PrestoPrime: Digital Preservation of Audiovisual Materials

Richard Wright, Preservation Guide, London UK preservation.guide@gmail.com

1. Introduction

Audiovisual archives are young institutions with young content: no major collection predates the twentieth century. These collections hold the sound, image and moving image recordings from a technology less than 200 years old: photography from the 1830s, sound recordings from the 1880s and motion picture film from the 1890s. The twentieth century brought audio and videotape formats, and then digital media: CD, DVD, DAT tape and minidisc. The proliferation of formats, the short life expectancy of the materials and the even shorter time before obsolescence (and the consequent need for continuous cycles of replacement) are major problems in preservation of audio and video media. The complexity of sound and moving image media cannot be ignored and should not be oversimplified: audiovisual content requires specialist knowledge.

2. Preservation of Sound and Moving Images

Preservation begins with asking what should be preserved, and why. The significance of sound and moving image collections will be covered, including the dramatic change to access (and therefore to the social significance of such collections) which comes with digitisation and Internet technology. The preservation status will be described, although in summary it can be said that the status is basically an emergency: virtually everything on shelves will have to be digitised to survive, that work still has a long way to go – and the prospect is that much analogue content will be lost. Finally we will cover the technology for preservation.

2.1. Importance of the collections

Europe has a unique resource: the archives of public service broadcasting. The general commercial significance of 'media industries' is presented, along with their larger public value. Broadcasting is radically increasing its use of Internet and mobile technology, leading to a great increase in archive access and a redefinition of broadcasters as publishers – with all their content potentially available on demand. Europe is unique in having public service collections built up over 50 years or more – content paid for by licence fees, and archives which are the largest collections of media anywhere (a dozen archives with over one million hours of content each). These collections have good documentation (generally) but their commercial and public value will only be released through a concerted programme of digitisation, rights management, metadata normalisation and conversion to multilingual access.

Broadcasting is just one source of audiovisual collections. Most European countries have a national institution devoted to film (cinema), and many have national sound archives with substantial amounts of non-broadcast content. Finally there is a vast array of small institutions and small collections holding archived sound and moving images. In the following sections Europe's audiovisual heritage is reviewed in terms of size, condition and public value. In small collections the documentation varies enormously,

and preservation status is generally poor. We present the case for investment in this content as having low marginal cost and high potential return. The cost is low in comparison to the investment which broadcast archives have already made, and the potential return is high in terms of increase in 'width and depth' of European content, and in terms of preservation of unique heritage in a cost-effective fashion.

2.2. Status of the content

A major survey of European sound and moving image collections was performed by the TAPE (Training for Audiovisual Preservation in Europe) project. The full survey[1] has a wealth of detail, but the summary agrees with smaller studies from the Presto and PrestoSpace projects: the majority of the analogue content is at risk, and something like 70% has an immediate need of some sort of preservation action. In volume, TAPE found 20 million hours of content, and UNESCO has extrapolated this to 200 million hours worldwide[2].

As an example, TAPE found over four million hours of audio content (excluding four broadcast archives, which have another five million hours!) which should be a good sample of the range of European institutions. The material was on the following formats:

Format	Percentage of Total
Wax Cylinders	0.02
Coarse groove replicated disks ('78s','shellacs')	1.00
Instantaneous disks of any kind	0.02
Microgroove disks (LPs)	10.45
Open reel magnetic tape	58.80
Compact cassettes	12.65
R-DAT	2.32
Replicated CDs, DVDs	9.86
Recordable and rewritable CDs, DVDs	4.23
MiniDiscs	0.31
Other	0.13

Two major issues are immediately clear:

1. Most of the formats are obsolete, from wax cylinders to Minidiscs. Only the CD and DVD formats can be considered current, and both are officially described by the International Association of Sound and Audiovisual Archives as too easily damaged to be suitable for preservation.
2. There is *no* mention of files on mass storage. All the content except for the 0.13% of 'other' is on physical carriers that sit on shelves.

Regarding file-based content, the TAPE survey took place during 2005-2007 when file-based storage was just beginning to make an impact. Even then, audio files were commonly written to DVD (certainly in the BBC) and so would have been included in the above list as 'rewritable DVDs', not as 'files on mass storage'. Subsequent work by PrestoSpace and PrestoPRIME has provided updated statistics on file-based content. PrestoPRIME (Wright,[3] 2010, pp.7–8) estimated (from the TAPE data) that a decade of digitization of analogue holdings (at 1.5 per cent per year, or 0.28 million hours per year) and new digital

intake (at 6 per cent per year for broadcast archives) would mean that in 2016 there would be 8 million hours of new digital content in addition to the 16 million analogue and 3 million digital found by the TAPE study in 2006. The analogue holdings (if removed from archives after digitization) would have dropped to 13.3 million hours, and the digital would have risen to 11 million hours. Unless the rate of digitization increases, at 0.28 million hours per year the analogue holdings will take a further 48 years to digitize. If only half the analogue content is selected for digitization, there would be about 25 years of digitization work remaining after 2016, and it is by no means certain that equipment, spares and operators (or budget) will be available for another three decades.

2.3. Methods of preservation

Within preservation there are two broad categories: conservation and preservation actions. Preservation actions are interventions, and conservation is what is going on all the rest of the time. Digitization is just one type of preservation action. A file format migration (from an obsolete file format to a current format) would be another preservation action.

Conservation addresses prevention or delay of deterioration and damage. Conservation does not prevent obsolescence, which for audio and video is the main driver for migration from old carriers (formats) to new ones. A preservation action is an intervention taken when conservation is not enough. The typical case is making a transfer (migration) to produce a 'new master'.

2.3.1. Conservation

Conservation of analogue content can be divided into the following concerns.

- packaging, handling and shelving (storing): the immediate environment of a physical item;
- environmental conditions: essentially control of temperature and humidity, and the stability of that control, but also protection from anything harmful in the environment, such as dust, pollutants, magnetic fields, excess light or infrared and ultraviolet radiation;
- protecting the masters; and
- condition monitoring.

Physical conservation is described in detail in the PrestoSpace wiki[4].

2.3.2. Migration

Keeping the originals is a basic principle of archiving but does not solve the main preservation problems of obsolescence, decay and damage. In consequence, archives periodically make new masters, on the same technology (dubbing an old audio tape to a new one; making a print of a film) or on new technology (making a CD from an audio tape, digitizing a film).

For three decades for audio, and for at least two decades for video, archives have been digitizing their content. The technical problems are not the digitization itself. The problem areas are:

- successful playback of the originals, in order to get an optimal signal to digitize;
- standards: what compression level, encoding method and file format to use; and
- efficiency: digitizing the existing analogue materials fast enough and economically enough to cope with the size and urgency of the problem.

Playback of audiovisual carriers: there are two main problems for playback of all analogue media: 1) lack of equipment, spares, calibration materials and experienced operators; and 2) condition of the originals, which may be fragile or already damaged or decayed.

There is a particular problem for tape, common to audio and videotape (and also common to both analogue and digital tape, including data tape): sticky shed. This is the shedding of the oxide coating of the tape during playback, leading to clogging of the playback head and resultant loss of signal. Initially the signal loss is usually brief and intermittent, but eventually the playback machine can jam completely.

Efficient digitisation: Large amounts of analogue content are being digitized, with much more remaining. This digitization effort has focused on an efficient process, the *preservation factory* approach developed by the Presto project and promoted by PrestoSpace.

The basic requirements for efficient digitization are:

- a collection level approach: making a budget for efficient digitization of a collection, rather than just looking at the time and cost of digitizing individual items;
- enough material to set up an industrialized workflow (typically at least 1,000 items in a single format or category according to the IASA guideline[5] (2009, Ch 9)); and
- division of labour: with enough content, the work can be divided into tasks (equipment operation, metadata, logistics of moving and labelling material) with consequent increase in efficiency.

In general, a small collection cannot set up a preservation factory, but larger institutions and private companies have set up highly efficient workflows. Industry has also developed high efficiency workstations: NOA and Cube-Tec Quadriga workstations run four simultaneous audio digitisations; Front Porch produce the SAMMA line of robotic systems for video digitisation.

There are automated methods for detecting audio and video problems. Automatic monitoring is built into the Quadriga, NOA and SAMMA systems (and other commercial products), but these methods produce extensive logs of issues and potential faults that have to be manually reviewed. Automatic detection of impairments has been developed for video and film restoration technology, but again the detection only flags potential problems that have to be manually reviewed. There is a need for effective integration of signal processing technology with human checking in order to produce a really efficient method of quality control within a preservation factory approach.

3. Digitisation and the Presto Series of Projects

3.1. The Preservation Factory - Presto

In 1999, the Presto project was formed by a group of broadcasters with a common problem. They had several million hours of content to digitise, and not enough time or money to digitise it all using existing technology. The result of the Presto project was the definition and widespread adoption of the

preservation factory approach: using division of labour to reduce costs by an estimated 40 to 60% (the cost reductions achieved by the BBC and the French Institutit National de l'Audiovisuel).

3.2. A Preservation Factory for Everyone - PrestoSpace

The PrestoSpace project followed, dedicated to 'spreading the knowledge' so that all audiovisual collections could reduce the cost and improve the efficiency of their preservation programmes. The project also provided basic information on all the stages needed before starting digitisation: taking a collection-level approach to preservation, making a map of the archive holdings and using a triage approach to set priorities, using that information to estimate costs – all the steps needed to build a convincing business case and so obtain funding for preservation work.

PrestoSpace provided a huge amount of information which remains available: http://digitalpreservation.ssl.co.uk/ The information includes:

- a wiki which goes step-by-step through preservation planning http://wiki.prestospace.org/
- a set of cost calculators for estimating digitisation costs (for preparing a business case) http://digitalpreservation.ssl.co.uk/about/56/59.html
- a set of training videos for making 'preservation transfers' from obsolete formats http://digitalpreservation.ssl.co.uk/training/2460/2542.html
- a database of formats, including how to identify formats plus key information for preservation and migration http://prestospace.it-innovation.soton.ac.uk/prestospace/

The 'Best of PrestoSpace' has been sifted and sorted by the PrestoPRIME project, and so concise documents with clear summaries are now also available through the PrestoCentre.

A major outcome of PrestoSpace was the realisation that there already was an industry for making 'preservation transfers', namely the people who already made transfers (and did other technical work) for the broadcasting and advertising industries. These companies, often call facility houses, had the equipment and technical expertise. What they did not always have was knowledge of archive requirements, and an attractive price. Many of these companies now do work regularly with archives and have developed commercial versions of preservation factories that do make the economies of scale available to all audiovisual institutions.

3.3. Sustainable Practice – PrestoPRIME

Digitisation is possibly the biggest single and most expensive step an audiovisual archive may face, but it is still just one step in the life of the content of the institution. PrestoPRIME addressed two problems:

1. Keeping the content alive: after conversion from shelf-based to file-based content, there is a whole new set of issues around preserving the file, and keeping their content usable;
2. Keeping the support alive: after Presto, PrestoSpace and PrestoPRIME there had been a decade of activity around the general issue of audiovisual preservation. These projects had raised awareness and provided information and solutions. After PrestoPRIME ended there would be a gap, and people who had obtained value from the Presto series of projects would not have an active source of support.

The first point has been the target for the technical work of PrestoPRIME, and results are presented in Section 7, below.

The second point has been addressed by the creation of an organisation that is intended to be self-sustaining: the PrestoCentre. This is a formal legal organisation with support from the national audiovisual organisations of France and The Netherlands, and from the national public service broadcasters of Italy, Austria and the UK (RAI, ORF, BBC). It already has a wealth of online information and has run two major international conferences. More information on the PrestoCentre is in Section 8, below.

4. Coping with Digital Content

When PrestPRIME started running seminars, workshops and other training sessions on digital preservation, we rapidly discovered that the questions that people were asking were not confined to preservation itself. People wanted information on all aspects of file-based content, which fall into three broad categories:

- **Creating Digital Content** – moving into the file-based world;
- **Working with Digital Content** – workflows, training, policies and all related managerial issues around working with files instead of physical objects; and
- **Preserving Digital Content** – the actual preservation issues, centred around standards, wrappers, codecs and digital migration.

4.1. Creating Digital Content

There are two main problems for playback of all analogue media: 1) lack of equipment, spares, calibration materials and experienced operators; and 2) condition of the originals, which may be fragile or already damaged or decayed.

There is a particular problem for tape, common to audio and videotape (and also common to both analogue and digital tape, including data tape): sticky shed. This is the shedding of the oxide coating of the tape during playback, leading to clogging of the playback head and resultant loss of signal. Initially the signal loss is usually brief and intermittent, but eventually the playback machine can jam completely.

Specific information on problems with digitisation of audio, video and film are given in a comprehensive document from the Digital Preservation Coalition[6]. A particular issue which has no counterpart outside the audiovisual world is the problem of content which is digital, but which is NOT in files. The audio on CD-Rs, DAT tapes and minidiscs has to be 'ripped' to be put into files that a computer can recognise and deal with. A related but more complicated process is required for video from digital video tape. For the audio and the video, an unsolved problem is that there is no way to know what exactly is on the carrier (tape, CD or minidisc) as there are layers of processing (including error correction and error concealment) between the raw media and "the bits" that come out of the playback device.

4.2. Working with Digital Content

PrestoPRIME has found that the major preoccupation of people who have digitised their content (or at least started down that road) is not the preservation of the result, but the many changes in working practices that are required – or made possible as new services — by content in files. The whole workflow from acquisition through documentation to delivery is changed. Most of this issue is not the concern of PrestoPRIME, though all the Presto projects have looked at the change in access as a major consequence of digitisation (in combination with Internet technology).

Information and support for changing work practices is available from the relevant professional organisations, such as IASA[7], AMIA[8] and FIAT-IFTA[9], as well as from the audiovisual sections of IFLA[10] and the ICA[11].

4.3. Preserving Digital Content

Digital Preservation is a new field which has been developing its technology and standards only over the last 20 years, and for most people their exposure to digital preservation technology is much more recent. Much of it grows out of the digital library world and is based around national and research libraries, national archives and a few governmental organisations such as space agencies (NASA, ESA). This list does *not* include broadcasting (where the bulk of audiovisual content lies) and does not directly include the commercial information technology world (except for that part of it which supports the libraries, archives and space agencies). The consequence is that formal digital preservation technology is still limited to a small subset of the organisations that actually have digital preservation problems. For the rest, which includes most of the holders of audiovisual content, digital preservation is not well understood. Instead, most organisations rely upon IT departments doing 'best practice'. Unfortunately the IT industry is too young to have very much 'best practice' on the sort of time scale (decades or centuries) that matter to archives! To make matters worse, most of the technology of so-called 'information technology' becomes obsolete on an ever shorter timescale (three to five years) than the already short timescales for obsolescence of audiovisual carriers, file formats and encoding systems.

The next section summarised current information technology related to digital preservation.

5. Standard Approaches to Digital Preservation

As we move from archive content on shelves to file-based content on mass storage devices, we move into the world of information technology (IT). This world offers a range of tools for managing content, some which come from the library and archive world and some which come from the IT world which can seem alien and unfriendly. The following is an attempt to categorise the various kinds of software, and explain their strengths and weaknesses. We will cover the following:

- digital archives
- digital libraries
- digital asset management
- trusted digital repositories
- digital preservation systems

5.1.Digital Archives

There are basically two kinds of digital archive:

1. off-line storage: a place to put files that do not need fast access; example: email archive
2. an electronic version of a shelf-based archive.

Most IT companies use the first meaning, and most audiovisual collections use the second! The first is a place for data that is no longer serving a primary purpose. The second is itself a primary purpose: the management of valuable file-based content.

The distinction between these two uses of the word archive is important, because computer systems that were developed as off-line storage are now starting to offer themselves as IT solutions for primary archives, which can lead to a great muddle. **For audiovisual collections, or for any primary content, the vital issue is to use computer systems which create access, not those which reduce it.**

5.2. Digital Libraries and Digital Asset Management Systems

The major experience in digital archives lies with libraries: national, research, academic. They have been building digital collections, and developing *digital library technology*, since the mid 1990's[8]. More recently, libraries, archives and government bodies concerned with the permanence of digital materials have moved on to develop *digital preservation technology*, with the Open Archive Information System[9] (OAIS) model (made an ISO standard in 2003) and the development of the concept of *trusted digital repositories*[10] – meaning trusted to keep things long-term.

Unfortunately for audiovisual content, the work on digital libraries, archive and repositories and the related work on digital preservation have concentrated on the main business of archives and libraries: *documents containing (mainly) text and still images*. Sound and moving images have all the additional issues listed above, with requirements that have not been addressed by mainstream digital preservation approaches. The result is that many tools developed for file-based archiving and preservation only support document and image formats, and certainly don't support MXF[12] (the standard file format in broadcasting and digital cinema).

A related problem arises within broadcast archives. Their technical and IT staff and systems are largely unaware of the technology, standards and systems developed for digital archiving and preservation. There is a 'two worlds' problem: those who know and use MXF, and those who know and use OAIS. Broadcast archives are caught in-between.

What is the difference between an asset management system and a digital archive? In the author's view, the answer to this confusion is in functionality: ignore what the system is called, and ask what it does. Secondary archiving systems reduce access – putting content 'somewhere else'. Primary archiving systems (and asset management systems) treat the archive content as the primary content, and concentrate on tools for dealing with that content. Both primary digital archive systems and asset management systems would have metadata tools. One would expect an archive system to excel in metadata tools, and offer professional library tools. An asset management system would be likely to be weak on the librarianship side, but strong on the media manipulation tools.

5.3.Trusted Digital Repositories

There are two basic differences between a conventional library IT system and a digital repository:

- **a repository holds the content**, not just a catalogue and tools for acquisition, circulation control and other tasks;
- **a repository prevents loss of content**, or at the least tries very hard to prevent loss – by incorporating processes and technology specifically aimed at insuring the continued viability (persistence and currency) of the content.

Today, most broadcast archives are moving from tapes on shelves to file-based content, and using asset management technology to manage these files. In general this means a reduction in librarianship tools, at least for those archives that had library-type acquisition, classification, cataloguing and control processes.

All these areas tend to be weaker in asset management systems:

- files can get in and out without going through a formal acquisition or control process;
- files don't have to be classified using a controlled vocabulary or a hierarchical classification system; they may get *tagged* using uncontrolled vocabularies (and the errors of unrestricted text) or there may well be no formal indexing;
- file content doesn't have to have *cataloguing:* an analytical description of the contents.

However there are full-feature asset management systems that do support library functionality, and when they do, they provide features that standard digital libraries do not provide: metadata (keywords, analytical descriptions) tied to specific shots (segments, clips) within a video asset, or tied to specific time points in an audio stream. **The whole area of time-based tools is what distinguishes media asset management systems from digital library systems. Time-based tools are essential for time-based media. Media asset management systems have such tools, and digital libraries don't** – in general.

This situation leaves audiovisual collection managers in a difficulty: they can have the high-level metadata and overall acquisition and circulation control tools they need, or the time-based media manipulation tools – but not both!

5.4. Digital Preservation Technology

Files face a range of obsolescence issues, addressed by digital preservation technology:

- methods for ensuring that obsolete files can migrate to new standards and formats; PREMIS preservation metadata, JHOVE and DROID file identification tools, databases of information on file formats (PRONOM, US Library of Congress);
- methods for emulating old IT environments to extend the lifetime of obsolete formats; the major example is EU project SHAMAN and the Multivalent approach;
- criteria for evaluating the reliability of a digital repository; the TRAC and DRAMBORA work, which has now produced an ISO standard;
- and finally an overall methodology: OAIS.

A brief review of the digital preservation technology and its application to audiovisual content is available from the Digital Preservation Europe project[13].

Until the various projects and initiatives just listed develop software that enters the commercial world understood by standard IT staff, implementation of digital preservation technology will largely be limited to national libraries and other major libraries. These institutions are big enough to have their own IT staff, recruited and trained specifically to implement technology needed by libraries. The rest of us (broadcasting, media production and archiving, small collections without dedicated IT staff) will find it difficult to implement any formal digital preservation technology. A major role of PrestoPRIME is to provide information – and tools – to ease that difficulty.

5.5. Audiovisual Requirements for Digital Preservation Technology

The conclusion with regard to audiovisual files is that the simplest way to preserve digital audiovisual content is by use of **uncompressed data, fully described by technical metadata**. Preservation metadata as contained in the PREMIS standard gives a structure for defining the whole IT environment needed by a particular file type. For uncompressed data, any IT environment will do; any generic player of audio and video will do; and the signal can be easily moved, preserving all the bits as in the original file, from one generic wrapper to another. In short, uncompressed audiovisual data short-circuits the need for most of the complexities addressed by PREMIS, or by OAIS itself.

Quoting PrestoPRIME[14]: "*Problems arise because of complexity. Many encodings can share a common wrapper, so that, for instance, a '.wav' file can contain many different ways to represent an audio signal, ranging from non-linear allocation of bits in samples to highly-compressed data. The situation for video is more complicated just because there are so many file and wrapper formats, as well as so many encoding possibilities. Further, two files made using the same encoder, and wrapped in the same type of wrapper, can still differ enormously. They could differ in their* ***compression parameters****, so that one MPEG-2 file (for instance) could be broadcast production quality at 50 Mb/s, while another could be unsuitable for professional editing, being lower in quality and not allowing edit at specific frames.*

This whole situation is very unsatisfactory for long-term preservation, because key knowledge is embedded in media players (playout or rendering software), rather than captured in formal presevation metadata as provided by PREMIS. Hence the survivability of the content is dependent upon the survivability of the players. Audiovisual content is not unique in this respect – text files are equally dependent upon software that can 'render' their contents. However there are ways to reduce the dependence:

- ***better metadata****: successful rendering software can determine what kind of data it is dealing with, by reading and interpreting meta-information from the file; this information could in principle be 'pulled out' of the file and made explicit as formal technical metadata. The audiovisual industry would benefit from much more agreement on where and how to place metadata in proprietary file types, and on ensuring that all the decode parameters were part of that metadata.*
- ***simpler files****: most of the complexities of audiovisual content are to do with compression methods and interpretation of compressed data. Uncompressed audio is virtually self-describing (or needs no*

description, beyond: the following is a sequence of audio samples – just work out three parameters and it can play perfectly). Uncompressed video is more complicated but it also is 'just a sequence of samples'. Virtually nothing general can be said about the data in a compressed file, and attempting the playback of a compressed file of an unknown type could well prove futile."

6. Audiovisual Content is Special

There are many differences between audiovisual files and text files:

- **Complexity:** in addition to the wide range of file formats, a file can contain (wrap) many elements: video, multiple sound tracks, subtitling, time code, metadata. There are also many types and qualities of encoding (MPEG 2, MPEG 4, JPEG, JPEG2000, MJPEG, AVI, MOV, WAV, MP3 and dozens more[15]), the fact that multiple versions (production quality, transmission quality, browse quality) have to be managed as a unit – and finally the issues of Digital Rights Management (DRM).
- Audiovisual content represents an analogue **signal**, and so has dimensions of frequency response and signal-to-noise ratio that determine how faithfully a sound or image is captured or reproduced. In addition to "preserving the bits" there is the added dimension of whether "the bits" have preserved the signal.
- **Compression:** audiovisual signals contain redundancy, and so for decades people have pumped high-bandwidth signals through low-bandwidth channels by manipulation to remove the least significant (hopefully) information. Only lossless compression is used on text, but audiovisual content gets subjected to processes that throw away parts of the signal: lossy compression. The effect, years later, of repeated application of various kinds of lossy compression is a hazard unique to audiovisual content.
- **Size:** "above four gigabytes, everything breaks[16]" is not absolutely true, but it is true that storage, systems, networks and applications are stressed by large files. Four GB was the memory limit of 32-bit processors. Personal computers are now mainly 64-bit, but many embedded processors that control everything from coders to routers to storage devices are not 64-bit – and many applications have limitations of their own (analogous to the Year 2000 problem) which are independent of hardware and which can kick in at less than four GB. An hour of standard definition video at full quality (uncompressed) is about 100 GB.
- **Time:** audio and video have a time dimension. Metadata and applications need to understand that dimension, so documentation and access can 'point to the right place' rather than simply dealing with an audiovisual file as a unit (lump).
- **Resilience:** for decades broadcasting has coped with errors. A glitch from an analogue VTR (videotape recorder) could be smoothed out using a time-base corrector; a glitch in playback using a digital VTR could be concealed through repetition of an adjacent line, or even an entire frame. Systems for files use built-in error detection and correction, and if that fails, the file as a whole can be rejected, generating an error message saying something like "file cannot be opened".

Because video is highly structured (into lines and frames), there is potential for playback despite errors – if only the IT systems would pass 'the good bits', plus an indication of where (along the time dimension) the error occurred.

6.1. Access Requirement for Sound and Moving Images

As mentioned, a fundamental difference between audiovisual materials and text or still images is the dimension of time. These files have a time dimension, and so their documentation, search-and-retrieval and play out (rendering) will benefit if they recognise and use this dimension. There are four requirements for time-base access which are fundamental to all capture, storage, archiving, access and re-use of audiovisual content:

- *Granularity:* division into meaningful parts, so the item can be represented in some visual way (eg by key frames or a storyboard), supporting navigation (getting to the right place)
- *Navigation:* playback from a specified point, so the user can click on a key frame (or an audio equivalent) and immediately get to the desired place in the content
- *Citation:* creation of a pointer (to a specific time point within a file associated with a permanent URI) that can be put in an email, website – or PhD thesis.
- *Annotation:* adding commentary at a citation point. If the commentary can be accessed by other people, there is then a basis for building a community or social network around the annotations and the annotation process.

7. PrestoPRIME Results

PrestoPRIME is the European project supporting the *digital* preservation of audiovisual content, meaning how to keep things that are already file-based – and how to keep them working. All the formal papers from PrestoPRIME are on the project website[17], but the best source of PrestoPRIME information is the PrestoCentre website[18], where there is already a wealth of information and where all new information will be located once PrestoPRIME ends (in November 2012). PrestoPRIME results include:

P4: The overall product of PrestoPRIME is a complete *preservation platform*, a full system for holding and managing file-based audiovisual content, indefinitely. P4 is an Open Source framework for the technologies developed within PrestoPRIME which support the process of running a preservation service. As an implementation of the OAIS model, it offers the basic services and interfaces for the ingestion of and access to AV items. Software tools develped by the partners or third party tools can be integrated according to the defined Reference Architecture. Specific tools that integrate with P4 have already been released as open source, including an MXF D10 checksum tool, and MXF metadata extractor and an LTFS archiver (for use with LTO-5 datatape systems supporting the LTFS structure). These tools are online here: http://www.crit.rai.it/EN/attivita/opensource/index.htm

PrestoPRIME has also developed specific tools supporting audiovisual files, which integrate with the commercial **Rosetta** digital preservation system.

Analysis tools: The tools from ITI are all free and online here: http://prestoprime.it-innovation.soton.ac.uk/ including documentation, example videos, user guide and FAQs as well as access to the source code, bug tracking and feature requests.

Preservation planning tools from PrestoPRIME include: iModel, a simulation tool for storage, transcoding and file-format migration of digital audio visual assets; iWorkflow, a simulation tool for digitisation/migration workflows of discrete assets (e.g. digital video tapes) and has been developed for a specific scenario at the BBC for their D3 project.

The storage and format migration tool (iModel) is intended to allow a wide range of questions to be considered when planning, selecting or operating a storage and access system. The tool focuses on the storage and access to digital content in files using IT systems. The tool does not include metadata management, rights management and other issues that are of course important to consider. However, the tool does allow the following questions to be investigated:

- When storing content, how many copies should be made, what technologies should be used, how much will it cost, what are the long-term risks of losing files?
- How does codec choice (e.g. compressed or uncompressed video) affect costs and risks?
- What are the pros and cons of just in time generation of access copies compared to creating and storing a full set of proxies in advance?
- When storing data, how often should it be checked to make sure integrity is intact, and when does this become counter-productive (e.g. act of checking causes more damage than it might repair)?
- When should media migration take place: regularly or at the point of obsolescence?
- What is the impact of ingest and access on shared resources for storage and data safety: what level of resources is needed to support both?

Service Management: Audiovisual archives are increasingly file-based and becoming an active element of the production, post-production and distribution process. Often archive systems are in-house, but increasingly parts are out-sourced and even off-site. How can archive IT systems be managed to ensure they meet key performance indicators for ingest and access whilst not compromising the safety of the assets? PrestoPRIME addresses this problem through policy-based automation applied to outwardly facing archive services and internal preservation processes alike, both defined through service level agreements (SLAs) and measured against metrics for performance, data integrity and availability.

MServe is an automated service management tool to monitor and manage the services required for audiovisual preservation, define SLAs for the different users of the system and maintain the quality of service defined in those SLAs.

Video Quality Assessment: Manual quality assessment of audiovisual media is a very time-consuming, and expensive activity. For automatic quality control of digital audiovisual media only some tools have been developed. Mainly syntactical properties of the digital files can be checked, e.g. stream encoding or file wrapper standards compliance. Tools for checking the quality of the video or film images itself in an automatic manner are rare today. PrestoPRIME has tools for automatically detecting various video

and film impairments including video breakup, noise and grain levels – and the interactive 'AV-Inspector Summary' tool for efficient human verification of automatically detected visual impairments.

Metadata Mapping and Validation: There is a large heterogeneity in terms of metadata formats between different audiovisual collections, e.g. broadcast archives, film archives, libraries holding a/v assets. Making contents accessible to wider communities in both B2B and B2C scenarios, e.g. on portals such as Europeana, increases the number of metadata formats involved, such as the MPEG-7 AVDP, the Europeana data model and the W3C Ontology for Media Resources on the Web. The variety of producers and consumers of metadata creates the need for mappings or crosswalks between the different metadata formats. PrestoPRIME has novel approaches for automating metadata mapping and semantic validation of metadata documents, that aim at overcoming the limitations of 1:1 mappings between each pair of formats. Visual Quality and Metadata tools are demonstrated at http://prestoprime.joanneum.at/

Management of User-Generated Metadata: Tagging games are explored as a method to collect user generated metadata. The Netherlands Institute for Sound and Vision deployed the video labelling game Waisda? to collect time-based tags for videos from their collection. PrestoPRIME studied the role of such user tags for audiovisual collections. and developed a workflow that allows the moderation of user tags by domain experts, with the possibility to semi-automatically link tags to concepts from the Linked Open Data cloud.

Rights Services: The goal is to improve the management of audiovisual rights information. From the analysis of a number of real contracts a glossary of rights terms has been obtained. We then defined an extension of MVCO (Mpeg Media Value Chain Ontology) which is both a model and a language for unambiguously representing the rights situation of AV intellectual property entities. The RAI RightsDraw services allow their users to create, view and edit the rights information related to audiovisual assets, or to specific queries.

8. The PrestoCentre Competence Centre

The PrestoCentre Foundation is a membership-driven organisation that brings together a global community of stakeholders in audiovisual digitisation and digital preservation to share, work and learn. As intermediary, PrestoCentre Foundation offers a valuable platform by connecting and facilitating interactions between these groups:

- For smaller archives, we provide services to connect, share experiences and learn from early adopting organisations;
- For larger archives, we offer structured ways to collaborate and share knowledge; For research and commercial organisations, PrestoCentre functions as a market and technology watch and helps in the evaluation and transfer of new R&D and commercial output;
- PrestoCentre also works with advocates, public services and professional associations, to forge closer ties with government, industry, and academia.

PrestoCentre's free tools and simple strategies save money and time, while improving long-term access to digital audiovisual collections. PrestoCentre works with experts, researchers, advocates, businesses, public services, educational organisations and professional associations to enhance the audiovisual sector's ability to provide long-term access to cultural heritage.

PrestoCentre is the outgrowth of more than ten years of work by several major European organisations and a main outcome of the PrestoPRIME project to:

- enhance collaboration between audiovisual content holders;
- facilitate coordinated action in the areas of digitisation, digital preservation of and long-term access to audiovisual archival content;
- and serve an international community of stakeholders in audiovisual digitisation and digital preservation through online and offline services, publications and training.

Members of PrestoCentre include: broadcast, sound and film archives; national libraries and national archives; regional, subject specific, small, mid-size, university, and corporate archives; creators of audiovisual materials; independent producers; providers and suppliers; funders; standards efforts, and other organisations concerned with archiving.

The PrestoCentre Foundation is a legal entity registered as a non-profit foundation that operates through a Board of Directors. PrestoCentre's governing board includes representatives from the British Broadcasting Corporation (BBC, UK); l'Institut National de l'Audiovisuel (INA, France); Netherlands Institute for Sound and Vision (Beeld en Geluid); Österreichischer Rundfunk (ORF, Austria); and Radiotelevisione Italiana (RAI, Italy). These organisations provide a basic support for the PrestoCentre and its activities.

The Founders of PrestoCentre have invested millions to develop new technologies and services that serve the public. PrestoCentre provides a structured way for leading organisations to share those results and accelerate research efforts in new media and education by forging closer ties between government, industry, and academia.

Most of PrestoCentre's activities will consist of collaborative work between co-workers placed in different in various member organisations. Competence in the AV domain is widely distributed. Providing access to that expertise requires online collaboration in the broadest sense. Online activities will be a primary focus for PrestoCentre and having PrestoCentre's community operate primarily in an online environment has many advantages.

---

[1] Klijn, E and de Lusenet, Y (2008) Tracking the reel world. http://www.tape-online.net/survey.html
[2] http://portal.unesco.org/ci/en/ev.php-URL_ID=2034&URL_DO=DO_TOPIC&URL_SECTION=201.html
[3] Wright, R. 2010. *PrestoPRIME Deliverable D7.1.3 Audiovisual Digital Preservation Status Report*. Paris, PrestoPRIME. www.prestoprime.org/project/public.en.html
[4] http://wiki.prestospace.org/pmwiki.php?n=Main.PreservationStrategy#Conservation

[5] IASA, 2009. Technical Committee, Kevin Bradley, (Ed.) *Guidelines on the Production and Preservation of Digital Audio Objects* (IASA-TC 04 Second edition) Canberra, IASA. http://www.iasa-web.org/audio-preservation-tc04

[6] Wright, R (2012) "Preserving Moving Pictures and Sound" Digital Preservation Coalition. http://www.dpconline.org/advice/technology-watch-reports
[7] International Association of Sound and Audiovisual Archives
[8] Association of Moving Image Archives
[9] International Federation of Television Archives
[10] International Federation of Libraries Associations
[11] International Council of Archives
[12] None of the standard digital library tools listed in the File Information Tool Set (FITS) supports MXF: http://code.google.com/p/fits/
[13] Wright, Richard (2008) "Preservation of Digital Audiovisual Content" http://www.digitalpreservationeurope.eu/publications/briefs/
[14] D2.1.1 "Audiovisual preservation strategies, data models and value-chains" (Section 3.2: Role of technical metadata in preservation) https://prestoprimews.ina.fr/public/deliverables/PP_WP2_D2.1.1_preservationstrategies_R0_v1.00.pdf
[15] FFMPEG supports 22 families of codecs, each with several varieties; http://en.wikipedia.org/wiki/FFmpeg#Codecs
[16] Addis, Matthew; project Avatar-m: www.avatar-m.org.uk/ IT Innovation, University of Southamption
[17] http://www.prestoprime.org/project/public.en.html
[18] http://www.prestocentre.org/