



Managing and Maneuvering Mass Digitization

Robin C. Pike, University of Maryland Libraries
rpik@umd.edu
MARAC, Rochester, NY, April 25, 2014



Delaminated glass disc (not from UMD collections) (photo by author)

RCP, MARAC S16 April 26, 2014



Our recorded history is at risk.



Part of my last large acquisition at a former job, featuring 16mm film, and U-Matic and VHS tapes (photo by author)

RCP, MARAC S16 April 26, 2014



In each of our repositories, on our shelves, sit boxes of jumbled cassette tapes mixed with sticky shed open reel tapes and 8-track cartridges. VHS and U-Matic tapes are stacked on open shelves. In the farthest corner of our shelving, away from any other materials, sit di-acetate films that reek with vinegar syndrome. We know preserving media is a problem, both financially and because most of the time the containers don't have enough descriptive metadata for us to gauge if the content is valuable enough to justify the cost of preservation and digitization. We're embarrassed because we don't want to admit that we can't do enough to save the history we're tasked to protect. And even institutions with a preservation plan and money are not immune to loss. Several large-scale studies by leading institutions have confirmed our worst nightmares—we have about 10-15 more years to preserve what we really want to preserve, and the media might not even last that long. So what did we do?



What can happen in five years, featuring 16mm film, and U-Matic and VHS tapes (photo by author)

RCP, MARAC S16 April 26, 2014



We started somewhere. Because doing something is better than allowing the sticky open reel tapes and stinky films to sit for another five years.

The Big Picture

- History of Digitization and Digital Projects
 - Early 2000s
 - 2005: Office of Digital Collections and Research
 - 2009: Digital Collections
 - 2012: Digital Conversion and Media Reformatting and Digital Programs and Initiatives



<http://hdl.handle.net/1903.1/1978>

RCP, MARAC S16 April 26, 2014



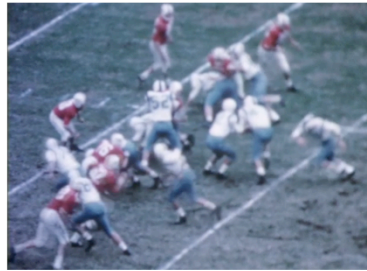
The UMD Libraries have been participating in digitization since the early 2000s. It started as ad hoc digitization in Special Collections to fulfill reference requests, tracked with home-grown metadata in an Access database. It later evolved to the Office of Digital Collections and Research in 2005, the first organized effort, which led to the design and implementation of Fedora, best practice standards, boutique digitization projects, and the creation of several collections and projects that would be added to over many years, such as AlbUM—images and film from university collections. Digitization operations were then moved back to Special Collections in the Digital Collections unit during hard financial times.

In February 2012, the Digital Conversion and Media Reformatting Department (or DCMR) was created in the Digital Systems and Stewardship Division to better serve digitization activities for all collections in the seven campus Libraries and Priddy Library at Shady Grove. DCMR seeks to support the Libraries' collection development goals and strategic priorities for preservation and access by working with collection managers and subject specialists to digitize collections of all formats through a centralized, production-based environment.

Digital preservation, repository development, and other digital initiatives were moved to the new sister department Digital Programs and Initiatives, also under Digital Systems and Stewardship, managed by Jennie Levine Knies. Together, we have been planning and implementing new digitization and digital preservation strategies for the Libraries' digital collections.

A/V Digitization

- History of Audiovisual Projects
 - Films@UM (900)
 - Football Films (775)
 - The Jim Henson Works (70)
- Ad hoc
- DCMR—February 2012
 - Centralized, programmatic



<http://hdl.handle.net/1903.1/10277>

RCP, MARAC S16 April 26, 2014



Most of the requests and boutique projects from 2000 to 2012 focused on digitizing what was easily achieved—images, publications, and manuscripts. During this time, the Libraries created three main audiovisual digital collections. 775 football films were digitized through a vendor from money raised by the hard work of the University Archivist Anne Turkos. Almost 900 films in the Films@UM collection are accessible on campus. These films were part of an early effort by our Library Media Services Department to make digital or streaming copies of educational films available to students. Most of the digital copies were provided by the production companies. We continue to add to this collection as our staff develop licensing agreements for additional film collections. The Jim Henson estate donated digital copies of close to 70 of his early works to Special Collections in Performing Arts.

While we have continued to add films to these three projects, they were not created as templates for future moving image digital collections; the metadata fields vary between all three collections. Aside from these three projects, other small audio and video initiatives existed in both our Performing Arts and Broadcasting libraries, but neither effort used consistent practices, had constant staffing, and did not contribute the results of what they digitized to the Digital Collections Repository. This inconsistent practice was not sustainable, due to the varying file standards, metadata, and digital storage used for the files. One of the primary reasons for the creation of DCMR and my hire was to coordinate and centralize all digitization efforts for all formats.

Planning, People, Policies & Procedures



- Read History
- Identify Stakeholders
- Short- and Long-term Plans
 - Staggered implementation of services
 - Included space and facilities redesign
- New FTE, New Students
- New Policies and Procedures
 - Digital Preservation Policy:
<http://hdl.handle.net/1903/14745>

RCP, MARAC S16 April 26, 2014



I started out by reading every document from every department that had performed digitization either on-demand or through projects, formally and informally, and meeting with more than 20 librarians and staff that I identified as digitization stakeholders—the people who had been participating in these processes. After analyzing what people were doing and wanted to be able to do in the future, I developed an immediate and long-term plan for digitization in the Libraries, focusing on a staggered implementation of growing services. We had been systematically digitizing only image and text materials but based on the stakeholder interviews, I realized that we needed to do the same for audio and video formats. I also realized that in order to implement this ambitious plan, I would not only need to add equipment and staff, but implement a space and facility re-design.

I hired a two-year contract librarian to develop audio and video digitization stations in August 2012. By the fall of 2012, we had moved the Digitization Center to a larger location, added one new flatbed scanning station, and the beginning of an audio digitization station with one audio digitization pilot project to test our workflows. One year later, we had two fully functional audio digitization stations in production for projects and requests, and the beginning of a video digitization station. Our video digitization station can currently digitize three families of formats, and we are finalizing procedures and workflows.

As we developed additional digitization capacity, I also hired a tenure-track librarian to perform quality assurance, assist with managing students, in-house digitization requests, and in-house projects, as well as to further document and develop procedures. For two years, whenever one student digitization assistant would graduate and leave, I replaced them with two. But I don't think of students just as students—I think of them as the potential hours that they work, which comes into play later in my plan. I also created a quality assurance student position to assist the Digital Librarian with quality assurance for our increased production.

I continue to meet with a growing list of digitization stakeholders in the Libraries (37 this year) to make sure they are aware of our expanding services, to collect projects for the coming year, and to help them plan and implement these projects.

Along the way, all of these new services were supported by updated best practices documents and service policies. I have been drafting an overall plan for digitization at UMD Libraries over the past two years, which is currently under review by our administration. After approval, it will be posted in DRUM, the Digital Repository at the University of Maryland, for download by anyone who wants a more detailed look at what I'm trying to build.

In-house v. Vendor

- Planning In-house Capacity
 - What we have?
 - What works?
 - Business case for digitization
 - Expertise
 - Equipment costs
 - Preservation issues
 - Rate and cost of digitization
- Formats:
 - 1/4" open reel tapes recorded at 15, 7 1/2, or 3 3/4 ips
 - ADAT
 - Cassettes
 - CDs
 - Digital Audio Tape (DAT)
 - Grooved discs (not acetate or lacquer)
 - Minidisc
 - DVCAM/MiniDV
 - VHS/S-VHS (Standard and Long-play)
 - Betacam/BetacamSP

RCP, MARAC S16 April 26, 2014



One of the most important decisions I had to make was what our in-house digitization capacity would be. I listened to and gathered lists of projects from all of our media collection librarians. I examined the lists of legacy playback equipment that we could use, what was operable or fixable for a fair price, and then I ordered the beginning of the hardware and software for three audio digitization stations and one video digitization station, which would fit our predicted needs over the next 5 years. Finally, I developed a list of equipment to include in the setups for our two-year contract Media Digitization Librarian. This list was based on a business case analysis for best use of our available resources. While we have additional legacy equipment, I intentionally decided that we would not pursue in-house digitization of particular formats. We will not digitize in-house where the expertise is beyond our means (styli selection for lacquer discs), equipment that is too difficult or too expensive to maintain (maintaining 2" Quad machines costs approximately \$20,000 every 500 hours of playback), and when the condition of the material is too complex (mold). We will digitize formats when we can keep the cost down with student labor. We can currently digitize the following audiovisual formats in-house.

Rate and Cost of Digitization



- Statistics
- Hours Worked
- Average Salary
- FY15 Project Production Timeline
- <http://dsuimd.wordpress.com/2014/03/13/project-planning-rate-of-digitization/>

RCP, MARAC S16 April 26, 2014



As I mentioned previously, I sometimes think of my students and staff in terms of hours and rates. While this may seem impersonal, it's ideal for budgeting and planning projects. Using the statistics we gather for the amount of files and metadata records of static and audio objects that we have been creating, and dividing these by the hours worked by my staff in a month, we get the average rate of digitization. By dividing this by the average of their salaries we get the cost of digitization. I am currently using these rates to plan an annual production timeline for all in-house digitization projects for fiscal year 15 (this July through next June). I am also using this timeline to help prioritize projects across the five campus libraries that are currently using our in-house digitization services. I'm also using this cost to calculate whether it will be cheaper to digitize a project in-house or through a vendor. For more information, please check out a blog post I wrote for our division's blog.

Metadata Workflow

- Description from EAD, MARC, Databases, Inventories
- Reuse, Export Metadata to Tracking Spreadsheets
- Multiple Checks
- Batch Ingest and Archiving

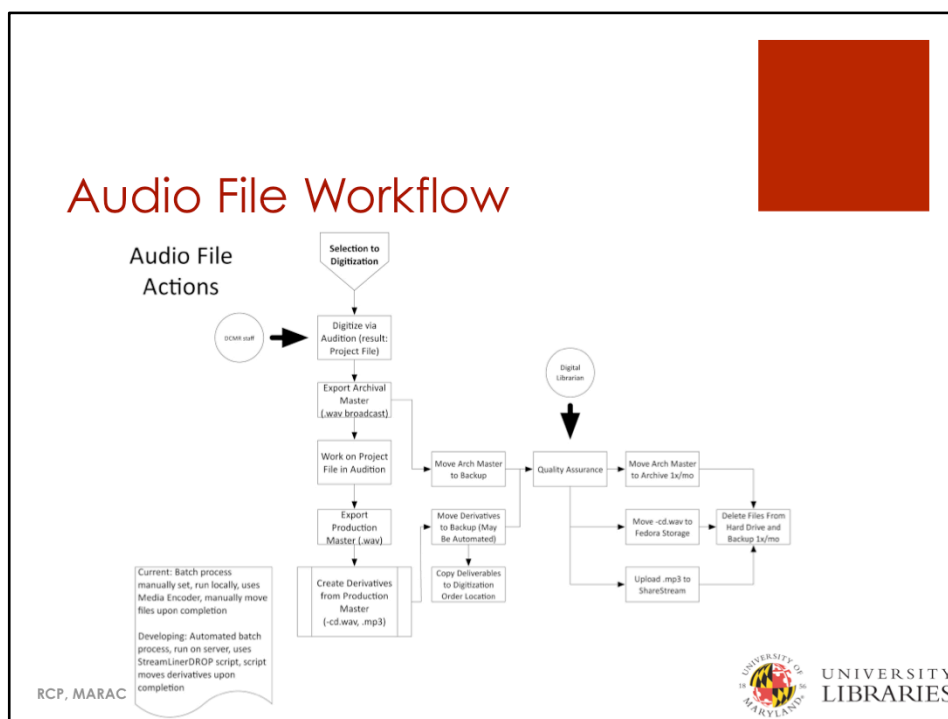
1	FileName	DurationMasters	DurationDerivatives	DigitizationNotes	DigitizedByDept	DigitizedByPers	DateDigitized	PlaybackDevice	Anal
				transformed at 48 kHz/24 bits; tape contained: stems; beant-024608-0001 = bass drum; beant-024608-0002 = snare drums; beant-024608-0003 = cymbals; beant-024608-0004 = drum overhead; beant-024608-0005 = electric guitar 1/bass guitar; beant-024608-0006 = bass guitar/electric guitar 1; beant-024608-0007 = electric guitar 2; beant-024608-0008 = vocals; beant-024608-0009 is a mid-down; beant-024608-0008 amplified by 10 decibels in					
3	beant-024608-0001	00:43.47	00:35.58	production master	DCMB	Savage, Felicia	2014-01-28	Alexis BlackFace ADAT "Bud's ADAT" BM	
4				transformed at 48 kHz/24 bits; tape contained: stems; beant-024609-0001 = bass guitar; beant-024609-0002 = vocals; beant-024609-0003 = electric guitar; beant-024609-0004 = n/a; beant-024609-					

RCP, MARC S16 April 26, 2014



Like most institutions, we have in-house on-demand workflows for requests and workflows for small projects, as well as workflows for activities we need to complete for projects sent to vendors. Material is digitized to nationally accepted, preservation-level standards, associated with metadata, and ingested into the Digital Collections Repository. Files are sent to patrons as needed via FTP, LAN, or physical media.

What I think is different about our workflows is that we are maximizing the amount of descriptive metadata that can be reused from our EAD finding aids, MARC records, legacy databases and spreadsheets. Before digitization, the collection managers are able to export some data to the tracking spreadsheets we use for audio digitization requests and projects. Our student digitization assistants enhance this metadata, and add technical metadata at the time of digitization. Our Digital Librarian finalizes this metadata during quality assurance and arranges monthly batches of audio for ingest and archiving. Dividing metadata generation responsibilities between the content experts and digitization experts, we are able to play to each other's strengths and make the best use of our time. The monthly batches are semi-automated, meaning a human initiates them, which saves on our upload time for larger files. This process did take a while to develop, so we are currently working through a three-month backlog of digitized files and metadata for final QA and ingest. Based on our current QA rate, we hope to be caught up by the end of May.

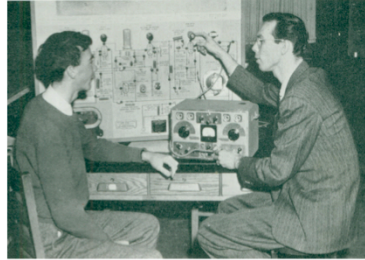


Likewise, our file management process also relies on batch processes. After digitization, a digitization assistant exports the Master File from the project file, as-is, and cleans up the project file to export a Production File, which eliminates large sections of silence, pops, hisses, squeals, and other ambient noises. Once we have a sufficient volume of Production Files (from the day or project), we create derivatives in batch. While this project is semi-automated with Adobe Media Encoder, our Media Digitization Librarian, Henry Borchers, is currently working on a script that will create the derivatives to our specifications, and then moves this batch of derivatives to the correct location for QA in a completely automated process. Finally, the files are ingested and archived in batch, in accordance with our Digital Preservation Policy (<http://hdl.handle.net/1903/14745>).

The important point to note about automation and semi-automation is that not only do they speed up the process, or at least human involvement in the process, they also decrease human error, which saves time by preventing re-work.

Vendors

- Expanding
- Consortial Opportunities



<http://hdl.handle.net/1903.1/1980>

RCP, MARAC S16 April 26, 2014



In addition to expanding in-house digitization efforts, I am also expanding our relationship with vendors. However, I also want to expand our buying power, meaning I want to be able to digitize more material for the amount of money we have. UMD Libraries is a member of Lyrasis, who has negotiated slightly lower digitization rates for their members with a variety of vendors. Lyrasis possesses better buying power with many members than UMD does alone. We have used the Lyrasis relationship with the Internet Archive for public domain publication digitization for the last four years. Because I am comfortable with this relationship, I am exploring using just these vendors for the majority of work we will do this year, including audiovisual digitization.

Digitization Budget



RCP, MARAC S16 April 26, 2014



And this is where some people may start to think, “It’s not going to matter if I can participate in consortial digitization rates. I don’t have any money for outsourced digitization.” The UMD Libraries does not currently have a digitization operations budget. For the last two years, any audiovisual materials we have sent out to be digitized have been paid either through the Preservation and Conservation Department’s budget, or someone’s gift account, which is money they’ve accumulated through good donor relations and fundraising to support their collection area. We’ve sent out a good amount of materials from approximately five different audio collections, which means I do a lot of asking. I’ve also been asking for an operations budget the last two years. The first year I got a consideration. The second year, I got a committee.

The Digitization Initiatives Committee is comprised of five members, one member from each Library division, including our Development Officer. Our goal was to solicit digitization project proposals for outsourced digitization, and to find money for them somewhere, primarily from current sources and gifts and endowments. We received 14 project proposals from across the Libraries to digitize a wide range of formats and subjects. And then we found enough funding to cover some projects completely, and other partially, creating pilot projects which the Libraries can present to donors or granting agencies to fund further in future years. Of the 14, 4 are audiovisual digitization projects.

Growth

- Digital Storage
- Since 2012...
 - 700 more in Digital Repository
 - 420 waiting
 - 7,000 coming
 - But wait, there's more!

RCP, MARAC S16 April 26, 2014



In planning for the cost of digitization, I'm also planning how much digital storage we need for these projects. Since 2012, we have added over 700 audio and video digital assets to our repository from many collection areas. This doesn't include the projects we've received from vendors in the last few months since we are still cleaning up the metadata before we ingest. Adding these will include approximately 420 recordings. We're also a partner in The American Archive project, a collaboration with the Corporation for Public Broadcasting, which will add approximately 7,000 more recordings, or about 9.2 TB some time this summer. While we may be dealing with a lot more TB, we're dealing with the same scalability issue of ingest, storage, and digital preservation as everybody else. However, because we're planning projects annually, we can also budget storage annually.

What's Next?

- Other UMD Initiatives
- Nation-wide Initiatives
 - LC NRPP:
<http://www.loc.gov/rr/record/nrpb/PLAN%20pdf.pdf>
 - NEDCC Survey
 - AVPS:
<http://www.avpreserve.com/avpsresources/tools/>
 - IUB: <http://www.indiana.edu/~medpres/>
- DCMR Blogging:
<http://dssumd.wordpress.com/category/digital-conversion-and-media-reformatting/>

RCP, MARAC S16 April 26, 2014



The Libraries are also participating in activities that support or will eventually lead to digitization. My colleagues have created a Preservation Plan and Assessment that, once completed, will provide them with more in-depth information to assess the preservation and digitization needs of audiovisual collections.

To put all of these plans and processes in perspective, the Library of Congress and CLIR released the National Recording Preservation Plan, an action plan for organizations and institutions to develop tools and training to address rapidly deteriorating audiovisual collections. To assist with this, NEDCC has circulated a survey to gather knowledge about the state of audiovisual collections nation-wide. AVPS, has been building tools to help people assess collections, as well. And I will be joining other institutions like Indiana University Bloomington, who has done a great job publicizing their efforts (<http://www.indiana.edu/~medpres/>) to create transparency around complex audiovisual collections by making my policies and procedures as public as possible, and by regularly discussing DCMR's work on the Digital Systems and Stewardship blog:

<http://dssumd.wordpress.com/category/digital-conversion-and-media-reformatting/>

This is how we've built a scalable process over two years. While it's not perfect and will continue to improve, we will keep forging ahead, digitizing as many collections as possible before they disappear.

Contact Information

Robin C. Pike

Manager, Digital Conversion
and Media Reformatting

rpika@umd.edu

@AVArchivist



<http://hdl.handle.net/1903.1/1910>



UNIVERSITY
LIBRARIES

RCP, MARAC S16 April 26, 2014

Thank you!