

DEVELOPING AN MXF AUDIOVISUAL PRESERVATION FILE WRAPPER SPECIFICATION IN THE FEDERAL AGENCIES DIGITIZATION GUIDELINES INITIATIVE

Carl Fleischhauer, Library of Congress, USA

This paper is based on a presentation made at the IASA/AMIA 2010 joint conference in Philadelphia, November 4, 2010.

Introduction

This paper describes a project to develop a file-wrapper specification for the archiving and preservation of audiovisual content. The specification under development is an implementation of the Material eXchange Format (MXF), a standard from the Society of Motion Picture and Television Engineers (SMPTE). The initial phases of this effort have been organized by the Audio-Visual Working Group of the U.S. Federal Agencies Digitization Guidelines Initiative.

MXF can contain various encoded moving image bitstreams, e.g., MPEG-2, JPEG 2000 frame images, uncompressed video streams, and many more, as well as soundtracks, time-code, closed captioning, metadata, and more. This wrapper function may be compared to the TIFF and WAVE file formats, both of which may contain a number of different representations of raster image data (TIFF) or sound waveform data (WAVE). The virtues of MXF include standardization by SMPTE, a robust structure for the containment of various related essences, and a clear mission to support professionally produced moving image content. It is also the case that the format is complex and that its adoption within the industry has been slow, albeit steady.⁵

The Federal Agencies Digitization Guidelines Initiative was launched in 2007 under the auspices of the National Digital Information Infrastructure and Preservation Program (NDIIPP) at the Library of Congress.⁶ The overall initiative is a collaborative effort with participation from more than one dozen U.S. government agencies, including the National Archives and Records Administration, the National Gallery of Art, the Voice of America, the National Library of Medicine, the Smithsonian Institution, and several others. It also has active participation from the Library of Congress Audio-Visual Conservation Center at the Packard Campus in Culpeper, Virginia.

Current activities and planning

The Working Group was motivated to develop an MXF archiving and preservation implementation because several member agencies face a pressing need to reformat videotapes. These agencies have extensive holdings of obsolescent magnetic recordings and wish to transfer them to a file-based format. At this time, three members of the Working Group are doing some digital reformatting of video. They have purchased SAMMA devices, a product of the Front Porch Digital company.⁷ The Library of Congress has done the most work thus far, while the U.S. National Archives and Records Administration and the Smithsonian Institution are starting to carry out projects of their own.

5 For more information see, the background paper here:

http://www.digitizationguidelines.gov/guidelines/FADGI-AV_AppSpecProj_Bkgd_101007.pdf

6 The initiative's goal is to develop guidelines that are comparable from agency to agency, for the sake of uniformity and to make it easier for the vendors who provide equipment and services. The main emphasis is digitization, i.e., the conversion of analog originals into digital form. There are two working groups. The first is concerned with the still images that result from the scanning of books, photographs, manuscripts, maps, two-dimensional art works, and other documents. The second is concerned with audio-visual materials, i.e., sound and video recordings and motion picture film. Both groups have a secondary interest in the preservation of born digital content. See <http://www.digitizationguidelines.gov/>.

7 <http://www.fpdigital.com/Solutions/Migrate/>

SAMMA devices can be set up for a number of different output formats. The Library is using SAMMA's best-known setup in a workflow that produces a stream of video-frame images, each encoded in lossless JPEG 2000. This picture data, together with the soundtrack and timecode, is wrapped in MXF. Files in this format serve as archival masters for preservation in the moving image collections at the Packard Campus. File sizes for standard definition video run from 25 to 50 gigabytes per hour, depending on variables like bit depth. The Library prefers 10-bit-deep sampling and thus is creating files of the larger size.

At the same time, others in the Working Group — notably the U.S. National Archives — are interested in essences that consist of uncompressed video streams. In this, they echo specifications in use at Stanford and Rutgers universities, as well as at the BBC.⁸ File sizes for standard definition video run from 75 to 100 gigabytes per hour. Each member of the uncompressed trio has selected a different wrapper: QuickTime at Stanford, AVI at Rutgers, and MXF at the BBC. To some degree, the two universities' QuickTime and AVI approaches can be seen as provisional, i.e., as a practical thing to do now, while better ideas and technology mature.

Although the Working Group's current concern is focused on the reformatting of old videotapes, we also wish to develop an implementation of MXF that is extensible in at least three ways. First, there is a desire to embrace the output of film scanning activities. Second, one agency has expressed interest in using the MXF format to wrap "audio only" essences that result from the reformatting of sound recordings. And third is the matter of born digital video.

Regarding born digital video, the Working Group has frequently heard from organizations about their growing collections. Some of these are "non-memory" operating agencies, e.g., the National Oceanic and Atmospheric Administration (NOAA), where marine biologists and other scientists are producing extensive video footage. Another example is the Voice of America, where the staff wish to retain current broadcast productions for future repurposing. Meanwhile, files from agencies like these are destined for future handoff to the U.S. National Archives or the Library of Congress for permanent custody.

The *native encodings* for some born digital files are promising for sustainability for at least a few years, before format obsolescence reaches the point where transcoding is necessary. Examples include MPEG-2 and file-form DV. Other born digital files are in formats that may not be equally sustainable and these will require transcoding in the near term. We would like our MXF implementation to be capable of wrapping born digital encodings. For the sustainable native encodings, this action would support near- or medium-term preservation. It is worth noting that some agencies are required to retain even the less sustainable native encodings for many years; for example, copyright legal considerations may require that an item is kept "as submitted". All of these factors have made us seek an MXF wrapper implementation that can be extended to embrace born digital essences.

Beyond all of this, members of the Working Group have identified other content elements that would be well served by an extensible format, ranging from digital cinema to film-strips-with-soundtracks. Some members also see the archiving and preservation wrapper as a useful way to encapsulate items that are associated with the main content element. Examples of associated materials include such things as documents found in the original container along with the videotape, printed matter associated with a sound or moving image publication, transcriptions of oral history recordings, and detailed metadata of special "local" interest.

8 The Rutgers specification URL is: http://rucore.libraries.rutgers.edu/collab/ref/dos_awwg_video_obj_standard.pdf. Information about Stanford's reformatting is here: <http://lib.stanford.edu/stanford-media-preservation-lab/moving-image-digitization>. A white paper containing a 2007 description of the BBC approach is at this URL: www.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP155.pdf.

About MXF, application specifications, and the Advanced Media Workflow Association (AMWA)

MXF is seeing increasing adoption in broadcasting and it is central to the digital cinema specification developed in Hollywood for theatrical distribution. These industries include the big customers for whom tools are built, and their broad-based adoption of MXF makes the standard attractive to us. It is also the case that SMPTE is the most important standards organization for professional broadcasters and movie-makers.

MXF is a broad-spectrum standard that features many options for packaging, metadata embedding, and essence encoding. The successful implementation of an MXF approach for any given application will be enhanced if users define a set of constraints. Well-defined constraints support the development of tools to validate files and encourage multiple vendors to provide conforming equipment. These actions will increase the degree to which practices are standardized and promote adoption. Those two outcomes will in turn increase interoperability, the feasibility of content exchange, and long-term, preservation-oriented data management.

For users of the MXF standard, formal constraint statements are called *application specifications*. These can be compared to JPEG 2000 profiles or to the profiles and levels that characterize MPEG video content, e.g., the MPEG-2 *Main Profile @ Main Level* (MP@ML). Like the profiles, each application specification is tailored to a particular *application*. The incubation of MXF Application Specifications is the special province of the Advanced Media Workflow Association⁹ (AMWA), an organization that provides a meeting ground for professional moving-image users and vendors. The Working Group is working with AMWA in the development of the application specification.

One of the virtues of working with AMWA is the inclusion of both users and vendors in the discussion. The AMWA will form a special technical committee to finalize the specification and we anticipate that the committee's membership will include archivists from memory institutions, archivists from broadcast and motion picture production organizations, and representatives of the companies that manufacture the technology used by broadcasters and archives. We believe that this mix of participants will increase the likelihood that multiple vendors will build systems that meet the specification, which in turn will increase the likelihood that truly interoperable content will be archived for the future.

What might you find in an Application Specification?

At a high level, an application specification will cover the following:

- Aspects of the underlying MXF structure, e.g., which *operational patterns* are permitted, the use of *partitioning*, and the handling of *embedded metadata*, the use of *clip wrapping* and/or *frame wrapping*
- Picture — the permitted essence schemes (encodings) and other elements
- Sound — the permitted essence schemes and other elements
- The handling of timecodes
- Closed Captions and other VBI — about the elements in the Vertical Blanking Interval of the source signal and how they are to be retained in the digital copy
- Associated content elements — about the approach to be used to wrap in associated items like still images, documents, and texts
- Inclusion of file-integrity *hash data* (“checksums”) to be used when monitoring digital content over time.

9 <http://www.amwa.tv/>

The 24-page application specification AS-03, published by the AMWA in 2010, provides a good sense of the structure and level of detail in this type of document.¹⁰ AS-03 is “a vendor-neutral subset of the MXF file format to use for delivery of finished programming from program producers and program distributors to broadcast stations”.

About JPEG 2000 and our proposed specification

JPEG 2000, like MXF, is a broad-spectrum standard with many options. It was developed by the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC), and the family of standards documents carries the general identifier ISO/IEC 15444.¹¹ At last count, twelve parts have been published. The JPEG 2000 compression approach employs what is called the wavelet transform. When using JPEG 2000, one notable option is whether this transform is applied in an irreversible manner — resulting in lossy compression — or in a reversible manner — producing lossless compression. For our preservation-oriented application, the most desirable JPEG 2000 profiles are those that feature the reversible transform.

Various members of the moving image community have been developing digital cinema and broadcast profiles for JPEG 2000. The most recent publication is amendment 3 to part one of the ISO/IEC 15444 standard, titled *Profiles for Broadcast Applications* (ISO/IEC 15444-1:2004/Amd 3:2010). Two of the seven profiles in amendment 3 feature the reversible wavelet transform, i.e., lossless compression. We plan to reference this standard in our MXF application specification.

In the course of our exploration of the intersection of JPEG 2000 and MXF specifications, we have encountered one technical matter that is not fully resolved by the various existing documents: the handling and labeling of interlaced video. Experts in the field have said that there are a number of different ways to encode and label interlaced picture data in this context. At a high level, there is the question of whether the image stream consists of a series of *fields* or *frames*. At a finer level, there is the matter of how to indicate which field is dominant and how or when to employ what the MXF standard calls *frame-wrapping* as compared to *clip-wrapping*. If different approaches are adopted by different production systems, and if the labeling is inadequate, there is a risk that files will not interoperate, i.e., that a file produced on system A will not play back on system B.

How should interlaced image frames or fields be wrapped and labeled? What body should deliberate this question? One technical expert in the ISO/IEC JPEG 2000 standards community noted that this issue is not specifically about picture encoding: “It has to do with how you put the stream in a file,” he said, “and that makes it an MXF problem”, adding, “it will not be part of our work as we continue to develop broadcast profiles”.¹²

Our Working Group agrees and we were pleased to learn of recent SMPTE plans to specify an approach (or approaches) for interlaced picture data in a to-be-drafted amendment to SMPTE ST 0422-2006, *Material Exchange Format - Mapping JPEG 2000 Codestreams into the MXF Generic Container*.¹³ This standard was originally drafted to support the digital cinema specification. Since digital cinema consists of progressively scanned images, ST 0422 does not include recommendations for the handling of interlaced picture data. For the time being, we plan to include our own wording on interlacing in successive drafts of our MXF application specification even as we track the revision of SMPTE ST 0422.

10 http://www.amwa.tv/downloads/specifications/AMWA-AS-03-Delivery-Spec-1_0.pdf

11 Listings of all ISO standards may be found here: <http://www.iso.org/iso/catalogue.htm>.

12 Personal communication.

13 <http://store.smpete.org/product-p/st%200422-2006.htm>



Ancillary data

It is worth noting that there is a similar wait-and-see situation regarding closed captioning and other ancillary data and, by extension, motion picture subtitles and other forms of adjunct content. On the video side, there are number of engineering nuances but roughly speaking this is about the data found in analog video signals in the vertical blanking interval (VBI) and about the similar data in the digital realm referred to as VANC. There is also horizontal ancillary data or HANC. Some of this pertains to North American standards; there are also multiple embedded-data specifications in play in other parts of the world, including the European Broadcast Union's Subtitling Data Exchange Format (EBU STL).

Why worry about retaining this embedded data? We believe that it includes information that is needed to properly understand and manage video content objects for the long term. In some cases, the embedded data may be construed to be an essential part of the original item that must be migrated forward in order to create an authentic and complete copy. In some cases, the embedded data contains information (closed captions or subtitles, other descriptive information) that — once ingested and indexed into a search system — will support researchers who seek to discover relevant materials. The embedded data is also likely to contain technical information that will support the management of the item and may also shed light on production-method or provenance topics of interest to researchers.

Our exploration has highlighted some options to consider, which I will describe in simplified form. The first option embraces *current practices* and trades on the use of what are called *ANC packets* in SMPTE standards.¹⁴ The second option pertains to an *emerging practice* that should accommodate both US and European formats. This option would employ the structures being standardized in SMPTE-TT Timed Text standards (the ST 2052 family) being published during 2010 and 2011.¹⁵ Meanwhile, there is a European Broadcast Union project to standardize the carriage of EBU STL inside MXF files. Finally, there is a current practice among broadcasters to create sidecar files (separate data files, not embedded in the main MXF file) that contain VBI and related data. This may be practical for the near term but we would prefer to see the adoption of a standardized, embedded-data approach.

Metadata

Our federal agencies discussions thus far have highlighted a dual approach to the embedding of metadata in MXF archiving and preservation files. On the one hand, we have talked about a minimal header (or header-like) element. In spirit, this would be akin to the Broadcast Extension (bext) chunk in the Broadcast WAVE file standardized by the European Broadcast Union. In practice, we would like to see something less constrained, with room for a bit more data, not the least of which would be multiple tagged identifiers. The data in the header-like element would be basic, consisting of metadata needed to identify and manage the content object as an object, e.g., in a preservation storage system.

On the other hand, to support more complete representations of descriptive, administrative, and technical metadata, we are discussing finding places “deeper in the file”. Oliver Morgan, the Working Group's expert consultant, has called our attention to what are called *generic stream partitions*. These elements within MXF are standardized in SMPTE ST 410-2008, titled *Material Exchange Format - Generic Stream Partition*. These partitions were established for a variety of

14 Engineering wording for the first option may be found in the Section 5.1.7 of AS-03, op.cit. "If present, CEA 608 line 21 (CC and XDS) data shall be carried in a SMPTE ST 334-1:2007-and-ST 334-2:2007-compliant ANC packet within a SMPTE ST 436:2006-compliant VBI/ANC GC Element, using 8 bit encoding. If present, CEA 708B DTV captioning data shall be carried in a SMPTE 334-1:2007-and-ST 334-2:2007-compliant ANC packet within a SMPTE ST 436:2006-compliant VBI/ANC GC Element, using 8 bit encoding. Caption language shall be specified using AMWA AS-04."

15 The introductory standard is SMPTE 2052-0:2010: <http://www.smpte.org/standards/st2052-0-2010.pdf>

applications, one of which was to contain various classes of data streams, such as extensive blocks of “metadata that cannot suitably be stored in the Header Metadata (e.g., specialized preservation metadata)”.¹⁶ A corollary Recommended Practice document from SMPTE is RP 2057-2011, titled *Text-Based Metadata Carriage in MXF*, which “defines how to carry text-based metadata with a specified text MIME type encoded using either Unicode UTF-8 or UTF-16 character encoding (such as XML) in a MXF file”.¹⁷

The Working Group is interested in technical metadata, the moving image equivalents to a pair of standards from the Audio Engineering Society: (1) AES57 (forthcoming; the draft form was labeled AES X098B), titled *AES standard for audio metadata - Audio object structures for preservation and restoration* and (2) AES-X098C (still only in draft), titled *Administrative metadata for audio objects - Process history schema*. The former provides a description of a given file’s technical characteristics — not unlike the instantiation elements in PBCore¹⁸ — while the latter offers a description of the process that created the file, what is sometimes called *digital provenance*. In the federal agencies effort, there is a subgroup devoted to technical metadata and their work is still under development.

The federal agencies Working Group does not anticipate offering significant recommendations regarding descriptive metadata. The group includes representatives from both archive and library organizations, with practices for resource description that vary in significant ways. With all types of materials, libraries favor *bibliographic data* while archives prefer *finding aids*. In the end, we are likely to have more to say about *where* a chunk of agency-produced descriptive metadata might be embedded than about *what* it should look like.

We also do not anticipate offering recommendations regarding structural metadata, beyond the structural option inherent in MXF itself. Different agencies and even units within agencies take a variety of approaches to content packaging — the *binding* or *bundling* of multiple related files. We have, however, discussed the idea of having the MXF application specification for archiving and preservation include a way to wrap collections, i.e., sets of items.

Conclusion

As we proceed, the Working Group is well aware that we are at an early stage in this process, with comparatively little experience. We believe that there is value in drafting a thorough application specification — an important step in the direction of standardization. But we will wait until we have more experience under our belts before making a firm recommendation to other archives.

This thought has led us to explore some form of provisional advisory: “What shall we do in the meantime?” We have heard from colleagues who, thinking of their videotapes in dire need of reformatting, are drawn to the practices at Stanford and Rutgers — mentioned earlier in this paper — that create files with uncompressed video streams in widely used wrappers like QuickTime and AVI. Meanwhile, for born digital content, other colleagues have sketched provisional plans to hold some newly arrived materials in their native video encodings, reckoning that these will remain playable for a few years. The Working Group is interested in provisional solutions like these and hopes to assemble an advisory in the foreseeable future.

16 <http://store.smpete.org/product-p/st%200410-2008.htm>

17 <http://store.smpete.org/product-p/rp%202057-2011.htm>

18 <http://pbcore.org/v2/elements/pbcoredescriptiondocument/pbcoreinstantiation>