

# **Bridging Analog Media and Digital Asset Management with Automated Migration Systems**

Gilad L. Rosner, Systems Integrator  
Media Matters, LLC  
August 2005

## **Abstract**

Automated audiovisual migration systems are the only way to migrate large volumes of magnetic media to digital files. DAM systems historically had no way to ingest analog assets en masse. By combining these two technologies, archives and commercial organizations can acquire digital versions of valuable content in a uniform, replicable manner. These legacy media assets can be monetized and re-purposed, providing a lasting return on investment. Automated, consistent metadata collection further eases the transition from analog to digital, and enables quick discovery of assets within a DAM system. The evolution of audiovisual migration and restoration to automated systems positively impacts the DAM industry by leveling the field of assets, obviating the distinction between analog and digital, affording new business opportunities for DAM vendors.

## **Introduction**

Magnetic media has been the most common way to record audiovisual material for the last several decades. Moments of historical and cultural importance, commercial and industrial productions, and news and entertainment footage have all been captured on magnetic tape. The value of all this audiovisual material is incalculable, yet the material is in great danger. Magnetic tape cannot be physically preserved beyond a certain point – components of the tape breakdown, causing the degradation and eventual destruction of the recorded content. The only way to preserve the content is *migration*, the process of copying the recorded content to a new tape or, more recently, to a digital file, while correcting for signal degradation.

Though Digital Asset Management (DAM) software has focused on the management and manipulation of born-digital and already encoded assets, it has been historically difficult to ingest and handle analog assets. Until very recently, encoding and metadata creation could not be performed en masse; this was particularly true for legacy media formats, such as the U-matic videotape. The skill sets required to encode magnetic media assets with high quality, consistent results were uncommon, and were rarely found in combination with traditional audiovisual restoration skills.

The solution to the problems of migrating legacy analog media to digital files – bringing these media assets to the front door of DAM systems – is automation. By transforming inefficient, expensive, human-centric processes into repeatable, consistent, automated processes, valuable analog assets can be turned into digital assets accompanied by appropriate metadata. The wealth of material locked away on countless hours of magnetic tape are ripe for ingestion into modern DAM systems, resulting in high quality content that is ready to be monetized and re-purposed.

## **Underused Assets**

Magnetic media has been a valuable technology to anyone with the need to record or produce audiovisual content. Whether recording the performance of a famous dancer, creating corporate industrial films, or interviewing newsworthy people, magnetic media has been used to preserve the audiovisual products and heritage of two generations. Magnetic media's ubiquity and low cost have come with a price: a comparatively short lifespan and difficulty in its preservation. Many factors can influence a tape's lifespan such as manufacturing quality, storage conditions, or number of plays. Twenty to thirty years is a generally agreed upon age at which to migrate a tape's content, though it could certainly be shorter. By that time, a piece of magnetic media could show signs of mild to severe degradation – the most common being the degradation of the adhesive binder that holds the metal oxide particles to the plastic backing of the tape. Physical degradation means signal degradation and the loss of content.

The content contained on these tapes is valuable not only for its cultural or historical qualities. For media, entertainment, and broadcast companies, audiovisual content is their lifeblood. The countless hours of material trapped on cassette and open reel magnetic media represent underused assets. Many archives do not know what is contained on the tens of thousands of feet of magnetic tape on their shelves, nor is there necessarily a good set of records to inform them. The only way to learn what is stored in a collection is by reviewing the tapes and recording metadata about each item, a laborious process to say the least. By allowing this audiovisual material to degrade or become effectively lost, archives and their parent organizations are turning their backs on content that would otherwise be monetized, repackaged, and reused to their benefit. Any system that could shoulder the burden of digitizing analog material and assist in the application of consistent and accurate metadata would have a substantial return on investment. Further, if such a system had increased productivity over an equivalent manual process, the pain of choosing which material to save and migrate would be diminished since more assets could be converted for the same amount of investment. In the deployment of such a system, the initial assets chosen for migration should be high-value content that could be quickly monetized or used to reduce costs elsewhere to prove the value of an automated migration system.

## **Traditional Migration**

The archivists and librarians charged with preserving the content in their collections have had few choices. The traditional way to preserve large numbers of tapes was to send them to a video restoration and migration facility to transfer the content to new tapes, usually on a more modern format than the original masters. The process was laborious and expensive, requiring specialized knowledge and equipment. The result of the process was more tapes; these tapes would themselves need to be re-migrated at some future date.

The choice of which tapes to migrate in a collection further complicates the task of preservation. While some tapes show obvious signs of degradation – visible particles, liquid staining, odor – many tapes do not. Therefore, archivists must make migration choices based on criteria such as tape age or cultural value. Most large archives have limited resources, further complicating these choices. The result is that many tapes that

should be migrated are not, resources are potentially misspent, and high-value content is potentially lost.

To discuss the shape and substance of an automated migration system, it is important to review the history of audiovisual migration. Traditionally, the process of migrating magnetic media to fresh tape stock has involved a set of skilled technicians working in teams of 2 to 3 people, with other specialists becoming involved at different stages of the migration process. In basic, the steps are:

- Shipping – tapes are sent from the collection holder to a specialized audiovisual restoration and migration company.
- Cleaning – all tapes must be cleaned upon receipt. This is a manual process where a staff member continually checks the condition of the cleaning tissue to ensure that the tape is not being “overcleaned,” which could harm the tape and its content.
- Migration – tapes are copied from an original master to a blank tape. Technicians clean up the signal as best they can on a tape-by-tape basis. Metadata about the process is collected, at best, on a limited, piecemeal, and inconsistent basis.
- Quality Assurance – tapes are reviewed to ensure that content has been migrated correctly.
- Shipping – masters and new re-masters are returned to the client.

Each step of the process involves a different set of staff members at the restoration company. Tapes that are too fragile to be cleaned are sent to yet another team for more aggressive restoration, such as baking tapes in an oven to re-glue the binder to the oxide and plastic backing to get one last playback from the tape. Metadata about the various stages of the process is paltry, and the process itself does not scale – that is, the more tapes in a job, the more labor required to do the job. There is no benefit in cost or labor to increasing the size of a project. In short, it is a process rife with inefficiency. It is expected that teams of 2-3 technicians working 40 hours per week could migrate no more than 5,000 hours of content per year.

### **Towards an Automated Migration System**

The solution to the problems cited above is a robotic tape handler and automated cleaning equipment, capable of running continuously, with software to support intelligent decisions about tape processing in order to streamline migration. This system would take advantage of economies of scale and drive down costs while improving tape and signal quality. The process should occur on-site, at the archive, with mobile equipment designed to be installed only for the length of the project and then removed. This idea of a ‘factory approach’ has been put forth by the European Union’s PrestoSPACE<sup>i</sup> preservation project as the necessary solution to the problem of scalability<sup>ii</sup>.

A robotic migration system controlled by customizable, intelligent software has numerous advantages over traditional methods. Running 24 hours a day, 7 days a week, a robotic system can process huge quantities of tapes. Sophisticated analysis tools can record valuable metadata about the content being processed. Metadata collected on individual tapes and across the entire collection gives insight into the health of an

archive's tapes that was never available in the past. Real-time analysis based on consistent, objective metrics allows for sorting and removal of fragile or damaged tapes that require extensive restoration work. Instead of treating an entire collection as if each tape needed extensive work, tapes are segmented into less costly batches that can be automatically migrated and more costly batches requiring human intervention.

Increased productivity can be realized by performing several preservation functions at once. Migration to tape stock, migration to digital media file, and collection of metadata should occur simultaneously in the system. Robotic systems allow for migration on a massive scale and can be customized to the particular needs of individual archives.

By reducing the amount of human labor involved in audiovisual migration, advantages of scale and consistency are introduced. A system in which computers are driving the playback, recording, and encoding processes will realize a significant increase in output over traditional migration practices. Robotic tape handling systems are a well-established technology born of broadcast television automation and allow for reliable cassette manipulation. Multiple deck-to-deck or deck-to-encoder configurations enable multiple playback "streams." Each stream can be routed to an appropriate destination, i.e. another tape format or digital file, or both. Each stream can also be monitored individually for signal issues that, if serious enough, could cause the system to abort a specific tape's migration. Each resultant destination format can be checked for quality at the conclusion of a migration, alerting system operators if the final tapes or files have problems.

Current mobile robotic tape handlers can contain from 3 to 5 to playback decks, depending on the format<sup>iii</sup>. A system built to accommodate the Beta family of cassettes would allow the use of up to 5 decks in single robot; if this system were migrating tapes to digital files only (i.e. no tape copies), such a system could theoretically migrate 120 hours of content per day<sup>iv</sup>. At 7 days per week of operation, this yields a potential of 43,800 hours of migrated content per year – nearly an order of magnitude increase over traditional methods, which are, in any case, focused on tape-to-tape migrations, rather than tape-to-file.

### **Case Study: Yale University**

A prototype of an automated migration system outlined above was launched at Yale University in February of 2005. Among Yale's numerous video archives is the Fortunoff Video Archive for Holocaust Testimonies. With over 4,200 testimonies recorded on U-matic tapes, the Fortunoff collection represents a tremendously important historical and scholarly resource. Testimonies began to be recorded in 1979 and have continued to be recorded to the present day. A set of 2,250 hours spread across 2,750 tapes was chosen for migration because of the tapes' age. Media Matters' System for the Automated Migration of Media Assets, or SAMMA, was deployed at Yale to migrate the Fortunoff tapes from one U-matic master to two re-masters on BetaSP tape plus MPEG-2 digital media files at a bitrate of 2.9 Mbps. In the Yale configuration, a maximum of 30 destination tapes could be loaded into the system, allowing for a maximum of 15 source tapes per migration cycle.

Before the SAMMA system could be installed, the Yale tapes had to be inspected and prepared for automated migration. When tapes are migrated manually, each tape is

treated as a “problem tape” – that is, each tape is inspected, cleaned, and migrated as if it required special handling, whether it needed it or not. Training a person to inspect all tapes prior to processing can eliminate this costly inefficiency. The Yale tapes were separated into two batches: those that could be processed by the automated system, and those that must be submitted for costlier manual migration. Tapes ineligible for automated processing were those with physical defects, such as a broken cassette shell, moldy tapes, tapes with severe breakdown of the adhesive binder, and tapes exhibiting excessive “shedding” of oxide particles. Since automated migration could potentially harm these severely degraded tapes, or harm the system equipment, the tapes were removed from circulation. Archivists could then make a decision about sending those tapes out for manual restoration based on their curatorial value.

Tapes that passed the inspection process were then barcoded with a 2-dimensional barcode the size of a postage stamp. The SAMMA system “sees” tapes by scanning the barcode affixed to them. To keep track of tapes and their state (inspected, Pass/Fail, cleaned, etc.), each tape is assigned a unique 64-character string<sup>v</sup>. Traditional 1-dimensional barcodes (such as the UPC symbol) would be too long, exceeding the width of a tape, with a string of this length. A 2-dimensional barcode standard called Datamatrix was selected because the symbols could contain over 2,000 alphanumeric characters. After barcoding, tapes were then returned to their cases to await processing by SAMMA.



After Yale’s archival staff prepared all the tapes, the SAMMA system was installed. Yale’s business constraints limited the number of daily reloads to two: one in the morning, and one at the conclusion of the business day. This limitation reduced SAMMA’s actual operating hours to 16-20 hours of content migrated per day. To realize the full potential of SAMMA’s operation, an archive would have to load and unload SAMMA at least 3 times per day for tapes averaging 30 minutes of content, or 1 load every 15 hours for tapes with 60 minutes of content.

SAMMA migrated Yale’s tapes from U-matic to BetaSP and MPEG-2 simultaneously. The MPEG files were stored locally on a 1 TB RAID. At the conclusion of the migration, the files were automatically moved from SAMMA’s storage to Yale’s storage system via Gigabit Ethernet. A checksum operation was performed to ensure the integrity of the file transfer. At this point, the new digital assets could be ingested into Yale’s network through whatever established workflow was most appropriate to Yale’s needs. This is the finish line for SAMMA – once the files have been moved to the client’s storage system, it is the client’s responsibility to ingest or manipulate the files. Specialized DAM functions such as speech-to-text, image identification, keyframing and shot segmentation, or closed caption transcription, can be performed by pre-existing customer DAM systems without regard to the process that created the digital files. This was a business choice; rather than replicate functions that are already done well by several DAM vendors, SAMMA was designed specifically to migrate magnetic tape-based assets to digital file, regardless of file type, storage technology, or method of asset management.

## **Lessons Learned**

Several lessons can be derived from the migration work performed at Yale University, with applicability to both the archival and commercial industries. Firstly, it is crucial to have high-level executive buy-in to accomplish a project of this scale. Though automated, a system of this kind requires a significant commitment in financial and personnel resources, as well as alterations to the prevailing workflow. Each archive is different – various workflows, infrastructures, funding sources, and uses for archived content – which means that any migration system brought in-house must be customized. Choices such as a destination tape and file format, network transport and security, personnel scheduling, and a myriad of other system characteristics must be decided upon prior to the beginning of a migration job. The best approach is consultative, where stakeholders and vendors are brought together to tackle large and small issues through in-person meetings, conference calls, and email. Since this type of automation is very new, both vendors and clients should plan for a good amount of reciprocal education; clients must be educated on the nature of the technology and limitations of the system, and vendors must learn each archive's extant workflows, business constraints, and expectations. Non-profit and governmental archives have a noticeable lack of knowledge about digital media technology, and of Digital Rights Management in particular. Commercial entities should ascertain their staff's level of familiarity with these technologies when discussing the potential use of an automated migration system.

One surprising lesson learned was that the value of metadata is in the eye of the beholder. The original system design for SAMMA included a tremendous amount of metadata that had never been available before from traditional migration practices. Metadata about tape cleanliness, tape defect type and location, audio and video signal characteristics, signal errors, silence in the audio track, information about the transformation from analog to digital, and many other types of information were to be gathered by the system during the inspection and migration process for the operational needs of the system and use by the client. However, since the availability of this information is very recent, many archivists were unsure of how to use it. The conclusion was that during the consultative process vendors and clients should discuss what metadata is available, what is desired, and what is extraneous.

## **Advancing DAM**

The issue of metadata is particularly important to the art of Digital Asset Management. Since there was no way to migrate magnetic media to digital en masse prior to the creation of automated systems, there are no standards for the capture and use of metadata that results from the migration. However, a DAM system is an ideal repository for the large quantity of metadata that can be obtained from a well-designed automated system.

Information about the transformation from analog to digital – so-called “provenance” metadata – is important to both archivists and managers of digital assets. For archivists, provenance is important to understand the collection history of a media item and ensure that future scholars who use it have complete a picture about the item as a source for research. In the commercial world, where a media item can be delivered via diverse channels such as cable television, the Web, portable media devices, and wireless

handsets, provenance information is important to the technicians responsible for conforming the media to the specifications of the delivery medium and ensuring that its quality is not compromised. Digital media characteristics like codec, bit depth, number of audio channels, and bitrate can be automatically “discovered” by an automated migration system and then handed off to the recipient DAM system to be appended to the associated database record for a media item. The same is true for closed caption transcripts, which can then be linked to time code, allowing keyword searching that relates to specific segments within a digital video file. The consistency that an automated system affords helps to advance DAM use, providing uniform metadata that assists content users in their search and retrieval of media assets.

Automated systems help DAM vendors by putting digital assets and analog assets on a level playing field. The low-hanging fruit for DAM vendors is the set of born-digital assets produced by their customers. The more costly business of transforming legacy media into digital assets is closer at hand with the advent of automated systems. DAM vendors can now make customers out of organizations whose media collection is mostly tape-based, or can offer solutions to their existing customers who have a significant amount of analog assets.

The union of an automated media migration system and Digital Asset Management can lead to standardization of the transformation of analog assets to digital assets. Archivists, digital asset managers, librarians, content users, and systems designers all have a role to play in the creation of best practices – long-term use and storage of formerly analog assets will yield the lessons needed to create standards and consensus. Efforts such as the European Union’s PrestoSPACE, a pan-European, multimillion-euro project to explore a “factory approach” to preserving Europe’s audiovisual heritage, is advancing the field of migration and restoration, and will have an impact on the DAM industry. In the metadata realm, the METS standard<sup>vi</sup>, under the guidance of the Library of Congress and the Digital Library Foundation, is advancing the methods of describing digital library objects, and has generated international interest and discussion. In general, momentum in the area of digital audiovisual migration will have positive effects on the DAM industry, and vice-versa. Automated systems are an important piece of a rapidly evolving field, and will drive down the costs to ingest aging, valuable analog media assets.

---

<sup>i</sup> <http://www.prestospace.org/>

<sup>ii</sup> PrestoSpace Annex I, September 15, 2003

<sup>iii</sup> This estimate includes reserving room in the robotic handler for an automated cleaning deck.

<sup>iv</sup> Professional video decks can only playback in real-time. Therefore 5 decks X 24 hours = 120 hours of playback.

<sup>v</sup> A “string” is an alphanumeric sequence. The string was generated using a standard from the Society for Motion Picture and Television Engineers (SMPTE) called UMID, or Unique Material Identifier. This UMID is the key identifier that links the SAMMA system to the pre-existing database of bibliographic information about each tape.

<sup>vi</sup> <http://www.loc.gov/standards/mets/>

# The System for the Automated Migration of Media Assets

Turning Analog Assets into Digital Assets for Use in Digital Asset Management Systems

