenced, some major decisions had to be made. Decisions regarding exactly what to digitize, and then how to digitize the chosen materials. Could a vendor be found to take on the out-sourced portion of this project and specify appropriate digital procedures in a contract? Did the digital project directors have sufficient expertise the time to manage the project? What standards and specifications would be best for both in-house and out-sourced scanning? How would the resulting digital objects be described? What metadata standards should be followed? How would the digital objects and their descriptions be made available to users?

The title of this project, *Crime in New York 1850-1950* was deliberately broad, potentially encompassing large portions of our collections. But the intention of this diminutive project was to test the waters by digitizing a small sample. This project was a starting point - an opportunity to increase the digital knowledge-base amongst the librarians, and begin to establish a digital infrastructure for the library. All the while keeping an eye upon the feasibility of larger grant-funded digital projects in the future. Three achieved goals of this project were to:
1. Assess the staff's ability to work with and maintain a project using open access content management software;
2. Compare staff involvement in the administration of in-house and vendor-supplied digitization and metadata;
3. Understand the overall impact of digital library development on existing library resources.

Content Selection:
The macro decisions of content selection for this pilot project were based on use analysis and potential for effectiveness as digital objects. Materials that were unique and free of copyright and privacy issues were particularly attractive. Physical considerations included finding items robust enough for scanning and which offered immediate visual appeal.

Multi-page materials - such as books, pamphlets or newspapers - result in 'complex digital objects' which require complicated linkages to produce digital surrogates of the original formats. To simplify end processing, single-sheet items were chosen for in-house scanning. Vendor-scanned items needed to result in manageable files. Overall, materials in scholarly demand were

sought which would allow a range of digital formats, technologies and solutions to be assessed and experienced.

For many years it had been hoped to provide better access to and more publicity for the collection of *Trial Transcripts of the County of New York 1883-1927*. This collection consists of 3,200 verbatim transcripts of trials, mostly from the Court of General Sessions of the Peace, which was Manhattan's criminal court at that time (Faber and Rowland 1994). The transcripts document the diverse social, criminal and political history of the city, and have already been used by a growing group of social and criminal historians, genealogists, and graduate students, resulting in the publication of a number of books, articles, and theses.[1]

The transcripts had been acquired by the library in great quantity and poor condition from the Clerk of the New York Supreme Court in 1972.[2] More than a decade later, in 1983, the Library obtained an NEH Preservation and Access Grant, which provided for the arrangement, microfilming and indexing of this collection.[3] The index was created on a mainframe computer database of individual transcript records containing entries for names, charges, dates, and other access points. The resulting index was printed and bound and served as an efficient finding aid to the collection, though only accessible in-house. Unfortunately, the mainframe computer database was not retained. Two copies of 425 reels of 16mm positive microfilm are now held in the Lloyd Sealy Library, and the negatives were deposited with the New York State Library in Albany. Apart from a small sample, the original paper transcripts were discarded at the end of the microfilming project. A major purpose of the *Crime in New York* project is to facilitate access to this unique resource.

Although the transcript microfilm reels are available by interlibrary loan, determining the reels desired required use of the paper index within the library. Digital conversion of the *Index to the Trial Transcripts of the County of New York 1883-1927* seemed essential to provide better access. The *Index* would provide the descriptive metadata access points, and thus the intellectual framework for the digitized transcripts collection, and allow for more controlled searching than free-text searching could provide. As the transcripts typically number hundreds to thousands of pages each, resources were not available to consider allocating metadata to every individual page.

This small grant could not provide for digitizing the entire collection of 425 microfilm reels of transcripts. Criteria were needed to select a representative sample. With no in-house capabilities for scanning microfilm - even large libraries outsource microfilm scanning - a vendor had to be found to complete this part of the project.

The transcripts are a fascinating and unique resource, but as a black and white text-based collection they do not provide immediate visual interest *(figure 1)*. It was also desirable to assess the feasibility of creating digital objects in-house. The project was broadened to digitize photographs and other materials from two other, complementary collections. Along with the transcripts, this material would illustrate and document the criminal, social and political history of New York City and create a robust web-resource.
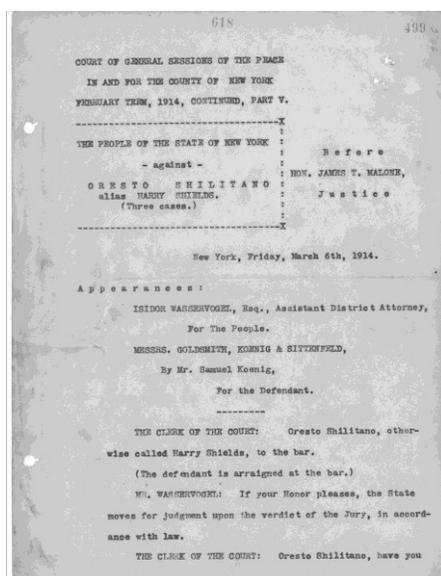
figure 1: Page from the 1915 Murder trial of Oresto Schilitano (or Shilantano), alias Harry Shields, alias 'The Paper Box Kid'. *Digitized from the Trial Transcripts of the County of New York.*
figure 2: Photograph of Schilitano (sic), convicted murderer. *Digitized from the Papers of Lewis Lawes.*

The Papers of Lewis Lawes, Warden of Sing Sing Prison from 1920 to 1941 attract a steady stream of inquiries, and have provided material for a few recent books (Blumenthal, 2004; Goewrey 2005). Sing Sing Prison was infamous for being the site of many state executions, although Lawes was famously against the death penalty. This collection holds nearly 90 mug shots of condemned death row prisoners. An exciting discovery of this project was that the Special Collections also holds the trial transcripts of many of the prisoners photographed.

The prison photography system is described in an article on the condemned prisoner Arthur Waite, a Manhattan dentist convicted of poisoning his in-laws, whose 1916 murder trial transcript is also held by the library. *(figure 2):*

> *"Ossining, N.Y. June 5* – For the first time since reaching Sing Sing Prison, Dr. Arthur Warren Waite, the prisoner, was let out of the death house today. He was taken across the yard to the Rogues Gallery to be photographed. He posed for two pictures, on a front and other a profile view. "I think I am getting thin," Waite said to one of the officers. "the strain is showing on me." He was locked up again after being photographed."
> *New York Times*, 1917.

figure 2: Photo of Arthur Waite, convicted of murder in Manhattan, 1917
figure 3: Photo of Martha Place, convicted of murder in Brooklyn, 1899
Digitized from the Papers of Lewis Lawes.

The Lawes Collection includes mug-shots of the first and last women to be executed at Sing Sing - Martha Place, executed in 1899, and Ruth Snyder in 1928.(*figure 3, 4*) Snyder and her co-conspirator Henry Judd Gray *(figure 5)* were famously fictionalized in the film *Double Indemnity* (Landis, 2006) and were executed on the same day at Sing Sing.

figures 4, 5: Photos of 'Double Indemnity' defendants Ruth Snyder and Henry Judd Gray, convicted of murder in Brooklyn, 1928. *Digitized from the Papers of Lewis Lawes.*

These examples are a few of the many famous and elsewhere documented New York criminals who became the condemned Sing Sing prisoners of these photographs. Perhaps even more

compelling are the nearly 30 mug-shots of unknown prisoners, identified only by a number *(figure 6)*.

figure 6: Mug shot of Sing Sing prisoner number B75789, date unknown.
Digitized from the Papers of Lewis Lawes.

Covering a slightly later time in New York criminal history, a selection from the *The Burton Turkus Papers* was digitized for this project because it offered compelling imagery of Brooklyn syndicated crime at its height. Turkus' accomplishments as Assistant District Attorney and Chief of the Homicide Division of Kings County (Brooklyn) 1940-45 are famously chronicled in the book and film *Murder, Inc.* (Turkus and Feder, 1951). Items from this collection chosen for digitization included mug shots *(figures 7-8)* and rap sheets *(figure 9)* of key figures in the Brooklyn Mafia as well as crime scene investigation photos. The imagery and corresponding metadata complements recently digitized collections of newspaper photographs such as those from the *Brooklyn Eagle* (http://www.brooklynpubliclibrary.org/brooklyncollection/photo-collections.jsp) and the *New York World-Telegram & Sun Newspaper* (http://www.loc.gov-/rr/print/coll/130_nyw.html). It was felt that these photos would prove popular.
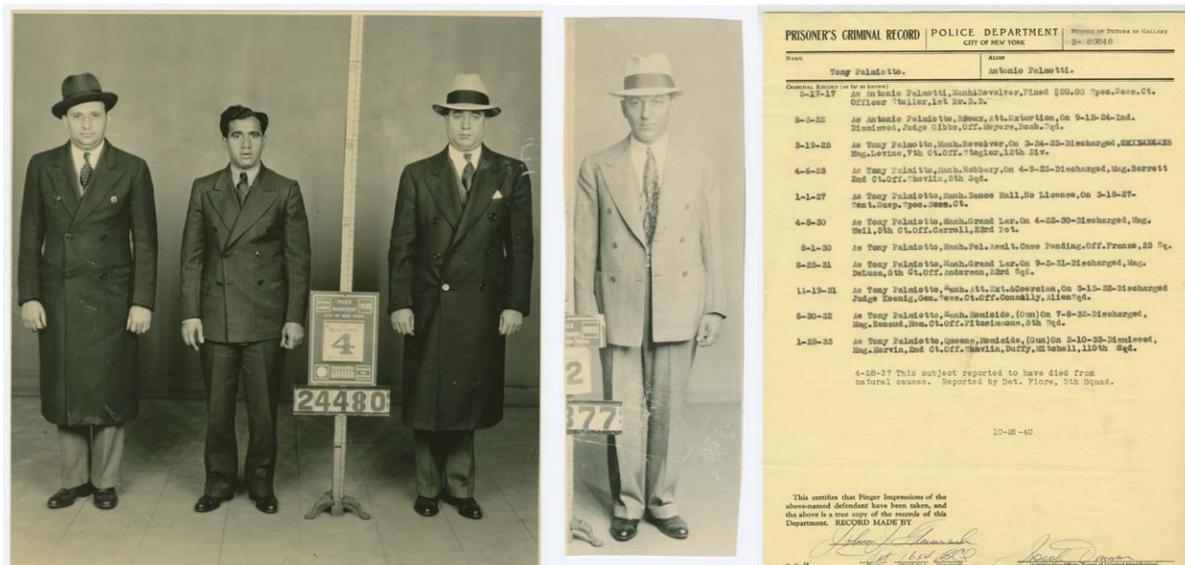
Figure 7: Police identification group photograph of Sam Bernstein, Abe Reles and Joseph Bernstein, Nov. 11, 1939
Figure 8: Police identification photograph of Louis Buchalter, alias Louis Lepke, June 12, 1933
Figure 9: Prisoner's Criminal Record of Tony Palmiotto, alias Antonio Palmotti. 1917-1937.
Digitized from the Burton Turkus Papers

With the digital content identified, and the grant secured, a year long project to create a digital library with the overall theme of a century of crime in New York City commenced.

In-House Digitization
Before the project, the project directors had gained some digital experience by scanning special collections materials requested by patrons, and electronic reserve materials for faculty. The project directors also had library school training and experience in digitization and preservation management. Given this collective experience, it was felt that there was sufficient knowledge to establish digitization specifications and metadata schema following current standards, in-house. During the project year the project directors significantly enhanced their practical knowledge by attending digitization and metadata workshops, and reading available literature on digital project management. This was invaluable while the project was underway, providing a chance for the absorption and immediate application of practical knowledge. The lack of available training opportunities related to content management software in general and the system chosen in particular did thwart the perfecting of the user interface for the digital images. This issue will be discussed later in the article.

Hardware infrastructure
The METRO digitization grant provided funds to purchase a large format good quality scanner with *Silverfast AI* scanning and calibration software. *(table 1).* Also purchased was an IT8 color calibration target to assure good quality digital masters. The library already had updated PC computers loaded with commercial software sufficient for digital object and metadata creation and storage. All workstations are linked to library-controlled and managed servers. This proved essential to the success of the project. Because of this network, work on digitization, databases, metadata and web creation could be carried out in offices or at the reference desk, while all scanning took place in the Special Collections Room. The library has full control over webpage creation and maintenance as well as data back-up.

Scanning specifications
Scanning specifications were based on *Western States Digital Imaging Best Practices Version 1.1*, (http://www.cdpheritage.org/digital/scanning/documents/WSDIBP_v1.pdf). A master 'archival' TIFF (Tagged Image File Format) file was produced at 600dpi from scanning in 48-24 bit color, with a resolution of 3,000 to 6,000 pixels across the original's longer dimension. Even though the originals were 'black and white' photographs, scanning was carried out in color because calibration controls gave the ability to create a digital surrogate as close as possible to the original. The scanned borders were set to be slightly larger than the original documents, allowing the edges of the original document to be recorded. Scanning went pretty quickly, as the material grouped into a few consistent sizes from 11 to 2.75 inches. Files were named, saved to the server, then the master TIFF file opened in *Adobe Photoshop* to derive a web-optimized JPEG with resolution of 150 pixels/inch, with 600 pixels across the longer dimension. Thumbnails were not made, as unlike some content management systems, *Greenstone*, the software chosen to web-deliver the collection, automatically creates them.

File conventions
File-naming conventions are very important for managing all digital object collections. Tracking and identifying digital objects is eased by following a logical file naming convention. Some programs only allow a certain number of characters in file names. In addition, carefully controlled file-naming conventions are useful for sorting like objects together in search and browse results. These conventions help in creating a persistent URL for digital surrogates. For this project a file naming convention was developed that resulted in 'readable' and predictable names, based on information about the collection from which the original artifact came, and indicators of the front, back and related original documents. For instance, bt00001a is an image of the front of the first scanned photograph from the Burton Turkus Papers; bt00001b is the back of the same document; bt0001c and bt0001d are the back and front of documents associated with file bt0001a.

Accomplishments
The scanning conventions were originally designed with a view to instructing a digital project assistant. However, trained and interested candidates for the posted temporary position did not emerge. One project director carried out the scanning, following the conventions, creating a total of 225 digital objects over the course of three intense days. While diverting attention from daily library duties, this provided the assurance that high quality digital masters were made, and that original documents were handled professionally. The specifications have since proved invaluable guidance for infrequent scanning requests from patrons, for publications and website development.

In House Metadata Creation:
Photographs
A modified Dublin Core metadata schema of eleven elements was developed to describe the digital objects created in-house (CDP Metadata Working Group, 2006). A variety of sources were consulted for authority and vocabulary control including OCLC, the library OPAC and the New York Times Historical File. Some original research was carried out to enrich the metadata. 'Subject' elements were derived from Library of Congress Subject Headings, including the Names Authority Index. After images had been scanned, metadata was entered and stored in Microsoft Access and Excel databases. The predictive text capabilities, and copy and paste functions of these programs were useful in maintaining language consistency. Images and metadata were later added to Greenstone.

Digitizing the Transcript Index
Metadata for the Transcripts collections was readily available from the existing *Index to the Trial Transcripts* collection. This *Index* is the product of one of the first computer projects in the library, but since the mainframe data was lost; ironically the first digitization task was to re-key the *Index*. An attempt at scanning the original dot matrix print-out and running the files through OCR software was unsuccessful. Instead student workers typed information from the printed *Index* into a specially designed database, using a Microsoft Access entry form connecting to a MySQL database. Quality control was undertaken by students and librarians, who produced a very useful and dynamic web delivered database. It was fortunate that the original *Index* had

been well thought-out and painstakingly created. Should additional information about individual transcripts come to light, the database can be amended accordingly.[4]

Out-Sourced Digitization: Trial Transcripts
A major cost of this project was contracting with an outside vendor to digitize from microfilm the text of the trial transcripts. It was planned that digitized transcripts in the form of PDF files be linked to the digital *Index* but was uncertain as to whether the library's technological capability to deal with numerous large files existed in-house. Funds to digitize the entire trails collection were not available, so a representative sample was selected.

At a METRO funded digitization expo, the project directors met with several vendors who seemed interested in serving a relatively small microfilm digitization proposal. Representatives of *iArchives* were present, and revealed that they were working on a much larger project with New York Public Library, so would be visiting the city periodically.   iArchives scanned a sample reel, which was reviewed at the Lloyd Sealy Library. After research and consultation, it was determined what type and size of files would be most appropriate, and a contract was drawn up. It specified that iArchives would provide archival TIFFs and searchable PDF files for access on the library website. An external hard drive was purchased to ship these very large files from Utah to the library in New York.

The grant provided sufficient money to pay for the digitization of 12 microfilm reels, containing a total of 150 trial transcripts or approximately 24,000 pages of verbatim criminal trial proceedings. (table 1). The criteria used to select 12 microfilm reels out of 425 included the dates of the trials, so as to have a sample that would represent the entire time period spanned by the collection.   The goal was to develop a dynamic resource that would grow in full text content over a period of years. As expertise grows, more reels will be digitized, either in small increments funded by grants, or by individual researchers. Now that the digital infrastructure exists, a future possibility may be to seek a large grant to digitize all remaining reels.

Web Delivery.
Mounting the photographs: Greenstone.
While the technological and intellectual duties of scanning and metadata creation were tasks with which the project directors were somewhat familiar, implementing a content management system was new. Content management systems are used to create, manage, organize and provide access to digital content. A search was carried out for an available system that was easy to use and affordable, but no easy solutions were found. *ContentDM* was considered - and was available free for two years as part of the grant from METRO - but there were concerns about the licensing costs beyond the initial two years. Instead, it was decided to use *Greenstone* open source digital library software, available for downloading free from the New Zealand Digital Library web site (http://www.nzdl.org).

No one at the library had any experience using *Greenstone,* and only minimal experience with any content management software. As with most 'open source' software, it did not come with 'out of the box' functionality. But published reports of its ease of use were plentiful (for

example, Witten, Bainbridge, Boddie, 2001; Witten, Bainbridge, 2003; Witten, 2004). Setting up and customizing *Greenstone* turned out to be more difficult than expected. It was decided to use grant monies to hire a consultant to set up and customize *Greenstone* according to design specifications, and deliver it on a CD. The library webmaster installed the customized *Greenstone* on a Linux server, onto which the digital objects and metadata were uploaded.

Creating a web interface for the transcripts:
The interface for the transcript texts was created by the library web master. A simple interface presents the users with options to browse or search the *Index* to the transcript collection. A user searching or browsing the *Index* retrieves the relevant record(s). Each record briefly describes the trial, with metadata pulled from the *Index*. The record includes a URL linking to the PDF containing the full transcript text (when available). The PDFs are stored on a server, and have been given permanent URLs.

The web designer designed the web interface to be powered by a MySQL backend database which allows synced off-line corrections and additions to the metadata. As the library becomes more familiar with these systems of open-source and other software, the web interface will be refined, and more data and digital objects will be added as they become available.[4]

The user interface:
A major challenge was to present the user with a seamless transition between two quite distinct web interfaces – Greenstone and the html interface to the *Index* MySQL database. A common design theme was used throughout to emphasize the unity of the collection. Colors used are red, black and white. Text is black on a white background, with red used as contrast. Red navigation buttons have been placed on the top right hand corner of each page, above a red banner running the width of the page. The banner announces the digital library title and includes illustrative thumbnail photographs from the collection.

The digital library has its home on the web, at (http://www.library.jjay.cuny.edu/crimeinny/). From here, users choose to go to the images collection page, or the trials collection page.

The trial transcript page prominently displays entry points to the *Index*. Users have browsing options or can use the search box to search any or all fields of each record. Search results consist of the full record describing each relevant case, with a hyperlink to the full text of each case when available.

The images page has a search box where the user can enter keywords. Photographs are displayed as thumbnails with limited metadata fields shown alongside. Hyperlinks lead to an enlarged photograph, and the complete metadata for each photograph. Metadata follows modified Dublin Core conventions, and provides information about the subject matter, digital image, and digitization specifications.

Unexpected challenges:
Out-Sourced Digitization of Microfilm: More labor intensive than initially planned
Digitizing the transcripts had been expected to be the least time consuming aspect of the project. Once the company was identified and chosen, and the reels selected, it was expected that the work would be complete. This plan, in retrospect, now appears naïve. Precise digitization specifications had to be established before the contract could be written. The authors discovered that they had a lot more to learn about the technology and terminology than originally thought.

PDF files, readable and searchable but still small enough to be delivered to a general user population through the World Wide Web had been requested from the vendor. The images achieved from scanning the 16mm microfilm reels were much more varied in quality, resolution and density than initially expected.[5] Cropping borders turned out to eliminate potentially important handwritten comments. The original transcripts had multiple paginations, a variety of typefaces, and a text on page format unfamiliar to the digitization company and their software. This project was tiny in scale and compensation compared to the vendors other digitization contracts. Compromise on the final product had to be made. In the end, the PDFs are quite readable although not as visually consistent as initially expected. After receiving some quite large initial PDFs, it was requested that instead of the gray scale PDFs initially specified in the contract, the company provide us with black and white PDFs, to reduce the file sizes. Single-paged TIFF masters were retained in case an opportunity comes up to achieve better quality user files in the future. Unpacking and storing the PDFs and TIFFs from the portable hard-drive and linking the PDFs to the *Index*, turned out to be more time consuming and technically demanding than initially thought.

Another option would have been to load the transcript images into Greenstone so they could be searched and viewed a page at a time, but the library did not have the expertise to do this nor the money to hire a consultant to write a load script. Single-page PDFs were received from the vendor (in addition to the multiple page PDFs) and a way may be found in the future to use them.

Photograph metadata:
Information about the subjects of some of the photographs was not available in the archival collections. Additional research was required, for which the *New York Times Historical File* was invaluable.

Rigid adherence to the Standards for Dublin Core (CDP Metadata Working Group, 2006) may have resulted in metadata more comprehensible to librarians than to the public. For example, the YYYY-MM-DD date formula called for in these standards is not intuitive. Similarly, use of the rather cumbersome language of the Library of Congress subject headings has resulted in a vocabulary that the average member of the public is unlikely to use in a search. For example the words 'mug shot' or 'rap sheet' are not accepted terms in LCSH or other thesauri (but these terms did get used in the description and title fields). Many of those depicted in these photographs did not have name authority records, and had multiple names and aliases with variant spellings *(figure 2)*. Establishing names for these individuals took research and guesswork.

Greenstone:
Choosing Greenstone open source software gave a tremendous initial monetary advantage in that it was free. But there was not sufficient expertise in the library to alter the default *Greenstone* appearance, and customize it as desired. The small support staff could not provide the programming and development support to make the system work as envisioned. The absence of hyper-linking on individual terms in *Greenstone* - available in some systems such as *ContentDM* - means that it will be difficult to gather images together by metadata terms. The small grant was only able to provide for a few hours for a consultant to alter the appearance and functionality, but not enough for complete customization of the system. This portion of the project is not considered completely satisfactory. The librarians intend to continue learning about Greenstone and developing sufficient in-house expertise for full control of the system.

True Project Costs:
Finding uninterrupted blocks of time to work on the project was often a real challenge. The work required to manage and finish this project was seriously underestimated. This was particularly surprising given the large outsourcing component of the project. Because a record was not kept of the time spent by the professional, technical and student staff on this project, it is not possible to accurately reflect its true cost. The many tasks and duties undertaken to complete this project are listed for consideration *(table I)*. In the end, the bills for the *Crime in New York 1850-1950* amounted to $256.00 over-budget.

Conclusion:
A robust framework, specifications and procedures for continuing to digitize resources at the Lloyd Sealy library was established. Perhaps most important is that valuable experience had been gained and many lessons learn, which can be applied to future digitization efforts, either grant funded or a part of daily tasks. Recommendations for a similar small digital project are given *(table II)*, but don't all match those for a large project. The authors have demonstrated that it is possible for a small library to put together a digital library on a small budget, and they encourage you to do the same.

Table I: Costs Associated With the *Crime in New York 1850-1950* Digitization Project

TASK	STAFF LEVEL	COSTS ABSORBED BY LIBRARY*	EXPLANATION	COSTS
Selection	Professional	Librarians' time	Criteria developed	*unknown*
Scanning (in house)	Professional	Librarians' time	Flat-bed scanner and calibration target.	$2,570.00
Scanning (outsourced)	Professional	Librarians' time for management of vendor relations Portable hard-drive Shipping costs	*iArchives* microfilm scanning at 20¢ a page, approx. $400.00 a reel	$4,867.00
Metadata (images)	Professional	Librarians' time	Research and composition of metadata	*unknown*
Metadata (transcripts)	Student & Professional	Student staff time & Librarians' time	Existing metadata typed by students and edited	*unknown*
Content Management (*Greenstone*)	Professional	Consultant & Librarians' time	Setting up, customizing and loading of content management software	$1,500.00
Content Management (transcripts)	Clerical & Professional	Webmaster's time & Librarians' time	Designing database, uploading and linking PDFs	*unknown*
Training	Professional	Workshop fees [approx. $1,000] & Librarians' time	Metadata and digital project creation and management workshops	*unknown*

**Not considered:* existing computer hardware and software already in place in the library.

Table II: Recommendations for Planning and Managing a Small Digital Library Project

SELECTION	- Select simple objects for digitization that will result in one or two images and metadata to manage and link. - Pick unique, popular, and compelling objects in your collections. - Choose a project title and scope that is broad enough so that other items in your collection could be added as resources become available. - Select materials of the same size described by similar descriptive terms.
SCANNING	- Choose a short, meaningful, and predictable file naming convention for digital object files. - When drawing up a vendor contract for outsourced projects, allow plenty of time for research, consultation and dialogue. - A few intense days by professionals experienced in standards and procedures may take less time and require less quality control than training others for a small project. - Scan original documents in color. Use scanner calibration tools. - Scan microfilm in grey scale but use smaller black and white PDFs as the web access files. - Follow appropriate standards, such as Western States Standards
METADATA	- Use existing indexes and finding aids as a basis for metadata, including those that would require re-keying by clerical staff. - Use existing thesauri, for example, Library of Congress name and subject authorities, for controlling language, but consider using popular terms in other fields of the metadata record. - Metadata creation by professional staff may take less time and require less quality control than training others for a small project - Follow existing metadata standards such as Dublin Core
CONTENT MANAGEMENT	- Plan a dynamic system that can provide access for continual correction and updating of metadata as information and resources become available. - Choose a content management system that puts you in control. - If open sourced software is to be used, be sure there are sufficient development resources to customize your system - Consider creating your own system with existing resources.
END PROCESSING	- Allow IT staff time to 'unpack', store, upload and link digital objects - Make sure you have a sufficient server space for storage of masters and derivatives and a back up system for all images, including off site storage of the 'archival TIFFs' - Allow time and resources for website content creation

Endnotes

[1] Recent publications that relied on our Trial Transcript Collection include: Donovan, B. (2005); Robertson, S. (2005); Chauncey, G. (1994) among others.

[2] These records were given to John Jay College in May, 1972; we thank Dr. Leo Hershkowitz of Queens College/CUNY for facilitating this transfer. Corresponding indictment records from the same group of court documents were transferred to the NYC Municipal Archives.

[3] This microfilming project, supported by NEH grant # RC-20481-83, amounted to $127,899.00 given to the John Jay Library in 1983-84

[4] For example, we hope soon to link to the 1911 transcript from the Court of General Sessions from the criminal trial of the owners of Triangle Shirtwaist Factory, which was recently digitized by Cornell. www.ilr.cornell.edu/trianglefire/texts/transcripts/introduction.html

[5] Scanning was done using our positive reader copies, which may have affected overall quality. For this small project, the project did not attempt to use the master negative on deposit with the New York State Archives.

References/ Bibliography

Bainbridge, D., Thompson, J., J. & Witten, I. H. 2003, 'Assembling and enriching digital library collections' in *Proc. of the third ACM and IEEE Joint Conference on Digital Libraries*, *Houston, Texas*, pp. 323–334.

Blumenthal, R. 2004. *Miracle at Sing Sing: How One Man Transformed the Lives of America's Most Dangerous Prisoners*. St. Martin's Press, New York.

CDP Metadata Working Group 2006. *Dublin Core Metadata Best Practices, Version 2.1.1*. Available at http://www.cdpheritage.org/cdp/documents/CDPDCMBP.pdf

Chauncey, G. 1994. *Gay New York: Gender, Urban Culture, and the Making of the Gay Male World, 1890-1940*, Basic Books, New York.

Donovan, B. 2005. 'Gender inequality and criminal seduction: prosecuting sexual coercion in the early 20th century', *Law & Social Inquiry,* vol. 30, no. 2, pp. 61-88.

Egan, N. 2007. 'The Lloyd Sealy Library of John Jay College of Criminal Justice: academic library, special library, or both?' *Behavioral & Social Sciences Librarian*, vol. 25, no. 2, pp.1-22.

Faber, E. & Rowland, E. 1989. *Trial Transcripts of the County of New York, 1883-1927; A Historical Introduction with an Index to the Microfilm Collection*, John Jay Press, New York. Available at: http://www.lib.jjay.cuny.edu/info/speccoll/Trialtranscriptspamphlet2.pdf

Goewrey, D. 2006. *Crash Out: The True Tale of a Hell's Kitchen Kid and the Bloodiest Escape in Sing Sing History*, Crown, New York.

MacKellar, L. 2006. *The "Double Indemnity" Murder: Ruth Snyder, Judd Gray, and New York's Crime of the Century*, Syracuse University Press, New York.

Roberston, S. 2005. *Crimes Against Children: Sexual Violence and Legal Culture in New York City, 1880-1960*, University of North Carolina Press, Chapel Hill, N.C.

Terpstra, J., Ongley, D., & Boddie, S. 2005. 'The Tundra Times newspaper digitization project', RLG DigiNews, 15 Feb. Available at: http://www.rlg.org/en/page.php?Page_ID=20522&printable=1&Article_ID=1706

Turkus, B & Feder, S. 1951.. *Murder, Inc.; The Story of "the Syndicate"*, Farrar, Straus and Young, New York.

Valero, W., Perry, C. & Surprenant, T. 2007. "History on a Postcard." *Library Journal NetConnect*, Jan. 15. Available at: http://www.libraryjournal.com/article/CA6404150.html

Western States Digital Standards Group, Digital Imaging Working Group 2003, *Western States Digital Imaging Best Practices, Version 1.0*. Available at: http://ww.cdpheritage.org/digital/scanning/documents/WSDIBP_v1.pdf

Witten, I. H. 2004, 'Creating and customizing digital library collections with the Greenstone Librarian Interface' in *Proc International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society (DLKC'04)*, Tsukuba, Ibaraki, Japan, pp. 97-104. Available at: http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp97.pdf

Witten, I. H. & Bainbridge, D. 2003. *How to Build a Digital Library*, Morgan Kaufmann, San Francisco.

Witten, I., Bainbridge, D. & Boddie, S. 2001, 'Greenstone: Open-source digital library software', *D-lib Magazine*, vol. 7, no. 10. Available at: http://dlib.anu.edu.au/dlib/october01/witten/10witten.html