Digitization Workflows: Streamlining the Digitization Process and Distinguishing the Peculiarities in Capturing Various Archival Materials

Marina Georgieva

University of Nevada, Las Vegas, marina.georgieve@unlv.edu

Follow this and additional works at: https://digitalscholarship.unlv.edu/lib_articles

Part of the Library and Information Science Commons

Repository Citation


https://digitalscholarship.unlv.edu/lib_articles/652

This Article is brought to you for free and open access by the Library Faculty/Staff Scholarship & Research at Digital Scholarship@UNLV. It has been accepted for inclusion in Library Faculty Publications by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
Overview

With thousands of archival collections prioritized for digitization and numerous grant opportunities for external funding, large-scale digitization is on the rise. That’s an exciting trend — from users’ perspective, it makes rare, fragile, historical, local and research-significant materials readily available online. Users nowadays want to find all their research materials on the web, and digital librarians are working hard to meet the demand for digitization and to provide access to archival collections.

Large-scale digitization is a great initiative; however, it brings challenges to the digital librarians. The main question that emerges is: “How do we streamline the process to make it efficient and optimal?” It also brings a subset of other questions that cause some anxiety: “What technology should be used? What’s the best workflow? How do we troubleshoot issues and address challenges?” and so on.

This article is far from being comprehensive on the topic; it’s rather a case study based on one librarian’s experience that shares some of the practices, issues, challenges and tips related to making the digitization process more robust. They are universal and equally helpful for large-scale and smaller scale projects.

UNLV Digital Collections and Digitization Stats

UNLV Digital Collections (http://digital.library.unlv.edu) is a department in UNLV Libraries Special Collections & Archives division (https://www.library.unlv.edu/speccol). The internationally renowned UNLV Libraries Special Collections & Archives houses over 11,000 linear feet of unique archival collections, over 32,000 rare books and periodicals, and over 1,800 maps. It is located in the main campus library — Lied Library (https://www.library.unlv.edu/about/quick-facts-about-unlv-libraries).

The Digital Collections department has 5 full-time permanent staff (faculty and professional), 6 part-time student assistants, 1 full-time visiting faculty and 3 full-time temporary professional staff. Each digitization project is unique and requires different team configurations. Sometimes, there is a project manager and 1 to 3 student assistants; other projects are entirely run by students and managed by a student supervisor; still others are digitized by faculty or professional staff as proof-of-concept and workflow improvement projects.

Digital Collections is a very collaborative department and staff not directly involved in digitization still contribute with their expertise — the metadata librarian and the visual materials curator are always available to share their knowledge, consult on project workflows and provide help if needed.

UNLV Digital Collections is frequently awarded various digitization grants and external funding. The grant-funded digitization projects are usually comprised of professional level project managers and specially hired project technician or student assistants.

It is hard to quantify the digitization turnover per year — it depends on the number of active digitization projects and available funding, as well as the digitization speed, which directly relates to the condition of the materials and the level of processing of the archival collections.

The capture station is powered by the mighty Phase One Capture One Cultural Heritage edition software (https://dtculturalheritage.com/capture-one-ch). With unique features like auto crop, auto rotation, and LAB color readouts, the software allows for instant capture of materials placed under the camera, quick global editing, and final output, producing digital surrogates of finest quality. Images are saved on an external SSD drive and seamlessly backed up on a separate large hard drive.

 continues on page 00
Furthermore, UNLV Digital Collections has two separate editing stations allowing for continuous capturing, even while other staff are preparing and outputting their final archival tiffs.

Types of Archival Materials and Major Peculiarities in Their Digitization

The Phase One Cultural Heritage capture station has vast possibilities for digitizing a variety of materials. The ease of use and the speed of capturing have turned it into the preferred digitization technology in our department. The Phase One is so universal and suitable for multiple formats that we have made it our top choice for most projects, and we keep an availability calendar, because it is in such high demand!

UNLV Special Collections and Archives has unique archival collections that are a blend of multiple formats and sizes. The Phase One allows us to rapidly digitize most collections regardless of the format and the size of the materials. Rarely, for oversized items that go beyond the table surface and cannot be captured at 300 ppi, we use a special large-format scanner.

Some archival material types and their peculiarities are outlined below. During capturing flat reflective materials, bound reflective materials and transmissive materials patterns emerge and twist the general digitization workflow.

- Reflective materials

  Reflective materials fall in two main categories: flat and bound. Flat materials are photographs and loose manuscripts such as correspondences, scripts, drafts, sketches. They vary in color, from black and white to full color prints. They also range in size from thumbnails to large posters.

  Bound materials are typed or handwritten books, manuscripts, photo albums, and scrapbooks with various content (anything from photos to newspaper clippings and 3D objects).

  Usually the smoothest digitization process is capturing flat reflective materials. When all materials are properly arranged and named, capturing happens rapidly. Unexpected issues are generally easy to troubleshoot. Some challenges include flattening out folded or curled manuscripts, taking photographs out of protective sleeves, capturing faded ink, or handling items with torn edges, etc.
Bound manuscripts (books and scrapbooks) vary in size and thickness. They follow the general reflective materials digitization workflow, and yet they need special attention due to the thickness: the camera focus needs frequent adjustment for optimal quality of the image. Half-torn pages or papers sticking together are common problems that require careful handling due to their fragile nature.

Another frequent peculiarity of bound handwritten manuscripts is bleed-through of text below from thin paper materials. Although it is an easy fix, it slows down the capturing process as the operator needs to separate the pages with white paper to enhance the text readability.

Occasional missing pages from bound volumes can significantly clog the pipeline depending on their number. To enhance the user experience and improve the metadata, the operator keeps a log of all missing pages and adds this information in the description fields of the parent object.

Faded ink of manuscript pages can be fixed by contrast adjustment during the image processing step. Although it is a quick adjustment, multi-page materials bring a challenge if some pages are faded and the rest are normal. This calls for manual contrast correction of selected pages and slows down the process.

- **Transparencies**
  
  Transparencies can be film positives or film negatives. Their main peculiarity is being transmissive. Types of transparencies include:
  
  - **Film (loose)** are loose positives or negatives of various sizes (medium or large format, or 35 mm) that originate from film strips. It can be black and white or full color.
  
  - **Filmstrips** feature several film frames at a time. They have the same features as loose film and can be digitized as strips or each frame individually.
  
  - **Mounted slides** are loose film shots mounted in special frames so they can be projected.

  Digitizing transparencies has one major peculiarity: distinguishing the glossy side from the emulsion side before digitizing. Occasionally this is very difficult and the operator needs to examine the image closely. It’s helpful if the shot has text — it determines what side goes up for digitization.

  Frequently, loose film stored in plastic sleeves brings the challenge of removing it without breaking it or leaving fingerprints on the surface. Some experts recommend handling film with gloves; however, our experience proves this poses more danger due to clumsiness and decreased finger sensitivity. The best solution for handling film is bare clean fingers touching the corners only and gently placing the shot in the digitization holder. An air puff effectively removes any dust.

- **Preparation**
  
  Preparation is a process that includes careful review of the materials in the entire collection — from examining the finding aid (if available) to inventorying the physical materials’ formats, sizes and condition. The appropriate grouping of physical materials aims to enhance the digitization efficiency as it sets the groundworks for a more streamlined process. While preparing the physical items for efficient and rapid capturing, we should bear in mind that all archival materials should be organized in their original order, so proper notations are recommended.

  Usually the preparation of large collections (containing mixed materials in multiple boxes) happens on a collection level. This method features digitization of items similar in format and size that need the same scanning settings, while skipping and notating all items that get other settings on the capture station. The notation process is on a digital and physical level: noting and color-coding the digitization documentation with item-level information and recommended scanning settings and sticky notes attached to the outside of archival boxes and folders.

  - **Large-scale vs small-scale approach**
    
    Selecting the most appropriate digitization approach for each collection is a critical decision point that greatly affects digitization efficiency.

    Large-scale digitization differs greatly from the small-scale approach and follows a separate workflow. Usually, the large-scale approach mirrors the archival collection: archival materials are scanned on a folder level resulting in multiple complex digital objects. The finding aid metadata is reused and the digital objects get parent level descriptions. This approach is less labor-intensive and allows capturing and describing materials at a fast pace leading to a streamlined production line.

    The small-scale approach features individual attention to objects; each archival item is digitized separately and receives a rich description. This approach is more suitable for smaller collections, collections with greater research value, or curated digital exhibits.

    Streamlining digitization and selecting the appropriate approach requires considering different factors, such as user needs, expected user interaction with the online collection, project specifics (timeline, scope, funding), and collection research value.

    Pausing for collection analysis and decision-making on the digitization approach is critical for the success and efficiency of each project. Both approaches differ significantly, so switching from one to another is never seamless as they follow different workflows.

    Proper analysis, selection, and planning around the appropriate digitization approach saves time and streamlines the project as it follows established workflows.

  - **Digitization**
    
    — **The process**

      On the surface, digitization looks straightforward (ready, set, go!), and yet it is an intricate system of workflow segments and procedures that involve many factors for consideration. Various archival materials require different digitization settings and workflows. For the sake of efficiency, knowing all workflow peculiarities and preparing for random unexpected issues is key, so it is a good idea to test workflows ahead of time.

      Digitization of flat reflective materials requires a specific lens and particular scanning settings. Potential random issues include curling paper, fragile (tissue-thin) manuscript pages, faded ink, framed photographs, etc.

      Besides the specific settings and lens, digitization of bound reflective materials requires more attention as the scanning progresses: as book pages get flipped, the thickness changes. This requires frequent camera focus adjustment to keep the pristine quality of the digital surrogate. Frequent pop-up issues that require troubleshooting are deteriorating books, fragile (ultra-thin) pages, sticky or half-torn pages, faded text and missing pages.

      Digitization of transparencies has its own peculiarities: it needs different equipment, i.e., a film kit (lens, lightbox, table and film holders, air puff) and a specific scanning session depending on the type of transmissive material (black and white or color). The film placement plays important role for the digital image quality: films get digitized with glossy side up. Films bring another array of unexpected issues: film stuck in plastic protective sleeves, deteriorating shots, and torn film, just to name a few.

      To keep the pipeline going and to achieve streamlined digitization, maintaining a troubleshooting log of potential problems and guidelines for quick resolution improves the performance. Staff involved in digitization should be able to quickly address and resolve problems, so occasional pop-ups should not clog the pipeline.

    — **The workflow**

      The general digitization workflow is simple and it applies to all collections regardless the technology and the type of the archival materials.
Well-processed archival collections (arranged by Technical Services and described by a finding aid) are generally smooth for digitization. They are less prone to pop-up issues. Established workflows and up-to-date documentation guarantee a streamlined and efficient process easy enough to be followed by newly trained staff, student workers, and experienced professionals.

— **Image processing**

Image processing follows the process of capturing raw images. It enhances the digital object quality, which makes it an important workflow segment. Some actions include cropping, straightening, file name corrections, and contrast/brightness corrections.

To achieve efficiency on the Phase One, digitization staff rapidly capture dozens of images per minute and leave the image correction for a separate workflow step because it significantly affects the capturing speed. For streamlined image processing, Capture One software allows automated actions that batch-process hundreds of images at a time. We have included this automated image processing in our reflective materials traditional workflow as it boosts productivity and is convenient for time-sensitive large-scale digitization projects.

The workflow of image processing color transparencies is less automated. Color corrections and enhancements are performed frame by frame. Our practice has proven to deliver color closer to real life, so this method is preferred for color film rather than the traditional automated batch processing after capturing. All remaining image processing steps are the same as for reflective materials.

— **OCR processing**

Textual materials (books, scrapbooks, postcards, photo inscriptions, etc.) undergo an additional step before batching. Optical character recognition (OCR) is an electronic conversion of digital images with typed or handwritten text into machine-encoded text. OCR'ed digital objects have a special metadata field “transcript” that contains the machine-encoded text. The “transcript” field enables users to perform full text searches within the documents and adds value to the digital resources by making them easily discoverable.

At UNLV Digital Collections, we employ one traditional OCR method: utilizing Abby Fine Reader software (as a separate process after export).

Large-scale digitization projects aim for speed, so our practice is to run OCR on all digital objects, regardless their type (textual or visual). Although this automated process may take longer processing time, it is still more efficient than selecting textual files only. Additionally, OCR accuracy correction requires human labor and is manually performed. It is not suitable for large-scale workflow, although, the manually corrected OCR is expected for signature projects. All large-scale efforts are streamlined through the automated OCR pipeline.

### Documentation

At UNLV Digital Collections, we love documentation for a reason — it keeps us organized, progress is trackable and measurable, workflow trends and patterns emerge, and it promotes continuity.

We cannot stress enough how important it is to keep documentation current. This effort pays off with increased productivity and boosted efficiency. Documentation varies by type and purpose; some projects may have more and/or modified templates due to their specifics. The ultimate goal is to promote an uninterrupted digitization process with an optimal timeline.

Some documentation we keep and find helpful include:

- Collection level master-file (tracks progress of all workflow segments and material formats/types)
- Cross-collection spreadsheet (tracks progress, types and priorities of all collections for digitization)
- Troubleshooting log and guide
- Digitization procedures and manuals
- Digitization schedule
- Training manual for new hires
- Technology procedures and manuals
- Cheat sheet for frequently occurring topics

### The Role of the Digital Librarian in Streamlining the Process

Digital librarians play a significant role in streamlining the digitization process including testing new iterations of workflow, documenting best practices, and tweaking existing workflows to achieve more efficiency for mixed materials collections. Unfortunately, there is no universal approach, no matter how hard we attempt to standardize the process. A huge part of the digital librarian’s job is the intellectual labor of making decisions to streamline each module of the digitization process; this requires familiarity with best practices and documentation as well as modifying workflows by combining solutions from other projects.

Digital librarians need to also recognize the various archival materials types and be prepared to find unique combinations of them in each collection. This knowledge empowers them to analyze and select the best approach for digitization that brings efficiency and high quality.

Lastly, the digital librarian should look at the collection from the user perspective to make it valuable for patrons. It is important to assess user needs and consider how users expect to see the archival items represented as digital objects.

### Conclusion

Streamlining the digitization process takes practice and iteration. Good knowledge of established digitization techniques, scanning technologies and troubleshooting strategies is critical. Additionally,
familiarity with archival materials types promotes better planning and collection assessment to customize general workflows as needed.

Iteration plays a key role in constantly improving digitization workflows and achieving products of the highest quality. It also reduces the time for completion and optimizes the team size to conclude the project. It includes testing new methods, trying new strategies on various types of materials, learning from past mistakes, and documenting best practices.

No process will ever be perfectly streamlined as we all know there is no “one size fits all.” As digital librarians, we strive for efficiency and boosted productivity, yet there are so many variables that bring challenges and clog the pipeline. Collection materials differ in size and format, technologies bring challenges, and some archival materials have more issues and require more attention than others. The team factor plays an important role as well — staff training, staff turnover, and equipment scheduling to maximize efficiency.

Although for some projects we have achieved a high rate of productivity and speed, we need to remain flexible and remember that the same approach may not perform so well on another project even after careful tailoring. We aim to be efficient and productive but we should always remain open to the unexpected and embrace every challenge as a learning opportunity; it will enhance our future performance and will equip us with more troubleshooting tools and problem-resolution techniques.

Acknowledgements

Special thanks to my wonderful colleagues for taking time to peer review and proofread an earlier version of the manuscript.

Aaron Mayes, Visual Materials Curator, University of Nevada – Las Vegas
Carrie Gaxiola, Nevada Digital Newspaper Project Coordinator, University of Nevada – Las Vegas
Cory Lampert, Professor and Head, Digital Collections, University of Nevada – Las Vegas
Emily Lapworth, Digital Special Collections and Archives Librarian, University of Nevada – Las Vegas
Kelsey Lupo, Library Technician II, Digitization Lab Manager & Student Supervisor, University of Nevada – Las Vegas