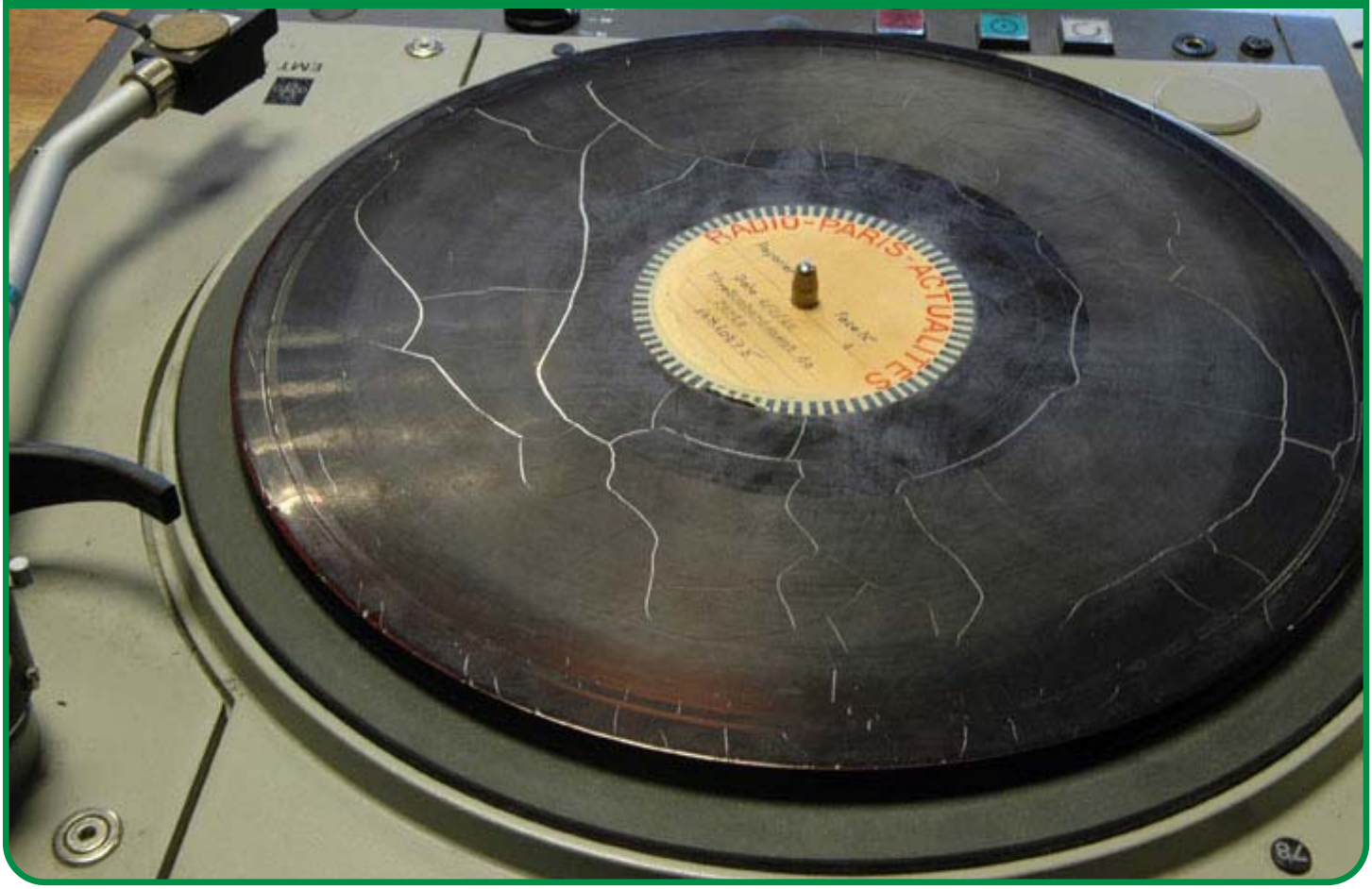
KEEPING AUDIOVISUAL CONTENT ALIVE

PRESTO CENTRE

Tutorial: Why Digitise AV material?



Version 1.01

Published by PrestoCentre, http://prestocentre.eu

Introduction

The time, effort and cost involved in digitising an audiovisual archive is a major investment. It's quite reasonable to ask for strong and compelling reasons to undertake this transformation. In this tutorial we explore the reasons for making this move from discrete and/or analogue storage, into a digital mass storage system.

We break the reasons down into sections: Preservation, Restoration, Metadata Access and Delivery, and Storage and Archive Management.

1. Preservation

The long-term preservation of media assets is a primary objective of most audiovisual archives. When an archive has a collection on discrete media, in either a digital or analogue format, this preservation work can involve a continual process of migration of content from old media to new.

Media formats become obsolete. This obsolescence puts pressure upon the archive to migrate content, often long before the media itself has worn out, so that the assets remain accessible. The obsolescence of media also makes it increasingly difficult and expensive to support the machinery needed to access the content, and with the need to migrate, this becomes more crucial over time.

Once media content is migrated from discrete individual specialised media storage formats onto a mass storage platform, the issue of maintaining equipment needed to 'play the media' is mitigated to an enormous extent, due to three main factors:

1. In moving to mass storage the link between physical format and data format is broken. This means that the process of transferring data between physical formats is far less complex, but still there is the risk of technologies approaching obsolescence.
2. The available physical storage media are no longer limited to specialised broadcast and professional audiovisual formats. Archives can now utilise the technologies available in the larger and more competitive IT storage market, reducing costs and offering greater flexibility.
3. Digital transcription between successive generations of media can, theoretically at least, be lossless. Unlike traditional analogue preservation processes, or even digital processes that involve decoding, decompression, recompression and re-encoding, the transfer of digital data files between media need not involve any loss of information. Quality of the content can be maintained indefinitely.

In short, preservation in a digitised mass storage archive is simpler, cheaper and more reliable, or can be if done well.

2. Restoration

Restoration of media assets is a major component of the work of many audiovisual archives. While assets remain on discrete analogue or digital media, the process of restoration is invariably focused upon the restoration of the media itself, with the assumption that this will lead to a greater retrieval of the information held on that media.

In the realm of film restoration, the restoration of old film stock can have a dramatic effect on the quality of the content, but it is an expensive and highly specialised task. With limited financial resources, it is an unfortunate fact of life that much material will be lost forever as media decays past the stage where it can be viably restored.

Digital restoration, however, is directed more at the content itself, the pattern of information contained within a sequence of images, or a sound wave. Software based restoration of images, both still and moving, can incorporate processes based on knowledge about both the media upon which

the information was stored, and also the scene photographed.

To give an example; if an entire frame of film is damaged, it is possible to double up a preceding or following frame without too much loss of quality. However, it is only digital process that can extrapolate multiple frames based upon information of the scene, and still retain an acceptable video quality.

The field of digital restoration is an advancing one, and is taking advantage of the rapid advances in computer hardware and digital signal processing software. These advances are making it possible to have restoration capabilities available on the desktop, at a relatively low cost, usable by personnel with less specialised technical skills.

Once content is digitised, restoration of its quality is a cheaper process, and one that can be carried out more quickly, and therefore upon more content. As technology advances the process can be repeated, to increase the final quality of the data, and successive generations of restoration can incrementally improve the content of the archive. Of course, to capitalise upon these digital restoration techniques, your initial digitisation must be of the highest calibre.

3. Metadata, Access and Delivery

Access to content is usually what justifies the existence of any archive - inaccessible content is not worth archiving. New innovative technologies and processes have been developed in these areas, capitalising upon the digitised status of archive content, and allowing quicker, easier and cheaper access to content, which in turn will be a powerful justification for your digitisation effort.

3.1 Metadata

The cataloguing of an archive is of critical importance. To accurately, completely and reliably annotate all the content of an audiovisual archive is an expensive process though. By using the digitised content as an information source, tools are now being developed that can greatly assist the process of cataloguing.

At the core of the new processes are speech-to-text and semantic analysis tools that can automatically generate appropriate searchable terms from a specified ontology, taxonomy, or keyword list. Key frames can be identified in the video stream by shot analysis, greatly assisting the segmentation of footage into catalogue sections.

> The obsolescence of media makes it increasingly difficult and expensive to support the machinery needed to access the content, and with the need to migrate, this becomes more crucial over time.

Once digitised, audiovisual assets can incorporate the metadata generated by utilising the MXF file format, and similar wrapper based encoding techniques.

3.2 Access

Finding relevant items in catalogues today, relies, in the best cases, on searching through a database, but such data sources differ widely in the standards they use to describe the content. This in turn makes searching across multiple archives a lengthy and difficult business. The use of standardised metadata, as researched in the Presto-Space project, enables federated searching across multiple archives.

The use of digitised content allows the possibility of searching for items by querying the content directly. While developments in this area are at

the very early stages, developments can already be seen in the use of rich metadata - key frames and textual content - as a searchable resource. Digitised assets, searchable as files, are therefore far more accessible to users of archives, and are set to become even more accessible over time.

3.3 Delivery

The distribution and delivery of audiovisual assets is an issue that taxes many archives today. As the assets exist as discrete high quality items on shelves, there are great costs involved in sending such content to a user: logistical stock management, copying, transcribing formats and loan management all tax the limited resources of an archive, and can set a limit on the amount of material available for loan.

Every stage of the distribution process is aided by conversion to digital mass storage:

- Availability: All items are always available, as only copies are delivered. The case of an item not being available, due to a previous or simultaneous loan, will not occur.
- Latency: The time needed to find and deliver the item is reduced dramatically, as there are no shelves to check, no copies to make and no conversions to carry out by playback on physical machinery.
- Compatibility: even in instances where the user requires the final content in a specific physical format, eliminating the need to ingest material for conversion saves time. Digitally stored files can more easily be converted than almost any discrete help asset. For an increasing majority of cases, copying onto a physical file format will not be required and file format conversion will be the most that is needed.
- Throughput: the transfer of content over IT networks eliminates the physical distribution of tapes, discs and other discrete media. Network speeds are increasing rapidly (in the last ten years, network technology developed from 10 megabits per second to a speed of 10 gigabits per second for the same costs), and in many situations the deliverable data rate over the network can only be exceeded by dedicated high-cost physical logistics operations. Even those few operations that can afford such support will soon see the performance of networks outstrip the physical distribution of material.

4. Storage and Archive Management

The storage of archive materials on discrete media is a task of forbidding complexity and expense. There are no easy, cheap options for the long-term care and storage of individual tapes, cassettes, or discs, and when volume is increased the stock management and housekeeping associated with large amounts of assets are considerable.
Moving assets into mass storage brings key advantages in the storage are because of a number of factors:

4.1 Density

When assets are held as files, there doesn't need to be a one to one mapping between assets and media items. Many files can live on a single disk, or striped and spread between multiple disks in RAID and similar technologies. Less space used results in lower costs. Further, direct manual access to the physical storage media is not required, obviating the need for extensive walkways and other areas for personnel. Even storage systems based on removable media can use robotics instead of manual logistics to reduce the space required by large amounts.

While discrete media will always require separate physical items for each stored asset, a minimum size for an archive holding a set number of items remains fixed. With a mass storage system, technology advances in storage density can deliver

continual reductions in the space requirements for your archive.

When we look at the physical storage needed for 700.000 hours of material stored on discrete media and we compare this with the amount of physical storage when all this material is ingested into the mass storage system we only need 0,1% of the physical space to store this content. This percentage drops every 8 years when migrating to more dense storage media (with approximately 75%)!

4.2 Media Cost

By moving to commoditised, IT industry storage equipment in a mass storage solution, archives can be released from the high costs associated with specialised audiovisual media. The IT storage market is large, fiercely competitive, and drives some of the world's most aggressive research and development. To be able to capitalise on the most cost effective products of this market, archives have to switch to mass storage technologies.

4.3 Disaster Recovery

The low cost, and ubiquity of mass storage facilities, and the network technologies that can connect these facilities, at last open the option to archives of maintaining full remote duplicates of their collections. Depending upon the value of the assets stored, the costs of duplication and remote storage may well be justifiable within budgets previously directed to the maintenance of large single discrete storage facilities. With Outsourced off site specialist Storage Service Providers, such recovery abilities should at last become affordable to smaller archives.

4.4 Stock Management

In an archive based upon discrete assets, the removal of items from storage is an unavoidable part of the process associated with access- items are removed from storage to be loaned out, to be dubbed onto a browse quality review copy, or to be duplicated or migrated. Items stored as files in a mass storage system need never leave the archive. Digital copying allows faultless duplication to any standard required, whilst the item remains within the storage system at all times. Stock management is a meaningless concept in such a system.

While discrete media will always require separate physical items for each stored asset, a minimum size for an archive holding a set number of items remains

4.5 Varying Value

You may call them the vaults, the back rooms or the dusty cupboards, but most archives have an area, a section of their collection that is less often accessed, less requested, and arguably less valuable than other areas. We hold on to these collections for a variety of reasons; a speculative investment in the future, a concern that these may be the last examples of their kind, or sometimes simply because we can. You may not wish to invest the same amount in the storage of the 'back room' collection as they might in their main, heavily used material. After all, that investment will be aimed at offering highly resilient, speedy access and delivery- not necessarily factors that matter much in this case. In fact lowest possible cost is almost certainly the main factor here.

Mass storage systems need not be utterly homogeneous though. It is quite possible, and in fact quite common, to have multiple tiers of storage in a Hierarchical Storage System.
At the top of such a hierarchy is the very fast reaction memory, a cache of frequently, recently requested items. This may be solid state memory,

relatively expensive, but very quick to deliver content.

The next layer down might be fast, high quality hard disk drives. Again, quite quick to deliver files, but the costs will be lower. As we move down the stack, progressively cheaper, slower storage mediums can be employed, to the point where, at the bottom of the stack, removable optical or tape based storage is viable, due to the acceptability of slow access times.

Within a properly constructed mass storage system, all these layers of storage can be available to store appropriate content upon, with a minimal impact to the user. Advanced file systems can even spread a single content item across multiple layers of the hierarchy, so that the most immediately required section is available to access whilst the rest loads up. Depending on your archives content, and the use to which your customers put that content, the storage system can be reconfigured to give the most cost effective use of storage possible.

Conclusion

The choices facing archives, libraries, and museums are now in some sense more complex (because there are more options), and in others more straightforward than when the Presto projects launched.

Digitisation of film, videotape, and audiotape is no longer experimental. It has evolved into established practice, both inside audiovisual archives, and at a number of service providers. Costs for digitisation have dropped significantly, as have costs for storage, and millions of hours of video material have become available online. Improvements in automated metadata extraction continue, and have been supplemented with crowd sourcing.

Digitisation may seem a daunting task, but the need for digital collections is growing every year. Starting to plan for digitisation allows planning for new options, audiences and uses of your collection as well. Using the advantages of digitisation in your plan will help you see the light at the end of the tunnel!


KEEPING AUDIOVISUAL CONTENT ALIVE
PRESTO
CENTRE


SEVENTH FRAMEWORK
PROGRAMME