

## **Federal Agencies AV Working Group**

Notes from the meeting held at the Library of Congress, May 21, 2010.

*Abstract: Five agencies represented. Discussion of video formats for preservation reformatting projects, a project to develop an MXF Application Specification in cooperation with the AMWA, and related matters.*

### **Framing the discussion of the video target format project**

Carl Fleischhauer (Library of Congress) framed the main topic for the meeting: discussion of the development of an MXF application specification for moving image preservation files. At a high level, Fleischhauer identified these desiderata:

- Standardized file “wrapper/container” for long term storage and data exchange, something supported by open specification; e.g., MXF, QuickTime, AVI
- Ability to embed some metadata (compare to the Working Group’s guideline for embedded WAVE metadata)
- Extensible specification(s); for standard definition video today, high def very soon, film scanning to come
- Something vendor-manufacturers can build to
- Validation tools can use the spec to validate
- Allows for multiple encodings, including JPEG 2000, uncompressed, and MPEG
- Reasonable synergy with broadcast and motion picture industries and also with other non-governmental archives in US and national organizations in Europe, Australia, New Zealand

### **Comparing MXF to other container formats, discussion of format pros and cons**

Although the Working Group's planned exploration has been framed in terms of MXF, some members are concerned about the extent of support and the range of tools available for this container format, as compared to QuickTime and AVI (in both cases meaning the container element, not the encoding), for which many handy tools exist.

The expert consultant Oliver Morgan provided a bit of history of MXF. As the broadcast industry moved from a tape-based to a file-based infrastructure in the 1990s, it became apparent to broadcasters that they would need an exchange format for moving video files between servers. AAF (Advanced Authoring Format) emerged in response to that need. Its subtype MXF arose from an EBU and SMPTE taskforce that looked to create an exchange format in the late 1990s. One of the goals of this taskforce was to create an exchange format that was codec-independent, since organizations often use more than one compression family. Although any one organization may select a single codec for use (such plans have been stated for the Library of Congress), the fact that MXF is codec agnostic can be a great boon to archival collections in a collective sense. MXF is not manufacturer-specific and has an open metadata system wherein each repository using MXF can integrate its own metadata schema into their MXF files. What about adoption? At the National Association of Broadcasters’ meeting in April, Morgan said that many asked him about MXF; “MXF remains a very hot topic with professional broadcasters,” he said.

In the discussion, some attendees asked about MXF's complexity in comparison to other, "light-weight" containers like QuickTime and AVI. One agency reports that it has found AVI to be an open, widely supported, and simple-and-direct means to format video data. Another participant reported that some archives feel that way about QuickTime file format (the container). One attendee noted that MXF is designed for a broadcast environment where video items that are moved around a lot, shipped from one organization to another: this happens far less frequently in the work of some agencies. Attendees from other agencies, however, expressed support for continuing to explore MXF, saying that its adoption was something they were seriously considering.

### **Comments on the uptake of JPEG 2000 in the moving image community**

Morgan reported that producers of digital-cinema formatted content (for movie theaters) have moved their push to standardize a little further upstream, to the post-production and mastering zone, where they are working to specify JPEG 2000 "mezzanine" files. Broadcasters, Morgan added, are moving towards generating broadcast profiles for JPEG 2000 and said that work is being done to map the encoding onto the MPEG-2 transport stream. Meanwhile, as a consultant-adviser to the National Geospatial-Intelligence Agency (specifically the Motion Imagery Standards Board), Morgan said that he knows that military and intelligence communities are spending considerable money on JPEG 2000 systems.

There was mention of the new JPEG 2000 Alliance, a group that is focused on moving image applications that met for the first time at NAB 2010. This organization is in an early phase of development and we should all listen in to see what comes of it.

### **Closed captioning in one agency's implementation of MXF-JPEG 2000**

One agency is working with a vendor to refine the handling of close captioning information as MXF-JPEG 2000 files are created. In NTSC and SMPTE 259M digital system video streams, captions are encoded onto line 21 of the Vertical Blanking Interval (VBI). This data ought to move to the "right place" in the resulting digital file; the discussion at the meeting highlighted references to the standards EIA-608B and EIA-708B, and to SMPTE 436M (*MXF Mappings for VBI Lines and Ancillary Data Packets*).

### **About shaping an application specification**

As reported in the notes for the meeting for March 8, 2010, the Working Group plans to develop an MXF Application Specification (AS) that would be suitable for preservation reformatting. An AS is a set of constraints against the broad spectrum of possibilities offered by the option-laden standards MXF and JPEG 2000. The outcome can be compared to the establishment of profiles and levels frequently used to manage the use of MPEG-2-encoded video. Although MXF ASes may ultimately move to SMPTE for further standardization, it is customary to incubate them under the auspices of the Advanced Media Workflow Association (AMWA; <http://www.aafassociation.org/>).

An AS specifies a set of constraints on what might be "legal" in a broad sense, when using the MXF container and a given encoded essence (uncompressed, JPEG 2000, etc.) within the container. In AMWA lingo, a *shim* is a further set of constraints, additional narrowing of the

specification. An AS is a *community* product, shaped by the collective work of AMWA participants, while a *shim* is authored by an individual organization.

Morgan observed that the "art of creating an application specification" (AS) is the art of the possible or practical. He stressed the importance of avoiding a design that is so "gold plated" that no vendor can create products for it.

Attendees carried out a small amount of discussion of requirements--what do archivists wish to have "from" their archival master files? But this topic was not well developed on this occasion. Carl Fleischhauer reported that this topic had arisen at an AMWA Technical Steering Committee meeting on May 17, 2010, where the following provisional outline was typed up, more or less off the cuff:

- Preservation files ought to contain various content "sub-elements," e.g., if a videotape has been reformatted, some archivists will wish to include a still-image scan of the box or of associated documents: this is sometimes referred to as "ancillary items."
- Files may include video essences that are uncompressed, lossless or lossy JPEG 2000, and others.
- Files may contain the secondary or derivative viewing or "production master" (mezzanine-like) copies of the program, i.e., an MPEG-4 version side by side with the lossless JPEG 2000 version: this is sometimes referred to as "multiple versions."
- File should include file-integrity "checksum" data for essence monitoring over time.
- Files need to retain closed captioning and all other elements encoded in the VBI (and elsewhere?) in the original analog signal.
- Files ought to offer support for multiple languages, and/or other multi-option soundtracks.

In addition, attendees noted that requirements and desired outcomes for metadata would be especially needed.

## Test files

There was some discussion of test files. Some ought to be produced during the AS development. Although the meeting did not dwell on the following, or finalize what this Working Group will do, the discussion suggested that test files fall into three categories:

- **Input example(s) for encoder-and-container testing.** What is at stake is a given device's success at (a) encoding an incoming video stream and (b) creating an output file that contains the encoded essence data, the primary test element that is needed is the video stream, the input to the system.
- **Successful final product example as benchmark.** A well-formed, valid example of a final product that conforms to the specification. Used as a benchmark or yardstick for comparison to files produced by a proposed/nominated system.
- **Successful final product example for decoder testing.** To help decoder manufacturers determine if their product performs properly with this encoded example?

At the meeting, Morgan offered a sketch for files for the first category: input examples to test the success of a nominated device at encoding and container-making. Morgan said that we need to have "golden samples" – samples that are engineered to be good. We could have several basic

tiers of samples: gold (those engineered to be perfect); silver, bronze and copper ("real-world material"); and "arsenic" (problematic, bit-flipped, troubled samples). One attendee said that we should be looking to find "where the encoder breaks" with the low-end files and also looking to see how wide of a latitude of "good enough" we get with those files.

### **Metadata**

Morgan's work with AS-03 (which started life as the AS for PBS and public television distribution) had brought PBCore and Dublin Core to his attention. Would those be candidates for preservation? Fleischhauer said that it might be difficult to get a number of agencies on the same wavelength regarding *descriptive* metadata (libraries and archives do not follow similar standards) but there should be some consensus on *technical* metadata. Another attendee reported that PBCore is on version 1.3 right now with 2.0 expected in early 2011.

### **The AMIA/IASA meeting in November**

The Working Group hopes to have some materials ready for community discussion at the November IASA/AMIA meeting in Philadelphia, which will bring together representatives from a large number of moving image archives from around the world will be present in Philadelphia.