Against the Grain

Volume 27 | Issue 4 Article 8

2015

Audio Visual Preservation at the National Archives and Records Administration

Fenella G. France
Library of Congress, frfr@loc.gov

Audio Visual Preservation at the Library of Congress

by **Fenella G. France** (Acting Director, Preservation Directorate and Chief, Preservation Research and Testing Division, Library of Congress) <frfr@loc.gov>

The issue of unstable media is not a modern concept. While paper had been around for centuries, new technology in the mid-nineteenth century linked with the Industrial Revolution enabled a rapidly expanding demand for paper, and this paper produced from wood pulp was acidic. Once again, with the desire to mass-produce information for mass consumption with audio-visual formats and storage media formats from the early twentieth century to current times, we now see the same challenges. Manufacturers were more focused on being the first to market, since the more they could dominate the supply for new media formats the more profitable it would be. This was true for media ranging from various compositions of shellac and lacquer discs, to the more recent explosion of optical storage media — compact discs (CDs) and digital versatile discs (DVDs). Within optical media were different formats of CDs and DVDs, read only memory (ROM), recordable (R) and rewriteable (RW). Understanding the issues of various composite materials allows prioritization for transferring content from those materials most at risk to loss. This is especially important for media that are machine-dependent for readability of content, requiring a specific instrument for them to be playable, a risk that increases as the media ages and the machines themselves are no longer manufactured. This challenge expands to all storage media used currently to access and store digital information, from CDs, to thumb-drives, to portable hard-drives and magnetic tape formulations, that, although not necessarily all machine-dependent, are at risk due to the lack of attention or foresight from manufacturers to ensure longevity.

Many cultural heritage, academic and business institutions have large collections of audio-visual materials. The **Library of Congress** has over 160 million items, and the Library's **Packard Campus** for Audio Visual Conservation in Culpeper, VA houses 6 million items, including more than 3.5 million sound recordings and 1.7 million film, television and video items, representing over a century of audiovisual production. Since the mission of the Preservation Directorate is *"to assure long-term uninterrupted access to the intellectual content of the Library's collections, either in original or reformatted form,"* the Preservation Research and Testing Division (PRTD) undertakes research that will ensure preservation of all library and archive materials — traditional and modern — that comprise the Library's collections, by advancing knowledge and understanding of degradation mechanisms in these materials. This research is critical for allowing custodians of these modern media audio-visual (AV) materials to best assess risk to specific areas of AV collections and ensuring the ability to capture content before loss or irreparable damage occurs to these materials. This research is critical for other cultural and academic institutions as well as private practitioners working to save content before loss or damage to these often unique and irreplaceable AV collections.

When assessing the potential for loss of information, the **McKinsey** report on data storage (2009) suggests anywhere from 966 Pedabytes in manufacturing to 619 Pedabytes in banking, to 269 Pedabytes in academia, before the current explosion in data capture, databases and the use of data analytics are required to process and interpret the large volumes of data. The current deluge of data and the ability to store large volumes of data (including all those digital photos people keep) at very low cost has led to institutions and the general public relying on many types of storage media to retain this data. As technologies have advanced rapidly, many media quickly become out of date, and many people may not recall or indeed used 3.5" and 5.25" floppy discs or zip drives. In the early stages of development of new technologies, manufacturers are racing to gain the lead in the market and are not focused on the preservation of materials, so often the quality and longevity of early formats is less than those created when the technology is well established. While the future options for access and storage of AV data are challenging and expanding, the need to address the high risk to more historic AV formats requires the need for informed research to allow concerned institutions to best address these *at risk* formats.

There are three major categories of AV storage media. Magnetic media includes hard disk drive, networked drive (multiple, active hard disks), electronic tape, sequential access (reel, cartridge) and floppy disc. Optical media includes optical disc (CD, DVD, Blu-Ray) and optical tape, while the third category of solid-state media includes flash drive (USB flash memory drive), flash memory cards and solid-state hard drives. Many institutions still utilize CDs, and there has been a recent focus on the longevity of compact discs (CDs) and DVDs. The **Library of Congress** has large CD-audio and data collections dating from the earliest days of CD manufacture. CD-ROMs are a convenient media for access and temporary storage of data, and many other institutions and members of the public have data stored on CDs as well as audio collections. As with former types of storage media, CDs are machine-dependent, so access to the data on these materials is dependent on the equipment needed to read the disc, including working hardware and software. Additionally, these media are subject to deterioration and the impact of the environment, just like any other material. CD-ROMs are composite materials made of many different layers, and the materials used in these layers can change over time, depending on when the CD-ROMs were made and how they age. Depending on susceptibility, wear, and environment, various layers of a CD may undergo oxidation, hydrolysis or mechanical stress (impact of oxygen, moisture or just being used), leading to damage sometimes referred to as CD "rot."

Differences in composition impact longevity and cause of errors, for example whether the data layer is *stamped* into the disc or *burned* into photosensitive dye. For a *CD-ROM,* the basic design includes a *p*olycarbonate substrate, a metal reflective layer and a lacquer coating. Data is molded into the disc as microscopic pits in the first layer. The second layer, a metallic coating, reflects the laser to read the data in the pits, and the top layer is a clear lacquer applied to seal and protect the metal layer. While most users focus on damage

continued on page 14

Dr. France is Acting Director of the Preservation Directorate and Chief of the Preservation Research and Testing Division at the **Library of Congress**, researching non-destructive imaging techniques and prevention of environmental degradation to traditional and modern heritage collections. Her current focus is the development of spectral imaging techniques and increasing links and access between scientific and scholarly data. She received her Ph.D from **Otago University**, New Zealand. After lecturing at **Otago**, she was the research scientist for the Star-Spangled Banner Project at the **Smithsonian Institution**. An international specialist on polymer aging and environmental deterioration to cultural objects, she focuses on links between mechanical properties and chemical changes from environmental damage and treatment protocols. **Dr. France** has worked on projects including the World Trade Centre Artifacts, Pre-Columbian mummies and textiles, the Ellis Island Immigration Museum, and lighting standards for the preservation of cultural heritage. She serves on a range of standards and professional committees for cultural heritage preservation and maintains close links and collaborations with colleagues from academic, cultural, forensic and federal institutions.

to the lower layer, scratches in the substrate layer disrupt the light from the read laser, while scratches in the lacquer layer can damage the metal reflective layer. CD-R and DVD-Rs have organic dyes under the metal reflective layer, and for longevity the data layer is the most significant factor. Phthalocyanine dyes are most stable, while gold and gold-silver alloys for the reflective layer are most stable. The greater the BLER, the greater the loss of information or data stored on the disc. There is often confusion about the difference between CDs and DVDs. While there is a difference in the way these media are constructed and the data stored, the main difference is in the volume of data that can be stored on the media. Since the smallest capacity DVD is able to store about seven times more data, the potential for loss of data from minor damage is much greater with a DVD than a CD.

The Library has undertaken nearly two decades of research into optical disc longevity. For CD-ROM media this included both *Natural Aging* and *Accelerated Aging* studies, where the accelerated aging allows research to take place in a shorter time frame but is closely analyzed to ensure it matches real-time natural aging degradation processes. Natural aging research models the cumulative effect of ambient temperature, relative humidity (RH), use and handling. Storage conditions for discs in these studies were 70-74 °F and 40-60 %RH.

The effect of damage may lead to errors in signal playback, and the methods to assess and measure damage to a disc include birefringence, optical skew, transparency and "block error rate" or BLER, the latter being the most common. BLER is the most common measure of disc failure *(also called "frame" error rate in the ISO standardized test 10149)*. The measure of failure to define disc damage in ISO 10149 specifies frame error rate < 3%, or 1X BLER of 220 /s. This means that a CD-ROM is allowed a BLER of up to 220 before it is considered a "bad" disc.

The first *Natural Aging Study* assessed the average maximum BLER for set of 125 discs at each test interval and found the average increase in maximum BLER (or damage over time) from 1997-2010: ~ 13. This equated to a total failure rate of approximately 4%. The second *Natural Aging Study* allowed a snapshot of the condition of the Library CD-Audio collection. In this study 1271 discs spanning 20 years (early 1990s through 2010) were assessed with 389 discs having undergone more than 10 years of natural aging. The results indicated only 19 discs with BLER > ISO limits, equating to approximately a 1.5% failure. Accelerated aging studies to ensure that predictive testing relates to natural aging (so damage can be prevented before it occurs) was undertaken with a Datarius CDCS-4.2L Compact Disc Quality Control System using the conditions in the following table. Results from testing gave a 4% failure rate, comparable to Aging Study 1.

Table 1: Aging Conditions for Accelerated CD Testing.

Condition T C/ %RH	No. of CDs	Incubation (Hours)	Total Time (Hours)
80 / 85	10	500	2000
80 / 70	10	500	2000
80 / 55	10	500	2000
70 / 85	15	750	3000
60 / 85	30	1000	4000

CD degradation can occur both on the upper polycarbonate substrate layer and the lower reflective layer. Damage to the reflective layer is often referred to as CD rot. Type 1, showed in Figure 1 below, indicates damage to the read layer, where pinhole defects in the metal reflector caused by oxidation of the aluminum layer allow this area to become transparent to the read laser. On the label side, degradation of the seal coat is apparent above the hole in the reflector.

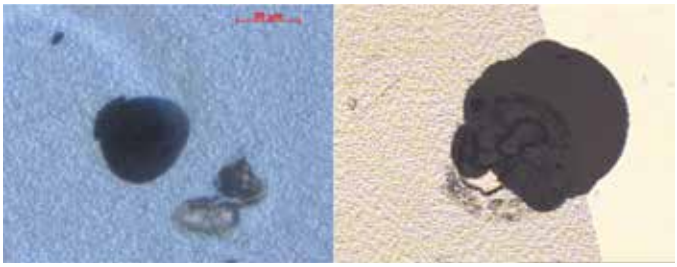

Figure 1: Left: Damage to read layer. Right: Damage to seal coat.

The second type of CD rot is characterized by oxidation of the reflector at outer edges of the disc, sometimes referred to as edge rot, while the third type of rot, bronzing of the reflector, is a result of corrosion of a silver reflector from exposure to sulfur from pollutants and atmospheric moisture.

An example of extreme aging is indicated in the oxidation of the metal reflective layer in Figure 2 below. After 500 hrs (80°C / 70% RH) the reflector on the CD on the right oxidized, while its twin on the left, aged under same conditions, retained its stability. This could be due to the discs having been stored or impacted by exposure to different environmental conditions over their lifetimes.

Figure 2: Aging of Two Identical Discs.

The impact of adhesive labels on disc longevity is compelling. In the two discs shown in Figure 3, a surface map of the disc error rate (BLER) after 1000 hours of aging shows the disc on the right with a BLER > 220/s, indicating significant damage and rendering the disc unplayable. The difference between the two originally identical discs is that one had an adhesive label attached and the other was unlabeled. Damage to the reflector from security labels, as illustrated in Figure 4, resulted in the library rejected the use of such devices on CDs in the collection. Studies into the effect of laser engraving were undertaken on 160 discs CD-ROM (Audio), where 80 were laser-engraved and 80 were unlabeled. Laser engraving had little effect on disc stability, and based on this study, the library proceeded with its plans to property mark discs in the collections.

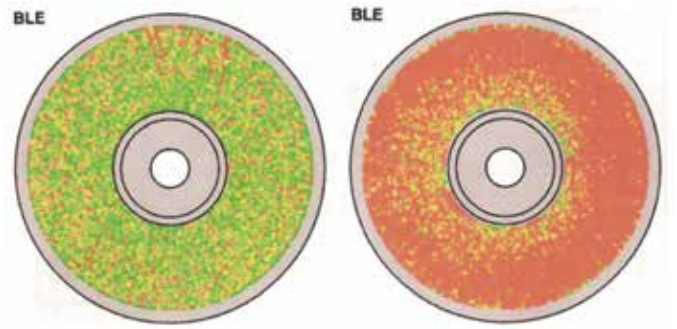

Figure 3: BLER Surface Map of Unlabeled and Labeled Discs.

As noted above for recordable media, CDs are more stable and better for archival purposes than DVDs, partially due to the volume of information on CDs being lower, so less information will be affected by damage

continued on page 16

or degradation than on a DVD. Studies above led to the development of new standard test methods for recordable DVDs: ECMA 379 and ISO 10995. As an overview of optical media, it has been shown that CD-ROMs are fairly stable but range in quality depending on their original manufacturing composition and environmental history. With "Natural" Aging of CD-ROMs, the BLER does increase over time, but the change is small, while discs that start with a high BLER degrade faster. Accelerated Aging of CD-ROMs indicated that approximately 30% have a service life less than 100 years, while around 45% have a service life less than 200 years. Adhesive labels on CD-ROMs significantly increase BLER, whereas laser-engraving as labels had little effect on BLER. While the specific manufacturer has an impact, most CD products earned a high probability of prolonged lifetime relative to their DVD counterparts.

In 2007, the library began conducting research into the degradation of magnetic tape, sometimes referred to as *sticky-shed syndrome*. This phenomenon was most common in analog audio and video tapes manufactured using polyester-urethane (PEU) binders beginning in the 1970s. While PEU was the main binder, more than 20 other binders were used. Understanding how the degradation occurred was challenged by the variety of other components used in the tape, a range of lubricants, additives and plasticizers. PEU is known to degrade at the ester linkage (reversible), with degraded tapes often containing lower molecular mass polymers chains than non-degraded tapes. The loss of information from these tape formats could impact the **Library of Congress** collection of over 300,000 audio magnetic and 450,000 audio visual tapes.

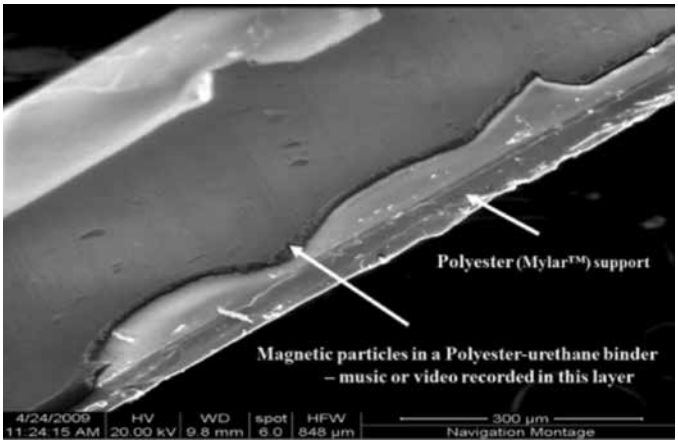

Figure 4: Scanning Electron Micrograph Illustrating Potential loss of Tape Content.

This project became collaborative research with the **University of South Carolina**. Chemometric statistical analysis of infra-red data from magnetic tape formulations allowed separation of degraded from non-degraded tapes. The research is ongoing. A further challenge with magnetic tape degradation is how to treat and capture information from tapes that are degraded. Baking of the tape allows a one-time capture since exudate on the surface of the tape is dried out, and allows the tape to be played. However, the Library is undertaking research into the assessment of potential changes in the fidelity of AV tapes due to various baking conditions and procedures.

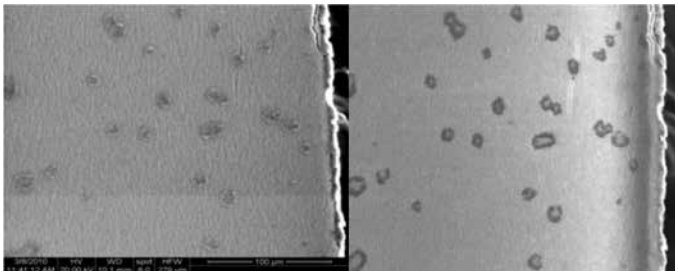

Figure 5: Scanning Electron Microscope Images: Before and After Baking.

The major challenge with better understanding the degradation of magnetic tape has been due to the difficulty in re-creating the damage through various methods of accelerated aging. This led to collaboration between the library and an industry partner, **Fuji Film**, Japan, who have been undertaking initial testing and characterization of specific historic tape formulations before and after accelerated aging in a variety of environmental conditions. This is allowing research to advance in relation to predicting and re-creating deterioration of older media formats so institutions can better understand what magnetic tape formulations might be at most at risk, and what the optimal storage conditions for preventing further damage should be.

While migration and transfer to newer storage media are current methods of preservation, the advent of new technologies has allowed us to develop non-invasive means of capturing sounds from historic audio formats when playing by traditional methods such as a stylus would cause damage. High energy physicists from **Lawrence Berkeley National Laboratory** began research approximately 12-15 years ago, applying optical metrology, pattern recognition, image processing and data analysis methods from physics research to issues of sound preservation and restoration. **Dr. Carl Haber** and **Earl Cornell** collaborated with **Peter Alyea** from the **Library of Congress** to create an instrument using optical methods to capture audio from historical sound recordings at the **Library of Congress**. Surviving original media can be unique or extremely rare, and deterioration from use, breakage or chemical breakdown can be exacerbated by the utilization of *traditional* playback equipment that damages the media through physical contact.

The **Library of Congress** developed the IRENE system in collaboration with **Lawrence Berkeley National Lab** to image the surface of lateral- and vertical-grooved sound recordings. The acronym IRENE stands for "Image Reconstruct, Erase Noise, Etc." and was coined since the first recording captured was *Goodnight Irene.* A large variety of media can be accessed through these 2D and 3D imaging methods including to date, 78 rpm shellac discs, instantaneous cut lacquer and aluminum discs, vinyl discs, wax cylinders, commercial cylinders, dictation discs and belts and phonautograms. This collaboration resulted in the construction of image-to-sound systems that have provided access to collections by capturing non-contact sound from severely damaged and fragile collection items.

As outlined above, increasing our understanding of how AV media degrade permits intervention actions that mitigate loss, and allows transfer and capture of content before further damage. The risk matrix assessing loss of AV collections will depend on the specific institutional policy, but there are a number of considerations that should be taken into account. The value or scarcity of the collection item will potentially raise the risk, whether the media is machine-dependent for readability, regardless of the age and availability of machine-readable devices, the density of information on the media and file formats and backward compatibility. For a general assessment of AV storage media risk, the following table provides some guidelines:

Table 2: AV Storage Media Categories.

- **Optical**
  CD-ROM > CD-RW > DVD
  Relatively stable when high-quality materials
  Range of lifetime due to # of manufacturers
- **Magnetic**
  Tape > 5 ¼" > 3 ½"
  Modern formats can exhibit low error
- **Solid State**
  Flash drives / Hard drives
  Relatively stable if written to minimal times

The identification and knowledge of storage media types in collections are critical for determining risk to content. This risk can be wide-ranging depending on the environment where they are stored and the quality of original manufacturing techniques and materials or stability of material formulations. The development of assessment tools for libraries, archives and museums is important to allow collections

continued on page 17

Audio-Visual Preservation ...
from page 16

care staff to detect (*before damage*) whether media can safely be played for transfer and access to content.

The library undertakes additional research into a range of AV formats to ensure long-term preservation and access to content. Research projects include cleaning solutions for lacquer discs and characterizing the nature of the exudate that forms on the surface to ensure cleaning formulations do not remove disc substrate, determining the composition of wax cylinders and which formats are more prone to degradation, composition and quality assurance testing on film cans, assessing sound fidelity of magnetic tapes before and after baking, and forensic assessment of CD/DVD and hard drive content recovery. Project summaries are updated as new research is completed and can be found at *http://www.loc.gov/preservation/scientists/projects/index.html*. One of the challenges with AV collections is institutions truly understanding the current state and condition of these collections, including accurate numbers of various formats. Knowledge of the condition is an aspect that is complicated by unknown histories of storage environments, use and wear of items in these collections. The **Heritage Health Index** is a National Collections Care Survey first undertaken in 2004 to assess the condition of American heritage institutions and repeated in 2014. While this was more focused on museums, there was a section on AV formats, and these modern materials still remain high-risk items as man-made compositions degrade over time. The 2012 **Library of Congress** National Recording Preservation Plan discusses the need for a national research agenda, continued research into preservation aspects for AV and training and technology requirements. As part of this plan, the library continues to address and research issues with AV formats as they come to light and endeavor to make this information and research available to other institutions to assist in preservation of their collection. In addition, the Preservation Reformatting Division (PRD) is responsible for the review of endangered materials that need to be copied to more stable formats using both analog and digital approaches to meet this objective. PRD provides access to at-risk library materials through converting items to new formats including microfilm, facsimiles or digital reproductions.

Conclusions

Audio-visual materials represent a very special component of our cultural heritage, these "new" storage formats being cutting-edge at the time and allowing us greater advances in accessing and storing large volumes of information. In addition to heritage institutions, many archives and businesses use or retrieve information from some formats, so understanding how best to preserve the content assures prevention of loss of content in many areas of social, historic and business functions. While there has been a move towards digital storage, the storage media this digital information is on still remains the risk component, and manufacturers will continue to develop new formats as technology advances. Continuing to preserve our modern and historic AV formats and storage will engage researchers for many years to come. Creatively utilizing these new technologies to capture sound and video from historic formats that are machine-dependent will assure retrieval of hidden collections and preservation of our cultural heritage.

Organizations with Online Resources

Association of Recorded Sound Collections (ARSC)
http://www.arsc-audio.org/pdf/ARSCTC_resources.pdf

International Association of Sound and Audiovisual Archives
http://www.iasa-web.org/

Electronic Media Group, American Institute for Conservation
http://cool.conservation-us.org/coolaic/sg/emg/index.html

Society of American Archivists (SAA) Recorded Sound Roundtable
http://www2.archivists.org/groups/recorded-sound-roundtable

Audio Engineering Society (AES)
http://www.aes.org/

Society of Motion Picture & Television Engineers (SMPTE)
https://www.smpte.org/

References

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Byers, Fred. Care and Handling of CDs and DVDs — A Guide for Librarians and Archivists; NIST Special Publication 500-252; 2003.

Heritage Health Index (2004-2014) *http://www.heritagepreservation.org/HHI/*.

Library of Congress National Recording Preservation Plan (2012) *http://www.loc.gov/programs/static/national-recording-preservation-plan/publications-and-reports/documents/NRPPLANCLIRpdfpub156.pdf*.

The Tangible Media Program at the Library of Congress

by **Moryma Aydelott** (Special Assistant to the Associate Librarian for Library Services, Library of Congress) <mayd@loc.gov>

The **Library of Congress** has a wide variety of digital items in its collections. Tangible Media is the broad term we use for non-networked digital collection items that a user can hold. It includes floppy discs and thumb drives, CDs, DVDs, hard drives, and digital tape. There are over 300TB of known digital data across the **Library of Congress'** Library Services divisions on various types of tangible media. In 2011, Library Services began the Tangible Media Program to look at obsolete tangible media formats in **Library of Congress** collections and begin to explore what could be done to develop workflows that could be used on a variety of materials in the multiple curatorial divisions and provide a backup copy of the digital data stored on long-term storage.

Seventeen curatorial divisions were surveyed in 2013 for information about the tangible media in their collections. There are wide variations in the numbers of tangible media items by division — close to 85% of the reported tangible media are concentrated in 5 of the 17 divisions, with 60% of the reported items in just one of those 5. This confirmed that, while some of the divisions had enough materials to necessitate long or continuous digital preservation programs, the majority of the divisions did not.

A program specialist worked with each division to consult and develop the program. The specialist designed specific, customized workflows for each division including documentation and equipment maintenance and obtained any needed technical resources. By the end of the program, the specialist had assisted the divisions with preservation of over 2800 items totaling over 7.5 terabytes.