Original Article

Digital preservation of mass media artifacts: Technologies and challenges

Ramesh Singh
is a senior scientist at NIC, and has worked extensively in O.S, Database Management Systems and other System Software. He has studied at IIT Kharagpur and at IIT Delhi. He holds an M.Tech from IIT Delhi. His present research interests are in Mobile, Pervasive Computing and Human Computer Interface.

ABSTRACT There is a need to digitize the large reservoir of broadcast archives, at reasonable cost using modern digital technology. An archive sees itself as having four functions, namely Collection, Preservation, Organization and Dissemination or Access. This paper focuses on providing some guidelines for an integrated approach to producing sustainable digital assets with graceful access for better return on investment and distribution. The crux of the matter is access. A highly accessible digital artifact is much more valuable than a physical artifact locked in a shelf accumulating dust. With the digitization of the audio-visual broadcast archive, this paper addresses the issues of Preservation, Restoration, Metadata Access and Delivery, Storage, Archive and Information Rights Management.

Journal of Digital Asset Management (2009) **5,** 185–195. doi:10.1057/dam.2009.12

Keywords: digitization; preservation metadata; digital encoding; information rights management; mass media artifacts

INTRODUCTION

Mass media in its most generic form is used to denote a section of the media specifically envisioned and designed to reach masses, such as the population of a country. It became prevalent during the 1920s with the rise of nationwide radio networks, mass-circulation newspapers and magazines. The mediums for mass media are newspapers, television and radio broadcasting. The new addition to the concept of mass media is internet media. These internet media can include personal webpages, podcasts and blogs.

The context in which the media is used here points toward the well-organized means and ways of conveying or disseminating the news, public opinion, various kinds and forms of entertainment, and all other information relevant to citizens, through newspapers, magazines, advertising, feature films, radio, television, World Wide Web, books, CDs, DVDs and any form of publishing.

Correspondence:
Ramesh Singh
System Software Division, National Informatics Centre, A-Block CGO Complex, Lodhi Road, New Delhi 11003, India

The 'requirements for digital preservation' for the broadcasting industry appears to be a daunting task assuming greater dimensions day by day. The first objective of this paper is to provide in-depth knowledge of the technical devices, technologies and systems for digital preservation of all types of audio-visual collections or broadcast archives in particular. The second aim is to provide sufficient know-how to build up preservation centers to help implement digital media libraries in order to provide affordable services to all kinds of collections of broadcast archives to manage and distribute these assets. These historical, cultural and commercial assets are now entirely at risk of deterioration.

The various forms of electronic and print media are broadcasting through radio and television, discs and tapes for music, films and documentaries, internet, blogs and podcasts, publishing via books, magazines and newspapers, and also computer games, which have become a mass form of media through devices such as playStation3, Xbox 360 and so on.

There are critical problems of right size digitization, automation, metadata extraction, storage, network bandwidth, access control and good delivery mechanisms for public access. The real challenge lies in the uniform aggregation of digital archive descriptive metadata and related digital rights information. We need to move to tapeless production, which introduces a new environment for broadcast production and re-use of the broadcast archives.

PRESERVING THE PRICELESS CONTENT: REASONS GALORE

Radio and television broadcasts have forever altered the way citizens communicate, learn and live. From local and national news to TV serials, talk shows, sports, music, comedies and dramas, broadcast programs serve as living history, and continue to shape how we view the world past and present. For years, the voices and images generated from radios and television sets were recorded on analog tape, making them vulnerable to the effects of age and highly susceptible to damage through use and deterioration over time.

The obsolescence factor

Millions of hours of valuable multimedia content in the form of radio and television broadcasts constitute an irreplaceable heritage of a nation that is at risk because of aging media and technology. This content should be preserved for the benefit of citizens, in a cost-effective manner.

There is a specter haunting the disappearance of broadcast archives, and the extinction of more than several decades of recorded historical and cultural memory of a nation. Videotape came into general use 30–35 years ago. More than half of the broadcast archive holdings are now aging, and need preserving before they deteriorate beyond recovery. Equipment to play audio recordings on tape and video recordings on videotape is now also aging or obsolete. Spare parts and skilled operators are also fast disappearing. Aging media and obsolete equipment combine to produce a problem not previously encountered in broadcasting. The 'requirements for digital preservation' for the broadcasting industry appears to be a daunting task assuming greater dimensions day by day.

Types of obsolescence

Because of the overly specialized medium, that is, films or obsolete 2(2 Inches) videotapes and audio on 1/4 open-reel tape, huge amount of archive materials are rendered useless for access by the users. Deterioration takes place in the form of dye fading on color films, weakening of binders, and adhesives holding magnetic particles to the polymer tapes used for both video and audio recordings. This problem is further enhanced because of 1/4 tapes in audio archives and U-Matic material in video archives.

The rate at which recording formats in the form of discrete media become obsolete can be intriguing for any archiving practice involving audio-visual assets. Although most of the formats were physically qualitative and quite durable to the extent of couple of decades, but there had been problems accessing these formats, very frequently. In fact, the digital media accelerated this problem in few cases. The lifecycle of formats like CDs(g), Digi Beta(g) video tapes and DVD(g)s is limited.

Acetic acid is formed from acetate tape (vinegar syndrome). This acid, primarily from film soundtracks, can then attack both film and video (or even audio) material if held in the same room. Nitrate film: all archives have segregated this material, which is extremely combustible and can be explosive, but not all archives have copied the material onto safety film. Fragile media: a large part of the holdings cannot be released for access because the media are too easily damaged, for example film negatives, film prints except for those accessed by qualified professionals, all shellac and vinyl audio recordings.

The access factor

Access to broadcast content is extremely difficult, because of its availability on proprietary formats like film and broadcast-standard videotape, which need special players, certainly unavailable to the general public and often unavailable even to other national institutions. Most of the content cannot be allowed to circulate in general. The major objective of the digital preservation of the broadcast archive is the preservation for access.

The typical problem of request for copies of tapes is another challenge, particularly when they come from broadcast companies. A staff member would have to copy a tape to the desired format (DVD, Beta, CD or 3/4), package it and dispatch it for delivery via regular postal services. The task takes at least 24 hours – plus shipping time – making it impossible to accommodate same-day broadcasts without special provisions.

Jumping on the bandwagon, catch on the wave: Digitize? Digitize? Digitize?

There must be strong and compelling reasons to undertake this digital transformation of audio-visual archive with respect to the time, effort and cost involved. It makes sense to move from discrete and/or analog storage into a digital mass storage system. With digital technology, there is the opportunity to preserve media content for generations to come and make it widely available to more people, through more channels. With the digitization of the audio-visual archive, we address the issues of Preservation, Restoration, Metadata Access and Delivery, and Storage and Archive Management (SAM).

Some statistics for contemplation

By the end of the next decade, a million hours of media content of audio type should not cost more than Rs. 10 000. One petabyte or 1000 terabytes (TB) will hold 165 years worth of continuous broadcast encoded in 16 bits at 48 KHz. Storing the equivalent amount of media in analog form typically represents 2.5 million documents spread over 80 km worth of shelves.[1] It is expected that 15 years from now, that amount of content will be in our pockets, on iPODs.

The following are some of the very important key findings from the Preservation Technology for European Broadcast Archives (PRESTO) project under the European Commission's Information Science and Technology program.[2]

1. There is a widespread lack of condition assessment.
2. Some 70 per cent of the material is decaying, fragile and/or on obsolete formats.
3. Preservation projects (planned or underway) would transfer 250 000 items per year (about 1.5 per cent of total holdings). At this speed, the material at risk would not be transferred for 60 years.
4. Much of the material will not last 60 years (average usable life of a videotape is 20 years).
5. Even if the material is intact for 20 years, equipment and trained staff to transfer the material will be scarce and expensive.
6. The situation is not static – and new material coming into the archives exceeds preservation capacity by a 4–1 ratio,
7. The transfer rate of 1.5 per cent may be optimistic, as the budgets and resources are insufficient.
8. The survey found that archives had half the budgets and half the capacity they needed.
9. The project went on to outline scenarios for what might happen, based on the current situation continuing. These varied from a 'best case', where 40 per cent of existing material will disappear by 2045, to a 'worst case', where 70 per cent of material will be lost by 2025.

Motivations for digitization

Technology is a great enabler. Sometimes it changes the very perception of the job we do. For example, in the analog world, preservation of the master copy is an extremely important step, whereas in the digital environment, there is no difference between duplicates and masters. In this digital world, carriers have become virtual. Today's digital world offers storage technologies to handle heterogeneous mediums like Hierarchical Storage Management that can combine hard disk storage with tape libraries and DVD jukeboxes, which can store millions of broadcast content near-online. The most significant of all this is the virtualization of the carrier, for the digital preservation and the distribution of the content.

Emergence of new ecosystem

With the advent of new standards, such as broadcast wave formats, Metadata eXchange formats, combined with enhanced streaming formats such as MPEG-1 Layer 3, are really setting the new paradigm for broadcast archiving. The new paradigm is more important from the distribution viewpoint than the preservation aspect. In this way, the archivist can ensure the right collection of metadata and facilitate search and retrieval much faster and accurately, in spite of the increase in the volume. The first step toward realizing the new ecosystem will be to start digitizing the day-to-day archival storage, ensuring thereby that tomorrow's content will not be stored on tapes. We should, in parallel, start to digitally transfer and properly catalog the decaying media.

A majority of organizations and institutions is still practicing shelved CDs as the preservation medium for audio archives. In a recent survey of 84 international archiving institutions, 45 per cent of collection owners use CD-R as the preservation format of choice, and 57 per cent still regard CDs as the appropriate media for access and distribution.[3]

At the outset, it may appear that CDs may be a very cheap option as a media for storage. But what is significant is the cost and labor required for maintenance of the lifecycle of such creations, namely the sequence from production to retrieval. This is definitely huge. The process of burning a CD, creating the index and preserving the medium on a shelf can be extremely time consuming, costly and labor-intensive operation. As per the estimate of a study, it may cost Rs. 250 000 to archive 9000 hours of audio recording at 48 KHz, 16 bit on CDs, out of which 80 per cent is pure labor, whereas storing 9000 hours of audio recording requires a total of 6 TB. The equivalent storage solutions based on high-performance, redundant hard-drives may cost just half the price of the earlier method and the same time if provided in an online environment.

Speed and access tradeoffs between digital file-based and tape-based models

Broadcast archiving has two facets, cataloging and retrieval. Both the facets are multi-step processes. Cataloging or preservation starts from ingest to logging and storage. Retrieval or distribution involves search, recovery and reconciliation.

The very process of broadcast mechanisms of cataloging and retrieval workflows suggests that there is a major productivity gain in adopting file-based model (Table 1), as against the tape-based model. The file-based model exploits the ubiquitous computer networks. This is further complemented by the day-to-day reducing cost of the storage systems. Real effort is needed toward creating a suitable metadata model, which can encompass the entire lifecycle of the broadcast system to capture related information at every phase of the workflow. This is how the archives can be transformed into assets and simultaneously minimize the retrieval time .

The tape-based model has several drawbacks ranging from gathering news from the field, dubbing, transfer to video server, reviewing logging material to producing a shot list and so on. The process is highly labor-intensive.

In a file-based environment, the ability to automate tasks is the primary source for productivity gains. Although the time required for feed ingest remains unchanged, the need for dubbing material to a compilation tape and for storing it disappears, thus saving time. Shortlisting can be partially optimized by the use of a low bit-rate frame accurate clone or

Table 1: Throughput advantages in shifting to file-based model

Processes or jobs	*Approximate job execution time in tape-based environment*	*Approximate job execution time in file-based environment*
Cataloging or preservation		
Ingest	3 min	3 min
Short listing and logging	12 min	10 min
Dubbing and storage	8 min	—
Total execution time	23 min	13 min
Retrieval		
Searching	16 min 30 seconds	7 min
Recovering	3 min 30 seconds	3 min 30 seconds
Reconciliation	6 min	0
Total execution time	26 min	10 min

proxy video file that can be browsed over the computer network. Implementation of productivity enhancers such as the use of foot pedals to optimize transport control of video preview can also provide significant reductions in the time needed for logging. In a file-based environment, the overall efficiency of the cataloging process can be enhanced by over 40 per cent.

The potential productivity boost that can result from moving to a file-based environment becomes even more striking when considering the retrieval or distribution process. In a tape-based world, searching for relevant material is a multi-step process: entering one or more keywords in an online text archive and obtaining a relevant hit list typically requires 1 min. Moving to a tapeless environment where all archived material is available online drastically reduces this retrieval process. The ability to type keywords and immediately browse the video files that meet the search criteria greatly enhances productivity. Journalists can use such a *visual search* to rapidly determine which shot is most relevant for their story. As they have no need to move from their workstation, they do not have to go tape-hunting across production and archive departments.

The cost factor for online storage for broadcast video

The benefits of adoption of a virtual, file-based carrier are significant. It takes less time to access a file over a network than to retrieve a tape from a shelf. Until recently, online storage for broadcast quality video remained too expensive, but this is no longer true.

There is a continuous increase in the capacity of hard drives; it is almost doubling every year. Maintenance and system administration costs tend to be lower for IT-based solutions, as local area data networking become easier to deploy, and Network Attached Storage (NAS) systems turn into plug-and-play appliances.

A file-based environment offers several options for newsroom kind of scenarios.

The first option requires an important upfront capital expenditure: a 5-TB online NAS appliance can be combined with a 120-TB IT-based tape library. The resulting system combines both online and near-online storage. The online NAS system stores a low bit-rate proxy of all archived content as well as a buffer for the most recent 3 months worth of archives in high resolution. The tape library offers a storage capacity equivalent to 10 000 hours of broadcast-quality video (or 8 years of archives). Journalists can browse all archives from their desktops, and have immediate access to the past 3 months of archives in broadcast quality; older archives will require a few min to be available in broadcast quality for editing.

The second option considers online storage as an operating expense. A 15-TB storage device is deployed to meet the next 12 months' need of archiving. Both high resolution and low bit-rate proxies are stored online. There is no need to use an IT tape library as a primary storage device. Journalists have immediate online access to all archives in both low bit-rate and broadcast quality. From then on, additional storage is purchased on a yearly basis. Considering that storage costs are divided by two every 12 months, digital archivists can ride the dropping price wave and adopt a step-by-step, incremental storage policy: storage capacity can be acquired on-demand, very much like tape was in the analog world.

A transition to a tapeless model allows archivists to truly take advantage of the fact that digital copies are in no way different from original digital recordings. Making sure that the archiving process does not give way to offline archives simplifies the generation of back-up copies stored at alternate sites, and ensures the implementation of a disaster recovery policy. What is more, it makes future migrations from one storage medium to another seamless, and constitutes a pre-emptive measure to technology obsolescence. This is all the more necessary, as the lifespan of digital data storage is much shorter than that of analog media. The life expectancy of polyester-based magnetic tape maintained in a proper environment is evaluated to be 50 years; for digital optical media, it is estimated to be between 30 and 200 years. However, for IT-based storage, it will not exceed 5–10 years.[4] From this perspective, the use of files as virtual carriers constitutes a major switch of paradigm for most players in the archive world. It recognizes that the quest for the ideal audio-visual storage medium is a dead end and that archives can no longer be defined by their physical attributes, but as a 'logical space independent of the production environment where records are protected from loss alteration and deterioration.'[5]

METADATA IS THE KEY IN CREATION OF THE ONLINE DIGITAL MEDIA LIBRARY

The present practice of broadcast archiving can be optimized across the organizations in order to address the issue of metadata creation and collection. IT-based storage may provide the ideal, virtual carrier, but it does not solve the challenge of properly indexing the content. The digital, online environment allows production teams to directly access archived material. News and feature producers can have immediate and around-the-clock access to the news library. As a result, they can stop maintaining their own private micro-archive islands, which typically litter corridors of most broadcast organizations. In this emerging context, the archivists and catalogrs will be able to change their role to *media* or *knowledge manager, and they can now move into the production space.* Although naming has its importance, the strategic issue is for archive professionals to move into the production space.

Much of the information required for indexing is already collected by journalists and feature producers. In many cases, *technical* and *descriptive* metadata can be aggregated with minimal changes to daily operations.

As an illustration, information from the newsroom computer system (NCS) can be automatically associated as descriptive metadata to corresponding clips. Scripts and rundowns can be periodically exported from the NCS. Based on the unique identifier common to a story and its corresponding clip, textual information can be parsed and scripts can be reconciled with the corresponding video files. As an extension, close-caption information can be gathered and associated to the original audio-visual material to enable full-text searches.[6] The use of automated, speaker-independent speech-to-text engines can be considered to generate rough transcripts for field material that lacks any associated textual information. Although insufficient for legal purposes, such automated transcripts can provide relevant metadata for searching large volumes of audio-visual data.[7] Whichever option is chosen, the resulting time code-based markers can serve as indexes for future searches, as long as the adopted metadata scheme is flexible enough to meet current and future requirements.

What and how of metadata

Metadata literally means 'data about data'. Any catalog – card or online – contains metadata. But today, the term is increasingly applied by information professionals to the value-added information that they create to arrange, describe, track and otherwise enhance access to information objects.

Metadata is used to describe, in a standardized way, the information necessary to locate a document. The aim is to provide a minimum set of metadata that is understood and used by everyone. A standard set of metadata will

- provide a standard way to describe network-accessible material;
- enable more precise queries to be made;
- help search engines to present 'hits' grouped by subject rather than as a random mix.

Metadata and the media lifecycle

It is essential to ascertain that technical, descriptive and *legal* metadata should be entered at every step of the media lifecycle. The ability to structure metadata and to dynamically adapt it to the type of archived asset as well as to user profiles is a key to successfully implementing a digital broadcast library. The future value of preserved assets depends on it. The European PRESTO project recommends 'the value of an item be more than four times the preservation cost in order to be financially justified on a commercial basis.'[8] To turn archives into such assets, archivists need to focus not only on the descriptive metadata, which is the key to future access, but also on the associated copyright information.

Metadata elements for radio archives

This paper defines a simple set of metadata elements that is adapted for use in radio archives (Table 2), but that is aligned both with the main metadata standards of the broadcasting industry (EBU/SMPTE/AES) and with the Dublin Core metadata (the general approach used by libraries and archives, as well as the World Wide Web[9]). The metadata set proposed in this paper contains

Table 2: Metadata definition for radio archives

Element	*ID*	*Definition*
Element 1	Title	A name given to the resource
Element 2	Creator	An entity primarily responsible for making the content of the resource
Element 3	Subject	The topic of the content of the resource
Element 4	Description	An account of the content of the resource
Element 5	Publisher	An entity responsible for making the resource available
Element 6	Contributor	An entity responsible for making contributions to the content of the resource
Element 7	Date	A date associated with an event in the lifecycle of the resource
Element 8	Type	The nature or genre of the content of the resource
Element 9	Format	The physical or digital manifestation of the resource
Element 10	Identifier	An unambiguous reference to the resource within a given context
Element 11	Source	A Reference to a resource from which the present resource is derived
Element 12	Language	A language of the intellectual content of the resource
Element 13	Relation	A reference to a related resource
Element 14	Coverage	The extent or scope of the content of the resource
Element 15	Rights	Information about rights held in and over the resource

15 individual metadata elements, with their qualifications, in accordance with the Dublin Core[10] and recommended usages.

SETTING UP THE PLAN AND PROCESSES FOR PRESERVATION AND DIGITIZATION OF BROADCAST ARCHIVES

Setting up a digital environment for preservation of broadcast archive involves proper designing of workflow concepts and business processes. This consists of four process descriptions related to the different activities.

1. Preservation Process
2. Restoration Process
3. SAM
4. Documentation and Publication Process (Tables 3–6)

DIGITAL TECHNOLOGY CONCERNS AND ISSUES IN BROADCAST ARCHIVING

Analog versus digital

Analog recording was the standard form of recording until recent years, and involves the representation of information as it has been recorded (Figure 1). When material is transferred from one analog device to another, the copy will not be exactly the same as the original, for example a sound recording will pick up background noise. This implies some loss in quality when a copy is made ('generational degradation').

Digital recording creates a binary, computerized representation of information (Figure 2), and uses this to re-play the audio-visual information. It involves no generational degradation.

Media and their types used to archive existing broadcast material

The following are the types and kind of media used for film, video and audio archives, in general throughout the world.

Film Archives: Most film materials in television archives are cellulose acetate-based. Film is stored on a range of different film media: positive or negative film, 16-mm or 35-mm film and magnetic sound film.

Video Archives: Analog video has been mainly recorded on magnetic tape. There is a variety of analog video formats, more than for film or audio. Within each format, performance partly depends on manufacture.

Audio Archives: Media on which audio archives are stored can be broken into three groups: magnetic media (0.25 inch tapes), vinyl-like media (78, 45, and 33 RPM vinyl) and other media (shellacs and wax cylinders).

A study of the present holding patterns of the broadcast material of organizations throughout the world indicates that their archives are mostly in their original *Analog Formats*. Hence, there are issues of storage and preservation.

Table 3: Step-1: preservation process

1. Referencing contents and media	*2. Preparation of batches*	*3. Shipping*	*4. Media cleaning or repairing*	*5. Playback and digitization*	*6. Quality control and data analysis*	*7. Updating archives*
1. Minimal description labeling and bar coding	2. List of media and contents for order forms and management	3. From archives to technical area facilities	4. Media condition assessment and physical preparation	5. Devices designed for archive material and supervision facilities	6. Checking integrity and quality level technical data management	7. Storage and description Information updates

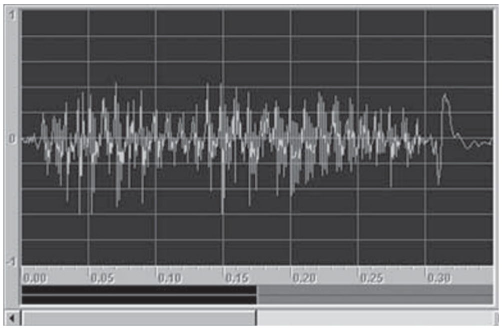

Figure 1: Analog waveform.

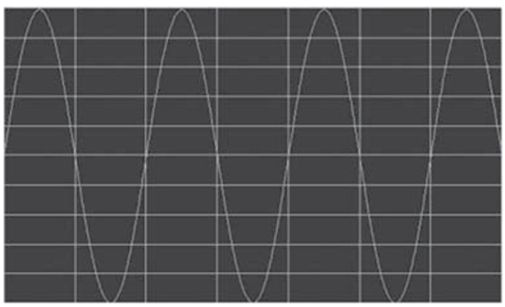

Figure 2: Digital pattern.

In general, the following statistics hold true on current analog archive storage:

- At least two-thirds of archived material cannot easily be used in its current form.
- About one-third of material is in some form of deterioration.
- About one-quarter of the material is too fragile to be released for access.
- The annual intake of new video and audio material is up to four times the rate of preservation.[11]

The famous PRESTO project of the European Union has identified the following Key Technology Links.

Videotape Transport/Playback – It is particularly important to have the best possible electronic processing in order to ensure reliable and stable signal detection to the greatest degree possible, using modern methods. This will minimize the dropout, and hence the transfer errors. This approach is better than relying entirely on dropout correction, which is not perfect, and is essentially a restoration technique, not a preservation technique. It is also possible to improve the mechanics of old transport, and improved handling is being developed for 1″ and U-Matic machines, particularly addressing the problem of tape motion impairment (sticktion). Dropout correction will be re-examined with regard to digitization, because if a video signal is being captured to a server, there are opportunities for signal correction that are not possible in a standard transfer (for example, using information from adjacent frames and from preceding as well as following lines).

Film Transport – A major problem in broadcast archives is news item films from the 1960s to the 1970s, often spliced item-by-item onto 'day reels'. This material often falls apart when running through film transport, and manual cleaning and re-splicing is 90 per cent of the preservation cost of this material. A fully automated method is being developed for this material. Alternative methods of film handling that do not require re-splicing are also being investigated.

Quality Control – This remains the main transfer cost, but with modern transfer operations requiring four or even five items being transferred simultaneously, full manual monitoring is not performed. Automatic methods to identify basic problems with audio signals will be developed, as a 'software toolkit', to back up the spot checks performed by the operator. This will allow the operator to

Table 4: Step-2: restoration process

1. Original media transfer to file	*2. Automatic defects and quality detection and restoration planning*	*3. Video restoration real-time pre-viewing and parameter adjustment Online processing* *4. Film restoration interactive previewing and parameter adjustment semi-automatic retouché offline processing* *5. Audio restoration real-time pre-listening and parameter adjustment offline and online processing*	*6. Preservation project management*	*7. Technical watch*
1. Flat transfer to intermediate on disk/file (Preservation Process)	2. Automatic and manual detection for expertise restoration plan	3–5. Defects restoration by restoration systems 3.4 and 5 Selection of restoration system according to media type, kind of defects, amount of defects	6. Export technical data to archives information database	7. Final restored version recording and delivery

Table 5: Step-3: storage and archive management (SAM)

1. Technology analysis	*2. Storage solutions and costs*	*3. Storage life cycle and planning*	*4. Preservation project management*	*5. Technical watch*
1. Analysis of technologies for audio-visual storage	2. Design and evaluation of dedicated storage system strategy	3. Archive quality assurance and migration plan	4. Preservation and access project plan	5. Updating technology analysis

Table 6: Step-4: documentation and publication process

1. Item acquisition planning	*2. Preservation process*	*3. Offline processing and manual enrichment*	*4. Query management rendering streaming*	*5. Repackaging and delivery*
1. Item ID allocation legacy data ingestion Choice of analysis tools and workflow setup	2. Digitization of content, performed by the preservation workflow	3. Non real-time processes, manual annotation, final validation	4. Query management rendering streaming	5. Repackaging of content, including essence and metadata, and delivery to customer

concentrate on properly setting up a transfer, ensuring initial quality, and then relying upon automation to warn of signal dropout or gross problems during the course of the transfer.

Signal Capture – Although video is easier than audio in terms of multi-signal monitoring, it has far more problems with signal dropout or transport errors (head clogging) during playback. The gross errors will appear on the monitors, and the trivial errors (single-line dropout) can be immediately compensated in hardware. In between these extremes, the operator will not know how much brief loss of signal has occurred. Improved monitoring of playback, with automatic logging, will be developed.

Error Logging – Many aspects of signal playback may have problems. Signal restoration projects have identified dozens of artifacts in archive signal: blotches, scratches, flicker, grain, noise and fading. Many of these artifacts can be automatically detected, with varying degrees of difficulty and success. Although preservation work is not the same as restoration, it is desirable when conducting preservation work to capture any useful information that may be needed for subsequent handling of the item. In audio, such a log is now an addition to the EBU Broadcast Wave Format[12] standard.

Encoding methods for audio and video –

- *Lossless Encoding* (audio and film/video),
- *Multi-level Encoding for film* (creation of a master format that efficiently supports conversion into all required presentation formats for both cinema and TV),
- *Multi-level Encoding for video* (efficient creation of various quality levels: key frame, internet, browse and studio).

The path to digitization

There is an up-curve toward converting the existing materials to the digital format, and the new materials are in general being created in digital formats.

The motivations for moving toward digitization are as follows:

- The older analog media degrade over time, and thus quality is reduced.
- As the devices needed to access and play analog media themselves become obsolete, acquiring and maintaining such devices becomes more expensive.
- Provided digital material is backed up and held on more than one site, it should be secure.
- Digitization of archives can potentially allow greater access to materials, and allow archives to become a public amenity.

Concerns to be set right before adopting digitization

Common standards for digitization

Standards always help in generalizations, and offer graceful interoperability. Although the archiving of new material is now likely to be in a digital format, multiple digital formatting options exist. Thus, the issues regarding choice of digital formats such as CDs and DVDs for viewing copies versus Digital Linear Tape and DigiBeta does exist.

The literature shows a debate as to whether such formats are an interim solution and whether television or other moving image activity is likely to migrate in the coming years to the Internet Protocol platform used by the World Wide Web. For audio archiving, while tape formats are still commonly used, audio material is increasingly being stored on computer hard drives. There is a concern about the long-term stability of such formats.

Requirements of common standards for metadata

A standardization of metadata (that is the cataloging information embedded in the material about a program or other archived item) will aid both the archiving process and the subsequent accessing of material. Such metadata allows archives to add additional layers of data according to special interests, for example it can include information useful to a producer (such as the breakdown of particular shots) or to a historian.

The advent of interactive television

Interactive television include Digital Television, Podcasts and Digital Video Recorders (DVR)/ Personal Video Recorders (PVR). The most widely known DVR/PVR is TiVo, which allows users to capture television programming to internal hard disk for later viewing, also called 'time shifting'.

There is an immediate concern regarding the impact of Digital Television, with its variety of applications including chat, e-mail, and games and shopping.

Copyright issues

As TV or radio material is stored digitally, it becomes easier to access remotely, and this raises the issue of material being listened to, or viewed, by large numbers of people, or being re-broadcast or re-purposed.

CONCLUSION

Digital Preservation as a discipline is in its infancy; yet it is here that the potential for audio and video data to survive past the life of discrete, market-driven carriers can be realized. The vast quantity of culturally valuable audio and video materials being held, the finite resources for undertaking the transfer, and the impact of format obsolescence and carrier decay make it critical that this preservation work is carried out once and in such a way that archivists can navigate the inevitable technological changes that will require them to migrate their data from system to system. The long-lived file, migrated from platform to platform, possibly from format to format, will see this invaluable data being used by future researchers in unimaginable ways.

REFERENCES AND NOTES

1 INA Official Website, www.ina.fr.
2 Archiving of Radio and Television Programmes in Ireland. Discussion Paper, November 2005.
3 Spence, J. (2002) *Survey: Dams & Digitization Preparedness*, IASA Conference, Aarhus, CA: ABC Sound Archives.
4 European Broadcasting Union. (2003) Archives in Digital Broadcasting. EBU Archive Report, September 2003.
5 Quinn, E. (2007) *SpeechBot, the First Internet Site for Content-Based Indexing of Streaming Spoken Audio*. Heidelberg: Compaq Computer Corporation.
6 Fleischhauer, C. (2003) *Audio and Video Preservation Reformatting: A Library of Congress Perspective*, 18th Annual Preservation Conference; March 2003, Tornoto, Canada: Library of Congress.
7 Morris, R. J. T. and Truskowski, B. J. (2003) The evolution of storage systems. *IBM Systems Journal* 42(2): 205–217.
8 Stefano, G. *et al*. (2001) Archive Preservation and Exploitation Requirements. Preservation Technologies for European Broadcast Archives, PRESTO IST-1999–2001, 3 June 2001.
9 Dublin Core Metadata Element Set, Version 1.1: Reference Description, http://dublincore.org/documents/dces/.
10 Dublin Core Qualifiers, http://dublincore.org/documents/dcmes-qualifiers/.
11 Wright, R. (2005) Broadcast Archives: Preserving the Future. BBC Information and Archives.
12 Lee, A. and Wright, R. (1997) Specification of the Broadcast Wave Format. EBU document T-3285, July, http://www.ebu.ch/pmc_bwf_ug.html.