

Digitally Archiving History: A game plan for large, unruly archival collections with limited staffing

Virginia A. Dressler; 1125 Risman Dr, Kent State University, Kent, Ohio, 44242

Abstract

In this paper, an outline of workflows currently in use for the digitization of a large archival collection at Kent State University is provided. A blend of both in-house digitization and outsourcing methods used to achieve the goals of the project with limited staffing and resources. Considerations of the format, as well as the capabilities and available equipment and software, are made throughout the process in regard to institutional capabilities. Selected workflows for digitization are embedded into the daily routine of the staff, reflected in a growing digital archive that is available for patrons regardless of location. Materials without copyright restrictions are uploaded into a openly accessible digital repository, while more problematic materials are stored offline as a preservation copy and the digital copies noted on respective finding aids. A small-scale digitization space has been created for the purpose of in-house scanning, and a process to select material from reference driven selections as well generating reports from a patron database to gauge the most highly used items has been refined to tackle the large task of digitizing a complex archival collection, one piece at a time. Further, the notion at the conclusion of the project is to apply these practices and workflows to other archival collections.

Project

At Kent State University, there is a single event which has loomed over the university and the town for forty-five years, namely that of the Kent State Shootings on May 4, 1970. To briefly recap the event, the National Guard were called into the city after a ROTC building on campus was burned to the ground on May 1st. At the time, the student protests were centered against the United States involvement in the Vietnam War, and invasion into Cambodia in late April of 1970. Three days after the ROTC fire, tensions quickly escalated and through a series of events and subsequent protests, the Guardsmen fired into a crowd of students, killing four students, and injuring nine. The shootings triggered a series of succeeding protests and events, in campus strikes and protests across many universities in the United States. The shootings also sparked what some historians believe is the turning point in the Nixon administration, creating strong political and social divisions in the nation.

At the Department of Special Collections and Archives at Kent State University, the May 4 Collection has proven to be one of the most researched and accessed collections in the department. The Collection also has many complicated issues throughout the more than one hundred subcollections, with complexities of copyright and usage restrictions. Over the past decade, the University Libraries has begun initiatives to create a digital archive of this collection, done so with a limited number of hours and staffing to dedicate to the task¹. The current digital collection

primarily has revolved around images and audio (oral histories), representing a small slice of the total collection. In this paper, tactics are presented which have been used over the years to strategically tackle a complicated archival collection and create a curated digital archive, using in-house production methods, reference request driven digitization and grant funding to complete the enormous task, with corresponding workflows. This paper will also highlight the decisions on which items from the Collection will be included in the digital archive and the rationale behind the selection decisions, as well as the scenarios with sticky copyright issues which have led to the offline storage of reformatted digital content for preservation purposes. Large archival collections can be quite overwhelming to undertake particularly for a small staff, though the approach our institution has taken thus far has been focused on how a small department can strategically, and realistically, initiate a modest digitization initiative successfully.

Background

First, an overview of the collection is provided to highlight the complexities, and also illustrate one method used to make selection for digitization. The May 4 Collection is an archival collection comprised of mixed media types including; documents, oral histories, newspaper clippings, magazines, photographs, art, flyers, banners, posters, manuscripts, moving image, reel to reel audio recordings, FBI reports, and various artifacts and memorabilia (Figure 1). The Collection is over 300 cubic feet, with varying usage policies, donor restrictions, copyright concerns, as well as an assortment of formats. The collection is heavily referenced by library patrons both on-site and off, and the department has utilized an internally designed patron database to determine the most frequently pulled subcollections and material, as to ascertain which particular items are of the highest usage. The reports generated from the database allow our staff to make selections from the May 4 Collection subcollections as to which items are of the highest interest to the patron base.

Next, we will highlight the existing accomplishments in regard to digital projects. The May 4 Digital Archive

<http://www.library.kent.edu/special-collections-and-archives/kent-state-shootings-may-4-digital-archive>.

is a digital image collection, pulling together photographs from five subcollections, which are free and clear of copyright concern.

These subcollections include: Chuck Ayers photographs, Ohio State Highway Patrol photographs, Frank Smith photographs, Ralph Solonitz photographs, and the University News photographs. These photographs represent a small portion of the overall photograph collection, though has also proved to be some of the most requested material from patrons. The approach of reference driven digitization in this scenario solves the issue of providing routinely

¹ Until the Spring of 2014, there was not a dedicated Digital Projects Librarian at the University (whose position is currently split between Special Collections and Archives digital initiatives and other digital projects at the library. There are five full-time staff and

faculty in the Special Collections and Archives department with two part-time positions (one permanent and one temporary), though none of these positions have primary job duties of digitization.

requested items, by capturing high resolution images while the staff fills a reference request, uploading the image to the digital repository and thus enabling use and access beyond the limited reading room hours and staff availability. This approach has proven to work directly into existing department workflows in the department with a few extra steps for digitization and ingestion processes.

Once an image is scanned for the purpose of a reference request, the image can be easily added into the digital archive at full resolution. We have created a basic workflow the project (Appendix A), scanning specifications and a metadata style guide (Appendix B). Once the image is uploaded, the finding aid is also updated with the permanent URL at the item level. Patrons will then be able to search directly for the image from the digital repository interface located on the department website, or alternately from the various finding aids on the department website (Note: both of these options are discoverable from web searches, as the repository is openly accessible, and the websites are completely indexed). In this process, we acknowledge that content is found by different avenues (digital repository/department website/web search, etc.), depending on the researcher and search method.



Figure 1. View of the May 4 Collection at Kent State University

The May 4 Oral History collection is comprised of over

<http://www.library.kent.edu/special-collections-and-archives/kent-state-shootings-oral-histories>.

130 oral histories from Kent State faculty, alumni, staff, administrators, National Guardsmen, police, hospital personnel and others, recorded over the past 25 years. The audio collection is almost completely transcribed by staff manually (95% complete at the time of this writing), and speaker release forms provided to allow the content to be shared openly online. The oral histories have been captured and managed in-house with regular additions throughout the year. Staff and faculty from the Special Collections and Archives department complete the entire process from start to finish; from conducting the interviews, to capturing the metadata, to the transcription of the audio files. Specific project workflows are created to complement each stage of the process (Appendix B). The process is slow and arduous, particularly for a small department, though the outcome is an entirely searchable base of oral history files.

Project specific workflows, as sampled in the Appendices, are created for the purpose of in-house digitization. Important to note, much of the metadata work in particular relies heavily on the subject

expertise of the professional library staff in the department to create local controlled vocabulary specific to this collection, and also a substantial awareness of patron needs. While this aspect is an element that is particular to the Special Collections and Archives unit at Kent State, some of the process-oriented workflows are more generalizable to apply to other collections. The digital archive to this point has been a reflection of work completed within a small amount of allotted resources (resources including personnel, time and equipment), and also exemplifies the coordination of a small in-house production unit, which is essentially a few higher quality commercial flatbed scanning stations and imaging software. As such, our team has also designed similar workflows to be applied to outsourced projects, as described below.

In many of the outsourced projects, the Special Collections and Archives unit is tasked with many pre-digitization tasks such as creating item level descriptive information (often taken from existing finding aids when available), collating original material and organizing the physical content for shipment. The Digital Projects Librarian is also tasked with providing technical specification to the outside vendor as well as post-production quality assurance, preparing files and metadata records for ingest and ensuring the archival master files are put into offline storage. The post-production digitization work can often be as time consuming as the primary scanning and cataloging work that is outsourced.

In the Spring of 2015, the department secured a small state grant from the Ohio Historical Records Advisory Board to outsource and digitize the reel-to-reel material from the May 4 Collection (119 in total). The reels are particularly fascinating examples of primary source material on an obsolete format, many containing the live coverage and footage recorded just before, during and after the events of May 4. The decision to select a particular format was in this case equally for both preservation and access reasons. Neither the department nor library had a reel-to-reel player to even provide in-person playback of the materials for library patrons. While the audio content was being preserved and housed properly, researchers did not have a mechanism to listen to the reels on-site and the staff and faculty of the department did not have a way to access the audio. Digitization of the material became a two-fold solution to a problem; creating a method to listen to the content, and providing access to aging media formats within the digital library.

Our project team identified an outside vendor with an excellent reputation for digitizing reel-to-reel material (George Blood Audio LP), since in-house production was not feasible for reel-to-reel material without extensive cost for the project set-up or professional media specific expertise. Our working group also determined that we did not have substantial collections of reel-to-reel media to justify the cost of the necessary equipment for capture and processing. Once completed, the audio files were uploaded into the content management system, which is also synced with an offline storage solution. The reels without copyright restrictions will be then published online and shared in a new digital collection, which was the majority of the project. Our team also created a set of in-house quality assurance (Appendix C) to run on the files once the vendor completed the pilot and remaining reels.

Also during the spring semester of 2015, the Head of Special Collections and the Digital Projects Librarian began an initiative to tackle other obsolete and aging audio-video formats from the collection. The first format identified for the project was magnetic tape media: namely, the VHS, Beta and U-Matic content. Over the course of the spring semester and summer of that year, the Digital Projects Librarian (with the assistance of student workers) completed this first phase of capture with some inexpensive hardware additions and supplies for a student workstation (namely, a Canopus video converter, a VHS machine and head cleaner) and created a new workflow (Appendix D). Since the library did not have a set-up to capture some video content, our team also utilized a Digital Lab in

the neighboring Library and Information Science department for the Beta and U-Matic tapes. In this scenario, the content was from a single subcollection (Films and videos related to May 4, Boxes 95-95C). The majority of the collection has heavy copyright and usage restrictions, so the purpose of this digital project was to create a

preservation copy for an offline dark archive according to Section 108 under United States copyright law. The small percentage that does not have restrictions will be published online and shared broadly in the digital library in the coming year (Table 1).

Table 1: Chart of current projects to date with tactic, staffing and timeline

Name of project	Number of items to date	Tactic, staffing and timeline
May 4 Digital Archive (primarily image-based)	572	Primarily reference driven; Student workers digitize in-house and staff create metadata. On-going, rolling basis to content management system since 2012 (digitization began in 1995, though without standards or a digital repository)
May 4 Oral History	133	Invited or requested sessions; Professional library staff conducts interview, captures audio, transcribes and creates metadata. On-going, rolling basis since 1995 (analog project began in 1990)
May 4 reel-to-reel audio	119	Outsourced with grant monies and matching internal monies; Staff reviews audio files for quality assurance, creates metadata and uploads to the content management system. March 2015-December 2015
May 4 video (VHS, Beta, U-Matic)	105	Identified video formats to digitize in-house; Student workers capture, Staff create metadata records and review files. January 2015-June 2015 **8.5% of total project is out of copyright and eligible for inclusion in the digital repository

In October of 2015, a project team submitted a grant to the National Records and Archives Administration for the Digital Dissemination of Archival Collections. This grant (if awarded) will be the largest push to date to digitize the May 4 collection. The grant application focused solely on print material from the most heavily accessed material from the collection (as indicated from the patron database reports), pulling content from over 30 subcollections. The proposal would revolve around an in-house digitization tactic with requested funds for additional equipment and temporary personnel. Since many of the sub-collections are heavily used in both in-person and outside reference requests, an in-house approach was the best solution to retain control over the collection while digitization is underway. Further, the grant is focused both on primarily material out of copyright, but also highlights the known collections of interest for patrons.

Currently, our project team is in the process of creating a plan to digitize the remaining audio-video materials from the May 4 Collection. First, the Head of Special Collections and Archives completed a survey of the collection and noted the quantity, type and location of the remaining formats (mainly, audio cassettes, LPs, 8 track, DAT tapes, and magnabelts). The plan to digitize items from the survey that we are able to complete in-house, namely the audiocassettes at this time. We will then create a project specific workflow and determine benchmarks for digitization. For the remaining materials, we will seek out grant funds (or internal funds, if available) to outsource the digitization of the other mediums, since in-house capture is likely not feasible, much like the decision to outsource the reel-to-reel material. Our team weighed factors such as the cost of additional equipment and physical condition of the original materials in the outsourcing decision. Additionally, our team determined the other media types would be better served if audio and video professionals with proper equipment and training captured these materials.

On a platform end, we have also recently initiated a plan to migrate the digital collections from a custom, in-house PHP designed solution to the open-source Omeka platform (Migration completed November 2015). This move will allow for inadequacies of the current solution; mainly: providing digital content as openly and fluidly as possible with an OAI-PMH enabled repository, having a comprehensive search functionality (previously limited to searching only within a single collection), and providing a platform which will adapt to increasing audio-visual digital collections and other formats (batch ingest options, streaming video interface potential and a method to easily create digital exhibits with current content). With increasing digital assets, it was paramount to have a system that adapted to the needs of the library and digital collections over time. And as a

final note, our library does have programmer personnel in place to address the needs of implementing an open source platform and tailoring the system to our institutional needs.

Conclusion

To summarize, the methods used to date at Kent State blend a variety of approaches to digitize a large, complex archive, including both in-house production and outsource solutions. Decisions are often on going to determine selection for digitization, and are done on a case-by-case basis in our regular working group. Additionally, our team has taken an inventory of existing equipment in the library, as well as taken into consideration the opportunities available in neighboring departments and units on campus to assist in completing project goals on limited budgets and personnel. The primary focus thus far has been on copyright free content or aging, obsolete audio-visual formats. In particular for the audio-visual material, there is a more pressing need to reformat content in the event of media or playback loss. Using this blended method of approach to a single, large archival collection, we have been able to digitize over 500 images using a patron driven digitization process with very limited resources. This success has been attained with very limited staffing by adapting to the potential of a small staff, blending production into regular processes like reference requests, and using grants to help supplement the cost of outsourcing. As well, our team has been able to capture over 130 oral histories since the early 1990s and fully transcribe almost all of the files with only a few staff on the project on an irregular, part-time basis. While our initiative may pale in comparison to larger digital projects at better-staffed and funded institutions, it does demonstrate that the hurdle to provide access to regularly requested items through a digital repository could take place with some time, patience and a minimal level of set-up and investment.

We continue to look systematically at the May 4 collection and create a sustainable model for in-house production based on a known user need and also with an awareness of the various formats and mediums. With limited resources and staffing, it is very possible to tackle a complex archival collection with success and blend these initiatives into daily operations and create a small but mighty, growing digital archive. Our hope and goal with the digital initiative is to share the new digital objects as openly and broadly as possible with our patrons, and in so, alleviating staff time for other initiatives and needs of the department. The strategy presented in this paper is applied to a small department within an academic library with successful results over time, and will be used on other collections.

Appendices

Appendix A

Basic Workflow for May 4 Digital Archive

- Review photographs to be scanned. Note any copyright or donor restrictions in making selections and check that image is not already in the digital archive.
- If the finding aid description is outdated or inaccurate, edit finding aid information
- Create digital object master file using scanning specifications
- Create derivative files (JPEG and thumbnail images are created automatically in the system; watermarked TIFF is created by staff)
- Create metadata record for digital object (use our metadata style guide for format to use for each field)
- Make image link edits to appropriate finding aid for collection
- URL format should be as follows:

http://www.library.kent.edu/drc/may4/item_detail.php?itemId=###

Appendix B

Kent State Shootings Oral History Project: Post-Interview processing checklist

Public Services Librarian: Transcribe the audio file, make notes for cataloger on cataloging worksheet and put in Special Collections Cataloguer's mailbox

Public Services Librarian: Update "Recent Interviews: Progress Chart," showing that interview has been transcribed

Public Services Librarian: Move transcribed audio file to "To Be Uploaded" folder

Special Collections Cataloger: Upload audio file into metadata input form, create the metadata, save as draft

Special Collections Cataloger: Inform University Archivist that transcript is ready to "go live"

University Archivist: Inform interviewee that his/her oral history is ready to "go live" (via email, phone, or letter)

University Archivist: Inform Special Collections Cataloger when notification process is complete

Special Collections Cataloger: Upload the transcript onto our website

Special Collections Cataloger: Complete metadata input form, and add URL of transcript and save as completed item

Special Collections Cataloger: Inform UA that interview is fully "live"

University Archivist: Create Facebook or other social media post highlighting new addition

University Archivist: Move files from "To Be Uploaded" folder to their permanent location

University Archivist: Update the "Recent Interviews Progress Chart"

University Archivist: Inform Department Head; she will highlight new interview on website

Appendix C

Steps for Quality Assurance on outsourced audio

- Check the Preservation audio file and ensure that the capture was done to the provided specifications (Broadcast WAV file with PCM encoding, 96 kHz/24 bit resolution)
- Check that the checksum file (MD5) and access file are present for each item from the vendor
- Play the entire audio preservation file from start to end (if time allows). This process will also be useful for capturing additional metadata, but may not be doable due to the size and timeline of the project.
- Tip- Listen for the overall quality and flow- Make any notes on the project spreadsheet and if needed, replay the original to see if the issue is present on the original copy, if playback mechanism is available.
- Additional playback test: Start the playback of the file, and then skip around to at least a few different places on the track to ensure that the capture is complete and audible.
- Notes on playback include:
 - Are there any skips or gaps that may have resulted from either the encoding or from the original capture? (Need original to compare)
 - Are there any issues with the audio
 - Ex.: Audio coming in from only one channel (note L or R), muted, unclear, equal volume levels, background noise, etc. Some reels may have been taped over, resulting in a undertones from the previous recording.
 - Note if there are two sides to the audio
 - Flag any potential issues and report to supervisor (Mismatched content to the container information, corrupt audio file)
- While listening, check the file information against the known metadata captured on the project spreadsheet. Update the file where needed with supplemental information.
 - There may be some introductions on narrators, interviewers, guests, etc. Be sure to note these names.

Appendix D

May 4 VHS workflow

1. Transfer tapes in batches of 20 from Special Collections and Archives to the locked storage area for Digital Projects
2. Log items into project spreadsheet using information from finding aid and note the storage location in the Digital Projects locked cabinet.
3. If any BETA or UMatc tapes are in the batch, alert the Digital Projects Librarian for assistance. These will be captured in the Digital Lab (School of Library and Information Science)
 4. Run the VHS Head Cleaning tape before each capture
5. First check that tape plays correctly in VHS player and ensure the tracking adjustments are made before moving on to capture.
 6. Refer to capture guidelines for the digitization specifics
7. Notify digital projects librarian if there are any issues during capture (tape pulling, unclear picture, etc.)
 8. Create derivative files once the batch of 20 is complete
 9. Return tapes and get a new batch

Author Biography

Virginia A. Dressler is the Digital Projects Librarian at Kent State University in Kent, Ohio. She holds a Master's of Library and Information Science from Kent State University, as well as a Master's of Art Gallery and Museum Studies from the University of Leeds. She has managed a large variety of digital projects and initiatives over the past ten years. The author can be reached at vdressle@kent.edu.