

2022

## Scaling Up Video Digitization at the University of Maryland Libraries: A Case Study

Elizabeth M. Caringola  
*University of Maryland, College Park*

Pamela A. McClanahan  
*University of Maryland, College Park*

Robin C. Pike  
*University of Maryland, College Park*

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/provenance>



Part of the [Archival Science Commons](#), and the [Dance Commons](#)

---

### Recommended Citation

Caringola, Elizabeth M.; McClanahan, Pamela A.; and Pike, Robin C., "Scaling Up Video Digitization at the University of Maryland Libraries: A Case Study," *Provenance, Journal of the Society of Georgia Archivists* 38 no. 1 (2022) .

Available at: <https://digitalcommons.kennesaw.edu/provenance/vol38/iss1/1>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Provenance, Journal of the Society of Georgia Archivists by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact [digitalcommons@kennesaw.edu](mailto:digitalcommons@kennesaw.edu).

# Scaling Up Video Digitization at the University of Maryland Libraries: A Case Study

By Elizabeth M. Caringola, Pamela A. McClanahan, and Robin C. Pike

## Abstract

In 2015, a team at the University of Maryland Libraries collaborated on a pilot project to digitize 100 VHS videotapes from the Liz Lerman Dance Exchange collection and, in doing so, established organizational workflows for video digitization and access. After completing the pilot phase of the project, staff who worked on the project published a case study in this journal that articulated a question echoed throughout that process: “Is this enough?” Enough descriptive metadata? Enough technical metadata? Enough storage space? This article will reflect on the pilot project, detail how the digitization specifications and workflows established during the pilot project have changed over the intervening years, and how they were scaled up to digitize and make accessible the remaining 1,125 videotapes in the collection under the auspices of a 2018 National Endowment for the Humanities grant.

## Introduction

In 2015, a cross-departmental team at the University of Maryland (UMD) Libraries collaborated on a pilot project to digitize a selection of video from the Liz Lerman Dance Exchange collection and, in doing so, established organizational workflows for video digitization and access. The initial project digitized 100 VHS videotapes containing rehearsal footage and performances and prompted staff to work through issues that are common when digitizing audiovisual material, such as establishing technical specification and deliverables, creating descriptive metadata before watching the media, and coordinating with staff across multiple departments of the Libraries to keep the project on schedule.

Since completing the pilot, the Libraries actively sought funding to digitize the remaining video in the collection and in 2018 received a National Endowment for the Humanities (NEH) grant for the project, entitled “Preserving and Presenting the Past, Present, and Future of Dance History: Digitizing the Liz Lerman Dance Exchange Archives.” The goal of the NEH grant was to digitize the remaining 1,125 videotapes; enhance descriptive metadata for these videotapes, as well as those digitized during the pilot project; and digitize 211 paper programs corresponding with the performance recordings. The original timeline for description and digitization was 18 months, with six additional months planned for outreach and programming.

After completing the pilot phase of the project, staff who worked on the project wrote an article that articulated a question echoed throughout that process: “Is this enough?”<sup>1</sup> Enough descriptive

---

<sup>1</sup> Bria Parker, Robin C. Pike, and Vincent Novara, “‘Is This Enough?’ Digitizing Liz Lerman Dance Exchange Archives Media,” *Provenance* 34, no. 1 (October 2016): 85-96, <https://digitalcommons.kennesaw.edu/provenance/vol34/iss1/11>.

metadata? Enough technical metadata? Enough storage space? This article will reflect on the pilot project, detail how the digitization specifications and workflows established during the pilot project have changed over the intervening years, and how they were scaled up to accommodate the large size of this project.

The COVID-19 pandemic struck during the second year of the grant, so it is impossible to describe the project without also describing how the pandemic and closure of UMD's campus affected workflows and the grant timeline. Despite those difficulties, the project team was able to create workarounds that will positively impact our workflows for future digitization projects.

## Background

Liz Lerman, a choreographer, performer, writer, educator, and speaker, founded the Liz Lerman Dance Exchange in 1976 in Takoma Park, Maryland, regionally close to College Park, Maryland, and the University of Maryland. The company has produced over 100 innovative dance and theater works and has toured throughout the United States and abroad. In a 40-year career, Lerman was one of the first to recognize the importance of advocacy through her work, which empowered senior adults and those who live with movement-impairing conditions to dance. She advocated that artists and the processes inherent in making art can benefit society. Lerman exerted important influence in the worlds of performance, arts-based community engagement, and cross-disciplinary collaboration. Her work has received critical and scholarly attention and has served as an important reference for other artists and choreographers. The Dance Exchange donated their archive to the Special Collections in Performing Arts (SCPA) at UMD Libraries for preservation, access, and to promote research in 2004.

For these and other reasons, the former SCPA Curator selected the Liz Lerman Dance Exchange collection, specifically the video series of performances, rehearsals, interviews, and promotional materials, along with accompanying paper programs, for digitization. In 2015, the former SCPA Curator worked with the digitization and metadata staff to digitize and create minimal metadata records for 100 videotapes to gain a better understanding of the videotapes' content and current condition. More detailed knowledge about the videotapes' content allowed us to make a stronger case when seeking funding to digitize the rest of the videotapes. A condition assessment of the videotapes helped us to estimate how long we had to digitize the remaining videotapes before they would degrade to the point of being unusable and therefore unable to be reformatted. The results of this pilot project demonstrated that we needed to watch most of the materials in order to create more useful metadata and promote greater accessibility for patrons outside the regional experts. We also learned that 15% of the collection had already degraded to the point where significant portions of videotapes (audio and/or video) had been lost. Due to the research importance of the collection; its need for preservation; existing partnerships between the UMD Libraries, the University, and the Dance Exchange; and a lack of internal funding, the former SCPA Curator and the Manager of Digital Conversion and Media Reformatting (DCMR) applied

for a NEH grant to digitize the entire video series. They applied for funding in 2016, 2017, and 2018, with the final grant application being successful.

Throughout this article, multiple personnel will be named in their roles on the project. A project of this scope was made possible through a team effort of multiple experts, most of whom already worked at UMD Libraries, with a smaller portion being paid for by the grant. The project team included (roles in parentheses):

- Grant Principal Investigators:
  - SCPA Curator
  - Manager of DCMR (overall project manager)
  - Head of the Michelle Smith Performing Arts Library (after SCPA Curator departure)
- Interim SCPA Curator (collection management/metadata creation)
- Performing Arts Librarian (outreach, online exhibit)
- Digital Projects Librarian (digitization project management, vendor liaison, portion of salary paid by grant)
- Head of Discovery and Metadata Services (training and managing the work of the Metadata Content Specialist)
- Metadata Content Specialist (an employee of Liz Lerman, LLC, who was hired as a subject expert for the project using grant funding)
- Archival Metadata Librarian (metadata cleanup and transformation for repository ingest)
- IT Systems Analyst (supported the servers and networked storage necessary due to working offsite during the pandemic closures)
- Systems Librarian (supported the archiving and ingest processes for all of the digital assets)
- Student assistants (paid by the grant, prepared the materials to be shipped to the vendor and performed quality review of the digitized materials and metadata)

This project was successful because within the Libraries we could tap into the expertise of curators, metadata and systems librarians, and IT staff, as well as hire additional personnel to fill gaps in our staffing through the grant funding.

The above list does not contain digitization personnel because this work was done through an external vendor, who was selected for their ability to perform preservation-quality work at this scale within the project timeline. In seeking the vendor statement of work, we asked them to build in flexibility to accommodate unknown tape durations, an unknown number of mixed videotape formats of unknown condition, and possible conservation work. We also asked for minimal intervention (such as digital enhancement) to 1) reflect the videotapes accurately, and 2)

to keep costs down, as this work is charged at an hourly rate<sup>2</sup>. The cost of digitization was the largest cost in the project.

### File specifications, size, and preservation

During the pilot project, project staff worked together with the digitization vendor to decide upon technical specifications and file deliverables for moving images. The selected format for preservation masters was uncompressed QuickTime files (.mov) with compressed MPEG-4 files for access copies. Standards were also determined for the container, extension, bit depth, chroma subsampling, framerate, timecode, audio channels, and audio quality. These file standards were used for subsequent video and film digitization projects for the next five years. The largest such project in the interim years was the Maryland Public Television digitization project that occurred in three phases from 2017 to 2019, totaling 677 videotapes. The Liz Lerman NEH grant proposed to digitize nearly double the number of videotapes. Therefore, it was imperative to consider how file size would impact video digitization workflows and long-term storage and preservation costs.

In 2019, staff in the Digital Services and Technologies (DST) division began exploring more economical ways to capture and preserve digital collections content, including digitized video files. Relating to video files, we explored what standards offered “enough” quality for preservation of the content. To inform the decision, DST personnel met in groups with collection managers across the Libraries and polled them:

- “How do users typically interact with digitized video content?”
- “What are typical usage cases when patrons request copies of files?”
- “Does the content of our collections merit high-definition and high quality, and would the Libraries or users lose fidelity and content if we selected a lower bit rate, standard definition, etc.?”

One of the specifications explored was higher-resolution uncompressed standards, moving from 10-bit uncompressed standard definition (SD) (4:3, 640x430) to 10-bit uncompressed high-definition (HD) (16:9, 1280x720), as some personnel wanted a higher resolution file for their content. Upon assessing our content, we found that the majority of our content is SD and creating HD files from SD original content unnecessarily adds information to the file, creating an artificially large file (also known as file “bloat”). Additionally, managing files of 500+ GB resulted in file management issues for the Systems Librarian and other IT staff (such as during file transfer, checksum generation, etc.).

---

<sup>2</sup> The statement of work read: “Create digital surrogates as they currently exist. The project does not include enhancement or restoration, such as speed and level correction or re-equalization...Item level quality control. Know some items were in poor condition from pilot including sticky shed syndrome and deterioration. Price includes baking if needed for any items and an “Extraordinary Intervention Allowance” of 5% for other conservation treatments. If the number of items exceeds 5% of the collection, additional conservation work will be charged at an additional hourly rate with approval of UMD before starting work.”

Curators also reviewed and compared sample files from recently digitized collections at 10-bit and 8-bit, at SD and HD, to assess how discernable the difference in quality was across these formats. Staff extrapolated how applying these standards would impact the total size of digital collections and the costs associated with digital preservation for video files. After the curators assessed what was possible, we agreed to specify in our vendor digitization technical specifications that SD should be digitized at SD and HD digitized at HD. SD would be acceptable for almost all digitized content because it retained the original aspect ratio of the recording and most viewers would be watching lower resolution streaming files where they would not benefit from the high-resolution preservation file. On the rare occasion that HD files were required for a specific project or patron request, the technical specification provided to the digitization vendor would include HD files in addition to the standard set of deliverables, or if previously digitized, UMD would create the black bars using post-production software, but UMD would not retain and preserve the HD file.

The more difficult decision was to downgrade bit depth from 10 bit to 8 bit. The consensus was that most digitized files look “true enough” to the original recording at 8 bit, though 10 bit could be used for collections that merited the higher quality and bit depth. Making the switch to 8 bit for the majority of digitized video and film collections will save 20% on storage space and result in keeping 90- to 120-minute videotapes under 200 GB to meet preservation file transfer constraints.

At the time, we also explored other preservation file formats. With our vendor, we explored newer technical specifications, such as FFV1, on a pilot of three videotapes. We decided against using FFV1 in the .avi container since the embedded metadata didn't fully import into the Adobe software we use, even though the .avi is cross-platform with Mac, the operating system used by DCMR. We also explored using FFV1 in the Matroska (.mkv) container but Adobe does not support Matroska.

At the conclusion of DST's investigation into more economic digitization and preservation standards for video, the only change that was made was to downgrade the default bit depth to 8 bit. Otherwise, the preservation and access file technical specifications remained unchanged from 2015 to 2019. See Table 1 for the Libraries' standard film and video technical specifications and for the standards used for this project.

When planning the specifications for the Liz Lerman NEH grant application, the SCPA Curator selected the higher 10-bit specification for the Liz Lerman Dance Exchange materials because it provided greater contrast and color range for the performances, which took place on well-lit stages while being recorded from the darkness of performance halls. The result of digitizing the Dance Exchange materials at 10-bit uncompressed standard definition was 91 TB of content on

15 8 TB hard drives. In a contrast of scale, the aforementioned Maryland Public Television 2019 digitization phase produced 25 TB of content on nine 4 TB hard drives and one 2 TB hard drive—a considerable leap in the scale and volume of files and the number of hard drives to work with and an adjustment for our IT support and Systems Librarian.

Table 1: 2019-2020 Digitization Technical Specifications

File	Container	Extension	Resolution	Bit Depth	Chroma Subsampling	Frame Rate	Timecode	Audio Channels	Audio
<i>For most videotape projects and requests:</i>									
<b>Preservation Master</b>	Uncompressed Quicktime File Format	.mov	Native	<b>8 bit</b>	4:2:2	Native	Native, midnight start	Original	PCM, 48kHz, 16bit
<b>Streaming</b>	MPEG-4, H.264	.mp4	Native	8 bit	4:2:2	Native	Native, midnight start	Original	AAC, 48kHz, 256kbps
<i>For Liz Lerman Dance Exchange Grant (and other selected projects or requests):</i>									
<b>Preservation Master</b>	Uncompressed Quicktime File Format	.mov	Native	<b>10 bit</b>	4:2:2	Native	Native, midnight start	Original	PCM, 48kHz, 16bit
<b>Streaming</b>	MPEG-4, H.264	.mp4	Native	8 bit	4:2:2	Native	Native, midnight start	Original	AAC, 48kHz, 256kbps

Note: If the video format is native HD, transferred at HD instead of SD.

Physical deterioration—did we act soon enough?

As reported upon the completion of the pilot project, 100 videotapes were sent to be digitized, and only two were not digitized due to content duplication. This was not the case in the full project. Of the 1,125 videotapes sent, 1,103 were digitized; 17 videotapes were not digitized because they were blank or contained only static, and five videotapes were not digitized due to irreparable physical damage and/or degradation issues. In the five years that passed since the original pilot project, either the videotapes had deteriorated past the point of saving the content or we had not selected the most severely degraded content to digitize in the pilot project.

In the pilot project, both the vendor technician and in-house staff performing quality control noted inherent issues (e.g., beginning cut off, end cut off, audio buzz, audio hiss, audio hum, audio low levels, and video picture breaks) and issues due to degradation impacting 15% of the videotapes of the 98 videotapes digitized (e.g., audio distortion and periodic dropouts). Of the impacted videotapes, only portions of the videotapes were unlistenable or unviewable as they were originally recorded—degradation did not impact the entirety of the tape. For the full project, we saw much higher numbers of inherent issues due to degradation, but except for the five videotapes that could not be digitized, again, only portions of the videotapes were impacted. The technician noted symptoms of physical deterioration as well as signal degradation, and unlike the pilot project, the vendor also noted issues with the video signal in addition to the audio signal. The vendor digitization technician noted the following issues with the videotapes: audio dropout over content (78.5%), audio distortion (24.8%), and video breakup over content (68.8%). Of the videotapes that were digitized, 10.4% needed to be cleaned and 1.4% required baking for sticky-shed syndrome to facilitate digitization.

The increase in degradation rate is backed up by the research performed by Indiana University and Harvard in the original Sound Directions project<sup>3</sup> and continued through Indiana University's Media Digitization and Preservation Initiative<sup>4</sup> and also has been documented through the research supporting the creation of the AVP Cost of Inaction Calculator.<sup>5</sup> Institutions are racing against the clock to preserve magnetic media, and every year we have the potential to lose content, even when it is stored in archival conditions that meet best practice. The project team agrees that we acted soon enough to save the majority of the content, particularly because it was the soonest the substantial funding needed for the project was available. However, there were significant losses to the content.

---

<sup>3</sup> Mike Casey and Bruce Gordon, *Sound Directions: Best Practices for Audio Preservation* (Bloomington: Indiana University; Cambridge, MA: Harvard University, 2007), [https://dlib.indiana.edu/projects/sounddirections/papersPresent/sd\\_bp\\_07.pdf](https://dlib.indiana.edu/projects/sounddirections/papersPresent/sd_bp_07.pdf).

<sup>4</sup> "Resources and Documentation," Media Digitization and Preservation Initiative, Indiana University, accessed December 14, 2021, <https://mdpi.iu.edu/resources%20and%20documentation/index.php>.

<sup>5</sup> Cost of Inaction Calculator, AVPreserve, accessed December 14, 2021, <https://coi.weareavp.com/>.

## Production during a pandemic

### Scaling up from the pilot

Undertaking the largest video digitization project for the Libraries to date meant tackling some challenges in adjusting the digitization production workflow to accommodate such a large number of items and the resulting digital files. After consulting with the digitization vendor, the 1,125 videotapes were split into three batches for digitization to make handling the physical and digital assets more manageable. Multiple batches also allowed staff on the project team to complete quality control review more quickly and move on to metadata enhancement, rather than having to wait for the entirety of the project to be digitized. The Libraries provided the vendor with 15 8 TB external hard drives, each labeled with an identifying number, to deliver the files and metadata. The Manager of DCMR worked with the IT Systems Analyst and the vendor to determine that 8 TB external hard drives would be the most efficient method for data transfer from the vendor to the Libraries because of the cost-effectiveness of hard drives and a relatively slow file transfer speed between the vendor's cloud storage to the Libraries' network storage, particularly for the larger preservation files.

Each of the three videotape batches had a corresponding metadata spreadsheet. These spreadsheets were created using the same data source as the pilot project, an item-level inventory of the collection's audiovisual materials that had been created by an intern from the Dance Heritage Coalition in 2005. The inventory, which relied on handwritten or typed labels on the media, included a title, a date, and usually a brief description of the content. During the course of the pilot phase, it became clear that the descriptive information on the media was not always accurate. However, in order to save time and to not risk damaging any of the physical media, the project team did not play back the media prior to digitization and relied on the inventory, knowing that any incorrect metadata could be fixed during the metadata enhancement phase after digitization by viewing the digital files.

The vendor agreed to send the digital files back in three separate batches aligned with the metadata spreadsheets. As the external hard drives began to return from the vendor in early March 2020, a student assistant from the UMD iSchool was hired to assist the project. They checked 100% of the files to ensure all files were delivered and in the requested format. Additional quality control was performed on a random selection of 25% of the files.

The student assistant originally performed quality control directly from the external drive with the use of a write-protector to ensure the files were not accidentally altered. The student assistant first went through each file and marked that it had been received in the metadata spreadsheet by listing the external hard drives' ID number. This also served to identify which external drive to select should we need to view a specific file later in this process, thus helping us manage and access the large amount of data. Then, the student assistant ran QCTools on the video files to

detect any image anomalies, and performed visual inspection on portions at the beginning, end, and mid-point of the randomly selected files. QCTools is open-source software that helps users detect corruptions or compromises in the results of analog video digitization or in born-digital video.<sup>6</sup> Additionally, the vendor provided condition reports that we consulted during the quality control process in order to determine if anomalies in the digital files were due to the condition of the physical material or due to an error during the digitization process

### Adjusting workflows due to the COVID-19 pandemic

When UMD and the digitization vendor closed down during mid-March 2020 due to the COVID-19 pandemic, the project was in the midst of receiving files from the vendor for review. Only the first two batches of digital files were completed and delivered earlier in the month. Additionally, the student assistant began work on March 10th and only had two days of training in the office before transitioning to full-time remote work. All of this led to changes in the quality control workflow.

Knowing that campus closure was imminent, the Manager of DCMR and the Digital Projects Librarian worked quickly to upload all of the streaming files and condition reports received thus far to a server that could be accessed remotely. Preservation files were not loaded as they would take up too much space on the server, but the quality control phase advanced with the assumption that if the streaming file, which was derived from the preservation file, was fine, the preservation file should be fine, too. They requested a laptop, loaded with all of the software needed to perform quality control on the video files, for the new student assistant and continued training through virtual meetings and email. The metadata spreadsheets were already loaded to a Google Drive and could be easily accessed and updated from anywhere. This meant that quality control for the first two batches and part of the third was able to continue remotely as soon as the campus closed. In addition to frequent communication via a chat tool between staff and the student assistant, the student was able to utilize the AV Artifact Atlas<sup>7</sup>, a tool maintained by the Bay Area Video Coalition, to identify many issues with the videotapes.

The digitization vendor reopened in June 2020 and was able to digitize and deliver the rest of the third batch. Since UMD employees were still working remotely and not able to receive mail onsite, Libraries IT staff created a server space where the vendor could use File Transfer Protocol (FTP) to remotely deliver the streaming files and condition reports. This allowed quality control to continue, and the student assistant completed quality control for the remaining batch of videotapes in July working from the files on the FTP server. The project's Metadata Content Specialist was also able to utilize the server containing streaming files in order to watch the digitized videotapes and continue metadata enhancement (discussed in the next section) during

---

<sup>6</sup> "Getting Started with QCTools," QCTools, MediaArea, accessed December 14, 2021, [https://mediaarea.net/QCTools/Getting\\_Started](https://mediaarea.net/QCTools/Getting_Started).

<sup>7</sup> AV Artifact Atlas, accessed December 16, 2021, <http://www.avartifactatlas.com/>.

the pandemic. The 8 TB external hard drives that the Libraries had provided to the vendor earlier in the project were still utilized to transfer the entire set of file deliverables needed for digital preservation. They were mailed to the home of the Manager of DCMR for safekeeping until she was able to return to campus and deliver them to the Systems Librarian for ingest to our digital repository and digital preservation system.

## Metadata standards and enhancements

### Embedded metadata

The final specification for preservation files that the Libraries examined during this time was a new requirement for embedded metadata. During the pilot project, project staff decided to receive XML sidecar files including basic descriptive and technical metadata. We archived these sidecar files with the preservation video files for additional context. We also extracted some of the XML metadata fields, such as the exact file duration, to include in the metadata record in our digital repository.

For the scaled up project, we wanted to include additional embedded metadata fields to more effectively track the content in an offline system to assist with file restore. This was desirable because, since completing the pilot project, the Libraries digital storage had migrated from a tape archive to a cloud-based archive for preservation files, and because staff were also experiencing an increase in requests for preservation files by external researchers. We also desired the ability to edit embedded metadata because, as we experienced in the pilot, many videotapes had incomplete or inaccurate titles and dates. We experimented with editing the metadata embedded in XMP in the .mov file and in the INFOCHUNK of the .avi file. Adobe After Effects worked well to read and edit the XMP metadata best, so we concluded that our decision of the .mov wrapper was still the best for our institution five years after our original technical specifications were selected.

The embedded metadata fields we chose to include were limited to prevent the need for excessive editing in the embedded metadata, as well as in the repository, should we discover the content is different than what the tape label purported to be. We included the following Dublin Core fields: Date, Description, Identifier, Copyright Notice, and Title. We have not yet perfected the post-metadata enhancement workflow to edit the embedded metadata and are doing so on an ad hoc basis. We may further examine the fields we include to prevent metadata editing after archiving. The embedded metadata maps to fields in the vendor manifest spreadsheet:

### *Video and Film Metadata:*

- PBCore XML file for all metadata created during digitization
- XMP Embedded Metadata:
  - Dublin Core Date: Original Recording Date

- Dublin Core Description: Description/Interview Summary
- Dublin Core Identifier: Object Unique Identifier
- Dublin Core Title: Title
- Dublin Core Copyright Notice: "To obtain permission to publish or reproduce, please contact the University of Maryland Libraries."

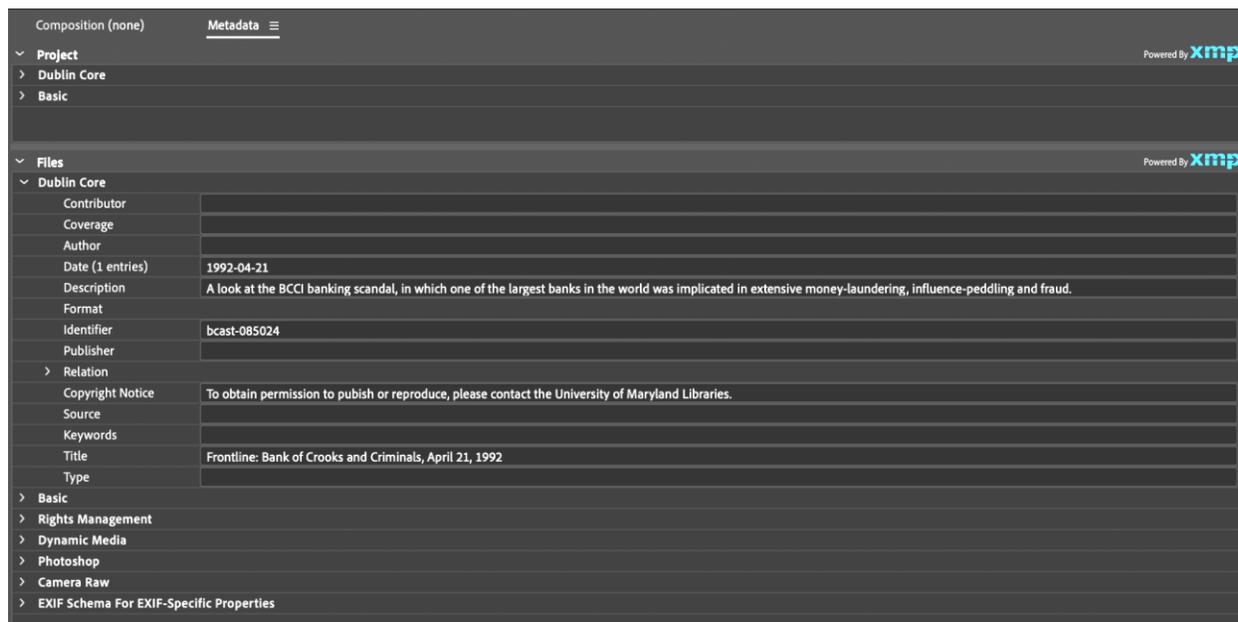


Figure 1: Screenshot of metadata display in Adobe After Effects, showing how the date, description, copyright notice, and title fields have been imported into the XMP embedded metadata in the preservation file.

### Road to “enough” descriptive metadata

As previously described, metadata from the pilot digitization project originated from an item-level inventory of the collection’s audiovisual materials, but this information was not always accurate. This, combined with the fact that more robust metadata was needed to aid in discovery of the materials, made it crucial to view the recordings after digitization and enhance the original metadata from the inventory. Pilot project staff suggested employing subject matter experts, such as master of fine arts students, or using dance-specific taxonomies to make the materials more discoverable to dance scholars without requiring assistance from SCPA staff.

The 2018 NEH grant included a subject expert, the Metadata Content Specialist, a staff member from Liz Lerman, LLC, who would watch the digitized recordings and enhance descriptive metadata for both the pilot project and the current grant-funded project. Because the Metadata Content Specialist was a subject expert, but not a metadata expert, the Head of Discovery and Metadata Services set project standards for metadata enhancement, trained and oversaw his work, and planned to do any data cleanup prior to archiving the files and ingesting them into our

digital repository. This partnership between an expert on Liz Lerman's work and a metadata librarian was meant to ensure that our metadata would "be enough" to provide access to the digital files, while also complying with our internal metadata guidelines and staying on the grant's schedule. However, unanticipated difficulties made this course a little less straightforward.

### Challenges to enhancement

The first challenge, unknown to the project team at the time of writing the grant application, was that UMD Libraries would select a new system for making digitized audiovisual content searchable and stream-able. After a pilot during the 2018-2019 academic year, the Libraries made the decision to migrate streaming media from ShareStream to Avalon Media System to facilitate public access. Although not part of the original project plan, Avalon provided features that were suited to this particular project, such as the ability to easily edit descriptive metadata in a staff interface after being ingested; the ability to add links to related files, in this case to the printed programs that were also being digitized; and the potential for captioning to be added at a later date, which would make the collection more accessible to anyone who is deaf or hard of hearing.

Due to the grant timeline, the Liz Lerman Dance Exchange project would be the first to load files and metadata into Avalon, and existing metadata guidelines and templates had to be adjusted to utilize Avalon's ability to batch ingest files and metadata. During the first year of the two-year grant period, the Head of Discovery and Metadata Services developed metadata guidelines for Avalon, including a template for batch import into Avalon, and created project-specific instructions for enhancing the Lerman metadata records. The project guidelines covered how to construct a title, the correct format for dates, the difference between a creator and a contributor, sample lists of acceptable genre and subject terms, and more. It also provided some guidance on what level of detail was appropriate for the project, especially given the grant's tight two-year timeline. These guidelines were tested by the Metadata Content Specialist as he worked to enhance the metadata created by the pilot project. He worked directly with the Head of Discovery and Metadata Services throughout this process to get clarification on any questions or issues that arose and to try to answer the question of: "Enough?"

For the pilot project metadata, the answer was: "Too much!" The Metadata Content Specialist meticulously added information to each of the 98 videotapes' metadata records, including the titles of the works, location, performers, contextual information about the works, timestamps for when each piece began, and credits for direction, musical composition, choreography, lighting design, costumes, set, and more. While this information is undoubtedly useful and demonstrates the power of combining subject expertise with a close watching of the videotapes, it was clearly impossible to scale up the same level of description to the 1,103 digitized videotapes of the full

project. The Metadata Content Specialist was advised to focus on summarizing the content of a recording when enhancing descriptive metadata for the full project.

Metadata enhancement for the full project was scheduled to begin in May 2020 and was slightly delayed due to earlier delays in digitization and quality control caused by the pandemic. Luckily, the workflows put in place by staff in DCMR to remotely review files also allowed metadata enhancements to occur remotely. The Metadata Content Specialist began enhancing metadata in June, one month behind schedule, using the streaming files on the FTP server and the metadata spreadsheets in Google Drive.

When the project team met at the end of October 2020 to regroup following the departure of two key project team members, the SCPA Curator and the Head of Discovery and Metadata Services, metadata enhancement should have been nearly finished, but it was estimated to be about 50% completed. The timeline for metadata enhancement was extended through the end of January 2021, with an emphasis on cleaning up titles and dates; the project team decided that any other metadata fields could be edited and enhanced after the files and metadata records were ingested into Avalon. Even with the extension and the decision to focus on titles and dates only, 364 of the 1,103 metadata records were not enhanced at all prior to ingest.

Clean enough?

Following the departure of the Head of Discovery and Metadata Services in September 2020, the Archival Metadata Librarian in Special Collections and University Archives joined the project team. She was responsible for ensuring that the enhanced metadata provided by the Metadata Content Specialist adhered to the established Avalon metadata guidelines and for mapping it to the metadata template for batch import into Avalon. Having extended the timeline for metadata enhancement, there was one month in the project timeline to complete this work.

Metadata cleanup consisted of addressing issues that are not unique to this project: fields not entered in the correct format; inadequate titles that would not facilitate discovery, or titles not formed consistently for the same events or performances; removing internal notes about the physical items that are not needed when viewing the digitized files; and removing metadata records for videotapes that were not digitized. With such a short timeframe, she focused on ensuring that fields were formed correctly and would not cause errors during Avalon ingest and making very basic enhancements to titles that may not have been reviewed by the Metadata Content Specialist. Any notations that were not relevant to specific metadata fields were retained in the original metadata spreadsheet but were not mapped to any metadata fields in the Avalon batch ingest spreadsheet.

Two issues emerged that were beyond the scope of metadata cleanup: 1) the presence of duplicate content; and 2) content not recorded or commissioned by Liz Lerman or the Dance

Exchange. From the pilot project, the project team knew that it was likely that duplicate content existed in the collection, but inadequate labels made it difficult to identify them prior to digitization. Also, without digitizing the content or playing back the videotapes, it would be impossible to select the best copy to be digitized. Furthermore, duplicates weren't always duplicates in the archival sense (i.e., an exact copy made from a master tape). Sometimes there were multiple cameras recording a single event or excerpts were copied to make a compilation or promotional video or to be used by other performers. Occasionally the Metadata Content Specialists noted which copy was best, but given the large percentage of videotapes not reviewed, it would have been impossible to identify all duplicates and pick a best copy prior to ingest, so all digitized files were ingested.

Additionally, some videotapes in the collection were not recorded or commissioned by Liz Lerman or the Dance Exchange. They were primarily recordings of documentaries and television programs that were used as source material for Lerman's work. Others seemed to be promotional works that may have been given to Lerman as a sample of a videographers' work, as well as some other videotapes deemed not relevant or having unknown relevance to the Dance Exchange. Neither the Liz Lerman Dance Exchange nor UMD Libraries have the rights to these 83 digitized videotapes. A rights statement of "Copyright Not Evaluated" was applied to these videotapes. If the Libraries is contacted by the copyright holder, permission to stream the files will be restricted to on-campus access only.

Metadata cleanup and mapping to the Avalon template concluded on time and was delivered to the Systems Librarian in March 2021. After working out a few discrepancies between the files and the metadata, the Liz Lerman Dance Exchange files were loaded to Avalon later in the spring. It was the first collection available when the public website officially launched in June 2021.<sup>8</sup>

#### Applying lessons learned during the pandemic to future projects

In August 2021, all vaccinated employees began to regularly work on the UMD campus again. Although in some ways digitization workflows could continue as "normal," some adjustments to digitization workflows made during the pandemic will be incorporated into procedures moving forward.

While DCMR staff previously worked from external hard drives for file checks and quality control, it is advantageous to continue working from FTP servers on future projects. The minimal fee associated with this extra request is justified. Building in this request from the start of a project will make files more accessible to staff and will allow for flexibility if staff are working

---

<sup>8</sup> "Liz Lerman Dance Exchange," University of Maryland Libraries Digital Collections, accessed December 16, 2021, <https://av.lib.umd.edu/collections/1g05fb60f>.

remotely more often in the future. It also decreases the number of times that the external hard drives containing the master files for preservation and streaming files for Avalon need to be accessed; they are accessed once to confirm receipt of files and a second time to begin the digital preservation process. While precautions such as a write-protector were used in the past, having staff perform quality control and metadata enhancement from an FTP server further reduces the risk of losing or altering the files that will be ingested into the Libraries systems.

The urgency to complete this project, despite setbacks, also had some positive effects for the Libraries. While the Lerman digitization project was at the top of the priority list for ingest to Avalon due to the grant timeline, other digitization projects, such as digitized basketball films from the UMD Archives and digitized audio from Westinghouse's Group W, were also ingested shortly thereafter. The work that the Head of Discovery and Metadata Services did in preparing metadata guidelines for Avalon guided the work of batch ingesting these additional collections and eventually will guide the migration of our audiovisual content in ShareStream to Avalon. The newly developed workflow for accessing streaming files and performing quality control and metadata enhancement remotely, though developed in response to the COVID-19 campus closure, can be used in the future to help staff work on these types of projects without working on campus.

In terms of metadata production and enhancement, working with subject matter experts always presents certain challenges. They may or may not be familiar with best practices for metadata creation, and they are prone to getting bogged down in a level of detail that might not be helpful to the majority of researchers, which can subsequently cause delays in the project timeline. Regular communication in the form of meetings or emails to quickly resolve questions and issues as they arise, as well as performing periodic checks on their work to ensure quality and pacing, is essential to preventing these problems. However, given the turnover in project staff and the complications that the pandemic caused in everyone's personal and professional lives, the ideal levels of communication were simply not achievable for this project, and the end result was that one-third of the metadata records were not enhanced prior to ingest to Avalon.

This isn't so much a "lesson learned" as it is a "lesson we already knew" in terms of the importance of communication when working with external partners and non-metadata professionals. Despite the thoughtfulness that went into planning a metadata approach that balanced description and discovery with consideration for project milestones and deadlines, circumstances outside our control prevented us from implementing metadata enhancement as planned. Although the migration to Avalon was also unplanned at the time of writing the grant application, one of the many advantages it brought to this project was the ability for collection staff to edit metadata in an easy-to-use interface without any mediation from IT. So although a significant amount of metadata wasn't able to be reviewed and enhanced as part of the grant project, it can happen at a later time as resources in SCPA allow.

## Conclusion

While the project didn't run as smoothly as desired, and we needed to seek a no-cost extension for the grant due to pandemic-caused delays and staff departures, overall, the project team was satisfied with the result of the project. We were able to alter many of our existing workflows to be successful at a much larger scale for video digitization and have been patient with the added time needed to ingest the volume of preservation files to our preservation repository. We were also successful in shaping metadata practices for Avalon, loading the large number of videotapes using Avalon's batch upload capability, and in making the materials available online. We agree that we acted in time to digitize the videotapes because we were able to capture most of the content. Most importantly, this project has helped us set internal best practices for "enough" metadata so we can balance searchability with project timelines.