



Programme Archives

– an EBU production seminar

Geneva, 26-28 January 1999

The issue

Television and radio archives have recently been confronted with a change in perspective. Whereas they were previously hidden in seldom-visited places or, at least, stored and managed outside the active production process, the recent evolution of media broadcasting has brought them to the front row of audio-visual operations. They are considered now as essential libraries for everyday production. Even more, they are valued as assets which can be a source of important revenues.

The problem is that the material in present-day archives is not very suitable for playing such a role. Film and videotapes degrade with time, both physically and chemically. They wear out mechanically. They suffer more and more damage as their use increases and, even for quite basic operations, their capabilities are very restricting. For example:

- ⇒ locating the material requires the intervention of dedicated staff;
- ⇒ searching (i.e. spooling) through film and tape takes a lot of time;
- ⇒ copying needs to be done in real time;
- ⇒ delivery of material needs to be on a physical medium;
- ⇒ and so on ...

So how can things be improved?

Given the promising developments in storage media and computer communications, the on-line archive is becoming a closer reality. Based on the use of disks and/or data tapes as the storage media, the idea is to make archives available at a distance – at least for consultation but also, when appropriate networks are available, for delivery direct to the production area.

The investments needed to achieve all this are considerable – investments in equipment, of course, but also in resources, in time and in training. New methods of working are involved. It is thus indispensable that the concept is analyzed critically to ascertain that it is not another myth of the millennium – particularly as the pertinent technology is often “leading-edge” and has not passed through a long experimentation period.

The event

With that aim in mind, the EBU Technical Department – with the active collaboration of the PMC – organized a two-and-a-half day seminar at EBU headquarters in January.

Five themes covered the issue:





EBU SEMINAR

- ⇒ users requirements and general architecture;
- ⇒ identification of content;
- ⇒ transfer of content;
- ⇒ storage media;
- ⇒ migration strategies.

These themes, and various items which they imply, were allocated to 23 speakers at the seminar. Each theme concluded with a table discussion, which made it possible to get a fairly extensive view of most issues.

The seminar was a great success. There were 202 registrations, the highest number ever achieved for such an EBU event. Some fifty organizations or companies were represented, coming from 30 countries.

As with the previous EBU seminars, the audience consisted of decision-makers together with specialists in their field. About 40% of the participants were archivists, 40% were engineers and the rest came from news services or were consultants.

The two groups of sessions covered complementary, albeit somewhat independent, subjects. The first group dealt with documentation and retrieval operations, which are the basic work of the archivist. The second group covered the technical facilities which directly influence the storage and provision capabilities of an archive.

The following is a report on the first group which comprised two sessions. A short article on the other, more technical, sessions will appear in the Summer issue of **EBU Technical Review**.

Abbreviations

DAVIC	Digital Audio-Visual Council	IPTC	International Press Telecommunication Council
DiVAN	Distributed Video Archive Networks	ISAN	International standard audiovisual number
DoD	(US) Department of Defense	ISO	International Organization for Standardization
EPG	Electronic programme guide	MPEG	(ISO/IEC) Moving Picture Experts Group
FIAT	<i>Fédération Internationale des Archives de Télé- vision</i> (IFTA in English)	PMC	(EBU) Production Technology Management Committee
IASA	International Association of Sound Archives	SMEF	Standard media exchange format
IEC	International Electrotechnical Commission	SMPTE	Society of Motion Picture and Television Engineers (USA)
IFTA	International Federation of Television Archives (FIAT in French)	UMID	Unique material identifier
INA	<i>Institut National de l'Audiovisuel</i> (France)		

Assistance to documentation and retrieval

Documentation – i.e. the operation of indexing and describing an archive's properties – is the most heavy task of an archivist and, yet, one whose limitations are difficult to overcome. So far, catalogues have been only text-based and therefore slow to consult. Their structure and vocabulary are generally peculiar to the organizations concerned. It is therefore very difficult, not to say impossible, to use another organization's catalogue. Even within a given organization, a detailed search cannot be fast, as indexing is very rarely available at the shot or scene level. Even if it were so, a lack of standardization in the use of terms within the organization often makes searching a tedious task.

Research has recently led to developments which should help to pave the way for faster and easier retrieval. Two European research programmes have established tools for an automated video indexing system: EUROMEDIA and DiVAN.

EUROMEDIA

The ESPRIT programme called EUROMEDIA – which comprised broadcast partners SWF/SDR (now SWR), ORF, the BBC and SVT – was the first to mature: it had just started to be used in production at SWR in January 1999 and, hence, the presentation at the seminar had a special appeal.

EUROMEDIA is an integrated project which can be implemented by stages within existing structures. It augments these structures by introducing new components, but interfaces have been designed to help with this. For example, a digitizer is available as an input tool. The documentation is based on automated video analysis, by which shot-detection and subsequent key-frame extraction allow video programmes to be represented by a "storyboard" – a succession of individual frames in the order of the shots – for fast preview searches. The frames are linked to factual content descriptions by means of timecode. The same tool can of course be applied also to the rushes. The content descriptions are text-based. However, the manual work in establishing the catalogue is made simpler than usual, as there is no need for a text description of the key frames. Catalogues are therefore easier to handle.

EUROMEDIA has concentrated on the needs of consultation. Digitization for preview is done using MPEG-1 compression. The main use today is in assisting editing. The browser client therefore offers conventional player functionality such as forward and backwards, frame by frame, but also it sets start and end markers with the possibility of annotations. There is no specification for the transfer of contents. The format and structure of the data to be exported can be individually selected. Interfaces for defined formats will become available. The commercial product, as used by SWR, is called "Media Archive" and is manufactured by the German company, Tecmath.

DiVAN

Another ESPRIT project called DiVAN (Distributed Video Archive Networks) is being developed and goes one step further than EUROMEDIA. The video indexing (at sequence or shot level) is also carried out by means of video analysis, with automatic feature detection. However, in the next step, a unique and language-independent identifier is attached to the material, which is retrievable through an object-oriented database. Searches that are above shot level benefit from more sophisticated tools which allow, for example, the limits of a scene to be defined by comparing the size and position of faces. Templates will be available which should provide automatic interpretation of the successions of shots. These successions, as well as the



shots, will become story units subjected to documentation. On this basis, a programme can thus be split into stories, which are split into scenes, composed of shots. Advanced user interfaces are planned for providing additions and corrections to the automatic process. CORBA architecture is used for the documentation database, which allows interoperability between various such databases.

The RAI – as well as INA and the Greek broadcaster ERT – is part of the DiVAN project, which is led by a Greek company called Intracom. The RAI will implement DiVAN for building a “multimedia catalogue” by which their stored video and sound material, and photographs, will be made available in a systematic way to production people and to interested bodies or persons in the public. The RAI has not waited until the DiVAN project is completed. Part of their material is already accessible (their most recent items), thanks to home-made software. The reconstruction of the existing catalogue is in progress. Present searching is still text-based.

The operations facilitated

Even in this first level of development, a clear time saving is obtained by using a computer database instead of the usual off-line textual search. Another benefit of computer-based retrieval comes from the possibility of navigating from one level of documentation unit to the other (from the episode to the contained segments) or between units (from a segment to the previous or successive ones). This, of course, is only valid if more than one fragmentation level is implemented.

The most powerful innovation brought about by integrated archives, however, lies in the automatic link which is provided between the catalogue and the library. The call-up of an item in the library is done through an appropriate index, associated with the searched programme segment. Indexes (frame numbers, timecodes etc.) are keys which access the programmes at the wanted granularity level. For the requester to get an idea of what he/she can expect to find in the library, the index will be associated with either a brief textual description of the content – the conventional “annotation” – and/or with keywords or keyframes.

Although the index is a key, it does not give the right to apply to the reached segment, any treatment that the requester may wish. Other sorts of information should also be available when the archived segment is planned to be re-used in the production process, or when its content is to be distributed outside the owner organization. These are the metadata.

Metadata

As is widely explained, metadata is “data about the data”, i.e. about the essential data which represents the basic programme components. A large number of metadata will be originated throughout the production process, which will characterize the various steps in that process. However, only a few of them will need to be kept with the finished programme when it is archived. One such subset is the one which starts with the very beginning of the programme’s life – even, perhaps, before a name has been given to it. This is the “Unique Material Identifier” (UMID). There is a proposed standard from the SMPTE, whose details were given at the seminar by Mr Jim Wilkinson, Chief Research Scientist at Sony, and Chairman of the relevant SMPTE ad-hoc group.

A succession of 64 kB of data is organized so that any unit of video (down to frame level) or audio (192 samples) programme material can be unambiguously defined. The structure comprises a reference number followed by data relating to the date, time, location and other char-



acteristics of creation (name of creator, of course) and whose combination, for a complete set of data, cannot be repeated for another programme segment. This series of bytes – comprising the reference number and the so-called “essential metadata” – would travel unchanged along the chain, while other metadata linked to the processing or modification of the segments would be added at the pertinent steps.

In future production equipment, the essential metadata would be generated automatically with known technologies. The method of generating the reference number – and an accompanying “instance” number for different resolutions, format versions etc – is yet to be defined.

The standard is being established by the SMPTE but its application can, of course, be worldwide.

The UMID is mainly intended to provide ownership details for reference and copyright. It works at creation level. The final, finished programme can also be defined with such a system. However, there are simple ways of doing the latter which are already in use. An example is the ISAN which is specified and administered by the ISO. Joint work is in progress with IFTA (FIAT) to extend its use within television production.

Other metadata are needed to assist in the production chain of operations, so as to allow each processing step to be performed confidently. Their full set will allow us to trace the history of the shots, pictures or objects which constitute the programme.

The BBC has worked on a standard media exchange format (SMEF) which is basically a dictionary defining data flows across systems and processes. A “data model” adds structure and hierarchy to the information, and opens the door to different levels of granularity. The philosophy of the work is based on the EBU/SMPTE Task Force findings. As Carol Owens, the Director of the BBC project, told the seminar, the SMEF could be “the next layer up the protocol stack”. Its implementation will be an essential contribution to the testing of such concepts within real television production. The EBU has decided to set up a project group on the same matter (P/Meta), to monitor the BBC experiment and, if possible, to validate it through implementation in other European organizations. The aim is also to study the feasibility of creating and adopting common exchange formats for “essence¹” and metadata.

Even in the world of media, metadata is not a new concept. Newspapers and news agency journalists have been using it for many years when sending texts, sound recordings and photographs to their bases. Standards to this effect have been published by the IPTC whereby any form of “data object” can be transferred together with editorial and technical information in the same envelope. For sending video items, additional standards are necessary and the IPTC has wisely decided to wait and rely upon the after-Task Force studies on this matter.

In the EBU/SMPTE Task Force, the US Department of Defense has played a key role in pushing towards the interoperability of different metadata concepts. A presentation by Stephen Long, Chairman of the DoD Video Working Group, was very welcome, even if only to take stock of the situation now that the Task Force report has been issued. One additional concept is that of “Key-Length-Value” by which the core concept for metadata could be used for any data type, be they metadata or data essence. This would enable using the same transport streams for all

1. “Essence” designates either video, or audio or data which is part of the programme, such as teletext. Essence and metadata together is “content”. It is worth noting that this vocabulary was settled as early as 1992, for digital libraries of printed matters, when the Library of Congress in the US started a large-scale project to convert historical collections to digital form and make them available on the Internet.



data activities. Drafts for an appropriate protocol have been established. Also a draft Metadata Dictionary is ready for submission to the SMPTE.

With an even wider scope, DAVIC is studying a general concept of metadata, intended for guiding the final user of a programme, whether available from a broadcasting channel or from the Web, to search and select any desired content through tools such as EPGs.

It is thus quite obvious that the coming use of metadata will have a profound impact on all links in the future television or radio chain.

“ It is thus quite obvious that the coming use of metadata will have a profound impact on all links in the future television or radio chain. ”

* * * * *

The Seminar title intentionally used the generic term “Programme”. Both video and audio were covered. A well-listened presentation was given by the Chairman of the Technical Committee of IASA (International Association of Sound Archives) on the issues which constrain the transfer from analogue to digital recordings. During the round table discussions, it appeared that some EBU Members have already started such a conversion. As an example, YLE has set up a plan for digitizing their radio archives, with the delivery and installation of the new system starting already this year.

The subject is a hot issue for both television and radio. It is hoped that the setting up of appropriate EBU project groups will help in managing the transition period. In fact, the EBU has just set up a group on “Future Television Archives” within the framework of the PMC’s activities. A widespread co-operation is planned with other concerned international bodies, and recommendations are expected – depending on the subject items – by the end of this year or in the year 2000.

Jean-Jacques Peters

