Federal Agencies
Digitization Guidelines Initiative

Digital File Formats for Videotape Reformatting

Part 5. Narrative and Summary Tables

December 2, 2014

The FADGI Audio-Visual Working Group

http://www.digitizationguidelines.gov/audio-visual/

Digital File Formats for Videotape Reformatting
Part 5. Narrative and Summary Tables

Contents

What is this document?

This is one of five documents that, taken together, compare a variety of digital file formats that are suitable targets for the reformatting of older video materials, generally physical videotapes.

The four companion documents are:

- Part 1. Detailed Matrix for Wrappers (unified large table)
- Part 2. Detailed Matrix for Wrappers (multi-page)
- Part 3. Detailed Matrix for Encodings (unified large table)
- Part 4. Detailed Matrix for Encodings (multi-page)[1]

---

[1] The URLs for the four documents are:
(1) http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p1_20141202.pdf
(2) http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p2_20141202.pdf
(3) http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p3_20141202.pdf

Introduction

This document concerns digital file formats that are suitable targets for the reformatting of older video materials, generally physical videotapes. These may be either analog or digital signals, but for this report, comparison is limited to examples with a 4:3 aspect ratio and the smaller resolutions that define standard definition (SD) content. Some specific examples are represented in the extensive holdings of 1-inch, U-matic, VHS, and Betacam tapes in many federal collections; digital formats like Digital Betacam and the less common BetacamSX and D2 are also scattered throughout many collections.

Federal agency preservation specialists and members of the wider archiving community share an interest in this topic. We have seen frequent exchanges about the pros and cons of various digital target formats for reformatted video on archiving listservs. There have also been exchanges about preservation target-format options for born-digital, file-form video and for the output of motion picture film scanning. The latter two topics are subjects of their own, and the FADGI Audio-Visual Working Group hopes to address them in future investigations and reports.

Guiding concepts

Three principles guided the FADGI team that assembled this matrix: first, the importance of producing an authentic and complete copy of the original recording, as evidenced by the attention paid to multiple timecodes, captioning, and soundtracks. This led us to pay close attention to functional characteristics, which we label as *Settings and Capabilities*. Not all wrappers offer settings and capabilities that support the production of authentic copies of certain types of source materials. Some archives compensate for this shortfall by maintaining data in associated elements, e.g., separate SubRip srt files that carry captions, or collection-management databases that contain important item-level technical metadata.

The second principle has to do with quality of reproduction. In general, the team preferred formats that maximized quality in both picture and audio reproduction. We favor uncompressed or losslessly compressed essences. At the same time, we know that some non-federal organizations with extensive broadcast holdings employ lossy compressed encodings for their archival master files (often called preservation masters), and we have included one lossy encoding to stand for this class of formats in our comparison set. We also recognize that no one can be "pure" in this context. For example, almost all old videotapes contain *composite* video recordings, e.g., NTSC in the United States. Meanwhile, all of the digital encoding formats we considered employ *color-difference component* video. This fact of life means that when an old tape is played back, before it can be encoded and written to a file, it must undergo an irreversible transformation from composite to component color models.

The third principle is the goal of producing archival masters that support the creation of access and access-support elements. When reformatting a recording with closed captioning, for example, selecting a format that permits the movement and storage of a copy of this textual captioning data as, say, XML, means that the archival master contains a resource that can be

---

(4) http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p4_20141202.pdf

more easily extracted for indexing, just as having an OCR rendering of a book text means that the book can indexed in order to be more accessible to researchers.

The character of the matrixes

This document was inspired by a similar matrix for the reformatting of printed matter and similar textual documents published by the FADGI Still Image Working Group.[2] In the still image matrix, each column in the table represents a combination of a *wrapper* and a *bitstream encoding*, e.g., a TIFF file with uncompressed raster data, the JPEG 2000 file format with JPEG 2000 Core Coding data, and so on. Video, however, is not only relatively more complex but also offers more opportunities for mixing and matching. The various uncompressed-video bitstream encodings, for example, may be wrapped in AVI, QuickTime, Matroska, and MXF. Thus we present two tables: one for *wrappers* and one for *bitstream encodings*.

Wrappers are distinct from encodings and typically play a different role in a preservation context. FADGI defines the word *wrapper* as a "term often used by digital content specialists to name a file format that encapsulates its constituent bitstreams and includes metadata that describes the content within."[3] Basically, a wrapper provides a way to store and, at a high level, structure the data; it usually provides a mechanism to store technical and descriptive information (metadata) about the bitstream as well. An *encoding*, on the other hand, defines the way the picture and sound data is structured at the lowest level (i.e., will the data be RGB or YUV, what is the chroma subsampling?).[4] The encoding also determines how much data will be captured: in abstract terms, what the sampling rate will be and how much information will be captured at each sample and in video-specific terms, what the frame rate will be and what will the bit depth be at each pixel or macropixel.

Our centerpiece is a pair of matrixes that compare wrappers and encodings in terms of about forty factors and subfactors. (Although best studied in their matrix form, for ease of printing and review, the same data has been reformatted one-row-to-a-page in an additional document.) These two extended matrixes are preceded by summary matrixes that highlight the key findings. The extended matrixes are followed by a discussion of the suitability of the formats for a number of illustrative video collection items. We suggest, for example, that certain classes of items--say, VHS tapes that contain video oral history footage--can be successfully reproduced in a number of the formats we compare. In contrast, a tape of a finished television program that may contain multiple timecodes, closed captioning, and four audio tracks, will only be reproduced with full success in one or two of the formats being compared.

Selection of formats

We tried to pick "likely" encodings and wrappers: (i) examples that we and our colleagues talk about on a regular basis and (ii) encodings and wrappers that fit the three principles outlined above. Although we favored uncompressed and lossless compression encodings, we have

---

[2] Three related still-image-format documents are linked from this page: http://www.digitizationguidelines.gov/guidelines/File_format_compare.html.
[3] http://www.digitizationguidelines.gov/term.php?term=wrapper
[4] See also http://www.digitizationguidelines.gov/term.php?term=encoding.

included the lossy compressed MPEG-2 encoding format. We were pleased to be able to respond to genuine interest in the FFV1 encoding and Matroska wrapper from advocates for open-source formats, even as we confess that we are not fully confident of our knowledge.

The matrixes presented here are neither comprehensive nor permanent. We have not tried to compare every encoding and wrapper, nor do we believe that the information we present is fixed forever. The list of available formats is certain to change over time, as will some of the assessments we report in the tables' individual cells. (To say nothing of inadvertent mistakes in our analyses! We have read the format specifications and other documentation carefully, but don't have the wealth of knowledge and familiarity that comes with daily hands-on experience creating and manipulating files in all of these formats.) In fact, the initial dissemination of this matrix on the FADGI Web site is accompanied by a plea to our preservation colleagues to send us corrections, comments, and suggestions. Version two may appear very soon!

MPEG-2 and other lossy encodings

Although lossy, we chose to describe and compare a very high-quality profile of the MPEG-2 encoding: 4:2:2 Profile@Main Level, limited to all I-frames. The all-I-frame limit means that this flavor of MPEG-2 encoding employs only intraframe (not temporal or interframe) compression and supports bitrates that are suitable for digital preservation of video, typically 50 Mbps for standard definition. It is worth noting that all profiles of MPEG-2 are very well-documented and offer good support for technical metadata and other modern file format features. MPEG-2 is widely adopted: part and parcel of the U.S. ATSC digital television specification and well-supported by the vendor community.

We did not, however, select MPEG-4 and lossy JPEG2000 encodings for comparison. While some MPEG-4 profiles have the ability to support high bitrates and intraframe compression we have not seen the same breadth of adoption for reformatting that older formats like MPEG-2 have achieved. Although lossy JPEG 2000 is in extensive use for digital cinema distribution and formats like AVC-Intra (an MPEG-4 subtype) are offered as a recording format by some new video cameras, we have not encountered instances in which these lossy JPEG2000 profiles are employed for reformatting.   This may change in the future and, as mentioned above, this document will evolve and change along with the prominence and support of different file formats. We invite our readers to send us comments on this matter.

MPEG-2 presented us with one wrinkle, different than any of the other formats we compared. When is it a wrapper and when an encoding? In ISO/IEC and SMTPE professional engineering circles, *MPEG-2 Transport and Program Streams* are referred to as *wrappers*; in part they serve as containers of content and associated information held within or associated with the underlying elementary streams. Our use of the term *wrapper*, however, is different from ISO/IEC and SMPTE usage related to MPEG-2. MPEG-2 Transport and Program Streams are, as the name suggests, also *streams* of data that multiplex audio, video, and other information into a single data stream, but do not specify file or data storage formats like the other formats in our wrapper matrix. In addition, MPEG-2 streams can be wrapped (our usage) within other file types in our matrix, e.g., MXF and AVI, or using its own ad hoc wrapper (with the filename extensions mpg or mpeg). For all these reasons, we treat MPEG-2 as an encoding.

Added explanations for selected comparison factors

Most of the scoring factors in the matrix have a rough-and-ready quality: the compilers did not feel that precise metrics existed for these and, in any case, the comparison is intended as a guide and not as a surgical tool. Our intended meanings for the factors are conveyed in the second and third columns in each matrix. The second column lists our *scoring conventions*, which often use simple, broad terms like "good, acceptable, poor." The third column is titled *considerations* and it contains brief questions or comments that indicate how we interpreted the factor in this row. When we drafted this narrative introduction, however, we felt, however, that the following three factors warranted some additional explanation.

Chroma subsampling and our preference for 4:2:2

Please refer to the appendix for a technical explanation of chroma subsampling and the meaning of ratio statements like 4:2:2, 4:2:0, 4:1:1, etc. The point of this section is to assert that the widely used 4:2:2 subsampling provides higher image quality than other feasible options. The 4:4:4 ratio does offer even higher quality and is occasionally used in moving image work, but practical considerations in terms of available equipment and interfaces generally preclude its use in video reformatting. The use of 4:4:4 also produces significantly larger files.

In addition to inherently better initial image quality, 4:2:2 also provides benefits if material is re-reformatted over time, in what is sometimes called a *cascading* scenario. For professional broadcasters, a *cascade* may be encountered in a chain of connected broadcast elements with the same risks of quality loss as in a cascade over time. A Web page from the Japanese equipment manufacturer NTT[5] offers an excellent set of tabbed images that illustrate quality loss in such a cascade, comparing 4:2:2 to 4:2:0 subsampling.

Bits per sample and our preference for 10-bit sampling

The explanation of chroma subsampling in the appendix identifies the elements being sampled in the digital image: *luma* data and two types of *chroma* data. Many digitizing systems offer the option of recording either 8 or 10 bits per sample. The compilers of this document generally encourage the use of 10-bit sampling for the sake of higher image quality.[6] Some archives use 8-bit sampling for certain classes of material in order to keep file sizes low. However, with 8-bit sampling, there is greater risk that imagery will show abrupt changes between shades of the same color. Image elements that feature natural gradients like blue skies or areas of (seemingly) solid tonality can show what is called *banding* or *contouring*. In these cases, not every change in the

[5] http://www.ntt-electronics.com/en/products/video/products/codec_system_solutions/hv9100/picture-quality-of-cascaded-video-codec.html.

[6] Some specialists argue, however, that there is no benefit for certain classes of material. Dave Rice, a digital-video expert at the City University of New York wrote, "We digitize Betacam SX tape to 8-bit UYVY but Digibeta to 10-bit V210 because these selections align with the nature of the data that is actually sent out over SDI from these tapes. SDI is 10-bit data, but when I piped the SDI video data from an SX tape to a binary display I could see the 9th and 10th bits were always zero. Thus by taking only the first 8 bits I could get all meaningful data. I have about 3,000 SX hours to preserve and choosing 8-bit instead of 10 saves me about 90 TB of storage" (private communication).

continuous gradient can be shown because there are insufficient bits to represent all of the shades. The risk of banding is reduced by increasing the number of bits per sample.

Video range and our preference to declare this fact in metadata

This factor concerns metadata, most likely to be associated with the wrapper: *Does the format clearly declare whether it contains broadcast safe range video or computer graphics video?* The differences between the two ranges are outlined in the appendix. The significance of the factor pertains to playback or future re-reformatting. In order to avoid the risk of misinterpreting or even clipping picture data, the playback or transfer device must know the range for the item at hand. However, at many archives, the types of source material or the business rules applied during reformatting may guarantee the uniformity of a given collection in terms of range, and this general knowledge can be used to guide future activities.

Clarification regarding terminology

In our discussion of "Cost Factors" and "System Implementation Factors" we use the term "free software" to describe some of the tools under consideration. For our purposes we use a rather specific definition of free software: Free software and its source code are able to be used, modified and shared by everyone equally. Additionally, free software is not restricted by any licenses nor does it require a fee to use or access it.

Matching Formats to Types of Material: Four Examples

How do organizations match classes or types of originals to formats? The following examples from the National Archives and Records Administration and the Library of Congress are intended to illustrate this. These reports sketch the characteristics of the original materials and of the output target formats, together with some comments about the fit of one to the other. The decisions represented here, of course, predate the creation of this comparison matrix but they serve to show how content features can be considered when selecting a format. There may be a future revision of this video format comparison package, and the FADGI working group welcomes comments and additional examples from other organizations.

1. NASA Onboard Shuttle Recordings, 1980s-1990s

National Archives and Records Administration

Source material:

- U-matic tapes
- Black-and-white footage
- 2 audio tracks
- No timecode
- No closed captions

NARA Selections:

Wrapper	**AVI**
Audio	8 channels[7]
Timecode	N/A
Closed Captions	N/A
Encoding	
Video	YUY2
Bit depth	8-bit
Chroma subsampling	4:2:2

Commentary:

- Wrapper:
  - Simple-structure source material like these NASA recordings can successfully be retained in a wide range of formats, including the AVI wrapper.

- Encoding:
  - NARA uses the YUY2 codec which relies on 4:2:2 chroma subsampling.[8] This aligns with FADGI's general preference is for 4:2:2 chroma subsampling when converting from composite signal to color-difference component.
  - NARA has determined that 8-bit uncompressed video will produce a sufficient preservation master of this material. 10-bit uncompressed video may produce a higher quality preservation master that captures additional tonal subtleties, but will require additional storage space and different hardware and software than NARA has been able to acquire[9].
  - Some practitioners may even argue that a high quality lossy codec like MPEG-2 at 50Mbps using the 4:2:2 Profile/Main Level will do a satisfactory job of preserving this material while significantly reducing storage and network costs.
  - Most likely, 8 or 10-bit video codecs will be the best choice for a video preservation master of this source material.

- Conclusions:
  - All of the encodings and wrappers under consideration can successfully preserve this material.
  - Practitioners may differ over which encoding is most suitable for their particular institutional circumstance.

---

[7] NARA reports that this is hard-wired in their conversion system otherwise they would capture only 2 or 4 channels as appropriate.

[8] http://www.digitalpreservation.gov/formats/fdd/fdd000364.shtml

[9] NARA is in the process of testing and implementing new video capture hardware and software and does anticipate having the ability to capture 10-bit video in the near future.

2. Forest Service PSAs (Public Service Announcements), 1990s

National Archives and Records Administration

Source material:

- Digital Betacam tapes
- Color NTSC footage
- 4 audio tracks
- Continuous LTC timecode
- Closed captions on line 21 (not as VANC data)

NARA Selections:

Wrapper	**AVI**
Audio	8 channels[10]
Timecode	Start value stored in timecode field
Closed Captions	Line 21 stored in raster
Encoding	
Video	YUY2
Bit depth	8-bit
Chroma subsampling	4:2:2

Commentary:

- Wrapper:
  - AVI stores a timecode start value and can support continuous timecode as a count-up from this initial value.
  - MOV, MKV and MXF are all able to store continuous timecode as well.
  - Captions on line 21, retained in raster: see encoding notes
  - Captions from line 21, converted and carried as binary or text-based data[11]:
    - Only MOV and MXF have dedicated storage mechanisms within the file wrapper to preserve closed captions (other than as data-in-the-raster); MOV uses a captioning track and MXF offers a variety of methods.
    - MOV can store CEA608 data in its dedicated caption tracks. It may also be able to store Timed Text data in external or sidecar files.
    - MXF is able to natively support both CEA608 and Timed Text captions. It makes use of Data Items to store CEA608 data and generic stream partitions to store Timed Text.
    - In contrast, the retention of captions with MKV and AVI wrappers requires the use of associated files: MKS files are the preferred file type to

[10] See above.

11 In this case, binary data would appear as CEA608/708 caption data. Text-based data would most likely occur as SMPTE XML-based Timed Text information.

use with MKV, while AVI is used with any of the well-known captioning file types, e.g., .srt, .scc, etc.[12]
  - With the right tools an MPEG-2 stream would also be capable of preserving the timecode and caption data from the original.

- Encoding:
  - NARA uses the YUY2 codec which relies on 4:2:2 chroma subsampling. This aligns with FADGI's general preference is for 4:2:2 chroma subsampling when converting from composite signal to color-difference component.
  - Digital Betacam tapes play out as SDI signal with 10-bit picture data. Uncompressed encoding would require v210 codec and lossless compressed encoding requires FFV1 or JPEG 2000.
  - The 8-bit codec that NARA uses is unable to capture all of this data which is a serious drawback. If the option were available, NARA would capture this content at 10-bit. It is not a drawback of the AVI wrapper (which can indeed support the v210 codec), but instead a shortcoming of somewhat dated hardware and software.[13]
  - Line 21 captions retained in raster: possible with the 720x486 raster; lost if the raster is reduced to 720x480, as is sometimes done when digitizing.

Conclusions:
- Although relatively simple in structure, this DigiBeta source material highlights the inability of AVI to carry caption data. Other wrappers will be more successful.
- Meanwhile, those with an interest in retaining line 21 caption data in the picture raster should pay attention to encoding specifications.
- In this case, AVI is able to preserve the timecode because it is continuous and the wrapper does not need to be able to support timecode breaks or interruptions.

3. Proceedings from the Floor of U.S. House of Representatives

Library of Congress Packard Campus of the National Audio-Visual Conservation Center

Source material:
- Date range: 1977 – present
- Formats: ¾" U-matic, 1" Type C, Betacam SP, Digital Betacam, DVCPRO, DVCPRO HD
- Color footage
- Mono on 2 audio tracks
- The first 15-20 years of the collection does not include timecode other than analog timecode recorded by the VTR. LTC and VITC timecode may exist for tapes recorded in the 1990s and later.

[12] If these captions had been stored as VANC data, the AVI wrapper itself would not have a mechanism to preserve them. It may be possible to transform the captions from VANC data into a format that could be stored in an associated file like those mentioned above.
[13] As mentioned above, NARA is in the process of testing and implementing new video capture hardware and software and does anticipate having the ability to capture 10-bit video in the near future.

- Closed captions may be present for tapes recorded in the 1990s and later.

NAVCC Selections:

Wrapper	**MXF OP1a**
Audio channels	Retained as native (mono on 2 audio tracks)
Timecode	Retained as native if present
Closed Captions	Retained as native if present
Encoding	
Video	**JPEG2000 Lossless reversible 5/3**
Chroma subsampling	Native

Commentary:
The Packard Campus receives analog and digital video collections through a variety of input streams including copyright submission and general collection acquisition. The entirety of the Library's audiovisual holdings will be digitized, creating both archive masters and access copies providing researchers with playback on demand in the Library's Capitol Hill reading rooms. In order to reduce the variability in these large and complex collections, video inputs are normalized to one standard format, JPEG2000 in MXF OP1a.[14]

- Wrapper:
    - MXF OP1a is standardized through SMPTE 377-1 and SMPTE 378M-2004. The MXF wrapper was specifically designed to aid interoperability and interchange between different vendor systems, especially within the media and entertainment production communities which are the primary content providers to Packard Campus collections. The file specification was standardized by the SMPTE (Society of Motion Picture & Television Engineers) & AMWA (Advanced Media Workflow Association) and allows different variations of files to be created for specific production environments and can act as a wrapper for metadata & other types of associated data including complex timecode, closed captions and multiple audio tracks.

- Encoding:
    - JPEG2000 is standardized in ISO 15444. The version of JPEG2000 adopted by the Packard Campus uses the mathematically lossless 5/3 wavelet transform. This compression is completely reversible and there is no loss of quality when the file is encoded and decoded. Other attractive features include that it does not have licensing issues, it can be wrapped in a standardized file wrapper (MXF) which promotes interoperability, and it can accommodate any color space and bit depth.
    - This collection, because it encompasses such a wide date range, includes both component and composite video. In order to capture the best signal off the tape, the source video is captured as component if it's component. Actions in the

[14] http://www.digitalpreservation.gov/formats/fdd/fdd000206.shtml

reformatting workflows had to adjust in order to convert from composite signal to color-difference component.

  - The Packard Campus currently uses 10-bit for video because this matches the specifications laid out for serial digital interface (SDI) as defined by SMPTE starting with ST 259M (although NAVCC is preparing to accommodate native bit depths beyond 10-bit). Additionally, 10-bit encoding is preferred over 8-bit as a harmonization encoding so that decoder software writers do not have to accommodate both.
  - The Packard Campus retains the native chroma subsampling.

Conclusion

- Collections at the scale of those held at the Packard Campus, which includes an archive with up to 5 PB in stored data including 200,000 archive files of digital and digitized audio and moving image recordings, require normalization in order to reduce variability and complexity. Video inputs are normalized on ingest to one standard format, JPEG2000 in MXF OP1a. The benefit of normalizing on ingest into the repository is that the toolkits are still (mostly) available for the submitted file formats and encodings. After a time, general production and toolkit availability will decrease and it may not be possible to access and transform the file.

4. Afghan Media Resource Center

Library of Congress Packard Campus of the National Audio-Visual Conservation Center

Source material:

- Content: Raw footage of 30 years of war in Afghanistan: internecine fighting, Soviet invasion, Taliban takeover, and the American invasion including original interviews with Mujahidin, Taliban, and resistance fighters against the Soviet invasion.
- Date range: 1979 – present
- Formats: 3500 hours of PAL & 625-line DV video
- Color footage
- 2 audio tracks
- No timecode or closed captions present

NAVCC Selections:

Wrapper	**MOV**
Audio channels	2
Timecode	None
Closed Captions	None
Encoding	
Video	**625 Standard Definition YCbCr ITU-R 601 standard**
Chroma subsampling	4:2:2

Commentary:
The AMRP is a joint project between LC and US State Department to digitize 35 years of video footage (primarily news coverage) from the Afghan Media Resource Center in Kabul, an organization that started before the Soviet invasion in 1979. The collection includes 3,500 hours of video, 10,000 audio recordings, and 20,000 B&W and color still images documenting the political events in Afghanistan. The digitization work needed to be carried out in Afghanistan prior to early 2014, before US Armed Forces left the country, because many officials feared that the material might be destroyed after the coalition forces had departed. This project is a test case to determine if an inexpensive (under $5000 each) system could be developed using common off-the-shelf equipment and software to produce industry standard files which could be migrated to the NAVCC Evergreen format at a later point. Moreover, it fulfills a core NAVCC mission to develop tools for all media archiving organizations that can produce archive quality digital objects at low cost while adhering to the "do no harm" principle whenever possible. This project can serve as a model for organizations with significant technical and funding challenges can move forward with archiving work in a responsible way.

- Wrapper:
  - Although proprietary because it is developed and supported by Apple, the QuickTime (*.mov) wrapper is very stable, well documented and widely adopted with a strong toolset available.
  - QuickTime has support for timecode and closed caption tracks although these are not pertinent to the AMRC collection
  - QuickTime converts easily to the NAVCC evergreen normalization format, JPEG2000 lossless reversible 5/3 in MXF OP1a.

- Encoding:
  - Both the PAL and 625-line video are reformatted according to ITU-R Recommendation BT.601, the industry standard for encoding SD interlaced analog video signals in digital video form. Rec.601 video is widely adopted with strong support through a large variety of toolsets.
  - Rec.601 video is easily convertible into the NAVCC evergreen normalization format, JPEG2000 lossless reversible 5/3 in MXF OP1a in part because the same chroma subsampling (4:2:2) and bit depth (10-bit) is implemented in both specifications.

Conclusion:
- The project deliverables, 625 Standard Definition YPbPr ITU-R 601 video in an MOV wrapper, can easily convert to the JPEG2000 lossless reversible 5/3 in MXF OP1a archive format implemented at NAVCC. This specification also closely aligns with the *Category 1: Analog Source* definition from the FADGI report, *Refining Conversion Contract Specifications: Determining Suitable Digital Video Formats for Medium-term Storage*.[15] One difference is the color space. The AMRC implementation uses YPbPr in accordance to the NAVCC Evergreen format which uses YPbPr for native 4:2:2 while the *Category 1: Analog Source* specifies YCbCr.

---

[15] http://www.digitizationguidelines.gov/audio-visual/documents/IntrmMastVidFormatRecs_20111001.pdf

Other Features to Consider

While the examples above have presented a variety of source material they certainly do not account for all types of information that may be stored on analog videotapes nor do they address the entire range of possible features required in preservation master files. Several of the topics listed below are discussed in *An Update on AS-07: MXF Application Specification for Moving Image Archiving and Preservation.*[16]

- Multiple timecodes
  - Some organizations may require the preservation of source timecode and the addition of a continuous timecode track; this is especially true if the source timecode is discontinuous. Playback and manipulation of files with discontinuous timecodes can present difficulties and it may be preferable at times to rely on a continuous timecode for purposes such as playback, editing and/or deriving running times.
  - Most of the wrappers presented will not be able to support storage of multiple timecodes.
    - AVI cannot fully support two timecodes and in fact presents challenges when storing a single discontinuous timecode track alone.
    - It is our impression that MKV is designed to carry a single timecode track and it is unclear whether it can support discontinuities.
    - MOV can carry a single discontinuous timecode track and is also able to accommodate a continuous timecode by storing a start value and a count rate. Beyond that, MOV does not seem to have support for additional timecodes.
    - In contrast, MXF offers robust support for storage of multiple timecodes.
- Multiple audio tracks
  - Some videotapes, especially those in broadcast archives, may contain as many as 8 channels of audio, all of which will need to be preserved.
  - Most of the formats discussed here will be able to support 8 channels of audio. MXF in particular has good support for complex audio configurations.
- Captions carried as Timed Text
  - Timed Text is an increasingly important means of carrying captions and subtitles. It is a relatively recently development. The W3C standardized a version for web-delivery (TTML) in 2010 and SMPTE built on this work to create a standard for broadband delivery to the home (SMPTE-TT) later that year. In 2012, the EBU also authored a standard for the implementation of Timed Text.
  - Of the wrapper formats analyzed in the matrix, only MXF has clear support for captions carried as Timed Text. Timed Text may be stored within in a MXF or as a sidecar file that accompanies the reformatted audio-video content.
- Carriage of associated materials, supplementary metadata
  - Some organizations may require additional materials or supplementary metadata be stored with the reformatted audio-video content. Possible examples may include transcripts or images of the case in which the original audio-video record was stored.

---

[16] See http://www.iasa-web.org/iasa-journal-no-42-january-2014 (for IASA members or for purchase).

- Some of the wrappers offer support for the storage of relatively simple associated materials. For example, MOV can store still images in addition to audio-video content. MKV is also fairly content agnostic, meaning that data in almost any form can be stored with the audio-video content. AVI and MPEG-2 do not offer clear support for data other than the typical audio-video content.
- Each of the formats above- MOV, MKV, AVI and MPEG-2- are able to store supplementary metadata in sidecar files. Depending on the software used to create the file, these formats may also support embedded supplementary metadata in the form of XMP or by making use of the INFO List chunk, in the case of AVI.
- MXF has good support for associated materials and supplementary metadata. Current drafts of the AS-07 (Archiving and Preservation) Application Specification indicate that it will support text-based files, still image and other audio-video content. Supplementary metadata is also well-supported in MXF and can be stored in an XML-based structure.

Appendix A: Summary versions of the matrixes

The summaries are on the following two pages. Detailed versions of the matrixes may be found at these URLs:

- Digital File Formats for Videotape Reformatting: Part 1. Detailed Matrix for Wrappers
  - http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p1_20141202.pdf
- Digital File Formats for Videotape Reformatting: Part 2. Detailed Matrix for Wrappers
  - http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p2_20141202.pdf
- Digital File Formats for Videotape Reformatting: Part 3. Detailed Matrix for Encodings
  - http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p3_20141202.pdf
- Digital File Formats for Videotape Reformatting: Part 4. Detailed Matrix for Encodings
  - http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p4_20141202.pdf

File Wrappers					
Attribute Category	**AVI**	**MOV**	**Matroska**	**MXF**	**MPEG-2 (ad-hoc wrapper format[17])**
Sustainability Factors	-Well-disclosed and moderately well-adopted -Transparent format, but lacks some self-documentation capabilities -Not likely to be impacted by patents or technical protection mechanisms	-Well-disclosed and widely adopted format -Fairly transparent with good self-documentation capabilities -Possible impact from patents and technical protection mechanisms	-Acceptable documentation and moderate adoption -Transparent format with good self-documentation capabilities -No impact from patents -Possible impact from technical protection mechanisms	-Acceptable documentation and moderate adoption -Fairly transparent format with good self-documentation capabilities -No impact from patents -Possible impact from technical protection mechanisms	-Poor documentation, but moderate adoption -Poor transparency and self-documentation -Possible impact from patents -No impact from technical protection mechanisms
Cost Factors	-Low implementation cost -Cost of software and equipment needed is low -Storage and network costs will depend on the encoding in use	-Medium implementation cost -Commercial software offers richest set of features and functions -Storage and network costs will depend on the encoding in use	-Low implementation cost -Low software and hardware costs -Storage and network costs will depend on the encoding in use	-Low to medium implementation cost -Costs of software and hardware vary widely -Storage and network costs will depend on the encoding in use	-Low to medium implementation cost -Low software and hardware costs -Storage and network costs will depend on the encoding in use
System Implementation Factors	-Low complexity -Wide availability of tools (except for validation)	-Moderate complexity -Wide availability of tools (except for validation)	-Moderate complexity -Wide availability of tools (except for identification and validation) -Many tools require advanced technical skills to implement	-High complexity -Wide availability of tools (except for validation) -Some tools require advanced technical skills to implement may not be interoperable	-Low complexity -Wide availability of tools (except for validation)
Settings and Capabilities	-Good support for high quality bit depths and chroma subsampling -Poor to acceptable support for most modern wrapper features	-Good support for high quality bit depths and chroma subsampling -Good support for most modern wrapper features	-Good support for high quality bit depths and chroma subsampling -Good support for most modern wrapper features	-Good support for high quality bit depths and chroma subsampling -Very good support for modern wrapper features	-These features are handled at the encoding level

[17] The MPEG-2 format is standardized at the stream (or encoding) level, the .mpg wrapper is completely ad hoc.

Encodings					
Attribute Category	**Uncompressed 4:2:2, 8-bit (UYVY and YUY2)**	**Uncompressed 4:2:2, 10-bit (v210)**	**JPEG2000 - Lossless**	**ffv1**	**MPEG-2 - 4:2:2 Profile/Main Level**
Sustainability Factors	-Acceptable disclosure and wide adoption -Transparent formats - No impact from patents or technical protection mechanisms	-Good disclosure and wide adoption -Transparent format -No impact from patents or technical protection mechanisms	-Good disclosure, but low to moderate adoption -Acceptable level of transparency and self-documentation -No impact from patents or technical protection mechanisms	- Good disclosure and moderate adoption -Acceptable level of transparency and self-documentation -No impact from patents or technical protection mechanisms	-Good disclosure and wide adoption -Good level of transparency and self-documentation -Possible impact from patents and technical protection mechanisms
Cost Factors	-Varying implementation costs from low to high -Low to medium cost for software and hardware needed -High storage and network costs	-Medium to high implementation costs -Low to medium cost for software and hardware -High storage and network costs	-Medium to high implementation costs -Low to medium cost for software and hardware -Moderate storage and network costs	-Low to medium implementation cost -Low to medium cost for software and hardware needed -Moderate storage and network costs	-Low to medium implementation cost -Low software and hardware costs -Low storage and network costs
System Implementation Factors	-Low complexity -Wide availability of tools (except for identification and validation)	-Moderate complexity -Wide availability of tools (except for identification and validation)	-Moderate complexity -Moderate availability of tools (except for validation), but lingering issues with interoperability	-Moderate complexity -Wide availability of tools (except for validation) -Some tools require advanced technical skills to implement	-Low complexity -Wide availability of tools (except for validation)
Settings and Capabilities	-Acceptable support for high quality bit depths and chroma subsampling -Poor to acceptable support for most modern wrapper features	-Good support for high quality bit depths and acceptable support for chroma subsampling -Poor to acceptable support for most modern wrapper features	-Good support for high quality bit depths and chroma subsampling -Acceptable support for most modern wrapper features	-Good support for high quality bit depths and chroma subsampling -Good support for most modern wrapper features	-Acceptable support for high quality bit depths and chroma subsampling -Acceptable to good support for most modern wrapper features

Appendix B: Additional technical information on selected comparison factors

Chroma subsampling

Generally speaking, the digital video streams we encounter represent image brightnesses and colors in *color-difference component* streams. There are two widely used terms used to name the three color-difference components in digital video picture streams: YCbCr and YUV. (In some non-broadcast contexts, there can be a fourth transparency component). At a very high level, the terms have the same meaning. However, the term YCbCr is usually employed in a relatively precise way while YUV is often used more loosely. The Wikipedia entry for YUV[18] (consulted May 28, 2014), notes that the scope of these and other similar terms "is sometimes ambiguous and overlapping."

The designation YCbCr comes from the broadcast profession. Careful writers will use Y' ("Y prime") instead of Y since this component represents luma, i.e., gamma-corrected brightness intensity data. (Strictly speaking, no-prime Y represents intensity in linear terms and is called luminance.) Cb and Cr represent chroma (color) components. As explained in the Wikipedia article about YCbCr[19](consulted May 28, 2014), "Y'CbCr is not an absolute color space; rather, it is a way of encoding RGB information. The actual color displayed depends on the actual RGB primaries used to display the signal. Therefore a value expressed as Y'CbCr is predictable only if standard RGB primary chromaticities are used." The designation YUV (or Y'UV) comes from outside the broadcast community, often used by workers who focus on data networks and computer-based activities.

The discussion in this document is intended to inform those carrying out the preservation reformatting of older analog and media-dependent digital videotapes. For this reason, they concern the widespread 4:2:2 chroma subsampling pattern, the most common referent for the term uncompressed video when used by professional broadcasters. In 4:2:2 subsampling, the two chroma components are sampled at half the rate of luma. Reducing the horizontal chroma resolution by one-half reduces the bandwidth of the uncompressed video signal by one-third with little visual impact.

Chroma subsampling is usually expressed as a three-part ratio (in this case 4:2:2) although it may also include a fourth part (e.g., 4:2:2:4), when alpha or transparency data is part of the stream. As explained in the Wikipedia article on chroma subsampling[20] the ratio describes the number of luma and chroma samples "in a conceptual region that is J pixels wide, and 2 pixels high." The three key parts of the ratio are as follows, omitting the alpha channel:

- J: horizontal sampling reference (width of the conceptual region). Usually and in this case: 4.
- a: number of chrominance samples (Cr, Cb) in the first row of J pixels. In this case: 2.

[18] http://en.wikipedia.org/wiki/YUV, consulted September 5, 2014.
[19] http://en.wikipedia.org/wiki/YCbCr, consulted September 5, 2014.
[20] http://en.wikipedia.org/wiki/Chroma_subsampling, consulted September 5, 2014.

- b: number of (additional) chrominance samples (Cr, Cb) in the second row of J pixels. In this case: 2.

For 4:2:2 picture data, the conceptual region consists of a block of eight pixels that "contains" 12 samples: 8 luma and 4 chroma.

The sampling rates for picture data were codified by the International Telecommunications Union Radiocommunication Sector (ITU-R) specification BT.601, published in 1987 and designed to provide a common digital standard for interoperability between the three analog video/TV systems (NTSC, PAL, and SECAM). ITU-R BT.601 enables their signals to be converted to digital and then easily converted back again to any of the three for distribution. Meanwhile, version 1 of BT.709 was published in 1990 and has seen a number of significant changes and extensions; version 5 was published in 2008. The sampling frequencies used for both standards are as follows:

- BT.601(standard for SDTV), Luma sampling rate = 13.5 MHz, chroma sampling rate=6.75 MHz (4:2:2)
- BT.709 (standard for HDTV), Luma sampling=74.25 MHz, chroma sampling=37.125 MHz (4:2:2)

Video range (Broadcast safe range or wide range/computer-graphics video)

Uncompressed video streams are encountered with two different sets of levels, one standardized and one ad hoc. The standardized levels are specified by the International Telecommunications Union Radiocommunication Sector (ITU-R) and are often referred to as "video range," "legal levels," or "studio swing." These levels carry values from 16-235 for Y (luma) and 16-240 for Cr and Cb (chroma), assuming 8 bits per sample (higher values if 10-bit).

The specification for "previous generation" standard definition picture is ITU-R Recommendation BT.601 (often called Rec. 601 or by its former name, CCIR 601). BT.601 encoding of North American 525-line 60 Hz and European (and other) 625-line 50 Hz signals (both interlaced) yields 720 luminance samples and 360 chrominance samples per line (non-square pixels).

The specification for "current generation" digital picture is ITU-R BT.709 and it codifies interlaced and progressive scanned picture at a variety of picture sizes and frame rates (square pixels in the specification's later versions). In professional video production, BT.601 and BT.709 signals are carried by the SMPTE-standardized serial digital interfaces (SDI, HD-SDI, etc.).

Meanwhile, ad hoc uncompressed video streams with "wide range" or "super white" levels (from 0-255, assuming 8 bits per sample) may be produced in desktop computer graphics systems.

In all cases--BT.601, BT.709, and "wide range"--the data for a pair of pixels are stored in the order Cb-Y1-Cr-Y2, with the chrominance samples co-sited with the first luminance sample.