Digital preservation of audio, video and film

Richard Wright

The author

Richard Wright is Technology Manager at the BBC Information and Archives, Brentford, UK. E-mail: richard.wright@bbc.co

Keywords

Digital storage, Audiovisual media

Abstract

Digitisation is used for preservation of audiovisual material. This preservation work is a major producer of digital collections – which then need digital preservation for sustainability. EC Project Presto surveyed the holdings and status of ten major broadcast archives – a significant portion of total European broadcast archives, including some of the largest individual collections. The main findings are that approximately 75 per cent of this material is at risk or inaccessible and that the collections are growing at roughly four times the rate of current progress in preservation work. This paper gives further results of the project, and gives practical guidance for preservation of audiovisual material. Presto demonstrated the effectiveness of the "preservation factory" concept for major broadcast archives – a way to reduce cost while still maintaining or even increasing quality. There is now a new EC project, Presto-space, which will make the preservation factory available to small and medium-sized collections.

Electronic access

The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0305-5728.htm

VINE: The Journal of Information and Knowledge Management Systems
Volume 34 · Number 2 · 2004 · pp. 71-76
 · ISSN 0305-5728
DOI 10.1108/03055720410550869

Overview

The bulk of audiovisual material is at present analogue, not digital. However this material in the main requires digitisation for preservation and for access. Audiovisual preservation projects are creating large digital collections, which then share with all other digital collections the problem of sustainability: maintenance and regeneration of the collections. Major broadcasters across Europe are spending roughly €100 million per year on digitising their archives. The purpose of this paper is to give information on audiovisual digitisation, so that other digital preservation projects can learn from that experience.

There are three categories of media that will be discussed: audio, video and film. Photographs are also audiovisual – but will not be covered as there is considerable information already available on that subject (Klijn and de Lusenet, 2000).

The overall conclusion is in three parts:

(1) There is a consensus on the preservation strategy for audio: digital files on mass storage media (of various sorts).
(2) There are two main schools of thought regarding video:
   - transfer to a current videotape format; and
   - transfer to digital files on mass storage media (of various sorts).
(3) Within broadcasting (film in TV is a special case) there is agreement on the long-term solution for film: high-resolution digitisation and mass storage. There is, however, at present no practical path from the current situation to that solution. Moreover, for film archives in general, digitisation has not been accepted even in principle.

Much more detail about current preservation practice is available in a public document prepared by an EC-supported project run by the BBC: Presto Deliverable D3.2: existing and emerging technology (Presto, 2001).

Problems with audiovisual media

The twentieth century was the first century with a record of its significant events – the sounds and moving images – on film, audio and video media, now held in broadcast archives. This record is at risk, as the recordings are deteriorating or on obsolete formats. Broadcasting did not develop as a "heritage institution". As a consequence, broadcast archives have no business or funding models designed to support preservation – and sustainable archives.

VINE: The Journal of Information and Knowledge Management Systems
Volume 34 · Number 2 · 2004 · 71-76

Presto project

Broadcast archives and technologists joined forces to better understand the issues and problems of audiovisual preservation, and to develop solutions and joint standards. This joint effort has been made possible by funding from the European Commission fifth framework programme in information society technologies: cultural heritage applications. Presto was a 24-month, €4.8 million project to develop broadcast archive preservation technology, which ended in August 2002.

Preservation is also transformation: to digital media of some sort, opening new possibilities for storage and access. The project also considered future methods of archive usage (and revenue generation) in order to achieve the true "best cost" when considering the possibilities, over the next 20 years or more, for usage of the preserved content.

The project was led by the BBC, and full partners are two of the other largest European national archives:

(1) INA = Institut National de l'Audiovisuel.
(2) RAI = Radiotelevisione Italiana.

There were also seven technical partners, and a user group of seven more (beyond BBC, INA and RAI) national broadcast archives.

Presto survey

Briefly, Presto data leads us to estimate that there is at least 100 million hours of audiovisual material, world-wide. For instance, the Swedish national archive alone has six million hours. About 70 per cent of all this material is in need of preservation work, for reasons of obsolescence (or the media or the players of the media), deterioration, or fragility of the media. The bulk of the problem is with analogue material, which is the most difficult to preserve because the transfer of analogue material is harder to automate than is the case with digital material (which can, in principle, be "cloned" by some sort of robotic machine).

Solutions

How to organise audiovisual preservation

For audio and for videotape, the cost of preservation work can vary enormously. The primary factors controlling cost are:

- Accessing the material: it costs a lot to search, fetch, and transport the old material, and then shelve the new material. If material is being accessed anyway (for issuing or research), then preservation work can be done at the same time (preservation on demand).
- Throughput: a conventional transfer consists of one person playing an item from an old format onto a new format, and then checking the result. This takes about three hours per hour of material, and is the least effective way to transfer a lot of material. A dedicated facility operated as a transfer factory, with each operator running not one but four or five simultaneous transfers, and with maximum automation of checking and labelling (the basic medata) – can save about two-thirds of the cost as compared to a conventional transfer.

The above two points are in serious conflict. Preservation on demand usually needs to be tied to a conventional, one-at-a-time transfer. A "factory" approach is required to copy a whole collection in a reasonably short time with maximum cost-effectiveness.

The most expensive way ("the obvious way") to do preservation is to access the material specifically for preservation work (not as in step 1 above, so no access savings), and then do the one-at-a-time transfers. Table I shows the Presto rough estimates of the costs of the three options.

The problem with Table I is that cost information is notoriously incomplete and imprecise. "Cost" ultimately is what a departmental or institutional budget has to pay – visible cost. If a department has the staff and the equipment and the time, the visible cost for the "obvious way" may be zero, rather than £150 per hour! In contrast, shipping it out the door to an efficient audio transfer factory would incur the very visible £50 per hour.

It is important to consider not just cost but time. If the material is already old and causing enough problems to motivate thinking about copying to a new format, then this copying may need to be done in months, not years. The total time as well as the total cost must be calculated for each of the options, and then the better choice is the cheapest one within the allowed time, not the cheapest overall.

Cost per use

The true cost of an asset is total lifecycle cost. The true benefit is related to the number of times that asset is used over the lifecycle. Although not every use has equal benefit, overall more collection access means more benefit. Therefore a simple way

Table I The Presto rough estimates of the costs of the three options

Preservation project method	Rough cost per hour for $\frac{1}{4}$" audiotape, £
"The obvious way"	150
On demand, when the material is coming off the shelf anyway	100
Mass transfers, factory method	50

VINE: The Journal of Information and Knowledge Management Systems
Volume 34 · Number 2 · 2004 · 71-76

to combine transfer cost, life cycle cost, and the significance of new service opportunities, is to translate those new opportunities into a predicted rate of item usage. Options for preservation can then be compared, in monetary terms, on a "cost per use" basis. A significant conclusion of the Presto survey is that archive preservation strategy should aim at the "lowest cost per use" over the life cycle of the new media, not at the lowest transfer cost.

Solutions for audio

There are many forms of audio media, with the following main formats:

(1) Analogue formats:
- wax cylinders;
- shellac and vinyl discs;
- 0.25 inch open-reel audio tape; and
- cassette tape.

(2) Digital formats:
- Digital Audio Tape tape (DAT – the first widely available digital format for audio, now obsolete);
- CD (audio CD; the CDs that work in a walkman);
- Mini-disc; and
- "computer media": floppy discs, CD-ROM, ZIP-drives, removeable hard drives, datatape of various sorts, e.g. Exabyte.

For analogue media, there is good reason to transfer both disc and tape formats to a modern format. Although shellac and vinyl have a very good shelf life (maybe centuries if not touched), shellac (78 rpm recordings) is brittle and easily damaged if handled, and both shellac and vinyl are virtually certain to be damaged if played. So for reasons of access, such material should be copied.

For tape (0.25 inch and cassette and DAT), there is a much shorter shelf life, perhaps as low as 20 years or even less if stored at high temperature and humidity. For all these media, including DAT, there is the additional problem that their transport – the equipment used to play the material – is obsolescent.

It is important for any tape-based medium to be stored at low temperature and humidity, if possible. The Audio Engineering Society (AES22, 1997) recommends storage at normal room temperature only if there is a low relative humidity (RH below 30 per cent). For every 10 percentage points increase in the relative humidity, the temperature should be reduced by 5°C. Storage at 15°C rather than at normal office temperature of 22°C should add about 50 per cent to the life expectancy of magnetic tape.

What to copy on to, for audio

The answer will vary according to the size and the purpose of the collection:

- Broadcast archives: across the archives included in the Presto survey, those reporting at the Presto workshop, and those reporting through the professional association International Association of Sound Archives (IASA), there is a consensus that the time has come for broadcast audio archives to digitise, to make electronic files, and to hold these files on mass storage for electronic distribution. This is a standard "digital library" approach, but the costs and infrastructure are now adequate to make this approach practical and cost-effective for audio collections. Broadcast archives are making.WAV files in a particular format: EBU Broadcast WAV format = BWF (EBU BWF, 2001). These files are held on various media: CD-ROM, DVD-ROM and datatape being the main choices.
- For smaller collections, or collections outside broadcasting, there may be fewer incentives to make electronic files. However, if part of the justification of a preservation process is to increase access, then some form of file format is definitely recommended. For Internet access, the most common formats are REAL, MP3, QuickTime or Windows Media. Of these, MP3 is in principle an "open" format because MP3 stands for MPEG 1 layer 3, and MPEG is an International Standards Organisation (ISO) format and therefore by definition publicly available, well-defined and in theory well-supported; such is not always the case with proprietary formats.

Writing to CD and DVD media

Best life expectancy is obtained by observing basic precautions. For a few hundred items, use of a professional service may be the most effective way to get a long-lasting result. For a large volume, the do-it-yourself alternative to a professional service needs special error-checking equipment (or professional advice) to insure that the blank media make a good match for the model of CD or DVD writer. Thorough testing is also required:

- for every new batch of blank media;
- for every CD/DVD as soon as it has been written, for proper read back; and
- selective media testing every six months, to monitor for degradation.

If written improperly, life expectancy of CDs can be measured in days. If done properly, a 20-year shelf life can be achieved. CDs are sensitive to mechanical damage (scratching), ultraviolet light and chemicals, and need proper handling and storage.

VINE: The Journal of Information and Knowledge Management Systems
Volume 34 · Number 2 · 2004 · 71-76

Solutions for video

There are even more formats for video than for audio. Again, this paper will concentrate on preservation of analogue formats. Within this category, there are two sorts of tape: the original open-reel formats (2 inch and 1 inch), and the later cassette formats (U-Matic, BetaSP). In addition, there are open-reel domestic formats, and the Philips domestic cassette format from the 1970s, and others.

Many of the logistical issues have already been discussed. If there is a large amount to be copied in a short time, it needs a factory approach or the costs will be tripled. However, the handling of cassettes is easier than for open reels, and for a very large project the cassette handling can be done in a robot – which would allow future transfers from old to new media to be fully automated. There is also a large difference between analogue and digital video: once material is digital, on any format, it will be far cheaper to copy to some future format. The big expense is getting analogue video material into a digital form.

Preservation strategy for video

The strategy here is mixed. Some major archives are committed to mass-storage with data files; others copy to current videotape formats (e.g. digibeta, DV-CAM) in the full knowledge that in the relatively near future these same formats will need to be re-copied. Even those archives committed to mass storage continue to also make current-format videotape, as backup or because not everyone has access to the digital files on mass storage, or because (for reasons of economy or for ease of delivery over existing low-bandwidth data networks) the digital files are not full quality.

The full automation of videotape preservation, following the factory model now accepted as standard for audio, has not been fully developed and adopted for video. The Presto project created a high-efficiency reference preservation process (a state-of-the-art process for transferring videotape, with all the efficiency and automation features currently available, including the new technology developed by Presto) for U-Matic material.

Outside of broadcasting, and for small collections, it will hardly be sensible to copy to datatape because then the only way the material can be accessed is via a dataplayer and some sort of computer system. Videotape can be directly played from a videotape player – and the cheapest videotape option will be the "semi-professional" DV (DVC, DVC-PRO, DVC-PRO 50, DVCAM) family of digital videotape. However as with audio every preservation project should consider making a compressed version for Web access. MPEG-1 is a rather high data rate (1 Mbits/second), and REAL, QuickTime and Windows Media encoding methods are the ones most commonly used. MPEG-4 is a non-proprietary alternative.

As well as outsourcing tape-copying aspects of preservation, it is possible to outsource the Web encoding, and the Web hosting.

The option of copying video to DVD should be seriously investigated. It will not have the same top quality as DV, but it may well:

- last longer;
- be easier to play, copy and distribute; and
- have a lower media cost.

DVDs can be written in MPEG II at a data rate (and corresponding quality) of 4 to 8 Mb/s. These DVDs will then play on conventional DVD players, and on computers with DVD drives. A data rate of 4 Mb/s is what digital TV is currently broadcasting, and commercial cinema DVDs are in this same 4-8 Mb/s range. Whether the quality of digital TV or commercial DVDs is adequate is a decision that requires analysis and advice.

For large projects that will take the leap to data tape, there is a choice: linear or helical. Linear uses low-demand technology, and writes the data in straight lines on the tape (hence the name), using fixed heads. Helical uses the technology invented for videotape: a moving head that writes tracks on a slant across the tape, allowing higher speed and higher density – but it is a much more technically demanding and costly technology. Writing to linear data tape is safer and cheaper than writing to videotape, because linear writing/reading is much less "near the edge" in technical performance than is the case for the helical writing/reading used for all videotape formats.

Solutions for film

Sixteen millimetre film was an audiovisual format 100 years ago, and will probably be a valid format 100 years from now. There is nothing comparable to film in terms of format stability. Modern advances mean that film made in the last 30 years, if properly stored, can be expected to last a century or more. Also film is very expensive to process: it costs about 20 times as much to make a new film master as to make a new video master, and about 10 times as much to make a new working copy (circulation copy or print). All of these factors work against doing anything with film beyond trying to hold it in the proper environmental conditions, for as long as possible.

However sometimes the sit-on-it approach simply is not adequate. Acetate-based film can turn to acetic acid (vinegar syndrome) and once the reaction starts, something must be done. There is a considerable literature on vinegar syndrome, so this paper will only discuss what to do about transferring material – whether forced by vinegar syndrome or by some other reason (such as

VINE: The Journal of Information and Knowledge Management Systems
Volume 34 · Number 2 · 2004 · 71-76

deterioration due to fading, age or inadequate storage conditions).

Vinegar syndrome attacks magnetic film sound tracks faster than ordinary film. Professional film archives will have this material, but small collections will probably only have "sound on film" material and so should have less of a problem with vinegar syndrome. Magnetic sound tracks can quite cheaply be transferred to CD – but that leaves the problem of how to use the CD when viewing the film. This is a specialist issue and people in the UK can refer to the British Film Institute for advice.

Preservation strategy for film

The actual current preservation strategy for film is simple and clear: sit on it. Archives are concentrating on storage conditions, not on digitisation. However, as covered in Presto Deliverable D5.2 Sustainability of film collections, there is rapid development in film technology coming from two directions:

(1) Increased resolution for video: high definition TV, already a reality in Japan, is the next technical step for the USA. This development will promote the HD format in TV production, and so the "film workhorse", the telecine, will inevitably be upgraded from 625 lines to over 1,000 lines; the only question is: when?

(2) Development of "digital cinema": the pressure is from the high cost of mass distribution of cinema productions on films – a major release requires 5,000 prints at a cost of $2,000 each. This fact is pushing the development of digital technology in the 2,000 to 4,000 lines area (the production standards now used in high-end computer graphics for cinema special effects).

For film material to be used in TV, it has to be converted from film to video: a telecine conversion. At present, the telecine process is medium rather than high resolution (roughly 600 lines per image, equalling standard definition TV) and so the film must be kept because the videotape is poorer image quality. As soon as higher-resolution scanning becomes as cheap as current telecine, more and more of a film archive can be scanned once-and-for-all. Archives are waiting for 1,000, and then 2,000 and finally 4,000 scanning to be as cheap as current telecine processes.

Presto developed an automatic film splicer, to replace manual labour in re-joining archive film with many tape splices. Such film is common in TV archives, where individual news stories were joined to make a "day reel" for each day's news. This device, and high-resolution scanning equipment, has the same price justification issues as for high-efficiency audio and film preservation equipment: it is expensive, it has to be kept in use or it is not cost effective. It can only be used on big jobs with a guarantee of sufficient material to warrant the investment.

Hence for film as well as for audio and video, the progress in efficient and cost-effective preservation is the province of the big archives only.

For small collections, and especially for collections of material that is not of top quality (not 35 mm, not professionally produced) one strategy is the same as for similar low-to-medium quality video: transfer to DVD, using MPEG II at 4 to 8 Mbits/second. This option will give "TV quality", and it is up to the archive to make a proper judgement about whether such DVD quality will suffice.

Problems with the solutions

The main problem with the Presto preservation factory approach is that it takes a considerable investment to set up and run such a factory. Only institutions with large volumes of audiovisual media could adopt the approach.

One could suggest that the large institutions open their factories to use by smaller institutions. The problem here is that all holders of audiovisual material are in a race against time for preserving their own holdings, and in general do not have spare capacity.

The preservation factory approach is the Presto answer to preservation for broadcast archives, and is suitable for:

- large collections;
- that earn money; and
- have sophisticated technical support.

What is needed is an audiovisual preservation solution for everybody else.

Preservation is a major issue, but cannot be viewed in isolation. The institutions that hold this endangered material perform services, and broadcast archives serve a highly technical and rapidly changing industry. Preservation strategy needs to consider – to foresee if possible – the future service requirement of multimedia collections for at least the next 20 years. These service requirements will increasingly be based on electronic mass storage and direct, networked end-user access – probably using Web technology.

A proposed solution is being developed by a new EC project: Presto-Space. The solution is to develop the preservation factory approach as commercial services, and to develop new archive access as the main method for acquiring funding.

VINE: The Journal of Information and Knowledge Management Systems
Volume 34 · Number 2 · 2004 · 71-76

Presto-Space: access funds preservation

Major audiovisual cultural collections in Europe have joined forces to propose Presto-Space, a sixth framework integrated project to provide practical methods for digital preservation of all types of audio-visual collections. The aim is to build-up preservation factories providing affordable services to all kinds of collection owners, including not just preservation but also the ability to manage and distribute the new digital assets.

The way to achieve the goal of "preservation for all collections" is with an integrated approach, to produce sustainable assets with easy access for much wider exploitation and distribution. The key idea is: an accessible item is more valuable than an item stuck on a shelf. An integrated process provides this access, generating revenues that will fund the activity and developing resources to finance collection maintenance. Presto-space will produce not just new media, but new business models involving electronic access to the new media. The preservation factories will deliver the full package: metadata, media, storage, Web site, asset management. The prospect of much better access is the mechanism for acquiring the needed funding: access funds preservation.

Conclusions

Broadcast archives are in the early stages of the biggest and most expensive media conversion they will ever face. The whole process of selection and digitisation of analogue media will take at least another 20 years. Without widespread funding and support, and without cost-effective and farsighted use of technology, the work will not keep pace with the deterioration of the material. EC project Presto has documented the problem, provided guidance for organising preservation transfer projects, and delivered multiple forms of new technology for reducing preservation project costs, and increasing their efficiency. EC project Presto-space (starting January 2004; http://prestospace.ina.fr) will develop commercial services to provide both the technology and the revenue for efficient preservation of all of Europe's audiovisual heritage, with special emphasis on film.

References

AES22 (1997), "Recommended practice for audio preservation and restoration – storage and handling – storage of polyester-base magnetic tape", available at: www.aes.org/standards/b_pub/aes-standards-in-print.cfm

EBU BWF (2001), "Broadcast WAV format", available at: www.ebu.ch/departments/technical/pmc/pmc_bwf.html

Klijn, E. and de Lusenet, Y. (2000), *In the Picture – Preservation and Digitisation of European Photographic Collections*, ISBN 90-6984-294-7, European Commission on Preservation and Access, Amsterdam, available at: www.knaw.nl/ecpa

Presto (2001), "Presto Deliverable D3.2: existing and emerging technology", available at: http://presto.joanneum.ac.at