

Preservation of Digital Audiovisual Content

The audiovisual (AV) record of the 20th century is at risk, and digitisation has been a solution, which has created a new problem: preservation of digital AV content. These files have requirements (size; specific formats) that are not adequately addressed by current technology. Best practice can be recommended, but three major changes are needed: 1) AV collections should use existing digital library and digital preservation technology; 2) technology should advance, to support time-based media; 3) mass storage and general information technology should advance, to support the specific requirements of AV files.

Problems of digital audiovisual preservation

Europe's AV holdings (in archives or other formal collections) have been estimated at 50 million hours of audio, video and film, most of it in analogue formats. About 70 per cent of this material is at risk now, and all of it will be at risk within 30 years – owing to obsolescence, deterioration and obsolete formats.

Major digitisation programmes have started: an estimated 10 million hours has been digitised in the last decade. While AV collections have been busy changing their tapes and gramophone records into files – as a preservation solution – the rest of the world has become aware that digital files present their own preservation problems.

Large collections of files are a technical management problem; the solution is *digital library technology*. Files need maintenance: they must be named, moved to new storage (frequently!), copied for access, encoded for changing access needs, checked for validity. They need metadata actions, ranging from cataloguing to automated harvesting (for standardised and global access). Manual maintenance is simply impossible – and too error-prone – once collections reach a certain size. Digital library technology supplies automation tools for creation, maintenance and access requirements of large collections of files. There are many guides to digital library technology.

Two worlds: Digital library technology comes from the academic library world. AV collections are largely outside that world. The biggest holders of content are broadcasters, and other major holdings are in film museums and other cultural and heritage institutions (one of the largest film collections in the UK is at the Imperial War Museum). Broadcasters vary, but it is common for the computer and technical staff of a broadcaster, and the management who decide and fund technology issues to know absolutely nothing of academic libraries and digital library technology.

The first hurdle faced in preserving AV files is to know about, understand, fund and use the existing digital library tools that can change a heap of files into a managed collection.

The second hurdle is recognising that digital library tools provide management (so files can be accessed and do not get lost) but do not cover preservation. 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, methods for emulating old IT environments to extend the lifetime of obsolete formats, criteria for evaluating the reliability of a digital repository, and finally an overall methodology: OAIS. AV collections have difficulty finding anyone on their IT staff who has even heard of OAIS, which rather limits support for funding and implementation. Fortunately, the EC project MEMORIES is developing OAIS and related procedures specifically for audio and video collections.

The third hurdle is that the specific needs of AV files are not fully supported by digital library and digital preservation technology, as discussed next.

References and Further Information

Audiovisual Status Surveys:

PRESTO: <http://presto.joanneum.ac.at>

PrestoSPACE: <http://www.prestospace.eu>

TAPE: <http://www.tape-online.net/survey.html>

http://www.tape-online.net/docs/audiovisual_research_collections.pdf

General Guides to Preservation:

<http://www.bbcarchive.org.uk/>

<http://digitalpreservation.ssl.co.uk/>

A general list of digital curation tools:

<http://twiki.dcc.rl.ac.uk/bin/view/Main/DevelopmentToolList>

JHOVE: <http://hul.harvard.edu/jhove/>

DROID: <http://droid.sourceforge.net/wiki/index.php/>

Introduction

PRONOM:

<http://www.nationalarchives.gov.uk/pronom>

National Library of New Zealand Library metadata

extractor: <http://meta-extractor.sourceforge.net/>

OAI: [http://www.openarchives.org/OAI/](http://www.openarchives.org/OAI/openarchivesprotocol.html)

[openarchivesprotocol.html](http://www.openarchives.org/OAI/openarchivesprotocol.html)

OAIS:

http://www.dpconline.org/docs/lavoie_OAIS.pdf

Migration:

<http://www.library.cornell.edu/iris/migration/>

Emulation:

<http://www.dlib.org/dlib/october00/granger/10granger.html>

Repository evaluation criteria reference:

<http://journals.tdl.org/jodi/article/view/199/180>

Projects and initiatives:

MEMORIES:

<http://www.memories-project.eu/>

European Digital Library:

<http://www.europeana.eu/>

Formats:

MXF:

<http://www.digitalpreservation.gov/formats/fdd/fdd000013.shtml>

WAV format specification:

<http://www.digitalpreservation.gov/formats/fdd/fdd000001.shtml>

Presentations:

Video Formatting and Preservation, Carl Fleischhauer,

NDIIPP DLF Forum, Philadelphia, 6 November 2007

<http://www.diglib.org/forums/fall2007/presentations/Fleischhauer.pdf>

Unique problems of digital audiovisual data and files

Because of the 'two worlds' problem, professional broadcast formats (MXF in particular) are unsupported by many digital library and preservation tools. Other 'standard' formats are better supported, but many (e.g. AVI, WMV) are proprietary, which is in itself a preservation problem.

The remaining problems relate to the actual content of the files.

- Most AV files are compressed. Whatever 'original quality' was lost in compression will remain lost. Preservation should maximise retention of quality, a capability that needs to be defined and added to current technology.
- Time-based content needs tools with a time dimension (cataloguing, navigation, edit).
- The files are complex. Indeed, the concept of a *wrapper* was developed to recognise the complexity of a typical AV file: multiple signals, multiple kinds of metadata – including time-domain (subtitles) and numerical (time code).
- AV preservation involves many related files: lossless and lossy encodings, multiple proxies (supporting access in multiple formats, e.g. Real, Windows Media, MPEG, AVI, Quicktime, Flash), various stages of edit and recombination, and a range of rights information: multiple interested parties, multiple collection agencies, non-uniformity from country to country. A complex of information representing signal, metadata and rights must be preserved.

Access

Libraries have a tradition of unified access: union catalogues based on standardised metadata, to provide an 'any book, anywhere' service. Many audiovisual collections have a tradition of being closed, or open only for professional or commercial access.

Digital libraries continue the tradition of expanded and unified access, often on a national or multi-national scale, as with the European Digital Library. AV collections need the technology of digital libraries to be accessible through major projects such as EDL. In turn, these digital libraries need to put more effort into understanding the problems of digital audiovisual data and files just discussed. In particular, digital libraries need tools for time-based access to both the AV signal and to the metadata (rights, for instance, can vary from moment to moment within a single AV file).

Much AV content is held by institutions with no history of working with libraries and who may prefer to limit access to 'their' content. Marketing, branding and rights issues impede a 'European Audiovisual Portal'. EDL may never include BBC content.

What to do

Despite the problems, some clear statements can be made about AV preservation:

- **Preserve the artefact:** Keep the 'original', even if compressed. 'Preserve the bits', whatever else is done. AV content has one advantage: there is a lot of it, in a relatively small number of formats. Methods to 'play the bits' may exist.
- **Decode to uncompressed** and save as uncompressed (*in addition* to keeping the original). This is a demanding requirement for video (100 GB/hr for 625-line TV), but storage is now very inexpensive.
- **Enhance the metadata:** A file extension (e.g. .wav, .avi is not sufficient). There are over 50 registered variants of encoding within the definition of .wav; MPEG-1 and MPEG-2 use the extension .mpg. Ideally, there will be a metadata extraction tool; otherwise, manual testing and documentation is needed.
- **You are not alone:** Use the file-type registries, software repositories, emulation platforms, and Preservation Guides listed in the references.