

# Video Formatting and Preservation

Carl Fleischhauer

National Digital Information Infrastructure and  
Preservation Program

DLF Forum

Philadelphia, November 6, 2007

Carl Fleischhauer  
Library of Congress

**[Introduction.]** My topic is video formatting, in the broadest sense, as it relates to preservation. I feel like a news reporter covering the 2008 presidential campaign. I see a number of worthy and appealing candidates, but the outcome is not yet clear. Preservation-oriented organizations are starting to develop approaches, but practices are not yet settled.

# Video Formatting Problem Spaces

- Reformatting
  - From old videotapes
- Born digital



From 1961: Ampex VR-1000-B  
2-inch quadruplex VTR

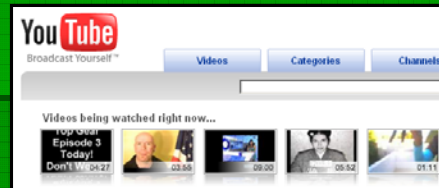
Image from Wikipedia

Carl Fleischhauer  
Library of Congress

Several organizations have tackled the problem of reformatting, i.e., copying old videotapes into digital file form, replacing the old practice of copying to a fresh set of videotapes. Fewer organizations have tackled the set of preservation problems that concern born digital content.

# Born Digital

- Professional zone: broadcasters
- Prosumer zone: YouTube

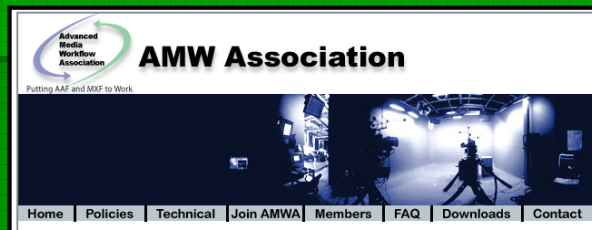


Carl Fleischhauer  
Library of Congress

I divide born digital into two classes. First, there is professional born digital: broadcasting and the professional end of documentary and independent production. Second, there is the often-inspired, prosumer work so significant today on Web sites like YouTube. My news gathering found *some* things to report concerning the formatting-for-preservation of professional born digital but virtually nothing about prosumer born digital.

# Allies in the Industry

- Producers, creators seek interoperability via standards
  - May be from “true” or industry standards bodies, MPEG-2, MPEG-4, MXF, AAF
  - May be relatively open and public, albeit proprietary, AVI, QuickTime



Carl Fleischhauer  
Library of Congress

Format standardization is important to preservation, and memory institutions have allies in the industry. Standardization is also important where video content is exchanged, during the production or distribution of broadcasts, or when content is archived. For professional video-makers, *interoperability* is an important keyword. Many feel caught between their desire for open standards and the desire of system vendors to be proprietary. Trade organizations like the Advanced Media Workflow Association represent content creators in the back-and-forth with companies like AVID over the implementation of standards.

# Formatting Elements

- Encodings
- Wrappers
- Metadata

Carl Fleischhauer  
Library of Congress

**[Three elements.]** I'll frame my report in terms of three elements: encoding, wrappers, and metadata.

# Formatting Elements

## Encoding

- Bitstream structures appropriate for our purposes
- Examples: MPEG-2 compression, JPEG 2000 frame-image representations

Carl Fleischhauer  
Library of Congress

By encoding, I mean the bitstream structures that are appropriate for our purposes. To define by example: MPEG-2 compression, JPEG 2000 frame-image representations.

# Formatting Elements

## Wrapper

- File formats that encapsulate one or more constituent bitstreams and include metadata
- Archetypal examples:
  - Broadcast WAVE, TIFF
- Complex examples:
  - QuickTime, MXF

Carl Fleischhauer  
Library of Congress

By wrappers, I mean file formats that encapsulate one or more constituent bitstreams and include metadata that describes what's inside. Archetypal non-video examples include Broadcast WAVE and TIFF. More complex (video) examples include QuickTime or MXF may contain multiple objects, e.g., one or more video and audio streams.

# Formatting Elements

## Metadata

- This talk: technically oriented chunks of administrative metadata.
- Metadata overlaps with wrapper encoding issues.
- To what degree is metadata embedded in the wrapper or even the bitstream?
- To what degree is such embedded metadata standardized?

Carl Fleischhauer  
Library of Congress

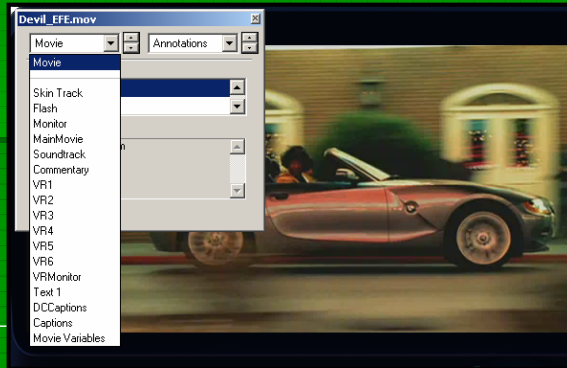
Regarding metadata, my talk will emphasize technically oriented chunks of administrative metadata. To what degree is metadata embedded in the wrapper or even the bitstream? To what degree is such embedded metadata standardized?

By the way, to reduce the length and complexity of this talk, I limit myself to picture information, and will not discuss sound.



# Profiles, Levels, and Application Specifications

- Many new specifications are complex, multipart
  - Examples: MPEG-4, JPEG 2000



*BMW movie  
QuickTime file  
with two  
soundtracks  
and a virtual  
reality feature*

Carl Fleischhauer  
Library of Congress

**[Profiles and more.]** Ending my last slide with “degree” questions sets me up to describe a factor that underlies the three elements: profiles, levels, and/or application specifications. Many published standards that pertain to video are complex and full of options, some of which will never be used.

# Profiles, Levels, and Application Specifications

- Which allowable elements will actually be used?
- Will *this* device play *this* file?
- Profiles and levels an important part of MPEG family from an early day

Carl Fleischhauer  
Library of Congress

These multiple choices inhibit interoperability. Will *this* device play *this* file?  
Long ago, video professionals profiled MPEG encodings.

# MPEG-2 Profiles and Levels

MPEG-2 Profiles					
Abbr.	Name	Frames	chroma format	Streams	Comment
SP	Simple Profile	P, I	4:2:0	1	no interlacing
MP	Main Profile	P, I, B	4:2:0	1	
422P	4:2:2 Profile	P, I, B	4:2:2	1	
SNR	SNR Profile	P, I, B	4:2:0	1-2	SNR: Signal to Noise Ratio
SP	Spatial Profile	P, I, B	4:2:0	1-3	
HP	High Profile	P, I, B	4:2:2	1-3	low, normal and high quality decoding

MPEG-2 Levels					
Abbr.	Name	Pixel/line	Lines	Framerate (Hz)	Bitrate (Mbit/s)
LL	Low Level	352	288	30	4
ML	Main Level	720	576	30	15
H-14	High 1440	1440	1152	30	60
HL	High Level	1920	1152	30	80

From Wikipedia

Carl Fleischhauer  
Library of Congress

For MPEG-2, the ISO/IEC specifications themselves spell out profiles (top chart) that define the structure of the encoded stream. They characterize the complexity of the encoding, indicating how difficult this signal will be to decode. Levels (bottom chart) influence quality--the big hint is in the last column: all other things being equal, the higher the data rate, the higher the quality.

# MPEG-2 Profiles and Levels

Profile @ Level	Resolution (px)	Framerate max. (Hz)	Sampling	Bitrate (Mbit/s)	Example Application
SP@LL	176 × 144	15	4:2:0	0.096	Wireless handsets
SP@ML	352 × 288	15	4:2:0	0.384	PDAs
	320 × 240	24			
MP@LL	352 × 288	30	4:2:0	4	Set-top boxes (STB)
MP@ML	720 × 480	30	4:2:0	15 (DVD: 9.8)	DVD, SD-DVB
	720 × 576	25			
MP@H-14	1440 × 1080	30	4:2:0	60 (HDV: 25)	HDV
	1280 × 720	30			
MP@HL	1920 × 1080	30	4:2:0	80	ATSC 1080i, 720p60, HD-DVB (HDTV)
	1280 × 720	60			
422P@LL			4:2:2		
422P@ML	720 × 480	30	4:2:2	50	Sony IMX using I-frame only, Broadcast "contribution" video (I&P only)
	720 × 576	25			
422P@H-14	1440 × 1080	30	4:2:2	80	Potential future MPEG-2-based HD products from Sony and Panasonic
	1280 × 720	60			
422P@HL	1920 × 1080	30	4:2:2	300	MPEG-2-based HD products from Panasonic
	1280 × 720	60			

From Wikipedia

Carl Fleischhauer  
Library of Congress

Put them together and you can associate profiles and levels with applications. An interesting wrinkle is that the *4:2:2 profile at High Level*--used for standard definition television and extendable to high def--was specified by the Society of Motion Picture and Television Engineers--SMPTE--after they found that ISO had not covered their needs.

# Target encodings

- Most reformatting requires playing back the videotape or file and transforming the existing signal into a *target* encoding format.
- For reformatting from existing videotapes; in time, may pertain to reformatting certain born digital files.

Carl Fleischhauer  
Library of Congress

**Now: about encodings:** First, there are *target* encodings. Many organizations are reformatting from existing videotapes, mostly analog, and occasionally digital, like DV. In time, we may begin reformatting some of the digital files that we receive. Most reformatting requires playing back the videotape and re-recording the output signal into a *target encoding format*.

## “Keeper” encodings

- Encodings for born digital that are sustainable (keepable) for several years as they stand.
- WGBH expert says “don’t waste effort transcoding now.”
- These encodings are not sufficiently appealing to serve as target formats.
- Examples later in talk.

Carl Fleischhauer  
Library of Congress

Second, there are “keeper” encodings. Some born digital content, on videotape or in files, employs encodings that are sustainable for a period of years, as-is: *keeper encoding formats*. My colleague Dave MacCarn at WGBH uses the elegant phrase “retaining acquisition bandwidth,” and he argues some encodings are good for several years—don’t waste effort transcoding now. But these encodings are not sufficiently appealing to serve as target formats.

## Component video


From Wikipedia, the free encyclopedia

*This article is about analog component video: for the processing of color components in digital video, see [digital video](#), [Chroma subsampling](#) and [YCbCr](#)*

**Component video** is a video signal that has been split into two or more components. In popular use, it refers to a type of [analog video](#) information that is transmitted or stored as three separate signals. Component video can be contrasted with [composite](#) video (such as [NTSC](#) or [PAL](#)) in which all the video information is combined into a single line level signal. Component video cables do not carry audio.

**Contents** [\[hide\]](#)

- 1 Analog component video
  - 1.1 RGB analog component video
  - 1.2 YPbPr analog component video
    - 1.2.1 Connectors used
  - 1.3 S-Video analog component video



Three cables, each with [RCA](#) [plugs](#) at both ends, are often used to carry analog component video

Carl Fleischhauer  
Library of Congress

**[Target formats.]** The discussion of target *encoding* formats generally begins with an unstated assumption: the signal arriving for reformatting is a component video signal. *Component video* is a bitstream in which luminance or brightness information is separate from color information.

## Composite video

From Wikipedia, the free encyclopedia

**Composite video**, also called **CVBS** (Composite Video Blanking and Sync), is the format of an [analog television](#) (picture only) signal before it is combined with a sound signal and [modulated](#) onto an [RF carrier](#).

Composite video is often designated by the **CVBS** acronym, meaning any of "Color, Video, Blank and Sync", "Composite Video Baseband Signal", "Composite Video Burst Signal", or "Composite Video with Burst and Sync".

It is usually in a standard format such as [NTSC](#), [PAL](#), or [SECAM](#). It is a composite of three source signals called Y, U and V (together referred to as [YUV](#)) with sync pulses. Y represents the brightness or *luminance* of the picture and includes synchronizing pulses, so that by itself it could be displayed as a monochrome picture. U and V between them carry the color information. They are first mixed with two orthogonal phases of a color carrier signal to form a signal called the *chrominance*. Y and UV are then combined. Since Y is a [baseband](#) signal and UV has been mixed with a carrier, this addition is equivalent to [frequency-division multiplexing](#).

Composite video



The [RCA connector](#) is the most common connector for composite video.

In contrast, *composite* video is represented by analog signals on older videotapes and the signals transmitted by broadcasters, until the great changeover to digital in February 2009. A composite signal blends the luminance and chrominance information.



## Component video


From Wikipedia, the free encyclopedia

*This article is about analog component video: for the processing of color components in digital video, see [digital video](#), [Chroma subsampling](#) and [YCbCr](#)*

**Component video** is a video signal that has been split into two or more components. In popular use, it refers to a type of [analog video](#) information that is transmitted or stored as three separate signals. Component video can be contrasted with [composite video](#) (such as [NTSC](#) or [PAL](#)) in which all the video information is combined into a single line level signal. Component video cables do not carry audio.

**Contents** [\[hide\]](#)

- 1 Analog component video
  - 1.1 RGB analog component video
  - 1.2 YPbPr analog component video
    - 1.2.1 Connectors used
  - 1.3 S-Video analog component video



Three cables, each with [RCA](#) [plugs](#) at both ends, are often used to carry analog component video

## Composite video


From Wikipedia, the free encyclopedia

**Composite video**, also called **CVBS** (Composite Video Blanking and Sync), is the format of an [analog television](#) (picture only) signal before it is combined with a sound signal and [modulated](#) onto an [RF carrier](#).

Composite video is often designated by the **CVBS** acronym, meaning any of "[Color](#), [Video](#), [Blank](#) and [Sync](#)", "Composite Video Baseband Signal", "Composite Video Burst Signal", or "Composite Video with Burst and Sync".

It is usually in a standard format such as [NTSC](#), [PAL](#), or [SECAM](#). It is a composite of three source signals called Y, U and V (together referred to as [YUV](#)) with sync pulses. Y represents the brightness or *luminance* of the picture and includes synchronizing pulses, so that by itself it could be displayed as a monochrome picture. U and V between them carry the color information. They are first mixed with two orthogonal phases of a color carrier signal to form a signal called the *chrominance*. Y and UV are then combined. Since Y is a [baseband](#) signal and UV has been mixed with a carrier, this addition is equivalent to [frequency-division multiplexing](#).

**Composite video**



The [RCA connector](#) is the most common connector for composite video.

Old videotapes are composite and their output signal must be transformed into component in order to use any of the encoding options that I'll describe in a minute. In fact, the composite-to-component transform has been part of video reformatting for years--it happens when you record to a SONY BetaCam videocassette.

**ADC-8032B - *Leitch Compatible***  
**ADC-8032B-S - *Leitch Compatible***  
**Analog Composite to SDI Decoder**

Superior quality analog-to-digital converter specially designed to handle tough microwave and satellite feeds as well as all general decoding requirements.

[What about the ADC-8032A and ADC-8032A-S products?](#)

```

graph LR
    subgraph Inputs
        Y[ANALOG IN: Y]
        M[ANALOG IN: M]
        C[ANALOG IN: C]
        S[ANALOG IN: S]
        L[ANALOG IN: L]
        R[ANALOG IN: R]
        EFR[EXTERNAL FRAME REFERENCE IN]
        CB[COLOR BURST REFERENCE IN]
        CL[COLOR BLANKING REFERENCE IN]
        CL_L[COLOR BLANKING REFERENCE LOOP]
    end

    subgraph Processing
        IFG[INPUT FILTER & GAIN] --> ADC[12 BIT ADC]
        ADC --> CAF[3D ADAPTIVE COMB FILTER]
        CAF --> LS[LINE SYNCHRONIZER]
        LS --> SDI[SDI OUTPUTS]
        LS --> TRF[15 TRACKING DELAY PULSES]
        TRF --> FS[FRAME SYNCHRONIZER 15 VERSION]
        FS --> TRF
    end

    subgraph References
        EFR --> MR[MASTER REFERENCE]
        CB --> MR
        CL --> MR
        MR --> CAF
    end

    TRF --> TO[TRACKING OUT]

```

- Advanced adaptive 3-D comb filter for near perfect composite decoding
- Oversampled 12-bit A-to-D conversion
- Designed to handle difficult, unstable signals such as off-air and VCR feeds
- On-board full frame synchronizer (ADC-8032B-S version)

I wish I understood the degree to which different transform methods yield different results. I have been surprised at how little this part of the process has been discussed within our community.

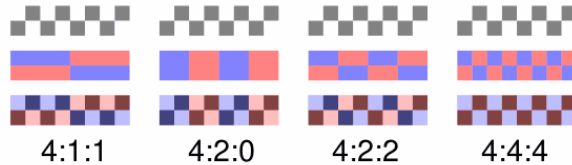
### Sampling systems and ratios

[edit]

The subsampling scheme is commonly expressed as a three part ratio (e.g. 4:2:2), although sometimes expressed as four parts (e.g. 4:2:2:4). The parts are (in their respective order):

- **Luma** horizontal sampling reference (originally, as a multiple of 3.579 MHz in the NTSC television system)
- **Cr** horizontal factor (relative to first digit)
- **Cb** horizontal factor (relative to first digit), except when zero. Zero indicates that **Cb** horizontal factor is equal to second digit, and, in addition, both **Cr** and **Cb** are subsampled 2:1 vertically. Zero is chosen for the bandwidth calculation formula (see below) to remain correct.
- **Alpha** horizontal factor (relative to first digit). May be omitted if alpha component is not present.

To calculate required bandwidth factor relative to 4:4:4 (or 4:4:4:4), one needs to sum all the factors and divide the result by 12 (or 16, if alpha is present).



The mapping examples given are only theoretical and for illustration. Also note that the diagram does not indicate any chroma filtering, which should be applied to avoid [aliasing](#).

From Wikipedia

Carl Fleischhauer  
Library of Congress

The sampling of component video can vary. A 4:4:4 specification means that there are equal amounts of brightness and color information. Most professional video systems work in a 4:2:2 mode, with half as much color information, and your consumer camcorder is very likely to be 4:2:0, with even less color information. Video data may also be at 8 or 10 bits per sample; the higher the better.

# Target encoding categories

- Uncompressed
- Lossless compressed
- Lossy compressed

Carl Fleischhauer  
Library of Congress

Generally speaking, existing preservation-oriented projects treat a 4:2:2 or 4:2:0 video signal, downstream from the composite-component transform, and their encoding falls into three broad categories: uncompressed, lossless compressed, and lossy compressed.

# Uncompressed video

- Stanford, Rutgers, NARA (early planning)
- 4:2:2 or 4:4:4 (NARA), 10-bit SDI stream
- About 100 GB per content-hour
  - Another source reported 70 GB for 8-bit video

## Recommended Standards for NJDH and RU-CORE Video Digitization

For preservation masters:

File format: Uncompressed, Full Frame Video (AVI file format)

Frame rate for analog Standard Definition (SD) video, NTSC: 29.97 frames per second, 640 x 480 resolution (assuming square pixels), 4:2:2 quantization, 30MiB/s data rate.

We recognize this sampling scheme as the minimum acceptable rate to ensure a good preservation master of analog SD video archives, and will be the most common sampling rate for objects that come to us as SD analog video. This standard is based on our experiences with digitizing S/VHS video objects.

RUCORE Media Standards Working Group:  
I. Beard, I. Bogus, E. Corder, N. Gonzaga, B. Nahory, R. Sandier

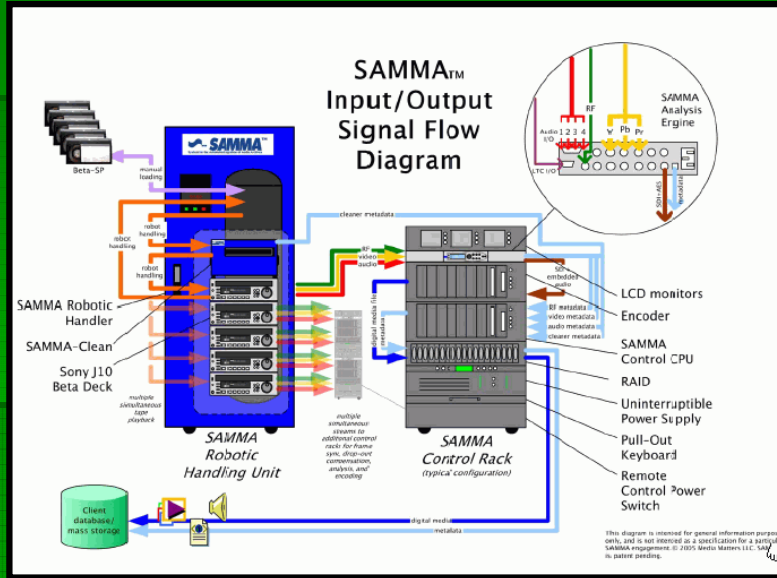
RUCORE and NJDH Standards Analysis for Moving Image Objects  
Draft Revision 3 - Last Modified 6 April 2007

Rutgers spec: [http://rucore.libraries.rutgers.edu/collab/ref/dos\\_avwg\\_video\\_obj\\_standard.pdf](http://rucore.libraries.rutgers.edu/collab/ref/dos_avwg_video_obj_standard.pdf)

Carl Fleischhauer  
Library of Congress

Projects at Stanford—overseen by Hannah Frost—and at Rutgers—overseen by Isaiah Beard—save the incoming signal without further compression. And I just learned that a planning group at the National Archives has made a very similar recommendation for their next phase of work. You could see this as the equivalent of saving uncompressed still image information in a TIFF file. My understanding is that this approach yields files on the order of 70-100 GB per hour of program time, depending on whether the incoming signal was 8 or 10 bits deep. Given that this is uncompressed data, my sense is that profiles are less critical. (Comment welcome.)

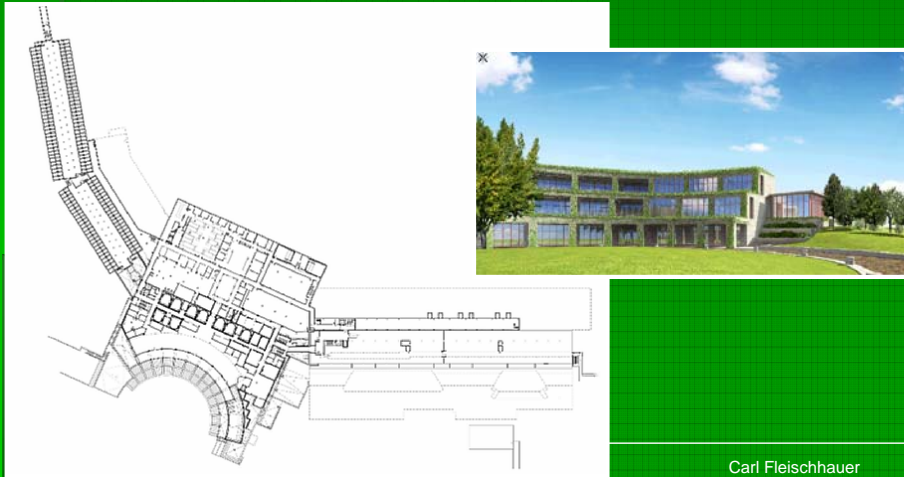
# Lossless compressed



Carl Fleischhauer  
Library of Congress

A second approach is to compress the picture using a lossless algorithm, the equivalent of saving your still image with LZW compression in a TIFF file. The leading proponent for this is Jim Lindner, whose company has developed an integrated system called SAMMA.

# Library of Congress Packard Campus, Culpeper



Carl Fleischhauer  
Library of Congress

The Library of Congress Motion Picture, Broadcasting, and Recorded Sound Division is beginning to implement SAMMA in the new facility in Culpeper, Virginia.

# Lossless compressed

- Each frame is a JPEG 2000 image
- Lossless (reversible) transform
- If 8-bit, 25-35 GB per content-hour
- If 10-bit, 35-50 GB per content-hour

Carl Fleischhauer  
Library of Congress

In this system, each video frame is compressed with the reversible (lossless) transform offered by the JPEG 2000 standard. Ian Gilmour, a member of the SAMMA team, reckons that 8-bit video will compress to something like 25-35 GB per hour; in one set of early tests, 10-bit came in at 35-50 GB per hour. For this type of compression, defined profiles would be very welcome but have not yet been developed.



# Lossy compressed

- Lossy JPEG 2000
  - Used by digital cinema
  - Provided by some new cam-corders, e.g., Infinity
  - No reformatting examples



\* Supports DV25, Infinity JPEG 2000, and MPEG-2\* compression schemes

Carl Fleischhauer  
Library of Congress

A third approach is to apply lossy compression to the picture information. You could do this using the irreversible transform in JPEG 2000, the approach used in the new digital cinema specification—movies for theaters. Although there is some uptake for lossy JPEG 2000 in born digital video in new cameras (like the Infinity), I have not encountered this encoding in archiving and have no estimates of possible file sizes.

# Lossy compressed

- MPEG-2
  - ITU-T H.262
  - In the ATSC digital TV standard
- MPEG-4
  - ITU-T H.263 and H.264
  - May come to play a bigger role as high-resolution increases

GUIDE  
TO THE USE OF THE  
ATSC DIGITAL TELEVISION STANDARD

5.1.2 Compatibility with MPEG-2

The video compression system does not include algorithmic elements that fall outside the specifications for MPEG-2 Main Profile. Thus video decoders which conform to the MPEG-2 MP@HL can be expected to decode bit streams produced in accordance with the Digital Television Standard. Note that it is not necessarily the case that all video decoders which are based on the Digital Television Standard will be able to properly decode all video bit streams which comply to MPEG-2 MP@HL.

Carl Fleischhauer  
Library of Congress

Today, the most frequently selected lossy compression encoding is MPEG-2 (aka H.262 in Europe). MPEG-2 has legs because it is part of the ATSC digital television standard, guaranteeing it a place in professional work for several more years. In time, one of the MPEG-4 schemes (H.263 or H.264 in Europe) may come into play. On paper, very high quality H.264 signals are possible but most applications today are lower quality, for mobile devices and home satellite delivery.

# SONY IMX, MPEG-2 @ 50 mbps

Sony's IMX Format  
by Alistair Jackson

This article first appeared in Digital  
Media World magazine October 2002

From: <http://www.edithouse.com.au/information/imx.html>

However, Sony has cleverly taken advantage of the fact that while an MPEG stream can be made up of a series of I, P and B frames, it doesn't have to be. The standard simply says that a GOP must start with an I-Frame, which can then be followed by P or B Frames. The Betacam SX format creates MPEG-2 GOPs of only two frames - one I and one B. The higher quality IMX format has only one picture to a GOP - a single I-Frame.

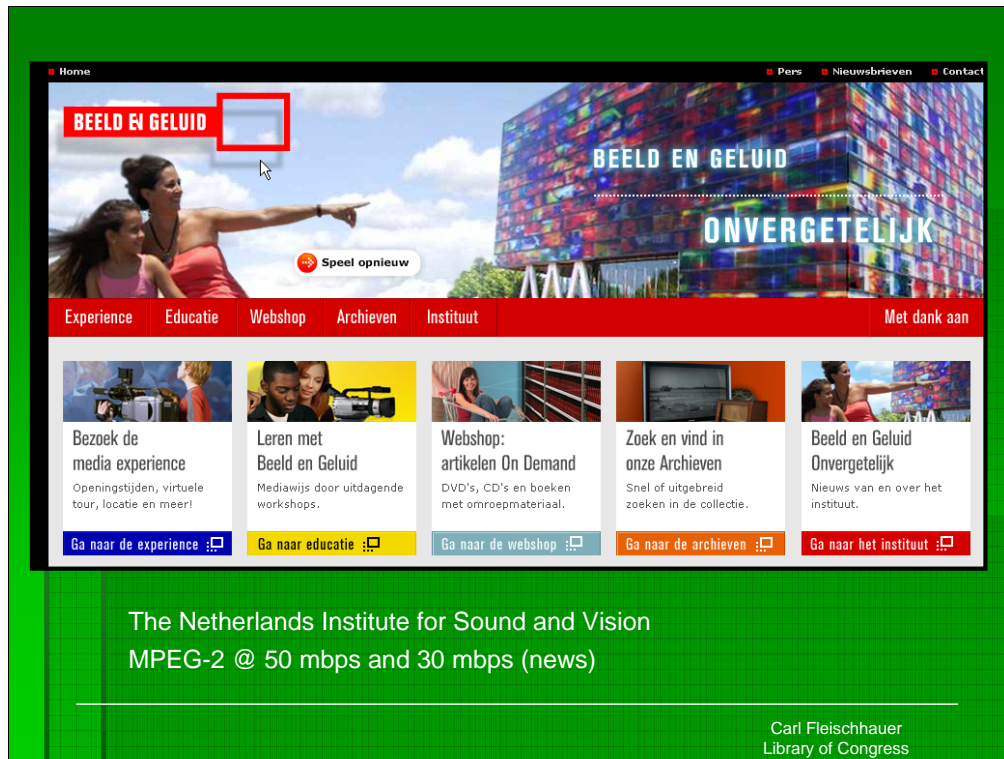
By only using I-Frames, IMX does not have an issue with edits. In fact, we are back in the same ballpark as DV based formats. However, in this case we have an MPEG-2 compliant stream. The idea is that you can load this tape footage onto a disk, and you end up with an MPEG-2 file. It is not as small as an MPEG-2 file that takes advantage of P and B Frames, but it is compliant with the standard.

IMX is seen by Sony as a key element for its MXF (Material eXchange file Format) vision for converging broadcast quality video into an IT infrastructure. A crucial part of this concept is the eVTR board, which allows IMX machines to interface to an Ethernet network. This allows for VTR control and for transfer of Audio and Video over a LAN, WAN, or even the Internet. The board buffers several frames from the tape, and if necessary pauses the tape until the buffer requires refilling.

- MPEG-2, all I-frames, 50 mbps
- File size about 28 GB/hour

Carl Fleischhauer  
Library of Congress

A number of broadcast organizations have selected MPEG-2 for archiving. SONY is a major player in professional circles and the widespread use of SONY IMX recorders has led to the acceptance of MPEG-2 with a data rate of 50 megabits per second as a benchmark. This encoding consists of all I-frames, meaning that each frame is fully represented in the data, and it yields files on the order of 28 GB per hour. 50 megabit MPEG-2 is often called a "contribution format" because producers use it to contribute content to a television network.



There is some use of MPEG-2 for archiving in Europe. For example, I just read an account of work at the Netherlands Institute for Sound and Vision. They have begun to digitize their extensive collection, with funding from the Dutch government, using MPEG-2 at 50 Mbps for much of their television material, and 30 mbps for news.

# Lossy compression

- |  |   |
|--|---|
| <ul style="list-style-type: none"><li>■ Lossy JPEG 2000<ul style="list-style-type: none"><li>■ Better quality</li><li>■ Less mature, less adoption</li><li>■ Few profiles</li><li>■ . . . may grow in the future??</li></ul></li></ul> | <ul style="list-style-type: none"><li>■ MPEG-2<ul style="list-style-type: none"><li>■ Lower quality</li><li>■ More mature, widely adopted</li><li>■ Profiles and levels well established</li><li>■ . . . may not grow in the future??</li></ul></li></ul> |
|--|---|

Carl Fleischhauer  
Library of Congress

How do specialists in the field compare the two lossy contenders, JPEG 2000 and 50 mbps MPEG-2? My sense is that they feel that lossy JPEG 2000 compression can provide better quality, but that its implementation is less mature and, except for digital cinema, moving image profiles are lacking. In contrast, MPEG-2 is very mature, widely implemented, and has well defined profiles, but with lower image quality. The pendulum may be swinging, however: increased high-def production may lead to more use of lossy JPEG 2000.

# Encoding preferences

- For high value, uncompressed or lossless compressed is very attractive.
- For second-rank content, some make a good case for modest-but-lossy compressed.

Carl Fleischhauer  
Library of Congress

How shall we vote on reformatting target encodings overall? For high value content, it is hard not to be drawn to uncompressed or lossless encodings, the latter adding complexity to the bitstream but reducing storage requirements significantly. For second-rank content, some will make a case for modest-but-lossy compression, to further reduce storage requirements or for other practical reasons.

# Born digital encoding



Coca-Cola:

MPEG-2 @ high  
streaming quality



Vanderbilt:

MPEG-2 @ 6 mbps

Carl Fleischhauer  
Library of Congress

**[Born digital encodings.]** On the born digital side, when transcoding is necessary, the same target options recur: uncompressed, lossless compressed, and lossy compressed. But what are examples of the *keeper encodings* I mentioned earlier? In the Motion Picture, Broadcasting, and Recorded Sound Division, two important acquisitions with digital elements began arriving during the 1990s: the Coca-Cola advertising collection and the Vanderbilt University television news collection. In both cases, Library staff conferred with the donors and the outcome was appropriately conservative for that time: MPEG-2 files at varying (but moderate) levels of resolution.



[About SCOLA](#)
[Technical Information](#)
[Educators](#)
[Program Schedules](#)
[Contact Us](#)
[Partners](#)
[Legal](#)
[Logoff](#)

[Home](#)  
[WORLD TV ONLINE](#)  
[SCOLA Channel One](#)  
[SCOLA Channel Two](#)  
[SCOLA Channel Three](#)  
[SCOLA Channel Four](#)  
[SCOLA Channel Five](#)  
[SCOLA Channel Six](#)

### SCOLA World TV Online and Week In Review

It is now **11 05 Central Time**  
[GO TO LIVE BROADCAST](#)

#### Channel Five Programming

SCOLA Channel Five features programming from continental Africa. More than 20 African dialects are offered from nations across Africa.

Tuesday 10/23/2007	Wednesday 10/24/2007	Thursday 10/25/2007	Friday 10/26/2007	Saturday 10/27/2007	Sunday 10/28/2007	Monday 10/29/2007
(Tuesday through Sunday are downloadable Archives from the previous week)						
Central Time	Country	Program				Downloads
0010--0045	Zambia (Bemba)	Daily News: Local Language News and Main News from Lusaka				<a href="#">download</a>
0045--0215	Tanzania (Swahili and Kiswahili)	Nightly News: Nightly News from Dar Es Salaam				<a href="#">download</a>
0215--0245	Uganda (Luganda)	Nightly News: Nightly News in Luganda from Kampala				<a href="#">download</a>
0245--0315	Nigeria (Igbo)	Evening News: from Abuja in Igbo				<a href="#">download</a>
0315--0715	Ethiopia (Amharic and Oromifa)	News: Awude in Amharic or the Oromifa Dialect from Addis Ababa				<a href="#">download</a>
0715--0815		SCOLA International Showcase: Varied Programming and Topics from a Different Country Every Day				<a href="#">download</a>
0815--0845		Kids Around the World: Children's Variety Programs from Around the World				<a href="#">download</a>
0845--1015	Ivory Coast (French)	Daily News: Le Journal in French from Yamoussoukro				<a href="#">download</a>
1015--1415	Sudan (Modern Standard Arabic-MSA)	Daily News: Daily News from Khartoum				<a href="#">Watch Now!</a>

[International Radio](#)  
[Foreign Text](#)  
[People and Places](#)  
["On The Street" Videos](#)  
[Specialized Word Video](#)  
[Search](#)

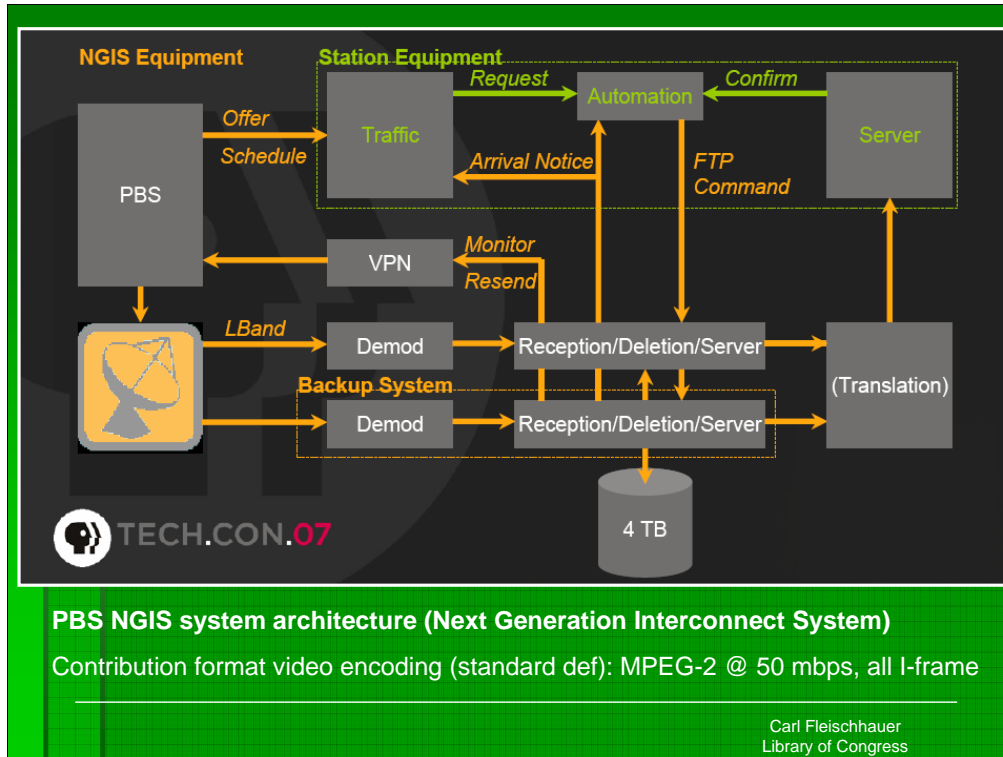
[Insta-Class](#)  
[Insta-Class](#)  
[Insta-Class Schedule](#)  
[Back](#)

SCOLA: MPEG-4, part 2 video compression @ 200 kbps  
 Planned upgrade in 2008: MPEG-2 @ 5 mbps

Carl Fleischhauer  
 Library of Congress

Meanwhile, in a project under the auspices of the National Digital Information Infrastructure and Preservation Program (NDIIPP), the Library is receiving large numbers of foreign news broadcasts with MPEG-4 "part 2" encoding, at Internet-streaming levels of quality. We believe that this encoding will also be sustainable for the next several years.





Another NDIIPP project connects us to a higher-end solution being devised in the public television community. Using PBS's new interconnect system, the producers of public TV content plan to contribute finished standard-definition programs as 50 mbps MPEG-2 files. The PBS team is developing an application specification for these files.

## From a report by Dave MacCarn, WGBH

This footage can be any number of digital formats. Since source material can also come from the Archives, this may include analog material, but for the purpose of this discussion, comments will be restricted to digital formats only. At the time this paper was written the physical formats most commonly used are (*see definitions below*):

- DV
- DVCAM
- DVCPRO
- Digital Betacam
- D3
- D5

Newer formats for standard definition:

- XDCAM

For HDTV it's:

- DVCPRO HD
- HDCAM
- HDV
- XDCAM HD

<http://www.ptvdigitalarchive.org/wp-content/uploads/2007/09/report-on-file-formats-and-packages-fin.pdf>

Carl Fleischhauer  
Library of Congress

The NDIIPP public television project is also looking at acquisition formats, for the footage recorded during program production, some of which are in keeper encodings. This is Dave MacCarn's list of the professional-quality encodings WGBH encounters at the acquisition stage. In conversations with me, Dave has said that most will be sustainable for several years.

Overall, this strategy seems promising: identifying native encodings that are safe to keep as-is (for several years) and distinguishing them from encodings that cry out for transcoding upon arrival.

# MPEG wrappers

- MPEG-2
  - No file wrapper established by standards body
  - De facto file format convention in wide use
    - .mpg extension (also for MPEG-1)
- *By the way: MPEG-4 has two standardized file formats, both based on QuickTime (as is JPEG 2000); .mp4 extension for both*

Carl Fleischhauer  
Library of Congress

**Now, about wrappers.** Some activities are proceeding in a no-wrapper mode, just storing files. We store the Coca-Cola and Vanderbilt content as MPEG-2 files, for example. There is no “legal” standardized file format for MPEG-2, although the de facto format is widely supported.

# MXF Wrapper

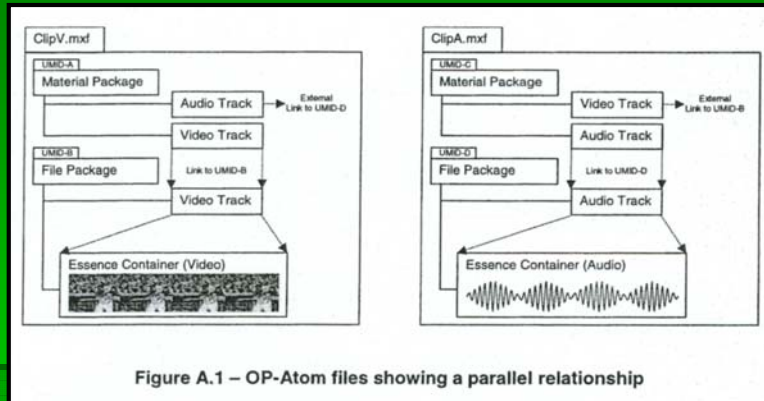


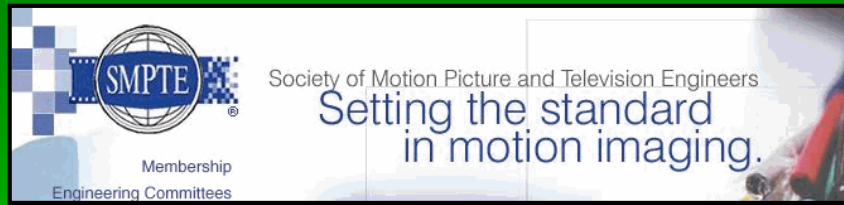
Diagram for the simple “OP-Atom” structure, from SMPTE spec 390M

BTW: AAF (Advanced Authoring Format) is a close cousin to MXF.

Carl Fleischhauer  
Library of Congress

In professional circles, the wrapper buzz these days concerns the Material Exchange Format (MXF), standardized by SMPTE. MXF is an object-based file format that bundles video, audio, timecode information, closed captions, and what amounts to an "edit decision list." Complexity of structure is categorized by what are called *operational patterns*.

## A few examples of MXF specifications . . .



<b>SMPTE 377M-2004</b>	Television Material Exchange Format (MXF) File Format Specification (Standard) \$90.00 - <a href="#">Purchase this Document</a>
<b>SMPTE 378M-2004</b>	Proposed Material Exchange Format (MXF) — Operational pattern 1A (Single Item, Single Package) \$26.00 - <a href="#">Purchase this Document</a>
<b>SMPTE 379M-2004</b>	Material Exchange Format (MXF) — MXF Generic Container \$30.00 - <a href="#">Purchase this Document</a>

Carl Fleischhauer  
Library of Congress

MXF is intended to support content interchange between creators and distributors, and to be implemented in cameras, recorders, and computer systems. It is used in the digital cinema specification. MXF is complicated and new: between 2004 and the present, SMPTE has published more than thirty specification documents.

The diagram illustrates the SAMMA (System for Archiving Moving Media) system components. It shows three main units: a front rack (left), a side rack (middle), and a rear rack (right). Each rack is 24 inches wide and 28 inches deep. The front rack is 42 inches high, while the side and rear racks are 42 inches high. The side rack is 42 inches wide, and the rear rack is 24 inches wide. A top-down view shows the system's footprint, which is 42 inches by 28 inches. The website screenshot shows the 'PRESERVING DIGITAL PUBLIC TELEVISION' project, a project funded by the Library of Congress. It features a navigation menu with links to Home, Purpose, Our Work, Events, Partners, Links, FAQ, and Contact Us. The main content area includes a section titled 'Current.org on archiving PBS programs' with a quote from a Current article, and another section titled 'The 51st State: New Finding Aid On-Line for 1970's Iconoclastic NYC News Magazine' with a quote from a New York Times article. The website is designed with a blue and white color scheme and includes a search bar and a 'search blog authors' button.

<http://www.sammasystems.com/>

<http://www.ptvdigitalarchive.org/>

Carl Fleischhauer  
Library of Congress

MXF is gaining momentum. In our archiving circle, there are two important MXF adoptions: in Jim Lindner's SAMMA system and by the NDIIPP public television team, each with its own encoding.

# From the working draft for the first PBS MXF application specification

## 1.4 Operational Pattern Constraints

### 1.4.1 Baseline Operational Patterns

AS-PBS files shall comply with MXF Operational Pattern OP1a.

Note: one exception may be where it is required to provide indication of SCTE-35 Splice Points or decoder precharge for seamless splicing of video in constraint sets other than SD Distribution. In this case OP2a shall be used.

### 1.4.2 Additional Constraints

#### 1.4.2.1 Container

AS-PBS Files shall use the MXF Generic Container Structure. The Number of Elements in each GC is defined for each

#### 1.5.1.5 Video

Video shall be MPEG-2 MP:ML 4:2:0 GOP 15, at bit rates of 8 up to 15 Mbps, in compliance with ISO 13818-2 Elementary Streams.

The video stream shall be carried in a SMPTE 381M-2005-compliant MXF GC Element.

#### 1.5.1.6 Audio

Audio shall be 48kHz, 16 bit PCM, with 2 or 4 channels.

Each pair of channels shall be carried in a SMPTE 382M-2007-compliant MXF GC Element within a BWF Container (not AIFF container).

#### 1.5.1.7 Closed Captioning

If present, CEA 608 line 21 (CC and XDS) data shall be carried in a SMPTE 334-1- and -2-2007-compliant ANC packet within a SMPTE 436M-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- and -2-2007-compliant ANC packet within a SMPTE 436M-2006-compliant VBI/ANC GC Element, using 8 bit encoding.

Carl Fleischhauer  
Library of Congress

The public television folks are attending to profiling: they have drafted one MXF *application specification* (“AS”) for their moderate-resolution *distribution* files, and PBS is about to draft an AS for the high-res *contribution* files.

## Other wrappers

- Motion JPEG 2000
  - ISO/IEC 15444-3 (part 3)
  - Not aware of use by broadcasters or archivists
- Relatively open, documented proprietary standards
  - QuickTime, used at WGBH
  - AVI, used at Rutgers

Carl Fleischhauer  
Library of Congress

Meanwhile, there seems to be little or no uptake for ISO's Motion JPEG 2000, a wrapper designed for use with JPEG 2000 frame encoding. The folks I talk to use MXF to wrap JPEG 2000 frames instead.

Another meanwhile: broadcasters and archivists sometimes employ proprietary wrappers, several of which have relatively open, mostly public specifications. WGBH uses the QuickTime wrapper while they wait for better tools to support MXF. For the time being, Isaiah Beard at Rutgers wraps his uncompressed files in AVI, an open spec from Microsoft and IBM.



# Video Object Metadata

- Compare to NISO image data, MIX XML schema
- Compare to AES (Audio Engineering Society) AES-X098B: Audio Object Schema (in final draft)
- Many definitions could come from SMPTE RP-210 registry of terms
- On the right track: PB Core Instantiation (box at right)

PB Core	
-- INSTANTIATIONS --	
25.00	<i>pbcoreInstantiation</i>
25.01	• <i>dateCreated</i>
25.02	• <i>dateIssued</i>
25.03	• <i>formatPhysical</i>
25.04	• <i>formatDigital</i>
25.05	• <i>formatLocation</i>
25.06	• <i>formatMediaType</i>
25.07	• <i>formatGenerations</i>
25.08	• <i>formatStandard</i>
25.09	• <i>formatEncoding</i>
25.10	• <i>formatFileSize</i>
25.11	• <i>formatTimeStart</i>
25.12	• <i>formatDuration</i>
25.13	• <i>formatDataRate</i>
25.14	• <i>formatBitDepth</i>
25.15	• <i>formatSamplingRate</i>
25.16	• <i>formatFrameSize</i>
25.17	• <i>formatAspectRatio</i>
25.18	• <i>formatFrameRate</i>
25.19	• <i>formatColors</i>
25.20	• <i>formatTracks</i>
25.21	• <i>formatChannelConfiguration</i>
25.22	• <i>language</i>
25.23	• <i>alternativeModes</i>
25.24	• <i>pbcoreDateAvailable</i>
25.24.1	• <i>dateAvailableStart</i>
25.24.2	• <i>dateAvailableEnd</i>
25.25	• <i>pbcoreFormatID</i>
25.25.1	• <i>formatIdentifier</i>
25.25.2	• <i>formatIdentifierSource</i>
25.26	• <i>pbcoreAnnotation</i>
25.26.1	• <i>annotation</i>

Carl Fleischhauer  
Library of Congress

**[Metadata.]** Let me close with a snapshot of three metadata subcategories. The first concerns the technical characteristics of the video *object* at hand, comparable to the NISO data set for *still images*, aka MIX, and to the *audio object* metadata specification from the Audio Engineering Society (AES). The closest video equivalent that I have seen is public broadcasting's new PB Core specification, which includes a section called *instantiation*.

# Video Provenance Metadata

- Compare to AES (Audio Engineering Society) AES-X098C: Process History Schema (in draft)
- Not aware of exact video example
- SAMMA logging metadata similar-but-different
  - Box contains 1 percent of one example

A screenshot of a video provenance metadata log, likely from a SAMMA system. The log is displayed in a window with a light gray background and a black border. It shows a series of XML-like tags and their values, including timestamps, source tape information, and processing details. The log is organized into sections, with some tags like <Log>, <Contents>, <ElapsedTime>, <SampleCount>, and <Data> being clearly visible. The data section contains a long list of attributes and their values, such as Index, Frame Count, Input Present, Timecode, Timecode User Groups, Color Present, RF Level, RF Level Warning, RF Level Alarm, RF Head Difference Warning, DOC Length, DOC Number, DOC Warning, DOC Alarm, Missing H-Pulse Count, Noise Level, Motion Detect, Luma Average, Luma Peak, Pb Max, Pb Min, Pb Average, and Pb Pr.

```
131 | <Log>
132 | | <Contents> Starting migration at 09:43:43 Thu Sep 27 2007 \v0D\v0A Rewinding source tape... \v0D\v0ASource
    | tape stopped \v0D\v0AMotion JPEG2000 cued for record OK \v0D\v0AMPEG2 cued for record OK \v0D\v0AEnabled
    | metadata \v0D\v0AStarted metadata analysis \v0D\v0APlaying source tape... \v0D\v0AMotion JPEG2000
    | recording... \v0D\v0AMPEG2 recording... \v0D\v0AINPUT threshold triggered normal End of Program \v0D\v0AStopping
    | migration \v0D\v0AStopped metadata analysis \v0D\v0ADisabled metadata \v0D\v0AStopped source
    | tape \v0D\v0AMotion JPEG2000 Stopped \v0D\v0AMPEG2 Stopped \v0D\v0ASource tape stopped \v0D\v0AEjecting
    | source tape... \v0D\v0ASource VTR cassette empty \v0D\v0AMotion JPEG2000 Transferring file... \v0D\v0AMPEG2
    | Transferring file... \v0D\v0AMPEG2 Post processing completed successfully \v0D\v0AMotion JPEG2000 Post
    | processing completed successfully \v0D\v0AMigration completed SUCCESSFULLY at 10:28:07 Thu Sep 27
    | 2007 \v0D\v0A </Contents>
133 | | <ElapsedTime> 00:33:57 </ElapsedTime>
134 | | <SampleCount> 57961 </SampleCount>
135 | | <Log>
136 | | <Data (X="Index" TC="Frame Count" IN="Input Present" TT="Timecode" TB="Timecode User Groups" CB="Color
    | Present" RF="RF Level" RW="RF Level Warning" RA="RF Level Alarm" HW="RF Head Difference Warning" DL="DOC
    | Length" DN="DOC Number" DW="DOC Warning" DA="DOC Alarm" MH="Missing H-Pulse Count" NR="Noise Level"
    | MD="Motion Detect" LA="Luma Average" LP="Luma Peak" UP="Pb Max" UM="Pb Min" UA="Pb Average" VP="Pr
```

For audio, AES has been working on what we sometimes call “digital provenance” metadata, to document what we did to make the file we have. There’s nothing quite like this in the video arena, although the SAMMA system collects extensive logging data and encodes it as XML.

# Video Preservation Metadata

- Same as or akin to PREMIS
- Being explored for video by NDIIPP public television project
  - Kara Van Malssen

Carl Fleischhauer  
Library of Congress

The NDIIPP-funded public television project will contribute on the metadata front (Kara van Malssen is on this panel), including ideas about special metadata to support long-term preservation, akin to PREMIS. There's a long history here --

# Video Preservation Metadata

- NDIIPP public tv project team includes Dave MacCarn
- Co-author of Universal Preservation Format documents beginning in 1996

**The Universal Preservation Format**  
A Recommended Practice for Archiving Media and Electronic Records.

## 2. Definition

The Universal Preservation Format is a data file mechanism that utilizes a container or wrapper structure. Its framework incorporates metadata that identifies its contents within a registry of standard data types and serves as the source code for mapping or translating binary composition into accessible or useable forms. The UPF is designed to be independent of the computer applications used to create content, independent of the operating system from which these applications originated and independent of the physical media upon which that content is stored. The UPF is characterized as "self-described" because it includes, within its metadata, all the technical specifications required to build and rebuild appropriate media browsers to access contained materials throughout time. Objects within the UPF are branded with a unique identifier that travels with that object throughout time. Any modification made to the content of the object must be reflected in its identifier.

-- in the mid-1990s, Dave MacCarn of WGBH co-authored documents pertaining to what was called the Universal Preservation Format, pioneering a concept very close to the OAIS archival information package. Today, Dave continues to promote the importance of preservation packages that incorporate encoding specifications and even content-playing applications.



*Thomas Jefferson Building  
Library of Congress*

Carl Fleischhauer  
Library of Congress

So . . . *metadata* seems to me to be like the other elements I have discussed: work is under way but there is still plenty to do!

Thank you very much.