

Spring 2018

Overcoming legacy processing in photographic collections through collaboration and digital technologies.

Terri Holtze

University of Louisville, terri.holtze@louisville.edu

Rachel I. Howard

University of Louisville, rachel.howard@louisville.edu

Randy Kuehn

University of Louisville, randy.kuehn@louisville.edu


Rebecca Pattillo

University of Louisville, rebecca.pattillo@louisville.edu

Elizabeth E. Reilly

University of Louisville, elizabeth.reilly@louisville.edu

Follow this and additional works at: <https://ir.library.louisville.edu/faculty>

 Part of the [Archival Science Commons](#), [Cataloging and Metadata Commons](#), and the [Collection Development and Management Commons](#)

Original Publication Information

This is an Accepted Manuscript of an article published by Rowman & Littlefield in *Collections: A journal for museum and archives professionals* 14.02 (Spring 2018): 133-150, Focus Issue: Case studies from the field of photographic preservation and collections management, edited by Juilee Decker, guest edited by Olivia Arnone.

ThinkIR Citation

Holtze, Terri; Howard, Rachel I.; Kuehn, Randy; Pattillo, Rebecca; and Reilly, Elizabeth E., "Overcoming legacy processing in photographic collections through collaboration and digital technologies." (2018). *Faculty Scholarship*. 382.
<https://ir.library.louisville.edu/faculty/382>

Overcoming Legacy Processing in Photographic Collections Through Collaboration and Digital Technologies

Terri Holtze, Head of Web Services, University of Louisville, Ekstrom Library, Louisville, KY;
terri.holtze@louisville.edu

Rachel I. Howard, Digital Initiatives Librarian, University of Louisville, Archives and Special
Collections, Louisville, KY; rachel.howard@louisville.edu

Randy Kuehn, Digital Technologies Systems Librarian, University of Louisville, Ekstrom
Library, Louisville, KY; randy.kuehn@louisville.edu

Rebecca Pattillo, Metadata Librarian, University of Louisville, Archives and Special Collections,
Louisville, KY; rebecca.pattillo@louisville.edu

Elizabeth Reilly, Curator, Photographic Archives, University of Louisville, Archives and Special
Collections, Louisville, KY; elizabeth.reilly@louisville.edu

Abstract

In the 1960s, a Louisville photography studio began donating its negatives, prints, and invoices to the University of Louisville Photographic Archives. The Caufield & Shook Collection remains a significant primary source for local history and a prime candidate for digitization.

Unfortunately, on its receipt non-archivists processed the collection with little documentation of original order or organizational decision making. Additionally, workflow choices were determined largely by the desire to maximize student labor. In 2017, the Digital Initiatives Librarian worked with in-house application developers and archives staff to create a workflow that has significantly sped up the process of making this valuable photographic collection accessible online. This article describes how archivists recovered from the poor processing

decisions, used technology to enhance the digitization workflow, and developed a list of best practices for future processing and digitization of large photographic collections.

[ARTICLE BEGINS HERE]

The largest distinct collection in the University of Louisville Photographic Archives (ULPA) is that of the Caufield & Shook studio archives. Comprised of nearly 400,000 items including negatives, prints, and associated records, the Caufield & Shook Collection is also heavily used by patrons.

[Insert Figure 1: Carter Dry Goods and National Grocer, Louisville, Kentucky, 1922, Image Number 043348, Caufield & Shook Collection, Photographic Archives, University of Louisville, Louisville, Kentucky.]

Founded in 1903, the Caufield & Shook commercial photography firm operated in Louisville, Kentucky for seventy-five years. Few aspects of life in Louisville escaped the lens of the company whose motto was "We photograph ANYTHING, day or night." The firm was hired by local architects and builders, wholesale and retail merchants, government agencies, public utilities, legal firms, sports organizations, theaters, restaurants and bars, as well as private individuals. Caufield & Shook was one of the first studios in Louisville to use flashbulbs and the first to shoot aerial photographs. The firm shot pictures for the three main local newspapers, which in the early 1900s had no photography staff, until a dispute over prices ended that arrangement. In 1924, Caufield & Shook became the official photographer of the Kentucky Derby,¹ and was also one of the first firms to shoot photographs for Ford Motor Company national advertising. In addition to paid jobs, some photographers went out and photographed sights around the region such as sunsets, river views, city park scenes, and rural landscapes to add to the firm's stock files from which customers could purchase prints.

A result of Caufield & Shook's wide-ranging work is an expansive visual catalog of twentieth century life in Louisville that receives heavy use by university and community patrons. Thousands of books, newspapers, magazines, and television programs have used Caufield & Shook photographs as illustrations, and copies of the studio's photographs adorn the walls of local businesses and homes throughout the city. Photographs from this collection are called upon regularly by patrons seeking information while doing renovations to their property or researching an old business, for example, and Caufield & Shook photographs are often enjoyed simply as pleasant visual memories of Louisville's past. The overwhelming size, scope, and quality of the Caufield & Shook collection ensure its photographs' popularity with local researchers.

The initial donation of Caufield & Shook archival material, approximately 145,000 negatives dating from 1920 to 1940, came to ULPA in two installments in 1967 and 1968, and constituted one of the first large collections in ULPA.² Over the next two decades, several more donations completed the Caufield & Shook collection, encompassing material to 1978, when the studio ceased operations.

Legacy Processing

Information regarding the Caufield & Shook collection was not centralized within ULPA, and much of its donation and processing history remained a mystery to the current employees working with the collection, including the Curator, Elizabeth Reilly, who arrived in 2011. Through a combination of researching legacy donation documents, collection files, and incomplete electronic files, as well as old-fashioned detective work and conversations with former employees, the staff have only recently been able to piece together some of the "hows" and "whys" of the collection as it has existed.

The material donated from Caufield & Shook consists of glass, nitrate, acetate “safety,” polyester, and color negatives of various sizes; original prints; and studio invoices. The negatives that came in 1967 and 1968 included nearly 145,000 8 x 10-inch nitrate and acetate “safety” negatives dating from around 1920 to 1940.³ The negatives arrived in their original cardboard film boxes, 100 sheet negatives per box, in order by number assigned by the studio.⁴ For much of their tenure, Caufield & Shook assigned a unique sequential number to every negative they created and maintained a record of the first and last numbers allocated each year. Hence, most Caufield & Shook images can be dated to the year they were created from their negative number.

Original invoices from the studio assignments identify the client name and address, date of job, the film size used, negative numbers created for that job, and sometimes more information about the photographed subject (See Figure 2). The studio kept these invoices alphabetically by client name and roughly sorted them chronologically.

[Insert Figure 2: Caufield & Shook Invoice #20402, Caufield & Shook Collection, Photographic Archives, University of Louisville, Louisville, Kentucky.]

Processing of the collection began not long after the initial two donations arrived, although most of the work was accomplished under the direction of a Photographic Archives Curator who served from 1972-2007. Initial processing efforts involved creating prints of every negative in numerical order, starting from the beginning. When the long-term Curator took over and observed many repetitive and boring pictures being printed, he ended that approach in favor of selecting negatives considered to be interesting or of potential value (in his opinion) to researchers, as well as those showing serious deterioration. From this work came the “Print File”—a large group of prints from original Caufield & Shook negatives made in the darkrooms of ULPA.⁵ This type of image selection had several flaws. The choices were not documented.

Further, such a strategy did not lead to significant processing decisions, such as removal of repetitive photographs. Moreover, “cherry-picking” done by a few people with their own biases and perspectives of the region’s history and culture, ultimately left many images unknown to researchers for decades.

The work to select negatives to print in the darkroom occurred when archives staff were “flipping through” boxes of negatives or were intrigued by the information on certain invoices. This selection process can account for some of the gaps in negative numbers in the Print File, but there are even larger gaps in the numerical sequence of the negatives. Reportedly, the gaps are due to entire boxes of negatives, containing both nitrate and acetate film, deteriorating to a point where the contents of the boxes melded together, making separation of each negative impossible and creating gaps of at least 100 negative numbers; fortunately, the remaining negatives in the collection were re-housed in acid-free paper sleeves around 1978. Other nitrate negatives had been lost when a pipe burst over the detached outside “bunker” that once housed them apart from the rest of the archives. The losses as well as the numbers of the negatives lost was not documented and was communicated to current staff only recently.⁶

Usually the only data that accompanies each image is its unique negative number, although occasionally a title or date was etched into the negative. In order to get more specific information about a photograph’s subject or location for description and categorization purposes, someone needs to refer to the Invoice File. But because the invoices were organized alphabetically by client name, it was difficult to find the invoice for an image without tediously browsing through hundreds of brittle invoice slips looking for a single negative number. As a solution to this problem, ULPA staff worked to create a record that would connect negative numbers with their invoices.

Around 1972, through a series of small grants, ULPA hired a student employee to assign a unique number to every invoice from the initial donation. Using a “Bates Numbering Machine,” each invoice was stamped with a unique sequential number in red ink (See Figure 2). Some invoice numbers were instead handwritten when invoices were found out of order. Other times the numbering machine failed to advance, thereby assigning multiple invoices with the same number and requiring the addition of qualifying numbers following a decimal. Once the initial set of invoices had numbers, a list of approximately 173,000 negative numbers was printed, with 500 numbers per page. Student employees began manually writing the corresponding invoice number next to each negative number (See Figure 3). This concordance allowed ULPA staff to quickly search for invoice information about specific images in the collection, but unfortunately it only covered the negatives from the initial donation, dating from approximately 1920-1940, and even many of these negative numbers lacked a corresponding invoice number. Without documentation of the project, it is difficult to ascertain whether these invoice number gaps are due to a missing invoice, a missing negative, or merely that this portion of the task was left incomplete as personnel changed, grants ran out, and other priority projects intervened.

[Insert Figure 3: Caufield & Shook Concordance, Photographic Archives, University of Louisville, Louisville, Kentucky.]

Another early approach to making the Caufield & Shook collection more accessible was through the creation of a “Streets Database.” In the mid-1970s a graduate student copied, in longhand, anything from the invoices that contained an address. This list was then typed into the text editor, a 1980s system that only allowed 132 characters per line, per record. Such constraints required the staff and students entering the data to use abbreviations. Unfortunately, they either

did not document or did not retain the documentation of the meaning of many of those abbreviations, which were also not internally consistent. Furthermore, the challenges of retyping six-digit negative numbers from the longhand notes inevitably created typos and transpositions of digits.

Other challenges with this database exist because some invoices referred to negatives that are no longer in the collection, either because they were not retained by the studio or because they had deteriorated and been discarded. Additionally, only the first negative number listed on an invoice was added to this database, leaving out a huge number of negatives that could be searched for by location. Similar to other undertakings with this collection, the process of creating the Streets Database was not documented and therefore not fully understood by current staff for many years.

Early efforts to process the massive Caufield & Shook collection were extensive and allowed for decades of use by ULPA, researchers, and community members; however, as student employees graduated and staff retired, their knowledge about the work they did left with them, creating a form of “technical debt.”⁷ Since processing was not completed and documentation of procedures—how and why certain decisions were made, and where work ended—was lacking, current staff did not have a full understanding of the collection.

Early Digitization Efforts

In the late 1990s, ULPA obtained its first flatbed scanners and digitized its first images. By 2006, when Digital Initiatives Librarian Rachel Howard was hired, scanning was already replacing the darkroom as the primary means of image reproduction for personal and commercial orders, and standards for formats, file size, resolution, and file naming were in place, although the

departmental server space was rife with image files that did not meet those specifications, including many from the Caufield & Shook Collection.

Prior to launching a CONTENTdm-powered Digital Collections⁸ site in May 2007, the Digital Initiatives Librarian had established a mission statement, adopted the NISO *Framework of Guidance for Building Good Digital Collections*,⁹ and initiated a digital collection proposal process. She also created project planning documentation including a data dictionary for metadata creation,¹⁰ standards for collection-level description, and technical specifications for scanning.¹¹ The Associate Dean then overseeing ULPA, along with ULPA staff, expressed a strong desire to prioritize getting the images from the Subject Files and the Streets Database added to the Digital Collections, but this task was deferred due to the complications of assessing permissions, processing status, and existing documentation for this large body of images.

As the only position dedicated to working with Digital Collections, the Digital Initiatives Librarian relied on a dispersed team who performed certain tasks either as part of their job description (for example, administering the CONTENTdm server as one of many Libraries-managed Windows servers) or out of personal interest. Colleagues from other units volunteered to create metadata for collections using spreadsheets or Microsoft Access databases, with training and resources on locating and applying appropriate subject headings and descriptive information provided by the Digital Initiatives Librarian, who also exported the metadata as a tab-delimited text file, performed queries to insert repeating information, and uploaded the metadata and images into CONTENTdm. Most of the scanning for ULPA collections was performed by student employees or had already been completed by ULPA staff. Later, a dedicated Metadata Librarian position was created, but the system of volunteer metadata creators

from other units continued, and student employees continued to do much of the scanning for digital collections.

Initial collections selected for inclusion in Digital Collections adhered to the NISO Framework, particularly Collections Principle 1: “A good digital collection is created according to an explicit collection development policy that has been agreed upon and documented before building the collection begins.” Because of the vast scope and size of the Caufield & Shook collection, plus ambiguities from legacy processing, it was difficult to be fully aware of the entire scope of the collection in order to make an explicit collection development policy.

Nevertheless, the continued desire by colleagues to broaden access to this popular collection, the presence of pre-existing scans, as well as the availability of a volunteer to create metadata led to shortcuts which veered from this principle.

The Head of Web Services, Terri Holtze, volunteered to create metadata for the first Caufield & Shook images to be included in the Digital Collections. Using the concordance to find invoices (and including information from these invoices), the Streets Database, and handwritten annotations on the back of prints, she wrote metadata for images that had already been scanned and stored on the departmental server starting with the lowest numbered images which corresponded with the earliest photographs. In April 2010, the Caufield & Shook digital collection¹² launched with a starter set of 500 images and a home page design customized by the Web Services Librarian to permit browsing by the image’s location as well as subject and time period.

Several issues were discovered during the metadata creation process, including a number of duplicate images, copy photos of works by other documented or undocumented creators, and images too damaged or faded to discern, particularly in the JPEG version viewable online. A

revised digital collection development policy was adopted for this and other large studio collections, allowing metadata creators the discretion to remove such images from the online queue while alerting users via the About the Collection page,¹³ about these overarching decisions.

The Web Services Librarian continued creating metadata for approximately 1,100 pictures total, most of which dated from the 1920s and 1930s. Because images from later decades lacked invoice numbers, she taught student employees how to go through the invoices, entering the negative numbers and customer names into a spreadsheet. This interim step aided the new Metadata Librarian's workflow, which also incorporated the command-line interface to batch import technical metadata such as file name, date created, and file size for the scans from the departmental server.¹⁴ Between 2013-2015, the Metadata Librarian added approximately 2,800 Caufield & Shook images to the digital collection.

Despite this work, the number of scans on the departmental server kept growing more quickly than metadata completion. Unbeknownst to the Digital Initiatives Librarian, student employees supervised by ULPA staff located on a different floor of the library were busily scanning images and storing them in separate directories on the departmental server. (Students did not have access privileges to the space where master files were stored on the server.) At times these student assignments were logical (for example, scanning deteriorated acetates before they further deteriorated); at other times, students simply selected the next sequential box of negatives and scanned them all, without communicating with the Digital Initiatives Librarian, who would have followed up to ensure that the scans met specifications, quality standards, and the digital collection development policy. Consequently, in addition to the already large backlog of scans, some of which did not merit inclusion in the digital collection, the backlog continued to

grow with errors in file naming, specifications, and selection, creating more mysteries to be solved and using more server space and resources. The lack of oversight of student employees, as well as lack of proper training and workflow documentation, was in part due to the failure to properly plan all aspects of digitization for the collection.

Rethinking Workflows

When the Metadata Librarian position became vacant in 2016, the Digital Initiatives Librarian began rethinking the piecemeal workflow that had barely made a dent in the Caufield & Shook backlog. She found scans and partial metadata squirreled away in unexpected corners of the server space, moved them to the appropriate locations, and merged the metadata with exported Caufield & Shook records from the Streets Database. With the help of the ULPA Curator, the Digital Initiatives Librarian added captions from various publications that had used the studio's images. The Caufield & Shook database grew to over 30,000 records, some of which were already in CONTENTdm, some of which had scans and minimal metadata, many more of which had been scanned but had no metadata, and some of which had minimal metadata but no scans. She then began a process of documenting which images had scans available and normalizing the available metadata (for example, expanding the Streets Database's 3-digit codes for neighborhoods, inherited from the text file with character limitations, into their controlled vocabulary terms).

Through a process of sorting the database vertically, column by column, the Digital Initiatives Librarian learned to interpret the code language inherited from the 1980s text editing program, and used "find and replace" to make updates. Spending time with a larger set of records than any of the previous metadata creators had worked with, she gained an understanding of the

variety and complexity of the collection itself which allowed her to formulate the requirements for a new and improved workflow. For example, while continuing the Metadata Librarian's method for capturing the technical metadata for scans, she wondered why those creating the scans, who also had access to the physical item, were not capturing digital and physical format and size. It would in fact be preferable to have them slow down on the scanning to include this information and double-check their file names. While checking off which database entries had scans, she discovered many duplicates, some of which had slightly different file names and some of which had been stored in separate directories, so the server's overwrite protections had not prevented them from being saved. Having the students enter metadata into an Access database as they worked only partially solved the problem as only one person at a time could work on the project. The ideal solution would provide a way for students to check which images had already been scanned, and allow them to input metadata without blocking other people from working in the database at the same time. The workflow already employed "resource leveling,"¹⁵ with digitization activities deconstructed into components performed by different people; what it lacked was a shared system of assigning and documenting those activities.

The Metadata Librarian position was finally reauthorized a year after it had been vacated, only to have the search put on hold due to a budget shortfall. The Digital Initiatives Librarian continued to normalize the data in the Access database, hoping to hand over a cleaner, more easily understandable set of information when the position could be filled. The Director of Archives & Special Collections (ASC), who oversaw both ULPA and Digital Initiatives, knew the effort that had gone into streamlining the metadata for this important collection. Inspired by other institutions who had made their budget woes more visible to their users,¹⁶ she proposed instituting "summer hours," reducing by two hours per day the department's reference services to

the public, and devoting that staff time to metadata creation for the Caufield & Shook collection. Eight ASC colleagues volunteered to assist with creating metadata and/or entering invoice information for the collection, what ASC staff referred to as the “Metadata Blitz.”

Having worked with so many volunteer metadata creators over the years, the Digital Initiatives Librarian knew that her colleagues could be trained to perform this task, but the problem remained of having one master database for their entries. Thus, she contacted her colleagues in Web Services with the idea for what became known as the “Metamachine.”

Developing Technical Solutions

In April 2017, the Digital Initiatives Librarian met with the Web Services Librarian and Digital Technologies Systems Librarian Randy Kuehn to discuss the possibility of developing a web-based system for collating the metadata for the Caufield & Shook collection. First, the system needed to accommodate staff and student employee access with differing levels of administrative access. Second, controls needed to be in place to prevent duplication or deletion of record data. Third, it was imperative to provide a streamlined approach to managing a multi-step, multi-user metadata entry environment.

The Systems Librarian developed the architecture of the site utilizing Rails Composer, an open source Ruby on Rails application creation tool¹⁷ with a MySQL database. He created temporary logins so the project developers could access the system as a student employee, staff member, or a system administrator, and see the details that each level of user would see. Using the existing spreadsheet of metadata fields, the Systems Librarian worked with the Web Services Librarian to envision the workflow and interface options. The anticipated workflow required that users be able to search the database (to find if something had already been scanned, and to identify descriptions from similar images for reuse) and edit individual entries; therefore, these

were the first functions developed, along with controlled vocabulary dropdowns for certain fields, including object type and physical file size (See Figure 4).

[Insert Figure 4.1 and 4.2: Caufield & Shook Metadata Entry Form, Archives and Special Collections, University of Louisville, Louisville, Kentucky.]

In continued consultations, the Digital Initiatives Librarian, the Systems Librarian, and the Web Services Librarian refined the functions and design of the system. Ransack, a Ruby on Rails search form creation library, was used to provide enhanced search functionality.¹⁸ The refined search function was configured to: 1) allow the user to limit the search to any of the fields; 2) find items where the search term is contained in, not contained in, greater than, or less than the entry in that field; and 3) add (and remove) extra lines for additional search conditions. This provided the flexibility to meet the many possible needs of the metadata creators and administrators.

One of the major design considerations involved displaying the item records in a user-friendly way. Each record included 28 fields, plus the display needed space for links to view the record and/or edit the record. One option was to limit the number of fields showing in the results screen and then display all fields when a particular record was chosen. On further discussion, though, the Digital Initiatives Librarian saw advantages to being able to scroll through results to find fields that were misconfigured (for example, records that had extra punctuation following the last subject heading); therefore, the decision was made to show all the fields in the results display (See Figure 5). Each record would be a row in the table with the fields being represented in columns. This made for a very wide table and necessitated a horizontal scroll bar which could only be added to the bottom of the table. In order to limit vertical scrolling, the default results

screen was set to ten items and navigation links were added to proceed to additional screens of results. To mediate an issue with lengthy fields stretching out the table and pushing the horizontal scrollbar off the screen of most monitors, a script was added on fields for description, location depicted, and subject to limit the display to 25 characters, and add a “more” link that would expand to show the rest of the content when clicked. In addition, some design considerations were implemented to improve the readability of the data. The Web Services Librarian set CSS alternate rows to display with a light, but noticeably different, background color to make it easier for the eyes to stay on the same record’s fields, and chose a serif font for record details so that it would be easy to differentiate between numbers and letters that looked too similar in a sans-serif font.

[Insert Figure 5: Caufield & Shook Search Results Table, Archives and Special Collections, University of Louisville, Louisville, Kentucky.]

The draft design received a test run in early June 2017. In the fine-tuning stage the project developers determined which elements would be visible to each level of user (for example, the option to delete a record only displays for administrators). Other changes included 1) adding a section in the item edit screen for “add to controlled vocabulary” so that new subject headings would be easy to find, and 2) a way to assign particular records to an individual staff member or student employee.

In the early stages of development, the Digital Initiatives Librarian worked directly with the Systems Librarian by providing functionality suggestions and continuous testing throughout each stage of the project. Open dialogue between the two during this early stage contributed significantly to the speed of the development and the success of the project. Two months after

the first project meeting, the Digital Initiatives Librarian uploaded the first set of 130 Caufield & Shook records created via the Metamachine into CONTENTdm.

Implementing Workflow

During the summer of 2017's "Metadata Blitz" ASC's dedicated staff made a small dent in the immense backlog of roughly 7,000 images, completing over 300 records. After a vacancy of almost two years, the newly hired Metadata Librarian, Rebecca Pattillo, began work on the remainder of the backlog, with the continued assistance of several "Metadata Blitz" volunteers, some of whom focus on transcribing invoice information in order to prepare records for metadata completion. The Metamachine expedited item-level processing of the large backlog for both the dispersed team and the new hire. The ease of creating queries powered by MySQL through the Metamachine's user-friendly interface made it quick to find images with commonalities, such as location, event, and customer, that otherwise would have required detailed physical processing to discover. For example, the collection contains multiple series of images of building construction. The studio photographer would return to the site throughout its construction, capturing progress photographs over the course of months or even years. The Metamachine search function made it possible to find these series of images by searching for similar language from the description or invoice information, thus making it possible to create metadata in bulk for a large set of images that otherwise were disconnected in the physical collection. Further, while technically the records are still item-level, it sped up the process by applying "More Product Less Process" (MPLP) theory to the metadata creation.¹⁹

Project Assessment

A post-production review of the Metamachine project, consisting of a short survey sent to the Digital Initiatives Librarian, allowed the Web Services Librarian to assess how the Metamachine was working: Was it functioning as expected? Did the workflow make sense? What needed improvement? As a result of the review, the Systems Librarian implemented changes, including 1) a way to save complex search strings, 2) clarified error messages to indicate the exact problem, and 3) the addition of the “copy record” option to reduce duplication of effort.

As metadata production increased, further insights into the collection became apparent and the collection development policy needed to be further updated. This included excluding portraits of individuals with no identifying information and sampling large sets of images that have little to no widely applicable research value. For example, a series of over fifty photographs of radiators may not be worth the time to scan every single one. Scanning one or several and including in the metadata record that the physical collection has x-number more would speed up the digitization process and still provide users knowledge of additional images in the series.

The Digital Initiatives Librarian has updated the data dictionary to accommodate the new features of the Metamachine. ASC staff will have to continue to evolve with the changing technological landscape. Digital Collections will migrate out of CONTENTdm into an open source system within the next few years, so the technical solution that works now will surely need upgrading or reassessment.

One year after the completion of the Metamachine, 3,000 records have been added to the Caufield & Shook digital collection, exceeding the number added by the previous Metadata Librarian over a three-year span. The current Metadata Librarian is on track to double the total number of Caufield & Shook images available online by her first anniversary on the job. Furthermore, the Metamachine database has invoice information ready for metadata creators and

documentation about images that have already been examined and determined unfit for the digital collection, which also contributes to decreasing the backlog. The Metamachine continues to ease the creation of metadata for the Caufield & Shook collection.

Best Practices

The particular scale and complexity of the Caufield & Shook collection led to *ad hoc* decision making. In order to avoid incurring technical debt over the long term, or encountering similar pitfalls with future collections, we recommend following these best practices:

1. Process a collection before embarking on its digitization. Weed duplicates, address preservation needs, and get a sense of the collection's content, condition, and organization before determining whether it—or a subset of it—is a good candidate for digitization.
2. Do not make assumptions about a collection's processing status. Even if a collection has been processed, seek documentation of your predecessors' decisions; document your findings, including those gleaned from institutional memory; and examine a large enough subset of the collection materials to verify that the documentation is accurate.
3. Determine the best pathway through the collection—Box 1 through Box X? Prioritize a certain series? Nitrate negatives before safety negatives?—and establish a scanning workflow accordingly.
4. Document digitization decisions, including selection criteria, metadata standards, scanning specifications, and file naming rules.
5. Train everyone working on the project to adhere to the appropriate standards for their assigned task(s), and verify their adherence by performing quality control (very

frequently immediately after training, and less frequently once fewer errors are discovered).

6. Aim to touch each artifact only once. The person doing the scanning should capture all information related to the physical item and the digital item while scanning. If separate documents such as invoices exist, the person reviewing them should transcribe or describe, as appropriate, everything on them so they do not need to be revisited.
7. Assess the skills and interests of the project team to make sure they're doing work they can handle and find rewarding. Consider making the process more fun, particularly for student employees, by gamifying processing or metadata creation.²⁰
8. Train project participants on the overall goals for the project, so that they can engage in critical thinking as they work with the collection. Similarly, train student employees in the overall functions of archives, even if their role only involves one project, so they can place their role within the larger context.
9. Provide a centralized system to track progress on all project activities. Our Metamachine serves this purpose well: everyone involved in scanning or metadata creation has a login to the system and can see what has already been completed for each item/record.
10. Communicate, communicate, communicate! Supervisors of student employees should check with project coordinators before assigning tasks to the students. Project coordinators should communicate to supervisors what sorts of tasks are needed. All should feel free to ask questions of one another when something is not clear or a new problem is encountered. All should use the centralized system (see #8) to track their work.²¹

11. Perform exit interviews with departing students and staff to ascertain their stopping point and tie up any loose ends (although following Best Practices #4-10 should minimize the likelihood of loose ends).
12. Reassess and adapt workflows if the available resources, technology, or the next stage of the project require it. Document those changes!

We have improved our productivity and our camaraderie by striving to follow these practices.

While we hope that other professionals have not found themselves with quite this large of a knot to untangle, we do hope that our experiences and the lessons we have learned can inform others on how to move forward with their complex collections.

Bibliography

- Allison, Sarah M. "Conducting a SWOT Analysis: Evaluating the Student Employee Program in Archives and Special Collections at New Mexico State University." *The Reading Room: A Journal of Special Collections* 3, no. 1 (Spring 2018): 23-38.
- Clair, Kevin. "Technical Debt as an Indicator of Library Metadata Quality." *D-Lib Magazine* 22, no. 11/12 (November/December 2016).
- Dorpinghaus, Sarah and Heather Fox. "Time Is On My Side: Harnessing the Power of Automation for Efficient Archival Workflows." *Kentucky Libraries* 79, no. 1 (Winter 2015): 20-23.
- Green, Mark A. and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist* 68, (Fall/Winter 2005): 208-263.
- Howard, Rachel. "CONTENTdm Cookbook: Recipes for metadata entry for UofL Digital Initiatives." accessed May 30, 2018.
<https://webservices.library.louisville.edu/contentdm/files/cookbook.pdf>
- Kentucky Historical Society. "KHS Campus to Adopt Winter Hours of Operation." December 3, 2012, accessed May 30, 2018.
<http://migration.kentucky.gov/Newsroom/history/KHS+History+Campus+to+Adopt+Winter+Hours+of+Operation.htm>.
- National Information Standards Organization. *A Framework of Guidance for Building Good Digital Collections*, Baltimore: National Information Standards Organization, 2007.
<https://groups.niso.org/publications/rp/framework3.pdf>
- Passehl-Stoddart, Erin. "Playfulness in the Archives: Enhancing Digital Collections through Card Sorting." *The Reading Room: A Journal of Special Collections* 3, no. 1 (Spring 2018): 39-54.
- Petersohn, Barbara, Traci Drummond, Melanie Maxwell, Kelly Pepper. "Resource Leveling for a Mass Digitization Project." *Library Management*, 34, no. 6/7 (2013): 486-497.
- Reilly, Elizabeth E. "The University of Louisville Photographic Archives: The First Fifty

Years.” *Kentucky Libraries* 76, no. 3 (Winter 2015): 4-9.

Notes

¹ The Kentucky Derby is a world-famous horse race that takes place annually at Churchill Downs in Louisville, Kentucky on the first Saturday in May. Begun in 1875, The Kentucky Derby is the first leg of the American Triple Crown and is often referred to as “the most exciting two minutes in sports.”

² The ULPA was founded in 1962 and became an official part of the University Libraries in 1967. Elizabeth E. Reilly, “The University of Louisville Photographic Archives: The First Fifty Years,” *Kentucky Libraries* 76, no. 3 (Winter 2015): 4-9.

³ Although the studio began in 1903, all of the negatives created before 1920, about 20,000, were lost in the devastating Ohio River flood of 1937 as they were stored in the basement of the Caufield & Shook studio. Fortunately, some glass negatives and about 1,000 original prints from those early years survived the flood because they were bound in stock file books for customers to browse while waiting in the studio.

⁴ James "Andy" Anderson and David Horvath, conversation with Elizabeth Reilly, April 20, 2018.

⁵ Subsequently photocopies of these prints were made and organized by subject, street, neighborhood, and theme into what is known as the “Subject File.” For decades, the photocopy Subject File had been the primary way for visitors to the archives to access images from this collection.

⁶ Ibid.

⁷ Technical debt is a software developing term that has been altered to describe issues with digital projects and metadata in the digital library community. Our department incurred this “technical debt” in several ways, but most significant in “knowledge distribution and documentation debt” as described by Kevin Clair, “Technical Debt as an Indicator of Library Metadata Quality,” *D-Lib Magazine* 22, no. 11/12 (November/December 2016).

⁸ University of Louisville Digital Collections, accessed May 30, 2018.

<http://digital.library.louisville.edu/>

⁹ *A Framework of Guidance for Building Good Digital Collections*, (Baltimore: National Information Standards Organization), 2007,

<https://groups.niso.org/publications/rp/framework3.pdf>

¹⁰ “CONTENTdm Cookbook: Recipes for metadata entry for UofL Digital Initiatives,” Rachel Howard, accessed May 30, 2018.

<https://webservices.library.louisville.edu/contentdm/files/cookbook.pdf>

¹¹ “About the University of Louisville Digital Collections: Images,” University of Louisville Digital Collections, accessed May 23, 2018.

<http://digital.library.louisville.edu/cdm/about#images>

¹² “Caufield and Shook Collection,” University of Louisville Digital Collections, accessed May 23, 2018. <http://digital.library.louisville.edu/cdm/landingpage/collection/cs/>

¹³ “Caufield & Shook Collection,” University of Louisville Digital Collections, accessed May 30, 2018. <http://digital.library.louisville.edu/cdm/description/collection/cs/>

-
- ¹⁴ Sarah Dorpinghaus and Heather Fox, "Time Is On My Side: Harnessing the Power of Automation for Efficient Archival Workflows," *Kentucky Libraries* 79, no. 1 (Winter 2015): 20-23.
- ¹⁵ Barbara Petersohn, Traci Drummond, Melanie Maxwell, Kelly Pepper, "Resource Leveling for a Mass Digitization Project," *Library Management*, 34, no. 6/7 (2013): 486-497.
- ¹⁶ "KHS Campus to Adopt Winter Hours of Operation," Kentucky Historical Society, December 3, 2012, accessed May 30, 2018.
<http://migration.kentucky.gov/Newsroom/history/KHS+History+Campus+to+Adopt+Winter+Hours+of+Operation.htm>.
- ¹⁷ Rails Composer on Github, <https://github.com/RailsApps/rails-composer>.
- ¹⁸ Ransack on Github, <https://github.com/activerecord-hackery/ransack>.
- ¹⁹ Mark A. Green and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, (Fall/Winter 2005): 208-263.
- ²⁰ Erin Passehl-Stoddart, "Playfulness in the Archives: Enhancing Digital Collections through Card Sorting," *The Reading Room: A Journal of Special Collections* 3, no. 1 (Spring 2018): 39-54.
- ²¹ Sarah M. Allison, "Conducting a SWOT Analysis: Evaluating the Student Employee Program in Archives and Special Collections at New Mexico State University," *The Reading Room: A Journal of Special Collections* 3, no. 1 (Spring 2018): 23-38.