A SYSTEM ARCHITECTURE FOR DIGITAL FILM ARCHIVES USING JPEG 2000 AND MXF

A. Nowak[1], S. Fößel[1], L. Nunes[2], E. Santos[2]

[1]Fraunhofer IIS, Germany; [2]MOG Solutions, Portugal

ABSTRACT

Moving image film archives can store their original films in their original image and sound quality for hundreds of years without decay provided the temperature and humidity is controlled. However film is inconvenient as a universal access medium, and expected to be replaced even in the cinema, whereas digital versions are widely and increasingly accepted. This paper introduces a concept of a **digital film archive system** that uses JPEG 2000 together with MXF. A highly scalable two-tier architecture provides a platform for long-term preservation of digital movie data in the highest possible quality and for automated access and dissemination of the stored material. All these technologies are based on open standards and therefore provide a strong foundation for our concept where future accessibility is one of the main concerns.

The different and highly customisable parameters of the two encoding and storage tiers make it possible to store the digital film images together with audio and metadata in their full original quality and to simultaneously allow for easy and automated access in. The stored content can be automatically converted to all currently used professional and consumer formats for dissemination such as DCPs, other project able cinema packages, HD DVD, Blu-ray Disc, internet streaming etc. The concept addresses non-commercial archives as well as commercial organisations and studios.

INTRODUCTION

The demise of film in favour of digital cinema projection within a few years is widely predicted, although when that will occur is still uncertain.

With this change will come several new imperatives already recognized by film archives but equally relevant to the modern film industry:

1. In the medium term, but perhaps very soon, there needs to be some universal open access route for storing our digital versions of film, able to generate whatever version is required for access. The package should also be able to output lower quality versions for all other access purposes.

2. Archives already hold many digital versions of their film holdings and need to be able to access all these in a common parallel and browse-able manner too.

3. The opportunity to directly connect metadata to content should not be missed. In parallel metadata will become the only link with visual and aural characters of original **film** cinema, and this may only be retained as “historic technical metadata”, which hardly exists in either film archives or commercial collections today.

Today no solution for digital film archives exists that includes the streamlined handling of many different types of digitised moving images for access and preservation. All the digital solutions that are available from international vendors concentrate either on search and retrieval solely on the metadata layer, or they are tailored to suit the needs of TV production rather than film archives. All existing systems for film archives mainly intend to support their work with customised databases optimised for the task.

The concept introduced in this paper goes one step further. It aims to provide a system solution suitable for digital preservation. This is expressed by the two main goals of the system architecture:

a) The system provides a solution to store film image sequences, audio and any accompanying material in the optimum digital quality for preservation and conservation. This includes digitally produced material, e.g. the output of the Digital Intermediate post-production process, as well as scanned and digitised conventional film.
b) The system provides fast, easy and reliable access to archived digital material and digitally stored access copies of conventional film within the film archive. Access to archived material is possible in various dissemination formats, resolutions and quality levels according to the requirements of the user or customer. The conversion to the different formats is done automatically.
c) The system provides a route to input existing analogue and digital video and other access formats already held by an archive, into the same store, and with the same fast access mechanism. This will be a widely used route initially although we expect it will be replaced by better more optimum digital version subsequently or when possible.

The proposed system architecture is developed within the European Commission funded EDCine project with the involvement of well known film archive experts. It is based on well-established models and international standards that are widely recognised in the archives and the film industry. It is modelled from the user requirements gathered in the course of the project and additionally tries to leave enough room and flexibility for future extensions to take care of requirements that are not yet known, or may change. Also these requirements and the system details are expected to change as a reaction to new insight and experiences as the project tests the processes with real archive material.

ARCHIVING MOVING IMAGES IN A DIGITAL WORLD

Digital archive systems cannot and are not meant to supersede existing traditional film archives, but new technologies can help to exploit the content stored in a traditional archive efficiently. In parallel a solution is also needed to preserve digitally produced movies in an effective way.

Today, access to archived films is often restricted to access to physical elements in the premises of the archive and the user can only view a copy of a film at viewing tables in the archive, or in the archive's cinema. The access copies can be in a high quality on film, but more often they only exist as low quality analogue or digital video tapes, such as DVDs or MPEG files.

A digital film archive system can store high quality viewing copies and use network access, e.g. via the internet, to facilitate access without needing to travel to the archive and waiting for a video tape to be retrieved from the shelves.

In general and at least in Europe, film archives already achieve the longest possible film life which can be up to several hundred years. Unfortunately, this is not true for digitally produced movies and digital data in general. Here, two main problems can be identified:

- obsolescence of storage media and equipment; only a few years after the production of a system the manufacturer ceases to provide the hardware to read the data back into a computer system, and
- obsolescence of logical data representation formats; new software versions may include new file formats and the old file formats become difficult, sometimes in impossible to interpret and render after only a few years. This is especially a problem with proprietary formats that are not openly documented and are only supported by one or a few manufacturers.

Any digital film archive system has to provide solutions for these two crucial problems. The obsolescence of storage media and equipment can only be minimised by data migration at regular intervals; copying the old files to new media formats and systems. Degradation of storage media is usually a smaller problem compared to obsolescence and that too can be minimised by migration. Thus until the storage industry presents new solutions for long-term preservation of digital data any digital film archive system must provide tools to facilitate migration processes.

Obsolescence of data formats can be overcome through the use of open and well-documented standards only. This enables programmers, even in the far future, to develop software that can interpret and render the data correctly, and, of course the file format must not result in quality degradation of the stored image and sound data. It is one of the main goals of this project to present an approach that solves this issue.

THE CONCEPT AND SYSTEM ARCHITECTURE

The system architecture has to meet the requirements of a very wide range of different film archives and underlying business models. It must be able to ingest and store material originating from a wide variety of different formats and qualities. As the size of film archives is variable in terms of the amount of stored material as well as the numbers of users and customers, scalability and expandability are key features of the system architecture.

The archive system must support existing workflows that have been established in conventional film archives and in the film post-production industry, which have proved to be reliable and effective, and there should be a smooth transition into the digital world. Existing Digital Intermediate post-production workflows need to be integrated into the concept because their product is often a starting point for digital archival.

The processes in the digital archive need to be automated wherever this is possible to reduce cost and to improve efficiency, specifically format conversion, data migration and dissemination procedures, but can also include metadata creation.

To fit these needs we present a two-tier storage model that allows simultaneously for highest quality long-term preservation and easy access to the stored items. A modular implementation of all key stages ensures scalability and provides interfaces to existing systems and the future addition of new functions.

The architecture provides two packages to store image, audio and metadata: creating a **Master Archive Package** (MAP), that is mainly meant for long-term preservation, and an **Intermediate Archive Package** (IAP), to facilitate access to the stored items. In the MAP all data is stored in a lossless compression scheme using the highest justifiable spatial resolution and bit depth for each item. This produces huge files that are difficult to handle. The MAP will probably reside on an off-line medium such as digital data tape. The IAP is

created from the MAP in an automated conversion process. Resolution and bit depth are scaled to a common format for all items and visually lossless compression is used to reduce the data to more practical range. The IAP usually resides on hard disks to allow fast and easy access. It is used to create the dissemination packages that are delivered on order to the users in a format best suiting their needs. It is also used to create preview streams for browsing and searching the archive.

A facility to ingest scanned film or pre-existing access material directly into IAP will also be available, to make immediate use of access material already held by an archive.

An overview of the proposed architecture can be found in Figure 1.

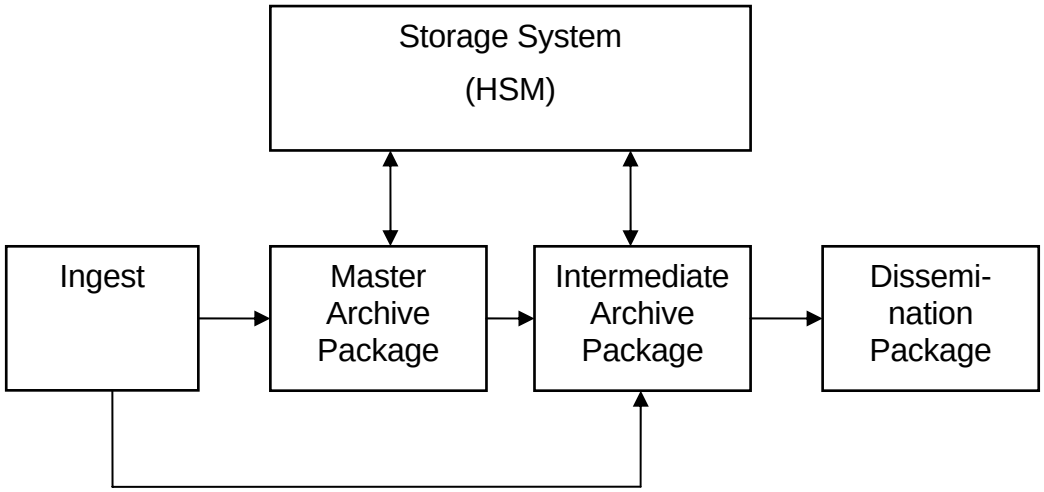

Figure 1 – Overview of the system architecture

The System Architecture

Generally, the following processes and activities can be identified:

1. Ingest
2. Cataloguing and metadata production
3. Long term conservation
4. Active preservation and restoration
5. Search and retrieval
6. Collection management
7. Access via analogue (currently theatrical screenings) and digital channels (currently mostly DVD, broadcasting and Internet streaming, and in the future D-Cinema)

Our system architecture has to provide functions to accomplish all these tasks. Therefore, we have chosen a flexible and scalable modular approach using a service oriented architecture (SOA). As described in (1.) this is a model for communication and data exchange between modules in an open and platform-independent manner that allows for future extensions and the integration of components from different manufacturers. The modules exchange data and commands through an XML interface over TCP/IP connections. The large amount of media data are exchanged via a shared file system. This approach allows us to build different user interfaces most suited for each task. Most parts will be accessible through a web-browser, especially functions that are needed by the users and customers, and for internal tasks there is also the potential to build special graphical user interfaces.

The Storage Format

Movies are not just sequences of images. They may be accompanied by audio data, metadata and sometimes other information such as texts or still images. The overall amount of data can be divided into two groups: essence and metadata. In many implementations these two are stored in different places. Essence data is stored in files in a file system while metadata is stored in a separate database. The “Reference Model for an Open Archival Information System (OAIS)” (2.) as standardised in ISO 14721:2003 describes a system for the archiving digital data that relies on an asset-based approach. This means that items can be seen as an asset consisting, in our case, of essence plus metadata that belong together and should be stored together. In this way the risk of breaking connection between essence and metadata is minimised, at the same time facilitating data migration or changes in data handling tools. The OAIS reference model is increasingly acknowledged by archives as the most appropriate universal and abstract solution for preservation of digital film data, and the EDCine Digital Film Archive System will use OAIS as a fundamental basis for its system architecture.

For each type of essence and metadata a format has to be found that meets the requirements of digital film archives. The main essence formats are audio data, textual information and image data, and in addition formats for metadata and the overall packaging are needed.

JPEG 2000 Image Coding

Compression of the image data is required to reduce the enormous amount of data to a better usable volume. For the presented architecture JPEG 2000 (3.) was chosen because it is an open standard that allows lossless as well as visually lossless compression and has the advantage of being a scalable representation of image data. It is a scalable, wavelet-based video-compression standard ratified as ISO 15444-1: 2004. Compression is performed on an intra-frame basis only. The use of wavelets produces a hierarchical code stream that can be scaled in terms of quality and resolution. This means that images can be encoded in the desired maximum resolution and quality. If a lower resolution or quality is needed for an application only a part of the JPEG 2000 code stream has to be read and decoded. Progression by resolution allows resolution scalability; a device does not have to decode the entire stream to reproduce a smaller resolution image. A handheld device, for instance, could decode only the portions of the code stream necessary for its small display, while the same stream could be fully decoded by a cinema decoder. Additionally progressions by quality, colour component and position are possible.

In a digital film archive system this feature can be used to easily extract lower resolution or preview streams from the Intermediate Archive Package for browsing and dissemination.

Audio Data

Historically, many different audio systems exist in the film archive world, but they mainly differ in principle in the number of simultaneous channels stored on the film and re-played. For storage in the archive audio signals are encoded in an uncompressed PCM format with the highest possible sampling rate and quantisation to retain as much information as possible for future restoration. All channels will be stored without alteration as they are on the source medium.

Textual Information

For some applications there is the need to store textual information with moving images that is conceptually not included in the associated metadata, for example subtitle information, text

images connected and synchronized to the moving images. This will be stored in a human-readable format, preferably in an XML representation where further structuring is needed.

Packaging using MXF

The Material Exchange Format MXF (4.) standardised by SMPTE and EBU is a file format intended for the interchange of audiovisual material and related metadata. It is a wrapper format that can contain several essence streams such as audio and video streams and can support several encoding and compression standards. We will use MXF to store the JPEG 2000 compressed image sequences together with accompanying audio and any other essence streams, and metadata.

Metadata Handling

Efficient and reliable handling of metadata is a crucial task in a digital archive. Metadata is stored in an object-based scheme within MXF files as descriptive and technical metadata directly associated with the essence. In addition, and new to archive metadata, mechanisms will be provided to store "historic" metadata that contains information about the origin and former processing steps an image or sound record has undergone.

Identification of each item is accomplished through the MXF Unique Material Identifier (UMID) that allows for worldwide unique identification of each item may provide references to external items (such as film reels that were scanned to produce digital item). A controlled dictionary can be used to provide keywords, especially useful for historic metadata that identifies film processing steps.

To facilitate search and browsing of the archived items the metadata from the MXF is automatically replicated and synchronised to a database. The database is used to store administrative data needed in the archive such as user profile, access control information, billing information, statistics etc.

The metadata model used will conform to IFLA's *Functional Requirements for Bibliographic Records (FRBR)* model (5.). This decision was taken because of the FRBR model's effectiveness in dealing with complex relationship issues between

- different works
- each work's multiple versions
- different objects derived from each version of the work

The actual implementation and interaction of this model with the present system concept using concrete metadata standards (as MXF DMS-1, SMPTE Metadata Dictionary, and the expected CEN metadata scheme for description of filmographic works etc.) will require further research.

INGEST

Several different routes are identified by which source material can find its way into the digital film archive, distinguished by the type and production history of the source material. Today, the most important routes are 1) from conventional analogue film sources, 2) from digital image sequences from a digital intermediate (DI) post-production procedure or digital restoration process of conventional film material, and 3) from pre-existing video versions. A special case is the ingest of digital material from inter-archival loans. In this case the starting point is not the same as for the other routes because the material has already undergone an archive ingest procedure at another archive. Thus, processing needs for this type of source should be minimal.

Images and audio data can be read in from various source formats. The images then are compressed and wrapped together with audio and metadata to form the Master Archive Package, or in the case of pre-existing video or where lower quality digital images have been produced, directly into the Intermediate Archive Package. Thus the Intermediate Archive Package can then either be directly created or may be derived from the Master Archive Package in a batch process. The Master Archive Package is then fed into the long term passive storage for preservation, if, or when one comes into existence. Optionally, the uncompressed source of the MAP can also be stored if this is required for any reason. The IAP is sent to the on-line distribution storage.

An alternative route exists in the case of digital cinema distribution. Usually for this process a set of files called DCDM (Digital Cinema Distribution Master) is created. This is then used to create a DCP (Digital Cinema Package). The DCDM can also be used to create the MAP and/or the IAP. In some cases an archive might only have access to a DCP. This is then usually directly converted to an Intermediate Archive Package. (Here the terms DCDM and DCP refer to the digital cinema concept as defined by DCI and currently with the SMPTE for evaluation as a US standard)

Many other routes are possible due to the modular design of the system.

STORAGE

The storage system is one of the most important parts in digital film archives. The two most relevant parameters are capacity and transfer speed to various parts of the overall storage. The capacity can quickly reach several petabytes with high demands for reliability. With these large amounts of data the issue of the time needed to access and transfer data quickly becomes crucial.

The underlying concept of the digital film archive system relies on the separation of 1) the content in terms of media data and metadata, and 2) the physical media on which the data is stored. Conventional film or video archives periodically copy deteriorating material onto new films or video tapes. Even, perhaps especially, with digital video this leads to quality loss because the data has to be decoded, decompressed, transferred and recompressed, possibly with a different compression system (and subject to managerial and finances pressures as well). If data and physical media are independent of one another the content can be migrated without any loss. Therefore the concrete media type on which the archived material resides becomes less important as long as all requirements in terms of capacity, data transfer rates and costs are met.

This concept also enables a film archive management the opportunity to consider the very long-term preservation of their holdings. This is a very large research field well beyond the scope of this project, but the EDCine digital film archive system is designed to work with all currently existing and future storage media types, and its modular flexible will directly benefit from new developments in the up-coming generation of hard discs, tape-based systems and optical media

The usual implementation of a storage system in this situation will make use of hierarchical storage management (HSM). This approach, widely used in information technologies, defines several layers of storage on different types of media with different characteristics regarding access time, transfer speed, media capacity and cost per terabyte. The EDCine system provides the means of moving data automatically from one layer to another depending on access frequency and other parameters.

According to HSM rules (e.g. a location in the hierarchy that corresponds to the access probabilities) a MAP will usually be stored on off-line media once the IAP is created, because that the MAP implies long handling times and more complex access. The IAP on the other

hand should be stored on on-line or near-online storage to allow fast and easy access, although IAPs may be transferred from on-line to near-online storage if they are seldom accessed.

ACCESS AND DISSEMINATION

A web-based application is used to allow access to all important functions of the archive, which includes an access control mechanism to give specific privileges and access rights to single users and groups, and the functions, tasks or content they can access.

The dissemination sub-system provides for search and retrieval of content to the users and customers and also produces the dissemination packages delivered to the users according to their specific needs, and privileges. Service modules would include database searching, streaming of preview clips, and automatic conversion of IAPs to create the dissemination packages and delivery. Multi-platform delivery will be permitted; a range of formats from low bit rate streaming of reduced resolution, through download of HDTV files for broadcast studio production, up to DCPs and other formats suitable for presentation in a digital cinema environment. This is achieved with a plug-in module architecture that provides automated transcoding and conversion services. The actual content delivery can take place over various channels.

The dissemination sub-system also comprises web-based user interfaces and mechanisms for ordering, shipping and billing.

SECURITY CONSIDERATIONS

In order to ensure future accessibility the storage system will usually not apply encryption to the stored data. Content security and protection is provided by physical control of media, servers and networks within the archive. Security measures are enforced at software level by the use of access control mechanisms that only grant access to the different stored items by authorised users and processes. For example, the dissemination sub-system is only allowed to retrieve a resolution layer from an IAP from the storage system to create a dissemination package and may not retrieve a MAP.

To ensure content security at the dissemination stage various DRM and encryption schemes may be employed, specific to each dissemination channel. As an example, for DCPs the Key Delivery Message (KDM) mechanism as proposed by the Digital Cinema Initiatives (DCI) (6.) can be used, if required, with AES encryption, whereas for preview and private (home) use OMA (7.) and ISMA (8.) tools may be used. Unencrypted data of the content in the archive is protected via access control and network security measures such as firewalls and gateways.

CONCLUSIONS

In a joint effort, involving the advice of archive professionals, a concept and architecture for a **digital film archive** has been developed that is flexible and scalable and will suit the needs of a wide range of different archives.

A remaining issue still waits for an answer: when film held in archives finally decays and film stocks for preservation is no longer available, the transfer of content, in the form of the MAP, to a digital version for long term preservation will be essential. At present the technological alternative to analogue photochemical film does not exist, although it is hoped the EDCine digital archive system may take one step towards finding an answer to part of this problem.

REFERENCES

1. http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=soa-rm
2. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, BLUE BOOK, January 2002, Consultative Committee for Space Data Systems, http://public.ccsds.org/publications/archive/650x0b1.pdf
3. http://www.jpeg.org/jpeg2000/index.html
4. SMPTE 377M and others, http://www.smpte.org
5. FRBR: Functional Requirements for Bibliographic Records, IFLA, Frankfurt am Main, 1999, http://www.ifla.org/VII/s13/wgfrbr/finalreport.htm
6. http://www.dcimovies.com/
7. http://www.openmobilealliance.org/
8. http://www.isma.tv

ACKNOWLEDGEMENTS

The EDCine– Enhanced Digital Cinema project is funded by the European Commission within the 6th Framework Programme FP6/2004/IST/4.1, contract no. 038454 EDCine.