

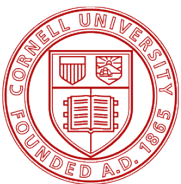
Scotch[®]

BRAND

recording tape



2017 Cornell University Audiovisual Preservation Initiative Survey Report



Cornell University Library



Authors:

Tre Berney, Director, Digitization & Conservation Services, CUL, Digital Scholarship and Preservation Services
Danielle Mericle, (former) Director, Digital Media Group, CUL, Digital Scholarship and Preservation Services
Desiree Alexander, Audiovisual Collections Assistant, CUL, Digital Scholarship and Preservation Services
Karl Fitzke, Audiovisual Specialist, CUL, Digital Scholarship and Preservation Services

Contributions by:

Melissa Wallace, Web Designer and DCAPS Lead, CUL, Digital Scholarship and Preservation Services
Marsha Taichman, Visual Resource Librarian, Cornell University Library, Fine Arts Library
Michelle Paolillo, Digital Curation Services Lead, Digital Scholarship and Preservation Services
Andrew Page, AV Systems Specialist, IT@Cornell
Kizer Walker, Director of Collections, Cornell University Library
Paul Zarnowski, IT@Cornell
Paul Allen, IT@Cornell
Rick Elliker, Media Systems Engineer, Macaulay Library, Cornell Lab of Ornithology
Bill McQuay, Audio Engineer, Cornell Lab of Ornithology
Melanie Lefkowitz, Staff Writer and Editor, CUL
Carla DeMello, Graphic Designer, CUL
Oya Y. Rieger, Associate University Librarian



Executive Summary

Cornell University has vast holdings of unique audiovisual (AV) assets vital to its mission “to discover, preserve, and disseminate knowledge.” These include but are not limited to lectures by Nobel laureates, heads of state, writers and artists, comprising over 12,000 hours of unique material. Cornell’s Lab of Ornithology is home to the largest collection of recorded natural sounds on earth, the only footage of the now extinct ivory-billed woodpecker and other original source recordings essential to research in biology and other sciences. The Kheel Center for Labor-Management Documentation & Archives houses over 40,000 hours of documented labor history, including video of the last known public appearance of influential labor union leader, Jimmy Hoffa. These assets represent significant institutional investment by way of staff time, infrastructure, and capital expenditure, and serve to document research and reflect the legacy of Cornell University. This institutional legacy now faces a very real and growing threat due to audiovisual media degradation and playback obsolescence which, if left unattended, will result in the loss of priceless media assets.

This state of affairs urgently calls for investment. What is needed is a commitment to the conversion of analog audiovisual materials and the preservation of the resulting digital content alongside born-digital materials. In a world of growing opportunity for and reliance on media for teaching and research, we anticipate unprecedented utilization of these materials. Safely accommodating this use will require reformatting. Moreover, in a rapidly changing IT environment, the digital media created through digitization of analog formats inescapably requires some level of migration planning to move content from one format and/or storage technology to another. Adding to the sense of urgency, government grants now require data management planning for research data, including preserving and sharing a grant’s intellectual output.

Cornell University Library’s dedication to maintaining scholarly collections is remarkable and we have nearly three decades of history in the area of digital initiatives and digital preservation, making CUL uniquely positioned to both execute and provide a scalable model for digitization, preservation and access to audiovisual materials. Over a period of 15 months, Cornell University Library conducted a campus-wide, collection-level survey of Cornell’s unique and/or rare AV items, resulting in the identification of 220,147 items. This includes archives and library holdings, pockets of research materials in faculty offices, and departments’ teaching and research data. In the context of research data, CUL’s Research Data Management Services Group’s 2012 survey, “*Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements*,”¹ uncovered that approximately 27% of research data being held on campus was audio or moving image material, formats not commonly not thought of as data. This all requires larger storage and planning for preservation and access.

¹ Steinhart, Gail, Eric Chen, Florio Arguillas, Dianne Dietrich, and Stefan Kramer. 2012. “Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements.” *Journal of eScience Librarianship* 1(2): e1008. <http://dx.doi.org/10.7191/jeslib.2012.1008>



In light of our findings, in order to functionally support preservation and access of audiovisual materials we recommend the following to the Library and partners:

- **A 3-year library funding increase of \$25,000 annually** to the current \$100,000 preservation fund for digitization of library collections and an 18-month position (Audiovisual Collections Coordinator) to create a prioritization plan with the library Collection Development Executive Group, curators, faculty and selectors. The Coordinator would clear and document rights and permissions for digital surrogates to be created as part of the plan and assemble and prepare metadata prior to digitization. The Coordinator would also help in the analysis of tiered storage options for content.
- **A 5-year prioritization plan** for the digitization of Cornell University Library's unique and rare audiovisual materials based upon a range of curatorial criteria, including intellectual and monetary value, uniqueness, fragility, and current storage condition. This would replace the annual Collection Development Preservation Grant cycle with a more streamlined approach to ingest pipelines in our Multimedia Preservation Lab, with metadata and preservation planning in place before digitization occurs. A small internal, discretionary fund will remain for library selectors and curators.
- **Analysis of preservation storage needs and requirements** for digitized and digital-born material created on campus. Audiovisual material is comprised of large data streams, driving the need for scalable storage and preservation solutions and functional access strategies. Also, large scientific datasets are large and complex to deal with, making large, storage a crucial need on campus. Collaboration is already underway between IT@Cornell, the Lab of Ornithology and library staff involved in digital production and preservation on investigating featured preservation storage for Cornell.

Cornell Audio Video Preservation Group Charge

Because preservation of Cornell's assets is imperative and directly linked to Cornell University Library's strategic effort to ensure access to the full scholarly record for the Cornell community², a strategic shift in thinking is required with regard to how digital preservation efforts should be integrated in our common work. The models we develop for digitally preserving our audiovisual material can inform creation, curation and life cycle management for all of our content. An investment of this kind will leverage centralized and scalable solutions for work done within Cornell University and also in complementary partnerships. Responding to this challenge, in 2012, a campus-wide group was formed to strategically address these issues in the wake of similar work already done at other institutions, as well as that of various audiovisual archival standards organizations. The Cornell Audio Video Preservation Group included key stakeholders looking for common approaches to the preservation and access of campus AV holdings. Representatives included members of the Library's Visual Resources Working Group, Digital Media Group, Library Technical Services and Research Data Management Services Group. A common thread of uncertainty helped bring partners from IT@Cornell, the Lab of

² Toward 2015: Cornell University Library Strategic Plan, 2011-2015
https://www.library.cornell.edu/sites/default/files/CUL_Strategic_Plan_2011-2015%28re-numbered%29_1.pdf



Ornithology, and Cornell Communications to the table to discuss areas of need, weakness and possible solutions (**see Appendix A**).

Our charge was threefold:

1. Conduct a preliminary campus-wide survey of audiovisual assets to determine scope and associated risks of collections at Cornell
2. Analyze existing infrastructure, staff expertise, and workflows to maximize economies of scale and minimize redundancy
3. Recommend tiers of effort with associated business models for moving forward

This report is organized into sections addressing the three steps listed in the charge. It includes an introduction to current research and similar initiatives to provide context for addressing this challenge and to help broaden possible solutions to involve University-wide partners. We discuss the findings of our year-long, collection-level survey of unique and rare media items on campus. Next, we describe the technical landscape and potential solutions we discovered both through the survey and in discussions with the larger group. Finally, we provide current funding models for digitization and preservation and illustrate how a modest increase in funding could be utilized to better serve the Library's mission and the University's collections.

Table of Contents:

[An Introduction to Degradescence](#)

[The Problem: A Small Window of Opportunity](#)

[Peer and Institutional Precedent](#)

[Campus-Wide Survey](#)

[Overview of Methodology](#)

[Results](#)

[Narrative Case Examples](#)

[Weill Medical Archives](#)

[Lewenstein and History of Technology](#)

[Johnson Museum of Art](#)

[Other Examples](#)

[Existing Expertise and Technical Infrastructure on Campus](#)

[Existing Expertise](#)

[Cornell University Library](#)

[Digitization Services](#)

[Metadata](#)

[Cornell's Lab of Ornithology](#)

[Academic Technology Center \(IT@Cornell\)](#)



[Technical Infrastructure](#)

[Cornell University Library](#)

[Storage and Microservices](#)

[Access](#)

[Existing tools](#)

[Cornell University Information Technology](#)

[Findings](#)

[Cost of Inaction](#)

[Key Findings](#)

[Recommendations](#)

[Internal Business Model/Costs](#)

[Increase of \\$25k in Preservation Funding at CUL for digitization of library materials](#)

[Short Term Library Funding Increase \(3 years\)](#)

[External Funding Partnership Options](#)

[External Grant Partnerships](#)

[Improve Technical Infrastructure for Digital Preservation](#)

[Library Storage Requirements](#)

[Metadata Requirements](#)

[Access Requirements](#)

[Implementation:](#)

[Prioritization Task Force](#)

[Coordination of Efforts](#)

[Risk Assessment and Prioritization](#)

[Storage Cost Analysis](#)

[Current Estimated Expenditures](#)

[Leveraging Economies of Scale](#)

[Digitization Cost](#)

[Preservation Strategies](#)

[Digitization at Scale](#)

[Appendix A: Charter](#)

[Appendix B: AV Preservation Survey](#)

[Appendix C: AV Streaming Policy](#)

[Guidelines for Streaming Audio-Visual Content at CUL](#)

[Available repositories](#)

[Copyright Guidelines for Audio-Visual Content](#)



I. An Introduction to Degralescence³

A. The Problem: A Small Window of Opportunity

Mike Casey, of Indiana University's Media Preservation Initiative, described the two biggest issues facing magnetic media loss with the portmanteau, "degraescence," a combination of degradation and obsolescence. Magnetic media has been widely reported to have a lifespan between 10 and 30 years, depending on storage conditions. According the Blue Ribbon Task Force report from 2010, "[Sustainable Economics for a Digital Planet](#)," we have less than a decade to both digitize and incorporate ongoing digital preservation support for this material before significant loss occurs. Most of the magnetic tape found in collections we identified was created more than 20 years ago. Most of the remaining materials date from closer to 30 years ago or more, reaching beyond the accepted safe lifespan of the medium. To adequately address this problem, we are shifting our workflows in our digitization labs to incorporate preservation actions earlier in the process of creating and capturing analog information.

Optical media like CD-Roms and DVDs are likewise increasingly problematic, causing yet another series of concerns for the coming decade. While we did not actively look for optical media materials in our survey, we did record the data that we found during the process.⁴ Film is also at risk. Recently, two films vended for digitization from the Division of Rare and Manuscript Collections at CUL were deemed "not able to be digitized" due to vinegar syndrome, a condition common to acetate film as it begins to break down, that can, in turn, contaminate the material around it. Despite the inherent vice of film formats, under proper storage conditions, it is a more stable medium than optical discs, giving us a longer window in which to prioritize and act.

B. Peer and Institutional Precedent

Cornell University Library is among a few academic and research libraries that have embarked on a systematic assessment of their media holdings. The following two examples of work done by leaders in the field have helped inform our work and subsequent discussions with these peer institutions have influenced our direction in the field of AV preservation.

Indiana University Library is a pioneer in the study of media formats and has advocated a strategy of proactive preservation, publishing a series of important papers and reports on the topic over the past decade that have become indispensable to the archival community at large. Most recently, their groundbreaking report, "[Meeting the Challenge of Media Preservation: Strategies and Solutions](#)," has been a valuable resource to our AV preservation group, and has largely guided our methods and our goals.

³ "Degralescence" coined by Mike Casey of Indiana University, IASA Journal 44 vol 3, "Why Media Preservation Can't Wait: The Gathering Storm"

⁴ We hope this report will lead to addressing the media landscape beyond magnetic media to include film and optical discs. The National Archives and Records Administration reports optical media are also vulnerable and subject to decay and obsolescence over the next 20 years.

<http://cool.conservation-us.org/bytopic/electronic-records/electronic-storage-media/critiss.html>



Additionally, last year The University of Illinois Library (Urbana-Champaign) published their "[Illinois Campus Media Census](#)" which covered roughly 9% of their campus with a targeted approach. Over time, our AV group has developed collegial ties with both of these institutions, and has greatly benefited from shared information regarding efforts in documenting need, digital production (digitization) standards and digital preservation.

II. Campus-Wide Survey

A. Overview of Methodology

Over a period of 15 months, we conducted a campus-wide census of unique or rare, at-risk AV material. Based upon peer institution and publication recommendations, as well as in consultation with initiative partners at Cornell and [AV Preserve](#) (an audiovisual preservation consulting and software development firm), we constructed a two-part inquiry model targeting all Cornell academic units as well as service units that may be storing such formats. The first part was a low-barrier, 10-question web survey that included basic questions about content type, relevance to discipline, format and fragility (**see Appendix B**). The second part was a follow-up, in-person interview. According to Mike Casey and Patrick Feaster of [Indiana University's Media Preservation Initiative](#), the in-person interview is absolutely crucial to gathering accurate data. Stakeholders were often not familiar with technical details of media types or formats and it also helped with identification of technology being used, content workflows, and pockets of expertise.

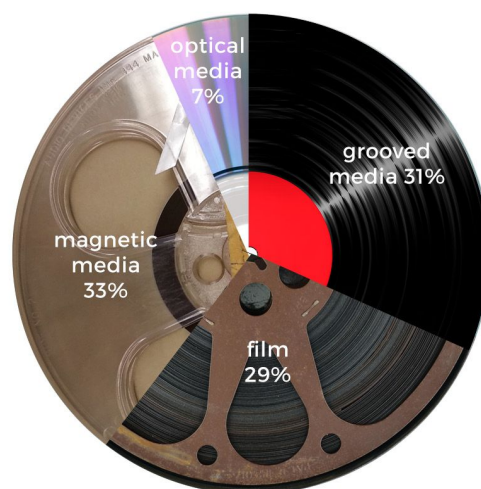
During this survey, we met with liaison librarians, curators, archivists, faculty, department heads and others to try to determine the scope of the audiovisual issues facing them and their units. Respondents' concern for the material ranged from low to very high based on previous or anticipated use and need. Those with low concern about the continued longevity of materials in their care were often unsure of the precise content or its value because of its inaccessibility; in some cases, they prioritized storage space over physical audiovisual materials. Those respondents with a high level of concern for their holdings were often motivated by the number of unique objects, their value as a resource or a sense of loss with respect to AV materials that were no longer accessible with current technology and would be expensive, if not impossible, to replace.

Because of the urgency of degrading media as well as the prescribed timeframe and our limited staffing, conducting the census at the collection level was the most scalable approach. We acknowledge that through this type of higher-level approach, more detailed information about the materials is sacrificed. This includes the type of content (i.e. interview, performance, or feature film) which varies widely from item to item. Likewise, we can only make a general assessment of whether a collection is unique or rare, even if comprised of commercially produced items that are no longer widely available. Many of the AV materials held by archives are one-of-a-kind. Though some materials, as in the case of industrial films (films produced to instruct employees or the community), may be certainly rare, they are not necessarily unique. In some cases we may only be in the position to make a general assessment as to whether a collection is believed to be rare or unique. The key issue here is that all of the materials we inventoried are at risk.

B. Results

With only 5% of the campus responding, we identified over 220,000 unique or rare, “at-risk” AV items held by key stakeholders. Cornell University has 18 unit libraries and 3 archives across the multiple campuses. Of these, 11 responded to the survey: a higher representation on average than other departments and units on campus. Cornell University Library (CUL), including University Archives, The Division of Rare and Manuscript Collections, and the Kheel Center for Labor Documentation and Archives, reported a combined 209,359 physical AV items, making the Library by far the primary stakeholder, holding 95% of the AV materials reported. Because CUL’s unit libraries and archives manage and care for the bulk of the University’s AV resources, it stands to reason that this high rate of response is due to a level of awareness and concern for their safekeeping. Preservation per se was not the only concern that CUL respondents expressed regarding the materials; access was also a driving factor in taking steps to quantify the problem. This is especially pressing, as space for students to access media content is in competition with other study space, both quiet and collaborative. As one example, the media room in Olin Library once housed several analog viewing stations but has since been removed, though no alternative that would allow continued access to AV holdings has been implemented.

The survey revealed a diverse range of physical formats among Cornell’s AV assets, including wire recordings, 20 different video formats, vinyl records, CD-ROMs and other optical media, and a variety of film formats. Each of these formats has its own special needs for storage and playback and can be grouped into one of four categories: magnetic media, optical media, grooved media and film (see **Figure 1**). We found magnetic media makes up 33% of holdings identified, followed by 31% of grooved media, which includes cylinder recordings and vinyl records, most of these held in the Sidney Cox Library of Music and Dance. Analog film holdings make up 29% and 7% of the media reported are stored on optical media, though it can be presumed this number will grow as time passes and more and more of this technology will be accessioned by archives and concern for its longevity is raised.



distribution of physical formats

Figure 1. Breakdown of major format types as reported by survey respondents

The census team found that of the materials reported, 56% are moving image resources, while audio makes up 34% (**see Figure 2**). This is significant to note when considering digitization, as moving image files are substantially larger than audio files and require significantly more storage and planning to maintain them. For example, if we assume that all of the videos identified in the census have average running length of one hour, once digitized at preservation standards, these items would amount to over 24.29 petabytes of digital content. The same number of audio recordings of the same length would convert to only .08 petabytes of digital content.

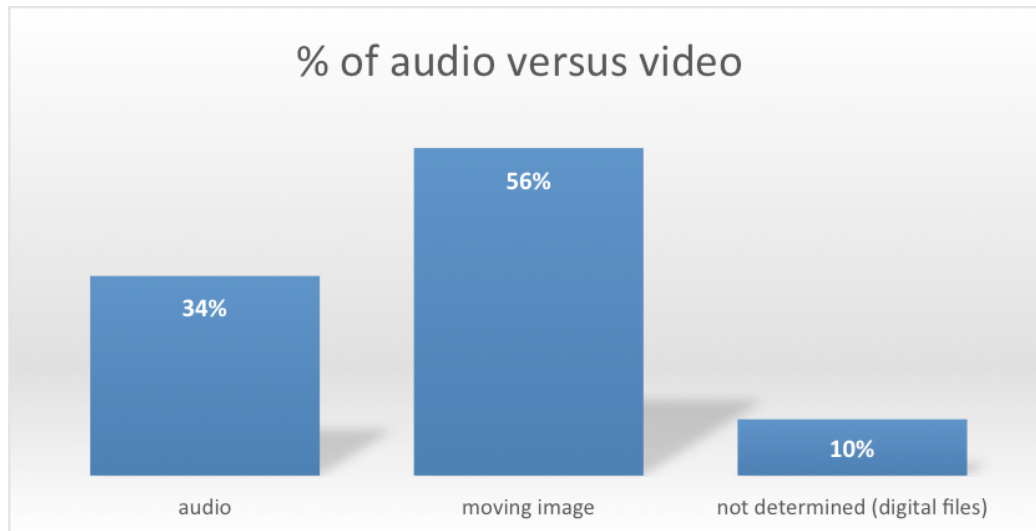


Figure 2. Based on number of items reported, this chart represents percentage of audio and moving image formats

Quantifying and categorizing digital audiovisual material proved more difficult than analog AV, as files and formats are often mixed. Despite this, we were able to identify 2.95 petabytes of unique or rare digital AV content, including the Lab Of Ornithology's Bioacoustics Research Institute and Macaulay Library of Natural Sound and Video. Notably, stakeholders expressed as much concern about continued access to digital AV files in formats no longer widely used as they did about format obsolescence for analog media.

The vast majority of digital AV items surveyed are stored in conditions that range from stable to poor based on best practice recommendations from the [Library of Congress](#), the [American Institute for Conservation](#), [International Association for Sound and Audiovisual Archives](#), and others. These include physical items stored at improper temperature and humidity levels, or in insufficient enclosures, as well as inadequate monitoring, validation and positioning of digital objects.

Generally, the three common threads that ran throughout our census interviews were:

- There was no plan to digitize the materials found.
- The barrier for use of these materials was the format and lack of digital surrogates.
- Stakeholders struggle with cost, program support and lack of technical understanding.



Beyond identifying at-risk materials, the survey uncovered a few different workflows for digitization of teaching materials and identified the strongest points of expertise in audiovisual media and delivery. While these areas of effort often utilized existing common service tools (CornellCast, Blackboard, Kaltura, etc.) via IT@Cornell, there are several smaller departments using their own platforms for delivery, such as the Language Teaching Center in Noyes Hall. Identifying where outlying material and workflows exist will ideally lead to a better understanding of the needs and uses of AV materials on campus, as well as ensure a preservation strategy where there was none. Likewise uncovered in the course of on-site interviews were legacy playback decks that are often disposed of when no longer of use to a department, but at the department’s discretion could instead be donated to the CUL Multimedia Preservation Lab to be repaired for digitization or used for parts: a boon to the unit as prices for obsolete equipment are rising as the market grows for audiovisual reformatting.

C. Narrative Case Examples

i. Weill Medical Archives

Upon visiting Weill Medical Archives we found unique and rare items in a non-climate controlled environment on the top floor of the New York Presbyterian/Weill Cornell Medical Center. This material is very vulnerable and the content has remarkably high usage data. It was stated that if the content were digitized, it would be used more easily and readily. This is an example of a collection facing obsolescence and degradation within the next few years.

Stakeholder	Format	Media Type	Quantity	Storage Condition	Unique/ Rare/ Widely Available	Commercial/ Non-commercial
Weill	Open-reel audiotapes	audio	290	unstable/vulnerable	rare	non-commercial
Weill	cassettes	audio	490	unstable/vulnerable	rare	non-commercial
Weill	films	moving image	303	unstable/vulnerable	rare	non-commercial
Weill	records	audio	29	unstable/vulnerable	rare	non-commercial
Weill	videos	moving image	1160	unstable/vulnerable	rare	non-commercial
Weill	DVD	moving image	96	unstable/vulnerable	rare	non-commercial

ii. Lewenstein and History of Technology

Bruce Lewenstein is a faculty member who documents the ways that public communication of science is fundamental to the process of producing reliable knowledge about the natural world. His materials highlight the history of technology and present an interesting case study, not only because changing technology has put them at risk, but also as material already accessioned by the University Archive. Lewenstein understands that though the archive can provide stable storage conditions and discoverability, these do not presuppose preservation. For that reason,



he reached out to the AV census team in an effort to identify the needs of these unique materials. These collections, created with National Science Foundation support, include examples of media coverage on specific scientific phenomena such as cold fusion and DNA fingerprinting as employed in high-profile court cases like the O.J. Simpson Trial, as well as over 400 hours of television broadcasts from 1999 with every mention of Y2K on three networks. Lewenstein's work focuses on how we communicate and construct theories. That story is documented on degrading audiovisual media, making this material a prime candidate for an initiative of the kind we are undertaking.

iii. Johnson Museum of Art

Founded in 1973 and housed in an iconic I.M. Pei building just off the Arts Quad of Cornell's Ithaca campus, The Herbert F. Johnson Museum of Art stewards more than 36,000 art works. Of these, 24,000 have been photographed or otherwise digitized for the purpose of documentation and access. The 70MB master files are stored on optical media, though the 16MB tiff derivatives are kept on Cornell's shared file service. 10% of these are uploaded into the museum's collection database.

Naturally, the museum staff is concerned about these digital assets, wanting to safeguard the investment they have made to document the unique items that make up the Johnsons collections. Additionally, there are materials that pose an even greater risk of loss, including lectures by visiting scholars, video art, and installation documentation on a variety of formats but stored in unfavorable conditions. Preservation and access to these resources potentially further the Johnson's reach as a teaching museum.

iv. Other Examples

Other examples across campus include the Language Resource Center, which is actively digitizing and delivering content to faculty through its own delivery system, operating outside of IT@Cornell's services and platforms. University Photography is an example of an organization on campus that is actively producing content on behalf of the University, much of which is slated for deposit in the University Archives. They are currently without a preservation plan and are looking for partnerships to try to build solutions for preservation and retrieval for their digitized and born-digital content. University Athletics has an extensive film collection that we were unable to survey, but which is split between the the department and the Library's Division of Rare and Manuscript Collections (RMC). Notably, Cornell sporting events are some of the most requested materials for digitization by RMC patrons.

CornellCast is a case in point for the fragility of digital and born-digital content. CornellCast is a highly-utilized platform for delivering streaming AV content and houses large amounts of University events, lectures, and other resources for Cornell. IT@Cornell partners actively produce University content delivered through Kaltura Media Server for CornellCast. IT@Cornell recently met with University Archives in RMC to develop a workflow for delivering this content into the Archives. For RMC, this translates into an increased need for digital storage, redundant backup, metadata management, ongoing file validation and other preservation mechanisms.

To further emphasize the point that this problem extends across the university, we note that the Research Data Management Services Group report, [Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements \(2012\)](#), stated



that approximately 27% of research data surveyed was audiovisual material. This is a significant finding and should be considered imperative in meeting national funding obligations at the university level. This has to be a shared goal with partnership-based, common solutions.

III. Existing Expertise and Technical Infrastructure

A. Existing Expertise

i. Cornell University Library

There is considerable support within the library system to handle the current flow of audiovisual reformatting, primarily based in the departmental units that handle digitization services, digital collection management, and resources on copyright. CUL's Audiovisual and Multimedia Preservation Lab handles a one-to-one workflow of fragile materials in-house in its expanded, three-room suite, as well as outsourced vending of some formats.

The Library Technical Services unit has an experienced Metadata Services team that participates in the cross-functional [Digital Consulting and Production Services \(DCAPS\)](#) unit. The Library's Metadata Services experts understand the challenges inherent to digital AV collections, including discoverability of assets, and the preservation of legacy metadata. This is integral to this work, as AV materials are often not cataloged via MARC (Machine Readable Cataloging) in CUL's Voyager catalog system. Our initiative hopes to leverage library expertise to achieve:

- Simple economies of scale
- Active connections with similar organizations with archiving needs
- Preservation-standard digitization capabilities
- Experience with a variety of curatorial and technical project considerations

Digital initiatives have been a key component of strategic planning in the Library for many years and CUL has developed digitization workflows to enable its staff to create and manage new digital collections efficiently. Established in 2003, the Library's DCAPS unit provides a framework for digital creation and management including requirements analysis, implementation, assessment, and archiving.

DCAPS is built on a holistic approach that recognizes interdependency among the various processes involved in creating and maintaining digital collections. Managing visual resource and metadata standards, user expectations, and resource requirements is a balancing act that demands solid business planning. This DCAPS service infrastructure operates on a cost recovery or service facility model, meaning our services come at a cost, both for the library and for clients on campus and beyond. The service rates are adjusted annually and are approved by the University.

a. Digitization Services

In DCAPS, we routinely consult on clients' options in preserving and providing access to their material. We are in a strong position to present stakeholders with a comprehensive set of considerations they should have in mind for any given digitization project.



Over the last three years, CUL has developed an expert multimedia preservation lab that can handle various legacy AV formats, deploying high levels of expertise and skill in the area of audiovisual preservation. We've been able to digitally preserve large, unique collections held by CUL with significant funds from the library materials budget that, since 2013, have been allocated annually specifically for digital preservation of Library holdings. We work closely with Library subject specialists and curators to identify materials that receive digital preservation treatment, coordinating with CUL's Collection Development Executive Committee to ensure preservation activities align with collection strategy and policy. In addition our work with Library holdings, our expert staff already provides consulting on fragile and legacy AV material and formats services for the University at large. Since the establishment of a dedicated audiovisual unit, the number of inquiries related to AV formats and requests for digitization of assets stored on obsolete media indicate a real need for this type of expertise and service. As a result, the Library has expanded its new media reformatting lab, increasing our throughput. We can also handle more formats in-house, extending our services into new areas, such as digital forensics and other data curation services.

The digitization of legacy audiovisual materials requires skills that are no longer widely available and equipment that is increasingly rare to find in working order. We play back analog material on the best equipment we can acquire and capture content to archival standards and in consultation with leaders in the field on best practices. We also consider how much wear and tear a given carrier (tape, reel, cartridge) takes when played on a particular deck, how easily a machine may be maintained and repaired, and what level of control over playback quality and/or monitoring a player provides. Because some media is particularly fragile or expected to become so during the next decade, it requires skilled engineers or technicians overseen by an engineering staff, who must prepare materials and pay close attention to the capture process to avoid significant and irreparable damage to vulnerable media and the machines used to play them. This type of single-object digitization workflow is referred to as "one-to-one" (1:1) transfer, where a single analog object is reformatted into a single digital file. In-house digitization is cost-effective in most cases for the Library's content, and because the work is done within the context of the library system, further benefits from the proximity of other areas of expertise that contribute to the digital content's lifecycle.

Materials on formats we do not digitize in our lab are outsourced to trusted vendor partners. This also can include materials we determine are more stable or that we prioritize as lower in value. It is important to consider whether or not Cornell holds enough material of any given format to warrant an investment in staff and equipment to capture it. For this reason, there will always be formats we continue to send to outside vendors. Working with external vendors can also be advantageous because these specialized operations are able to capture content in multiple streams, a process referred to as "parallel" transfer work. Parallel capture requires specialized software and equipment and may be prohibitively costly to implement in many institutions, including Cornell, so vendor partnerships are necessary and expedient.

The decision to outsource archival materials must factor the risk of damage or loss of irreplaceable objects as well as the real financial cost of project preparation, metadata assembly, shipping, insurance, quality control and assurance of the finished product. This still amounts to local labor hours and costs beyond the vendor's fee for service.



b. Metadata

Digitization of analog material aims to preserve at-risk content; successful preservation allows for faster access with improved intellectual control. The library's AV metadata requirements record functional details about a file and its creation in order to maximize long-term management of a digital object throughout its lifecycle. When our group takes on any given project, we require the client provide information such as bibliographic identifiers (where possible), call numbers, archival identifiers and accession numbers with associated metadata exported from already existing cataloging systems. We then expand that data to reflect further structural details inherent in the original object. We automate processes where possible to best map out the work, minimizing chance for error and confusion when we must manually make changes later. This is crucial for aging audiovisual material, as the digital surrogates we create become the new master content.

We capture media using commercially available software proven and accepted in the archival community (Wavelab) and signal measurement and testing tools with high-quality graphical interfaces (Spectra Foo, Audio Precision), automatically recording specific technical metadata. We also use a suite of emerging, open source tools to assure quality and metadata accuracy, employing batch processing where possible (BWF MetaEdit, QC Tools, FFMPEG). With this comes the ability to enter selected metadata into both our high-resolution preservation master media files and our lower-quality access copies. One of our focus areas is reducing staff effort and opportunity for human error in carrying out this work.

Best practice in digital production requires adequate metadata be associated with all digital media for identifying the format of the original material and technical details of its digitization and determining who should have what level of access to the resulting digital objects. In the interest of keeping metadata for a large collection manageable, the great majority of metadata for any individual project resides in a single database. For audio data files, we are currently complementing these stores of metadata with technical metadata and any descriptive information we have about the items that we embed within data files themselves. We are also investigating the use of Matroska video container with a lossless FFV1 codec, an open source file format that would allow us to embed metadata, transcripts, captions files and more into video data files as well.

The clear advantage to using embedded metadata is the ability to link the material to larger database records and provide information not necessarily in the database but unique to the file's creation, e.g., the make and model of tape on which a recording was made, the equipment used to play it back and digitize it, and physical item annotation. This data is stored according to an agreed upon standard and it travels with the file, making it easier to copy or move objects to other database systems later.

The metadata we include in audio files varies somewhat depending on the collection. Typically, however, we recommend including descriptive data that identifies the item and the collection that it is part of, its own item number, side number information (e.g., side 1 or 2 of a cassette tape), any identifiable text written on each side of the analog object, and notes fields for anything we are unable to read. These notes serve as a flag for file users who may be better prepared than we are to decipher the original item notation. We also include origination dates



that identify when the digitization process took place and an md5 checksum on the audio data alone (header and embedded metadata excluded) for use in confirming audio data integrity later. Finally, we include technical metadata specific to our in-house digitization process, identifying what equipment (hardware and software) we used to play the original material and convert it to digital media.

ii. Cornell's Lab of Ornithology

Cornell's Lab of Ornithology is a leader in digital content creation and asset management on campus and beyond. The Macaulay Library of Natural Sound and Video has been digitizing content for well over a decade to international standards, or better. They are the preeminent resource for natural sound and bird-related data across the globe. Media Engineers at the Macaulay Library have been integral to CUL's efforts in digitization of AV materials over the past three years. Macaulay's best practices and digitization workflows informed many of the processes that have been adopted at CUL. In 2013, Macaulay shared a digitization studio with CUL to begin digitization of a unique collection of Indonesian gamelan field recordings. Along with CUL, the Lab of Ornithology is investigating large, featured preservation storage for assets. The models being investigated are based upon current and projected growth of digital content.

iii. Academic Technology Center (IT@Cornell)

Academic Technologies handles basic digitization of legacy materials for use in the classroom upon request for faculty, staff and researchers. This is done on a small, "on-request" scale and does not meet preservation specifications set by standards bodies and organizations. Processes involve minimal quality control and assurance regarding file size and type, playback equipment, and signal testing. The result is usable and interpretable deliverables for immediate use, but with no descriptive or administrative metadata for preservation or discovery. A partnership on AV digitization between CUL and Academic Technologies could be beneficial and should include consultation on best practices, education on preservation issues involved in digitization work, and instruction on basic metadata practices within the Kaltura environment.

B. Technical Infrastructure

i. Cornell University Library

Digital audiovisual assets are essentially computer data. Systems for the use and preservation of these assets are always advancing, especially in this era of cloud storage and rising expectations around accessibility. To avoid loss through catastrophic events or technical obsolescence, attention must be paid to file formatting and long-term storage. Focused attention is also necessary to maintain the highest practical quality of service on a variety of delivery platforms. When assets are provisioned for adequately, we can minimize problems associated with migration between evolving formats and database systems. The Library has expert staff in information management and retrieval and skilled technicians whose work sets the stage for long-term accessibility of Cornell's unique audiovisual assets. This work includes processes of digitization, metadata management, storage, and migration.

a. Storage and Microservices

CUL's new Multimedia Preservation Lab will have local, networked storage that was designed and developed in collaboration with Cornell University Library IT (CULIT). The system will



provide a rapid, digital pipeline for newly created content, including quality control (QC), embedding of metadata, checksum creation and validation upon migration, derivative creation, final QC and a push to the IT@Cornell-based storage solution. Video capture and creation of preservation masters and access copies is more computationally intensive and complex than our typical audio workflow due to the sheer size of the data stream. The newly designed Networked Attached Storage (NAS) solution promises to improve video processing logistics as well, by giving us more flexibility and independent processing efficiencies at any given time.

b. Access

The return on investment on digitization goes well beyond preservation, in making material much more accessible, now or in the future. The ease of access one has to AV content on the internet can be misleading with regard to the challenge of digital migration of archive holdings. The reality is a staggering amount of cultural heritage and research data sitting dormant in archives. Our role as a research library and University requires us to preserve this material and it make accessible. With digital surrogates, patrons no longer have to be in the same place as the media itself, and access to restricted material can be legally managed in an automated way. Archival and other staff are freed up to work on improving the archive instead of dedicating some of their time to clearing users for access and assisting them with the use of audiovisual workstations. This is all accomplished through the use of a variety of delivery platforms and defining clear rights information for materials. Cornell's AV Streaming Policy Group has provided guidelines for internal Library staff that could inform larger, University-wide policies: <https://confluence.cornell.edu/display/CULAVS/FINAL+RECOMMENDATIONS>

c. Existing tools

Access for content can be provided through our campus-wide implementation of the video platform Kaltura, which provides the streaming of AV content into CornellCast, Blackboard, Mediaspace, and into CUL's eCommons repository. Access to content is managed by IP address and the Cornell University IP bands through shibboleth authentication. With limited metadata, content can be embedded via a player into websites and other discovery environments. There is work to be done regarding content restrictions and access and CUL's [AV Streaming Policy Group](#) has been working with Kaltura developers on metadata issues and implementation of token-style access provision, which will provide granular access control, on and off campus.

ii. Cornell University Information Technology

Cornell Information Technologies (IT@Cornell) operates several services that generate significant amounts of digital media. The AV Event Technical Support and Media Production Services unit supports and makes high-quality recordings of university events, such as Commencement, Convocation, Trustee Council Annual Meeting, Reunions, dignitary visits, lectures, seminars, and other special events. The service also supports and occasionally records distance learning classes, research groups, and routine administrative meetings.

AV Event Technical Support and Media Production Services also operates Bailey Hall. Archival recordings are routinely made of events at Bailey, including the music department ensembles (Jazz, Winds, and Orchestras), large enrollment classes (PSYCH 1101, Anatomy, and Oceanography), student performances, the Cornell Concert Series, the Concert Commission, and other special events.



IT@Cornell also offers a Video Streaming and Hosting service. This service offers three products that support classroom recording (“classroom capture”), media publication, and media management: Kaltura, Mediasite, and Panopto. These products support video streaming for CornellCast, CUL collection pages, Blackboard, and many colleges and departments.

IV. Findings

A. Cost of Inaction

A tool developed by AV Preserve, the [Cost of Inaction Calculator](#) is designed to assess the potential loss to an institution’s collection of research, teaching, and cultural heritage AV materials if no resources are put toward digital preservation. Using this tool in context with our census findings, we calculated the amount spent housing these analog materials over the years, the cost of our current rate of digitization, the cost of digital storage and most importantly, the percentage of our investments that could be saved if the library were to increase support in the preservation of these materials.

For this analysis, we decided to focus on AV materials housed in Cornell’s Carl A. Kroch Library, as we have more reliable data directly from the archives and facilities management. This assessment includes the Division of Rare and Manuscript Collections and University Archives, as well as the holdings of the Asia Collections. Coming up with usable data is problematic, since we were working with numbers of unique and many ostensibly “priceless” items whose worth, by definition, is difficult to measure monetarily. The value of archival material is determined by the curator and the marketplace, by acquisition and processing cost, as well as cost of storage and insurance. The very act of accessioning items for the Archives or RMC signals that they are to be preserved for the long term. In this scenario, we limited our analysis to magnetic recordings, averaging each magnetic item is worth **\$30**. While this price is arbitrary (and presumably extremely low), it is useful to focus the analysis on a smaller, subset of the collection with a fixed cost; this provides more reliable results that can inform our thinking on a larger scale. After consulting with AV Preserve, we moved forward with a very simple approach that is documented below.

We also decided to approach the physical storage costs simply in terms of annual utility cost, **disregarding** vault security, curator time, processing time, intrinsic value of the item, and insurance premiums. It is worth noting that this is a gross underestimation of overhead costs involved with both accessioning and maintaining these materials, but we wanted a manageable number that was easy to imagine and work with.

We gathered annual utility cost for Kroch Library (which houses the on-site collections of the University Archives and RMC, as well as CUL’s on-site Asia collection) from CUL’s Facilities team and divided this figure by the square footage of the building, separating out the total square footage of the the vault in the Rare and Manuscript Collections Division. This example assumes many variables that are very hard to calculate, including specific location of magnetic collections, size of various media types, kinds of storage methods (linear, condensed, multiformat collections, etc), and much more. The rough calculation is based on **20 years** of



operating cost at **\$51,596.27** annually for the vault. In terms of monetary investment, this as a general estimate and we feel this makes the broader case of value to the University and the library. This analysis does however provide a hard, quantitative case for the numbers of items that can be preserved through format migration, as well as the numbers of items that would be lost due to inaction. The data entered into the calculator is shown here:

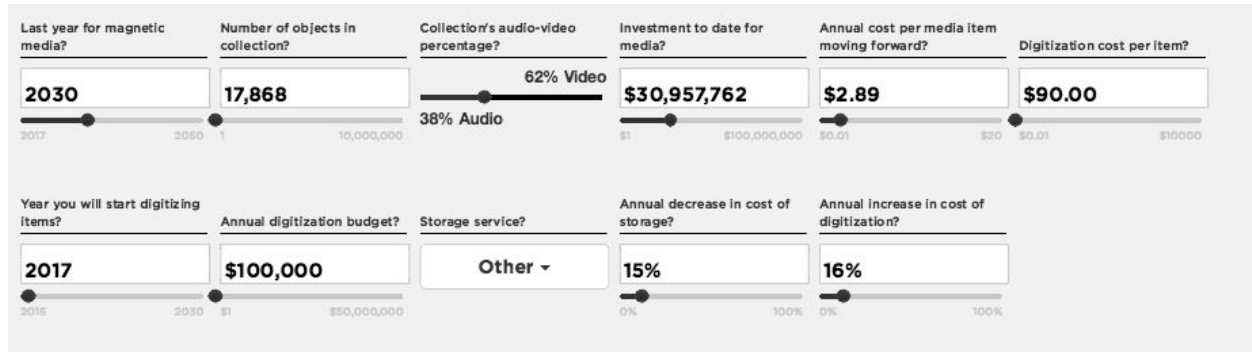


Figure 3.

If we assume annual funding for digital preservation is limited to current annual allocation from the library materials budget of **\$100,000**, we arrive at a forecast of migration, loss and general accessibility of prioritized AV content as represented in figures 4 and 5.



Investment Analysis

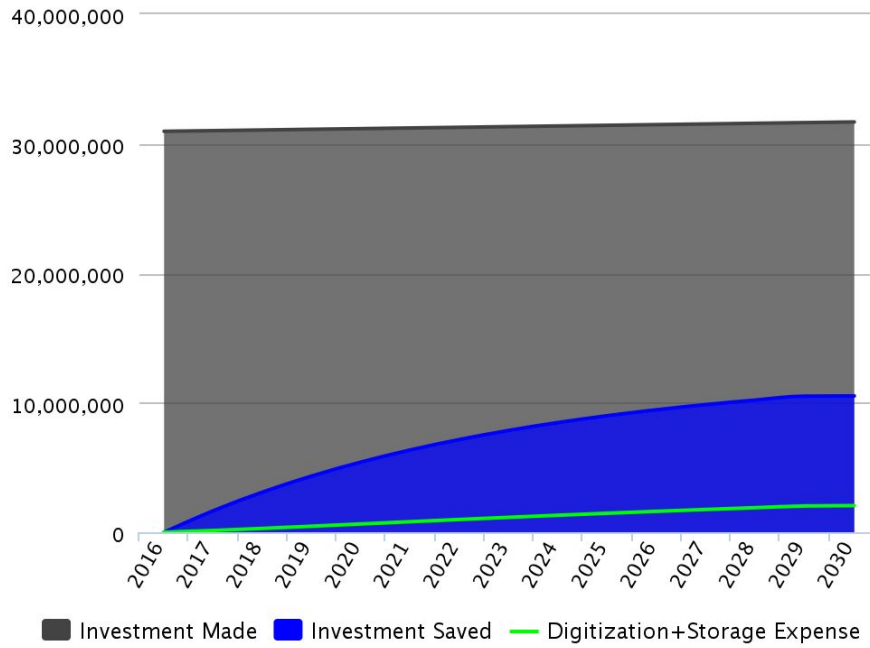


Figure 4.

Item and Accessibility Analysis

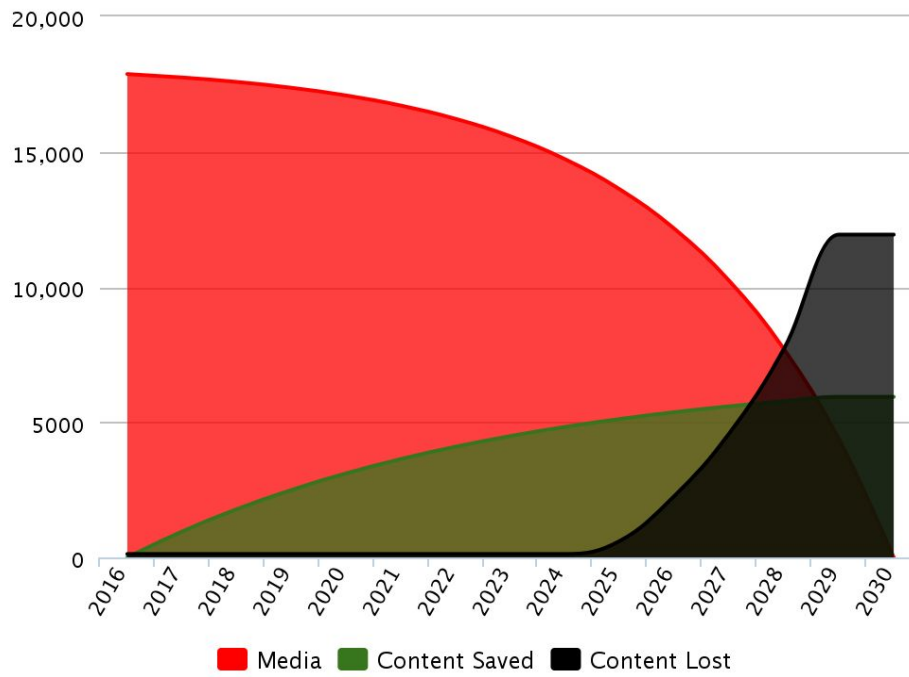


Figure 5



Accounting for both the time frame of magnetic media persistence (12-15 years as commonly accepted by experts) and the annual increase in the cost of digitization, at our current funding level for this work, we will likely lose 11,932 items, 69% of the magnetic items currently housed in RMC and University Archives.

Key Findings

Investment Saved per \$1 of expense	Return	Investment Lost	Content Lost
\$5.11	511.24%	\$21,156,059	11,932

Full Cost Parameters at Current Funding Rate

Year	Media	Content Saved	Excess Items Digitized	Investment Made	Content Lost	Investment Lost	Investment Saved	Digitization Expense	Storage Expense	Digitization + Storage Expense	Investment Saved per \$1 of expense	Quality of Selection
2016	17,868	0	0	\$30,957,762	126	\$30,957,762	\$0	\$0	\$0	\$0	\$0.00	100.00%
2017	17,742	958	793	\$31,009,401	126	\$29,347,072	\$1,662,329	\$100,000	\$37,441	\$137,441	\$12.09	99.30%
2018	17,578	1,784	609	\$31,061,039	126	\$27,960,514	\$3,100,525	\$200,000	\$96,701	\$296,701	\$10.45	98.37%
2019	17,361	2,495	427	\$31,112,678	126	\$26,767,502	\$4,345,176	\$300,000	\$167,175	\$467,175	\$9.30	97.16%
2020	17,076	3,109	240	\$31,164,316	126	\$25,741,625	\$5,422,691	\$400,000	\$241,809	\$641,809	\$8.45	95.57%
2021	16,702	3,638	38	\$31,215,955	126	\$24,860,073	\$6,355,882	\$500,000	\$316,042	\$816,042	\$7.79	93.47%
2022	16,211	4,094	0	\$31,267,593	126	\$24,103,151	\$7,164,442	\$600,000	\$387,050	\$987,050	\$7.26	90.72%
2023	15,565	4,487	0	\$31,319,232	126	\$23,453,851	\$7,865,381	\$700,000	\$453,203	\$1,153,203	\$6.82	87.11%
2024	14,716	4,826	0	\$31,370,870	126	\$22,897,483	\$8,473,387	\$800,000	\$513,679	\$1,313,679	\$6.45	82.36%
2025	13,602	5,118	0	\$31,422,509	612	\$22,421,367	\$9,001,142	\$900,000	\$568,196	\$1,468,196	\$6.13	76.12%
2026	12,137	5,370	0	\$31,474,147	2,285	\$22,014,548	\$9,459,600	\$1,000,000	\$616,816	\$1,616,816	\$5.85	67.93%
2027	10,213	5,587	0	\$31,525,786	4,595	\$21,667,568	\$9,858,217	\$1,100,000	\$659,813	\$1,759,813	\$5.60	57.16%
2028	7,685	5,775	0	\$31,577,424	7,730	\$21,372,262	\$10,205,163	\$1,200,000	\$697,586	\$1,897,586	\$5.38	43.01%
2029	4,364	5,936	0	\$31,629,063	11,932	\$21,121,575	\$10,507,488	\$1,300,000	\$730,589	\$2,030,589	\$5.17	24.42%
2030	0	5,936	0	\$31,680,701	11,932	\$21,156,059	\$10,524,643	\$1,300,000	\$758,643	\$2,058,643	\$5.11	0.00%

Figure 6.



While the magnetic tape housed in University Archives and RMC is a substantial part of the holdings on campus, it only represents **25%** of the magnetic tape material identified in our survey. This leaves **54,955** items unaccounted for in the analysis above. If we were to include the total number of audiovisual items we identified on campus it would amount to catastrophic loss at our current funding rates and structure.

On the surface, it may seem as if we could tackle this problem as we go, but given the nature of time-based media, running time is a challenge. Simply put, at an approximate average of 90 minutes running time per item there are **109,248** hours worth of magnetic media holdings in the library alone. This translates into **2,732 weeks** or **52 years** of running time if processed through a digitization station. We simply do not have time to wait. We estimate that with a team of three AV digitization specialists and capture stations, we can reduce the time to **17.5 years** given average vacation and health and personal leave time accounted for. While these spans may initially appear absurdly long, they are in fact accurate approximations and bear serious consideration. Our new Multimedia Preservation Lab expansion is designed for this level of functionality and also to accommodate the format percentage structure in both audio and video for the library.

V. Recommendations

A. Internal Business Model/Costs

i. Increase of \$25k in Preservation Funding at CUL for digitization of library materials

We recommend, due to the short life of this fragile media, the increased demand for access to analog and AV content, and the current relative low cost of digitization⁵, that we increase the annual materials budget allocation for digital preservation -- the funds that have supported the of the Collection Development Executive Committee (CDExec) Digital Preservation Grants -- to \$125,000 for three years. Additionally, we recommend replacing the annual grant-based funding model with a 5-year prioritization plan to be overseen by CDExec and CUL's Director of Digitization and Conservation Services, with broad input from subject specialists, curators, faculty, and others. This will ensure a more constant stream of incoming items into our expanded digitization pipelines. This requested increase in funding will employ the following efforts:

- 1. Prioritization Task Force and Storage Cost Analysis**
- 2. Increasing Digitization Scale**

⁵ Rebecca Chandler, AV Preserve, IASA Journal 46, "A Study on the Changing Prices of Audiovisual Digitization"



3. Investigating A Central Digital Preservation Solution for Cornell

ii. Short Term Library Funding Increase (3 years)

The 3-year, short-term annual increase of \$25,000 will help to fund increased digitization and vending of AV materials and a 18-month, temporary position (Audiovisual Collections .Coordinator) at .5FTE responsible for organizing and implementing a prioritization plan for the digitization of the “at-risk” media materials across campus. Working closely with library preservation staff, curators and collections stewards, faculty advisors, and technology specialists across campus, this position would coordinate meetings, conduct deeper assessments of neglected collections, facilitate metadata preparation, and rights management, in order to strategize a streamlined process for digitization. This is a natural next step, capitalizing on the work and investment that has been done to date in raising awareness, building partnerships and educating stakeholders on complex issues surrounding both digital preservation policy and audiovisual material needs.

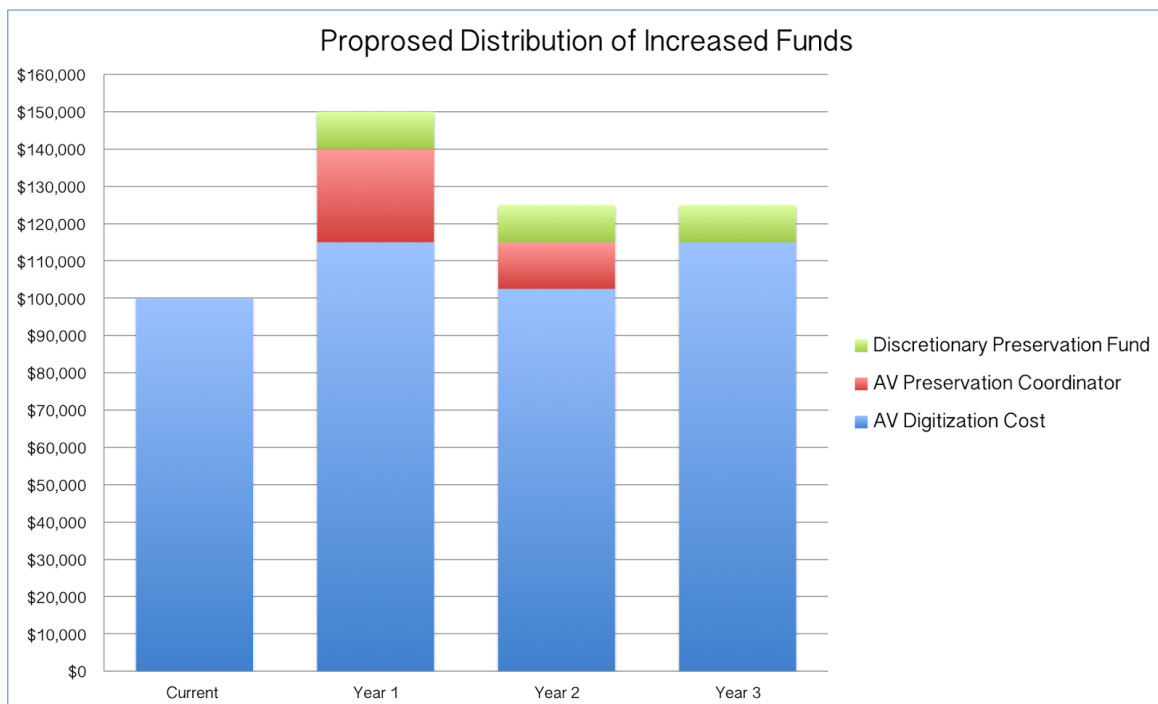


Figure 7. Proposed distribution of increased funds over three years

B. External Funding Partnership Options

i. External Grant Partnerships

Cornell University has a long history of successful and groundbreaking grant awards from National Endowment for the Humanities, National Science Foundation, and more. Audiovisual preservation is as much a research data management issue as it is a matter of preserving



cultural heritage. This is an area where Cornell should support increased collaboration, shared practices, and shared solutions among partners across the University, including the Library, the College of Agriculture and Life Sciences, College of Arts & Sciences, the College of Art, Architecture and Planning, IT@Cornell, and the Lab of Ornithology.

Recordings at Risk is a national regranting program administered by the Council on Library and Information Resources (CLIR) to support the preservation of rare and unique audio and audiovisual content of high scholarly value. Generously funded by The Andrew W. Mellon Foundation, the program will run four competitions from January 2017 to September 2018 and will award a total of \$2.3 million. A first call for proposals for preservation reformatting of magnetic audio media will be available on January 4, 2017.

ii. Crowdfunding Efforts

In January of 2017, Cornell University Library launched a successful preservation campaign for the digitization of vital Cornell history held on audiovisual formats. Working with the Office of Alumni Affairs and Development, we were able to meet and exceed our goal of \$15,000 to digitize a selection of University Archive's most fragile AV assets.

<https://crowdfunding.cornell.edu/project/3814>

C. Improve Technical Infrastructure for Digital Preservation

Digital preservation is an emerging and ongoing challenge across the University. Centralizing a service for the large amount of content being created and stored is imperative for the University. The University has a stake in the preservation of increasingly complex digital content, such as data sets produced as part of Cornell research. A plan for providing access to and preserving research data is mandated by most Federal funding agencies. The increasingly common requirement to include a data management plan in research grant applications has created a challenge for Cornell researchers. CUL has provided assistance in this area for several years through the Research Data Management Services Group, which is comprised of librarians, staff and faculty. Relevant data may include atmospheric and environmental data, a wide variety of research outputs, digital artworks, data analyzed in the context of digital humanities text as text corpora used in text-mining research, and much more. Central planning and funding will be required to achieve scalability and help reduce redundancy across the University over the next decade, not to mention protect and maintain these assets.

Cornell University Library continues to play a key role in how data and information is stored, organized, described and made more discoverable and easily accessible. Recent partnerships with colleagues at IT@Cornell and the Lab of Ornithology have produced and expanded services and improved workflow for the management of large AV data. The Library's Digital Asset Management Group, led by Michelle Paolillo (Digital Curation Services Lead) has produced a [three phase plan](#)⁶ to create a more coordinated digital ecosystem to support preservation and access. This work will inform the Library's digital production workflow and processes that the Library hopes will inform the University's handling of digital content. The

⁶ CUL Digital Asset Management Construction Planning Group, 2016
<https://confluence.cornell.edu/display/CULD/2016-05-17+Planning+Report>



open source digital preservation system Archivematica has emerged as a desirable software option because it is in wide use, can be customized, and adheres to the Open Archival Information System (OAIS) model for long term preservation of and access to digital objects.

i. Library Storage Requirements

At the current rate of approximately 100 TB per year, Cornell University Library continues to rapidly increase its content creation and digital holdings. At this rate we cannot continue to utilize IT@Cornell's Shared File System Storage (SFS), as the features of SFS storage do not encompass the scope of appropriate "preservation storage," as defined by the Digital Preservation Coalition (see <http://handbook.dpconline.org/organisational-activities/storage>).⁷ Below, in Phase 1 of our proposed work, members of the Library's Digital Scholarship and Preservation Services and Cornell University Library Information Technology units are working closely with partners in IT@Cornell on creating more affordable, secure tiered storage solutions, attending to the broader context of the University's needs.

ii. Metadata Requirements

In the first phase of [CUL's Digital Asset Management Group's planning report](#), CUL's Digital Media and Digital Curation Services Groups are working on simplifying and re-engineering our digitization process to better support archival workflows on campus and in the larger, archival community. Currently we are focusing on workflows for cataloged materials, but this should be expanded and modified to include uncatalogued AV materials in University Archives. Non-standard practices and workflows often lead to intensive metadata remediation and redundancy. Also, unique identifiers both for preservation masters and delivery files are necessary components of a coordinated digital ecosystem. The Audiovisual Collections Coordinator, working with Metadata Services in Library Technical Services will help coordinate this work, creating metadata plans for the prioritized material before digitization. This work is underway already, under DCAPS leadership and a collaboration between Metadata Services, Digital Curation Services and Multimedia Preservation Lab staff.

iii. Access Requirements

CUL is uniquely positioned to help classify permissions for digitized AV, as we have already determined access requirements in our work with the CDExec Digital Preservation grants and documented an access permissions framework in our [AV Streaming Policy Group](#)⁷. This will expedite the process of publishing content as it becomes available for delivery. The tiers of access restrictions are as follows:

- Open access
- Cornell community only (VPN/IP or Cornell Net ID)
- On-site restrictions (Library and library unit IP)

⁷ CUL AV Streaming Policy Group Final Recommendations, 2015, <https://confluence.cornell.edu/display/CULAVS/FINAL+RECOMMENDATIONS>



VI. Implementation:

A. Prioritization Task Force

i. Coordination of Efforts

The Audiovisual Collections Coordinator will recruit, organize and coordinate a small task force comprised of selectors, curators, technologists and faculty partners in order to devise a strategic plan to assess intellectual value of content, fragility of media, and long-term stewardship in order to prioritize digitization work. This prioritization plan would then be shared with a subcommittee of the Collection Development Executive group for final approval. Currently the Collection Development Executive group awards grants and funding annually to digitize content for preservation and access, based on proposals drawn up by library subject specialists and curators. The proposed task force would reach out to Cornell's wider academic community to help foster a direct link to the Library's key service preserving and making accessible Cornell University's assets. We will also utilize expertise and connections across the campus in promoting these services and engaging stakeholders. Our IT@Cornell partners will help in developing storage models that can handle larger deposits of digital content and take advantage of current services and new agreements with vendors such as Amazon Cloud Storage Services (Note: Preservation storage will be at additional cost and require funding sources).

ii. Risk Assessment and Prioritization

The selector and curator community at CUL will play an integral role in helping determine the intellectual value of content, securing rights information, and providing descriptive metadata for the selected materials. Utilizing this information in tandem with format-specific knowledge concerning media degradation, the Audiovisual Collections Coordinator will be able to assess risk to collections based on these variables and determine their prioritization in the digitization queue. As liaisons, subject specialists and functional experts in collection development, retention and management, the task force is critical in establishing and streamlining workflows to address the at-risk audiovisual assets identified in the survey. Documentation of decision criteria, workflow and data analysis will be a crucial part of process.

iii. Storage Cost Analysis

a. Current Estimated Expenditures

In FY16, the Library's storage bill was \$121,105 and costs are growing year-to-year. The main driver of storage size is audiovisual content, both digitized and digital born. Based on current funding and services using IT@Cornell's Shared File Service, for 2028, the year predicted by media experts to see widespread format degredescence, CUL's storage cost for digitized material can be estimated at \$696,361. The cost increases implied by our current approaches are not sustainable and we will require collaboration across the University to arrive at a new strategy. Our partners in IT@Cornell will be vital as the storage landscape evolves over the next decade as new services emerge. The first phase of work will include collaboration to identify cost savings in large storage that meet the Library's and the University's security goals and requirements.



Library and Library IT have identified requirements of featured preservation storage (appended) that includes fixity/validation, replacement, geographic redundancy and more. Storage needs to be appropriately coordinated with all available metadata, and unique IDs need to be assigned and exposed for coordination with other systems. This is the only way we can keep track of what is in storage.

Cornell University Library, in collaboration with IT@Cornell and the Lab of Ornithology have begun exploring featured preservation storage as a central service for Cornell. It is difficult to find a cost-effective solution that provides required features of preservation storage that are essential for digital life-cycle support. Several systems and possible solutions are being investigated and roles have yet to be defined as to the relationship between the digital object storage and the metadata management layer. Also under investigation are use cases for scientific data, as large data sets are growing larger as technology develops rapidly.

b. Leveraging Economies of Scale

In addition to the cost of digitization, one of the main drivers of audiovisual preservation is long-term digital storage cost. This is a necessary area for further exploration between the Library and IT@Cornell. The Library relies on IT@Cornell's essential SFS service for the bulk of digital storage. This service is good given what it does. The need for featured preservation storage (or archival storage) at scale is significant, not only for AV material, but for the preservation of grant-funded research data across the campus.

We recommend exploring a cost-effective, regional featured storage solution for not only the Library but for the University as a whole and possibly beyond. If the service was structured and leveraged properly, Cornell IT could provide this service to the state of New York's many colleges and universities. We could pursue grant opportunities to establish a service model that could be supported by member institutions based upon storage needs and requirements. This type of storage and preservation service is in keeping with the outreach and research mission of Cornell.

With this type of collaboration, we would continue to streamline efforts and in this case, provide internet 2 supported interfacing for preservation storage. We imagine this to be a collaborative effort, positioning both the Library's and IT@Cornell's expertise as a front door for preservation and subsequent access of collections, materials, research data and vital cultural heritage.

If approved, members of both IT@Cornell and the Library's Digital Production and Consulting Services would begin meeting in the Spring with possible clients on campus to identify needs, develop requirements and inform departments about concepts involved. Possibilities include pursuit of a grant to help with implementation of a digital preservation storage solution.

c. Digitization Cost

Digitization costs remain a big challenge globally, requiring significant resources dedicated to the reformatting of obsolete content. Digitization of an item without documenting context results in severe limitations to access and use. Curatorial effort, metadata preparation, and other digital lifecycle planning are necessary parts of digitization cost models. At CUL, our digitization services, while competitively priced vis-à-vis comparable services, are undertaken with



long-term preservation as a goal and our costs reflect the complex processes entailed in meeting that goal.

A recent article in the journal of the International Association of Sound and Audiovisual Archives states that digitization costs has never been lower and recommends that organizations take advantage of this state of affairs soon.⁸ The cost over time is predicted to go up significantly due to increasing obsolescence of playback equipment and the rising cost of expertise. The anticipated increases would affect all digitization services, whether in-house or by external vendors, and is reflected in the Cost of Inaction visualizations (Figures 5 and 6). Investment now in this effort will result in cost savings for the University over the long term, if Cornell is, in fact, prepared to address the preservation and access problems facing its AV assets.

d. Preservation Strategies

The goal of this report and the University-wide digital preservation effort it proposes is not to “keep everything.” Rather, we are recommending investment in a collaborative framework in which technical expertise and disciplinary knowledge can effectively combine with the principles of appraisal and selection that underlie library and archival collections in an effort to preserve essential audiovisual data, whether it be the output of Cornell research and scholarship, uniquely-held artifacts of cultural heritage, or documents of Cornell University’s rich history. Long-term preservation should be tied to usage and we propose that after digitization occurs, the life of the file and owner of the content will be responsible for the life cycle of the digitized content. Library owned materials will likely be assigned to selectors and curators for long-term management and curatorial decision making, as they are in the best position to speak of a collection’s value. Content digitized for external stakeholders will be returned on portable disc drives, shared storage, or LTO tape (recommended) and guidance will provided from CUL’s DCAPS unit as to next, best-practice steps for content. We hope for a scalable and elastic, preservation storage solution that might be a common service provided by a central body on campus, but until this is possible, the content owner will be responsible for digital objects created.

e. Digitization at Scale

At 1.5 years into this ongoing effort, we will have formed a Prioritization Task Force as a standing committee comprised of CD Exec members, selectors and curators, created a five-year prioritization plan for AV material, analyzed and identified more cost-effective, elastic storage, and will have digitized over a thousand of the University’s rare or unique items. Digitization work would continue based upon the recommendations from the task force for the full three-year increase in digitization funding, and would continue for the full five years. By this time we will have costed out storage estimates, identified scalable solutions, and met storage requirements. This will have been a partnership between CUL-IT, IT@Cornell and possible partners from the Lab of Ornithology. We hope to have looked at the broad landscape of large storage on campus and we are aware that the Lab of Ornithology shares many of the same challenges with CUL.

⁸ Rebecca Chandler, AV Preserve, IASA Journal 46, “A Study on the Changing Prices of Audiovisual Digitization”



Appendix A: Charter

Meeting the Challenge of Media Preservation @ Cornell: Establishing a Working Group for Developing Strategies and Solutions

June 2013

Cornell University has vast holdings of unique audio/video assets vital to its mission “to discover, preserve, and disseminate knowledge...” These include lectures by luminary figures (Nobel laureates; heads of state; writers and artists, and more) as well as original source recordings essential to research in biology, linguistics, art, and beyond. However, due to hardware and file obsolescence, media failure, and rights issues, Cornell University is now faced with an imminent and growing threat to these high value and unique audio and video assets.

The immediacy of the threat was made clear in Indiana University's 2009 report on media preservation, which stated that "there is a 15- to 20-year window of opportunity, less for some formats, to digitally preserve media holdings before it becomes impossible due to degradation or prohibitively expensive due to obsolescence." [1]#_ftn1] As time progresses, solutions become less effective, more costly, and more difficult to implement.

From preliminary investigations it is clear that the loss to Cornell will be significant if no action is taken. A recent survey taken across the Cornell University Library revealed literally thousands of at-risk materials within high-value and unique research collections. Among these are irreplaceable anthropological field recordings from the Music Library; important film and videos related to Cornell history from Rare & Manuscript Division; and large collections on international labor unions at ILR's Kheel Library. Other examples of campus-wide research collections at-risk include original lectures related to Carl Sagan's work with SPIF (Cornell's Spacecraft Planetary Imaging Facility); faculty generated collections, including rare linguistic recordings, artist performances, and classroom lecture material; and large historical collections at Cornell Athletics on sports at the University. Within the last month, two faculty members have approached the Library after having experienced significant data loss of their primary research and teaching materials (due to media failure of VHS tapes and external hard-drives). Even collections under excellent stewardship, such as Macaulay Library's extensive collection of scientific recordings related to animal biodiversity, are still faced with issues related to preserving assets in perpetuity.

These are just a few examples of the broad and deep audio and video collections that are threatened. As only the tip of the iceberg is in view, one of the first steps must be to inventory Cornell's holdings, identify the risk, and prioritize.

Beyond the immediate risk, there is also the issue of preserving ever growing research data collections in perpetuity or for a specified period of time to fulfill grant agency requirements.[2]#_ftn2] A recent survey across Cornell University conducted by the Research Data Management Service Group found that 15% of faculty had significant data sets in video formats and 10% in audio formats. There is also a new Office of Science and Technology Policy mandate from the White House requiring open access to research data, defined as the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings (which includes audiovisual material). Primarily a funding and data



migration challenge, there are added challenges in managing born-digital content, metadata standards, and access to the content. Without good policies in place, campus-wide standards, and University backing, we risk losing assets of vital importance to the University, and non-compliance with government funding agencies.

It is with an acute awareness of the above that a campus-wide group has formed to discuss and strategically address these problems, from selection priorities to digitization to long-term preservation and storage requirements. While individual efforts to preserve collections have sprung up around campus, we feel strongly that we will be more successful, efficient, and cost-effective if we work together to tackle these formidable tasks.

We are proposing to form a task force to work on the following objectives during April 2014-April 2015:

1. Material assessment

- Conduct a preliminary campus-wide survey to assess range of audiovisual materials in terms of formats, uniqueness, metadata, long-term viability, risks, obsolescence
- Identify and work with key stakeholders and collection owners on campus to:
 - Assess collections for uniqueness, value, and risk
 - Determine prioritization strategy
- Identify/contact key advocates

2. Infrastructure assessment

- Identify existing capabilities/operations on campus; associated costs if possible
 - Digitization facilities
 - Storage facilities
 - Preservation systems
- Identify existing expertise on campus (and overlaps)
 - Metadata and workflow
 - Intellectual property and copyright
 - Best practices in digitization and preservation for a/v
- Identify pros/cons of centralizing a service

3. Make recommendations

- Determine scope of initiative: what formats to tackle first; secondary and possibly even third phases of effort.
- Identify best practices for digitization, metadata, storage, unique IDs, preservation of original materials, etc.
- Address sustainability issues in means of managing and archiving digitized versions and associated files. Calculate rough storage requirements.
- Develop an assessment metrics and conduct a Strengths, Weaknesses, Opportunities, and Threats analysis to compare our options (determine scope of initiative- partial plan, comprehensive plan, no action)
- Recommend media types or subject domains for focus of initial digitization strategies.
- Articulate copyright and rights management issues.
- Recommend principles for collaborations and governance model



- Costs: Analyze financial aspects of different collaboration scenarios and identify direct and indirect costs
- Access: determine preliminary access provisions to content in compliance with preservation methodology.

The goal of the work will be both to increase our awareness of multimedia services at Cornell (with information about needs, existing services, and gaps) and also to propose a cost-efficient and scalable service model. The deliverable will be a position paper with recommendations for Cornell senior administrators in the related service areas to consider.

The following individuals expressed their willingness to contribute to the working group:

- Paul Allen, Director of Technology and Information Management, Lab of Ornithology
- Tre Berney, Digital Media Specialist Cornell University Library
- John Bowman, Director of Multimedia Productions, Lab of Ornithology
- Greg Budney, Curator, Macaulay Library, Lab of Ornithology
- Robert Carozzoni, Lead Enterprise Cloud Strategist, Cornell Information Technologies
- Rick Elliker, Media Systems Engineer, Cornell Lab of Ornithology Macaulay Library
- Karl Fitzke, A/V Specialist, CUL (formerly Audio Engineer, CLO Macaulay Library)
- Jason Kovari, Head of Metadata Services, Cornell University Library
- Dean Krafft, Director, CUL IT, Cornell University Library
- Barbara Lust, Professor, CHE Human Development
- Bill McQuay, Multimedia Producer, Lab of Ornithology (Co-lead on initiative)
- Danielle Mericle, Director, Digital Media Group (Co-lead on initiative)
- Liz Muller, Curator of Digital and Media Collections and Head of Archival Technical Services, Cornell University Library
- Andrew Page, Manager, Integrative AV Engineering, IT@Cornell (Co-lead on initiative)
- Chris Pelkie, Information/Data Manager, Lab of Ornithology
- Glen Palmer, Communication Manager, University Communication
- Oya Y. Rieger, Associate University Librarian, Cornell University Library
- David Shirk, Storage Systems Engineer, Cornell Information Technologies
- Mike Webster, Director, Macaulay Library, Lab of Ornithology
- Paul Zarnowski, Assistant Director for Storage Services, Cornell Information Technologies
- Wendy Kozlowski, Science Data and Metadata Librarian, RDMSG Coordinator, Cornell University Library
- ---[1]#_ftnref1] Information about the Indiana University's Media Preservation Initiative is at: <http://www.indiana.edu/~medpres/about.shtml>

[2]#_ftnref2] Cornell's Research Data Management Service Group (RDMSG) provides a network of Cornell librarians and other Cornell staff dedicated to facilitating researchers' access to the data management services they require (<http://data.research.cornell.edu>). It aims to present a coherent set of services to researchers. The group relies on the existing services. The issues identified in this report need to be addressed to provide research data services in support of AV content.



Appendix B: AV Preservation Survey

Respondent Name:

Respondent NetID:

Respondent Department / Unit / Division:

Do you have any audiovisual assets in any of your collections?

- Yes
- No
- Unsure

Please list kinds and approximate quantities of audiovisual assets in your collection. Please populate all that apply.

- Film _____
- VHS _____
- Cassettes _____
- Open Reel Audio Tape _____
- Betacam _____
- Betamax _____
- U-matic _____
- Hi8 / 8mm _____
- Digital Tape (DV, MiniDV, Digimax, DigiBeta, etc) _____
- File-based digital formats (MOV, WAV, AVI, MP3, etc.) _____
- Other _____

How important is it to digitally preserve your audiovisual content for future use?

- Not important
- Somewhat Unimportant
- Somewhat Important
- Very Important
- Extremely Important

How concerned are you about loss of audiovisual content?

- Very concerned
- It's not that important to me
- Who cares?

Who owns the copyright for audiovisual assets in your collection?

If your assets are commercially produced, would consider them to be:

- Unique (one-of-a-kind)
- Rare
- Wide-available



Which of the following best describes the working relationship between your department and Cornell's IT infrastructure?

- Minimal IT support
- A university owned computer; that's all
- Servers and storage owned and managed by the department or college
- Servers and storage owned and managed by IT@Cornell or another Cornell unit
- An external service provider: _____
- I'm not sure...

Do you have unmet needs in managing, preserving or making accessible audiovisual assets?
(Please use this space to further describe any collections or audiovisual needs that you have)

Appendix C: AV Streaming Policy

Guidelines for Streaming Audio-Visual Content at CUL

November 18, 2015

Available repositories

Cornell University Library hosts three delivery mechanisms that can stream audio-visual content: [eCommons](#), [CUL MediaSpace](#), and Cornell Digital Collections Portal. Below are general guidelines and contacts to help you determine which repository is appropriate for your content. CUL also has established workflows for pushing course reserve content to Blackboard (see below).

eCommons

eCommons is Cornell's general-level institutional repository. It provides long-term access to a broad range of Cornell-related digital content of enduring value.

- Supported content types: eCommons accepts a broad range of Cornell-related content, representing the intellectual or administrative output of the university.
- Access provisions: Open access; Cornell-only access; temporary access embargoes.
- Submission: Self-submission; library-assisted batch submission.
- Preservation: CUL has made a commitment to ensuring the long-term preservation and access to material deposited in eCommons. All items are assigned permanent URLs.
- Other considerations: Limited collection branding. Does not provide a set of visually-branded, associated web pages for supplementary material about the content.
- Additional information: <https://ecommons.cornell.edu/page/about>
- When to use: Good for Cornell-related AV content where wide and permanent accessibility is desired. Good for material that needs to be cited in published work.
- Contact: ECOMMONS-ADMIN-L@cornell.edu



CUL MediaSpace

CUL MediaSpace is a “YouTube”-like interface that provides short- to medium-term access to a range of AV content related to Cornell University Library. It is an out-of-the-box user interface for Kaltura.

- Supported content types: Stand-alone AV content; licensed content requiring access limitations; ephemeral content.
- Access provisions: Open access; Cornell-only access; access can be restricted to IP range, or individual IP or login.
- Submission: Self-submission; library-assisted batch submission.
- Preservation: CUL has not committed to maintaining MediaSpace, or preserving content within it. Items are stored in Kaltura; preservation system TBD.
- Other considerations: No collection branding. Does not provide a set of visually-branded, associated web pages for supplementary material about the content.
- When to use: Good for restricted-access AV content, content under development, or content with a limited life-span. If long-term access is desired, use eCommons.
- Contact: dcaps@cornell.edu

CUL Digital Collections Portal

CUL Digital Collections Portal is our Hydra-based platform for delivering a wide range of content originating from CUL’s collections.

- Supported content types: AV content belonging to thematically cohesive collection.
- Access provisions: Open access; Cornell-only access.
- Submission: Library-managed submission only (fee-based work).
- Preservation: Items are stored in Fedora, Shared Shelf and Kaltura; preservation system TBD
- Other considerations: Can provide contextualized websites integrated within Portal; allows for cross-searching other digital collections regardless of format (images, books, journals, etc.)
- When to use: Good for thematically cohesive collections of AV content that require or benefit from substantial surrounding explanatory material to provide context, history, etc., and which would benefit from cross-searching other thematic collections within the portal
- Contact: dcaps@cornell.edu

Blackboard

Cornell University Library maintains a link within all Blackboard courses to deliver course-related content.

- Supported content types: Course related content from a wide variety of sources
- Access provisions: Classroom only access
- Submission: no self-submission (requires Course Reserve library staff)
- Other considerations: Requires fair use waiver from faculty (to be renewed each semester material is required in classroom)
- Contact: okureserve@cornell.edu



Appendix D: Digital Preservation Requirements

Below are general guidelines regarding intellectual property issues regarding streaming AV material. When in doubt, always contact copyright@cornell.edu before proceeding with disseminating your content.

Material type	Permission to make preservation copy	Single stream access	Course Reserve use	Cornell Library premise (all libraries)	Cornell-only access	Open access
Cornell generated content/ no rights statement/ orphan work	yes	yes	yes	yes	yes	case-by-case
Cornell generated content/ Cornell owns copyright	yes	yes	yes	yes	yes	case-by-case
Cornell generated content/ third party retains copyright	yes	case-by-case	yes	yes/ has to be an obsolete format or media degradation or lost/stolen	yes, pending fair use review	case-by-case
Cornell generated content/ waiver signed; pre-digital	yes	yes/ pending review of waiver	yes/ pending review of waiver	yes/ pending review of waiver	yes/ pending review of waiver	yes/ pending review of waiver
Cornell generated content/ waiver signed; post-digital	yes	yes/ pending review of waiver	yes/ pending review of waiver	yes/ pending review of waiver	yes/ pending review of waiver	yes/ pending review of waiver
Cornell purchased content/ outdated	yes	case-by-case pending	case-by-case pending	yes	case-by-case pending	Unlikely



media type/ can't be repurchased		fair use review	fair use review		fair use review	
Cornell purchased content/ outdated media type/ can be repurchased	yes for unique or rare objects only	no	yes, when no streaming copy is available	no	no	no

Featured Preservation Requirements (CUL)

Defined by Michelle Paolillo, DCS, DSPS

The table below represents the manner in which “preservation storage” differs from ordinary storage. There are other attributes (access control, etc.), not represented here, that we assume both storage types to hold in common.

	Minimum	Better	Best
Elasticity (Storage can grow with our needs)	X	X	X
Redundancy (#copies)	2	3	>3
Separate threats		X	X
Diversity of media			X
SLA (must have one)	X	X	X
Include assurance for dissemination of original data and all metadata	X	X	X
Include “reasonable” availability TBD	X	X	X
Unique ID’s assigned, persistent and	X	X	X



available for reference			
Audit log available for changes		X	X
Must express aggregation	X	X	X
Must express inheritance		X	X
Fixity (storage must provide for or accommodate these activities)			
Calculate hash, store, and allow access	X	X	X
Verify calculated hash with supplied value	X	X	X
recompute hash value periodically, compare with stored value and monitor for changes		X	X
Remediate correctly when monitoring reveals fixity failure			X

References: These make good background reading regarding preservation storage.
NDSA Levels of Preservation, a maturity model for preservation services - http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf
DPC Handbook – Storage - <http://handbook.dpconline.org/organisational-activities/storage>
 (Note that there are vendors for preservation storage noted at the foot of this page)