DCP/A: Discussion of an Archival Digital Cinema Package for AV-Media

Peter Fornaro, University of Basel, Digital Humanities Lab , and David Gubler, Fachlabor Gubler AG (Switzerland)

Abstract

In this presentation we will discuss the motivation for a digital archival master format for motion pictures as well as some technical key questions along the road towards a possible standardization. Due to the large volume of uncompressed high-resolution motion picture data, we will discuss the quality, advantages and drawbacks of data compression algorithms. In addition we focus on issues like the ability of data formats to encapsulate preservation information metadata and the long-term stability of format definitions.

In the analog era of motion pictures normally an analog "projection copy" was handed over for archival purposes. Until recently, the Digital Cinema Package (DCP)[1] has been the digital copy used for projection. With the disappearance of analog film copies and the prevalence of countless digital media dissemination platforms we discuss whether the dissemination format, formerly known as Digital Cinema Package DCP, is appropriate to be used as archival information package (AIP). Further on we will discuss which features are needed to define an ideal data-format like "DCP/A" (in analogy to PDF/A).

Motivation

Motion pictures account for a constitutive share of cultural heritage as well as for contemporary multimedia output in social, scientific and economic ambits. However, after 150 years of tradition, not only analog film stock is endangered by ultimate decay: Also today's digital records are in urgent need of strategies and technological solutions for their conversion, long term management and distribution [2]. Any records management strategy requires archival concepts addressing "creation", "management" and "dissemination"[3]. Unfortunately, we find that for motion pictures and audio recordings these concepts are not yet well applied in practice nor are there commonly accepted best practices.

Archiving has fundamentally changed in the transition from analog to digital. First of all, in the digital domain any data appears physically the same, whether we speak about text files, photographs, digital audio or video content. All those objects are a sequence of bits, enriched by metadata to explain e.g. technical or contextual details. Moreover, digital files are much more fragile regarding their lifetime. Technological changes in storage systems make digital archiving a very active process. On the other hand if analog film is stored correctly one can expect that these analog artifacts still exist and will be readable after some decades. Maybe they face some kind of degradation but they are still there. In the digital domain this is completely different. If digital files are not taken care of continuously it has to be expected that the files won't be accessible in future or the file format can't be understood because of any or multiple technological obstacles. In general digital information is endangered if:

- the data carrier decays
- hardware to read and access data carriers changes with time and gets incompatible with earlier generations
- file formats become obsolete or they feature technologies that are contrary to the needs of digital archiving

All those points are independent from the actual content of the files. In particular, it doesn't matter if we speak about images, motion picture or text content. Hardware will have to be migrated after a short period of time but the file format should be as stable as possible to prevent file transformation. Therefore it is very important to choose a format definition that meets both requirements: quality and compatibility to archival needs.

Our motivation is to look to other fields of digital archiving to take advantage of already existing and successful approaches designed for and established in the digital domain. We will focus on data volume, sustainability of formats and metadata standards already proved by archives.

Problem

Digital involves huge data volumes and considerably high demands on image quality. Digital motion picture consist of typically 24 images per second, resulting in 86'400 images per hour. Depending on the spatial resolution and the quantization depth this requires terabytes of data for each hour of content. This immense data volume causes weighty costs even if prices of storage devices like hard-drives are declining towards zero. Typical image file formats for digital motion picture are DPX with a quantization depth of 10 bits logarithmic or TIFF with 16 bits linear data per channel. Both file formats are mainly used as digital intermediate to have access to the best possible image quality. Besides intermediate material "digital release prints" are stored as compressed files called DCP based on JPEG2000 encapsulated in an MFX container [4] to reduce transfer time and bandwidth requirements in cinemas as well as to ensure economic restrictions such as limitation of time and location of publication of the given media content.

In an analog workflow the final release print was more or less of the same technology as the intermediate material; both are typically state of the art chromogenic multilayer films. Such high quality release prints often were used as an archival master copy. However, the delivery of reels of film to public or governmental "legal depots" has come to an abrupt stop, as cinemas have become digital and analog film post-production labs have disappeared.

It would be straightforward to say a digital release copy – actually a DCP – could be used as an archival master file as well. Unfortunately, the DCP standard consists of some features that are not necessarily compatible with archival best practices: First of all, the fact that a DCP contains compressed image data has to be discussed more in detail in order to judge the advantages and disadvantages of such methods to reduce data volume.

Additionally, structural meta-information is necessary to ensure the storage of such content in proper archival manner. Last but not least DCPs allow encryption of their content that might result in a complete loss if the necessary if the necessary key became unavailable. The management of archival data is endangered by the lack of accepted standards and therefore impracticable amount of varying file formats, metadata schemes, proprietary encryption, etc. We face today the lack of established guidelines for digital archiving motion picture records. Technological progress and the loss of established players such as film post production labs or hardware providers are hindering the development of industry standards: Existing file formats are uniquely foreseen for digital dissemination but not necessarily to meet archival needs, nor do these formats meet archivists' crucial criteria of good documentation and the use of open standards.

Approach

In this publication we follow a twofolded approach: (1) Our archival methodology follows the reference model for an open archival information system, the established OAIS-model, DIN ISO 14721. Furthermore, respecting the generic methodological principles of the OAIS-model, we discuss (2) practical specifications that address the entire lifespan of digital records from creation to management and dissemination - specifically for motion picture records.

DCP/A

From a technological point of view digital archiving is not something that is, in most cases, strongly related to a specific type of data. Nearly all users of digital data have the same problems and therefore it can help to see what other communities did to solve typical archival obstacles. Taking a look at the PDF we notice a widely established file format that became the technical basis of the series of today's PDF/A [5] standards. The basic PDF Standard features interesting functionality like for example JavaScript and executable file launches or even encryption of content. However, some of the core functionality of PDF is contrary to the needs for digital preservation of electronic documents. PDF-files can become unreadable in the future or their content even might change depending on the time of access. Both are inacceptable for any archival use. Therefore the ISO standardized version PDF/A has been created. PDF/A differs from PDF by omitting features that are not suited for long-term archiving. In addition the use of standards-based metadata is mandated in PDF/A's. Therefore we propose a standard DCP/A, which follows the path of the well-accepted PDF/A. In a first step this approach requires a closer look at some of the properties and features of DCPs.

Data compression

How to cope with compression? In order to answer this question we have to discuss what data compression really is – in the context of images. Data compression is a method to reduce redundancy of data and/or to eliminate image content with no or little visual relevance for a standard observer. The first is called lossless, the latter lossy compression. JPEG 2000 can be used either lossy or lossless, whereas in a digital motion picture DCP a lossy compression is applied. The data rate is set to a maximum of 250 Mbit/s. In the context of archives lossy compression is not something that necessarily is a door opener. Anyway, the huge amount of data caused by uncompressed motion picture requires a closer look to data compression and some facts need to be clarified to understand what information an image carries and what limitations an analog workflow has. A digital 4K TIFF-image (an image with a width of 4096 pixels) has a data volume of about 72 MBytes depending on the bit depth and the image height chosen. If such a 4K uncompressed analog image is recorded to film by a laser film recorder one could expect that effectively 72 MBytes of information have been written to and stored on film. Recording image data to film is a standard procedure for the production of release copies and archival masters. If such an image on film is scanned with a resolution of 4K with a film scanner we gain a reproduced TIFF-Image with the same 72 MBytes of data volume. At first it seems like we have transferred the full information from the digital domain to film and back. In reality this is not true because film has a limited capability to store information. In early publications [6] it has been shown, that film can only store around 200 Kbytes per cm^2, resulting in a data volume of about 1 MByte per single image. This means that the original data volume is reduced by a factor of ~ 72 which has to be regarded as 'lossy data compression'. Of course the original 72 MByte image data do have a lot of redundancy but anyway analog film is the bottleneck in this scenario. The 72 MByte of the rescanned image again has some redundancy and besides the actual information, the scan is as well a capture of film properties, mainly grain. In other words: The typical procedure of recording digital images on film for archival purposes reduces the effective image quality and adds unwanted distortion. In addition to that the transformation to film and back to the digital domain is depends greatly on the machinery associated with the process. Noise and technical properties of the recorder and scanner have as well an influence on the image quality and they are not reproducible. We have observed that those properties of the image capture device create image artifacts that are at least as strong as artifacts caused by lossy JPEG2000 data compression. In a simple experiment we generated uncompressed TIFF-Images and stored them as JPEG2000 derivatives. The difference – actually the SNR – between the uncompressed and the compressed image is a quantitative measure for the difference of the two images. In a second step we captured two images with the same hardware and stored them as uncompressed TIFFs. A comparison of those two 'identical' images has been expressed as well as signal to noise ratio. The interesting point is, that the variability of the capture device makes a bigger difference than strong digital JPEG2000 compression. Two important statements are a) film has limited capabilities – data density – to store information and b) the artifacts introduced by the machinery to write and read film are not negligible. Therefore, a well-defined image data compression algorithm like JPEG2000 used in DCPs must be regarded as at least as good as the method of writing images to film. Let's have a closer look at DCPs and their capabilities as archival master format.

Archival Information Package (AIP)

AIP, as defined in the OAIS reference model, is an information package that is used to transmit archival objects into a digital archival system, store the objects within the system, and transmit objects from the system. An AIP contains both metadata that describes the structure and content of an archived essence and the actual essence itself. It consists of multiple data files that hold either a logically or physically packaged entity. The implementation of AIP can vary from archive to archive; it specifies, however, a container that contains all the necessary information to allow long-term preservation and access to archival holdings. The metadata model of OAIS is based on METS[7] specifications.

From a physical point of view the AIP contains three parts; metadata, essence and packaging information, which all consists of one or more files. Packaging information can be thought as wrapper information and it encapsulates metadata and essence components.

Deficiencies of the so far DCP standard and the cumbersome road of archival standardization

A Digital Cinema Package (DCP) is a collection of digital files used to store and convey Digital Cinema (DC) audio, image, and data streams. The term has been defined by Digital Cinema Initiatives [8], LLC in their recommendations for packaging of DC contents. General practice adopts a file structure that is organized into a number of usually multi-gigabyte size Material eXchange Format (4) files, which are separately used to store audio and video streams, and auxiliary index files in XML format.

One has to note, however, that this DCP standard evolves from the world of dissemination and it contains therefore a number of economically driven features such as encryption, limitations of the number or the time frame of play-backs, etc. However, these features are not at all compliant with archival needs! Rather will we recall a few cardinal requirements of archival formats: An archival format shall –among other criteria- be well documented, well distributed, shall not include any proprietary features and shall most certainly be free of any restriction towards the extraction (future readability) of its contents.

Let’s take a look at the reasoning discussions for establishing PDF/A in the year 2002: In order to establish an archival standard the PDF format has been chosen as basis for an, at that time yet to be established, archival format. Arguments in favor for PDF included structured objects (vector graphics, text, bitmaps), efficient compression, embedding of metadata and the independence of decoding environment (neither specific software nor hardware shall be required in order to render and represent the content). Thereupon the basis of PDF had to be cleared of some features (such as references into external files). At the same time the PDF/A standard has been supplemented with features that are mandatory, some are recommended, some listed restrictions. Also the PDF/A standard evolves and now includes various versions (for instance the barrier-free characteristics of the PDF/A-1a according to Paragraph 508 of the US Rehabilitation Act) and develops through an iterative process involving its community of stakeholders.

Sustainability

For archival purposes it is very important that a file format is chosen that is open and widely used. Both are true for DCPs. Any digital projection is today shown from a DCP file, which makes it a widely used file format. Besides that the format is well defined by the Digital Cinema Initiative (DCI). However there is a major drawback of most DCPs, there are encrypted. Encrypted DCP’s only can be played on a specific Digital Cinema server for a predetermined time. Therefore, a Key Delivery Message (KDM) is sent to the projection site to unlock the DCP for screening purposes. Of course it is not a necessity to have a DCP encrypted, it is just a method used by most studios to prevent digital motion pictures from being stolen and duplicated. For archival purposes encryption is not a good solution, because access to the files is endangered if the key is lost or corrupted. Therefore it is important to generate either DCPs need to be generated without encryption for archival purposes or archives need to have a KDM to decrypt DCPs in their own environment [9]. Of course this assumes that such a solution is operated in a secure infrastructure where unauthorized access is not possible.

Results

If we look at the image quality of JPEG2000 compressed images, it is straightforward that the measurable and visual quality is more reduced the higher the compression ratio is. The effects depend on the task, the image and the compression method [10]. However, measurements and the visual impression of the artifacts are judged in the digital domain only, e.g. on a display. In the case of archiving motion picture the resulting image is on film and therefore the compressed image has to be compared with the one that has passed the chain consisting of recorder, analog film material and scanner. In this case the distortions of the film have to be taken into account, which has naturally to results in a worse image quality from a quantitative point of view. From a visual point of view, the pass trough of analog film might even be judged as ‘nicer’ or more ‘film like’ looking. Such statements are the consequence of the adaption of a standard observer to conventional analog film projections. From a strict technical point of view such subjective visual habits should not be taken into account if only the visual effects of the compression shall be evaluated.

The sustainability of DCPs is strongly correlated to the availability of adequate KDMs, which is a political issue that needs an interaction between the parties. From the technological side the sustainability is given if e.g. the evaluation of the Library of Congress is looked at [11].

Another issue is the flexibility of DCPs. They e.g. only have limited capabilities to store different image dimensions. The aspect ratio can vary between ~ 1.78:1 (HDTV) and ~1.9:1 (full DCP container) any other ratio is impossible. Other image dimensions need to be scaled [12]. Of course it would be ideal if a potentially defined DCP/A would have more possibilities to store common aspect ratios and frame-rates, e.g. 4:3 or 35 mm anamorphic formats like 2.39:1 and e.g. a high flexibility in the number of frame rates.

Making use of outside industry experiences

As mentioned above, the AIP contains three parts; metadata, essence and packaging information. In the context of archival specifications the “essence” of motion picture or audio recordings is certainly based on specific industry familiarity and knowledge. However, towards the establishment of an archival standard the issue of metadata standards can build up on well-established and also field-tested solutions [13]. Within the author’s field-tested experience with large volume audio and film archiving projects we mention for instance the implementation of “metadata crosswalks” such as the Matterhorn METS Profile, registered at the Library of Congress on November 29, 2012. The Matterhorn METS Profile describes the core of the digital object model used to support digital archiving. This may be the first profile that describes the use of EAD within METS in any detail. As for the content or "essence" of the features the authors used a simple folder structure (occasionally combined with a .ZIP file), packed with JPG2000 files for the images, an .AIF file for audio and a METS PREMIS XML schema for all the metadata. As an alternative to the (traditionally with JPG2000) compressed image content we have also applied a H.264 codec for the dissemination packages (DIP), postulating non-proprietary decoding and precise specifications of the H.264 parameters. Given these assumptions, this would then be a very efficient combination of a valid archival file format (AIP) that simultaneously meets the requirements of fast (streaming) dissemination (DIP). The PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability. Above outlined real life projects derive a practical, reduced set of metadata and are examples of complete Submission Information Packages (SIPs), consisting of file containers, file formats, accompanying XML-files consisting of preservation metadata (PREMIS, METS, others) as well as of dissemination metadata (DCP structured SIPs).

Conclusions

DCPs are the standard file-format in digital motion picture dissemination; they have replaced analog film copies nearly completely within the last couple of years. For many film productions and motion picture archives it is seductively convenient to choose DCPs as an archival master.

However, it is absolutely urgent to discuss and establish archival standards for motion picture and audio files. Some issues such as image quality and descriptive metadata will require industry specific discussions. Fortunately, field-tested and reliable archival frameworks and patterns do exist and they are to be taken into consideration.

Furthermore, the cinema archives are called upon to gather momentum and tackle the cumbersome but necessary road towards an Archival Digital Cinema Package.

References

[1] The Digital Cinema System Specifications http://dcimovies.com/specification/DCI_DCSS_v12_with_errata_2012-1010.pdf

[2] The Digital Dilemma, THE SCIENCE AND TECHNOLOGY COUNCIL OF THE ACADEMY OF MOTION PICTURE ARTS AND SCIENCES, © 2007 Academy of Motion Picture Arts and Sciences (A.M.P.A.S)

[3] Reference Model for an Open Archival Information System (OAIS), Draft Recommended Standard, CCSDS 650.0-P-1.1 (Pink Book) Issue 1.1 August 2009, http://public.ccsds.org/publications/archive/650x0p11.pdf

[4] MXF, Material Exchange Format: The Society of Motion Picture and Television Engineers, http://www.edupdf.org/4316/material-exchange-format/

[5] ISO 19005-1:2005 – Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)

[6] SHAW (ED.), R. 1976. Selected Readings in Image Evaluation. Society of Photographic Scientists and Engineers.

[7] METS, Metadata Encoding & Transmission Standard, Library of Congress, Feb. 6th 2014, http://www.loc.gov/standards/mets/

[8] The Digital Cinema Initiatives, LLC, http://www.dcimovies.com/

[9] NOWAK Arne, Digital Cinema Technologies from the Archive's Perspective, Dep. Moving Picture Technologies, Fraunhofer Institute for Integrated Circuits, Germany, AMIA Tech Review Volume 2, October 2010

[10] BUCKLEY Robert, Using Lossy JPEG 2000 Compression For Archival Master Files, Prepared for the Library of Congress Office of Strategic Initiatives, Version 1.1 March 12, 2013

[11] Library of Congress, Sustainability of Digital Formats, Digital Cinema Initiative Distribution Package (DCP), Version 1.0, Last significant FDD update: 2011-12-27

[12] International Federation of Film Archives, FIAF Technical Commission, Digitisation for Film Archives– Assorted Complications, FIAF Technical Commission 2012

[13] Reorganisation of the Nestlé Historcal Archives, Switzerland, 2008-2015; AV Media Conversion for Syngenta Crop Protection, Switzerland, 2010 – 2015

Author Biography

Peter Fornaro is deputy head of the Digital Humanities Lab of the University of Basel. He is doing research in the field of digital archiving, imaging and computational photography. Besides his research he is consulting companies, archives and museums.

David Gubler is CEO of Fachlabor Gubler AG. He has rich experience in the field of digital records management and OAIS implementations. He is working as a consultant for many archives and museums.