[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.

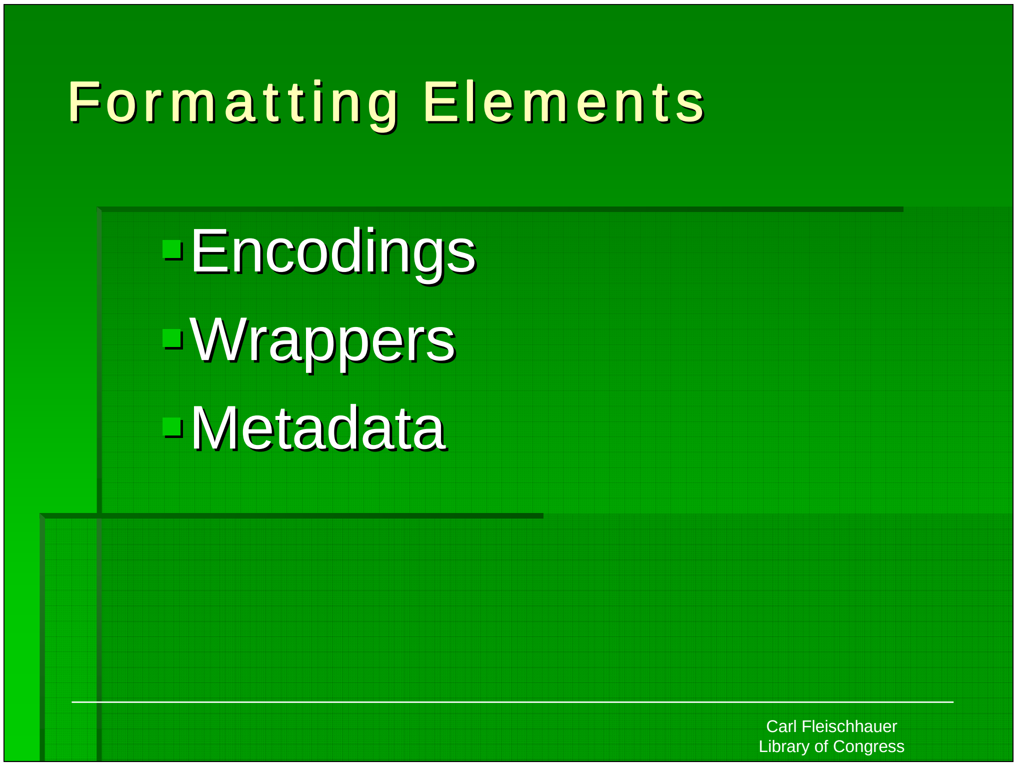
I divide born digital into two classes. First, there is professional born digital: broadcasting and the professional end of documentary and independent production. Second, there is the often-inspired, prosumer work so significant today on Web sites like YouTube. My news gathering found *some* things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

Allies in the Industry

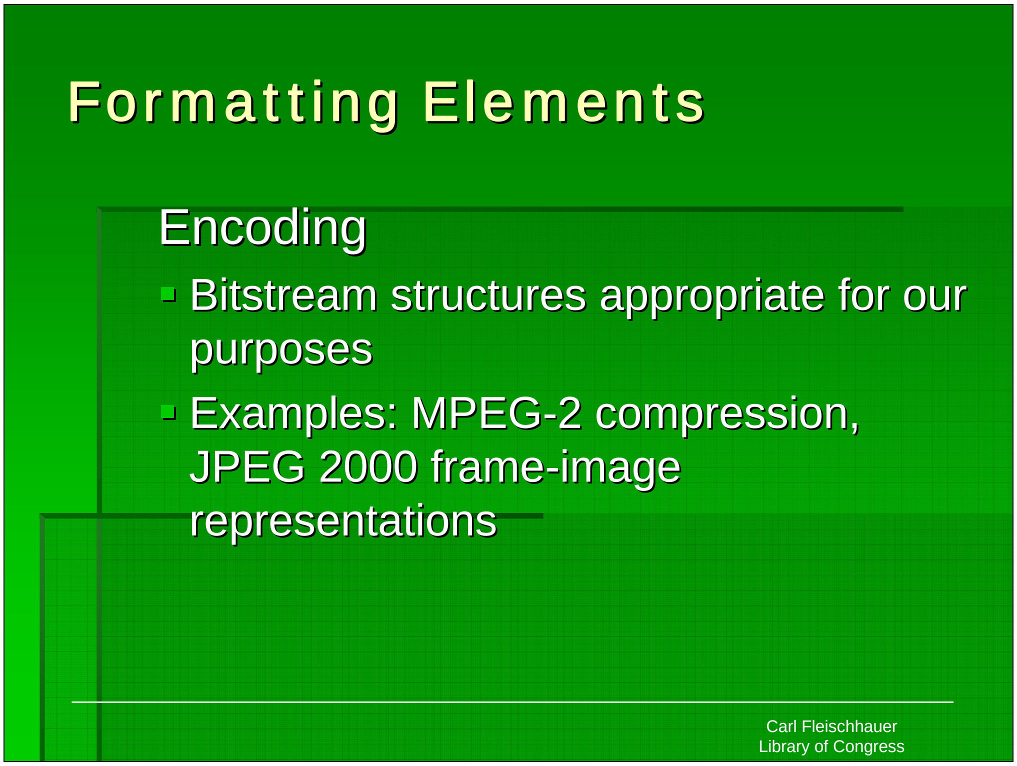
- Producers, creators seek interoperability via standards
  - May be from “true” or industry standards bodies, MPEG-2, MPEG-4, MXF, AAF
  - May be relatively open and public, albeit proprietary, AVI, QuickTime

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.



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



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.

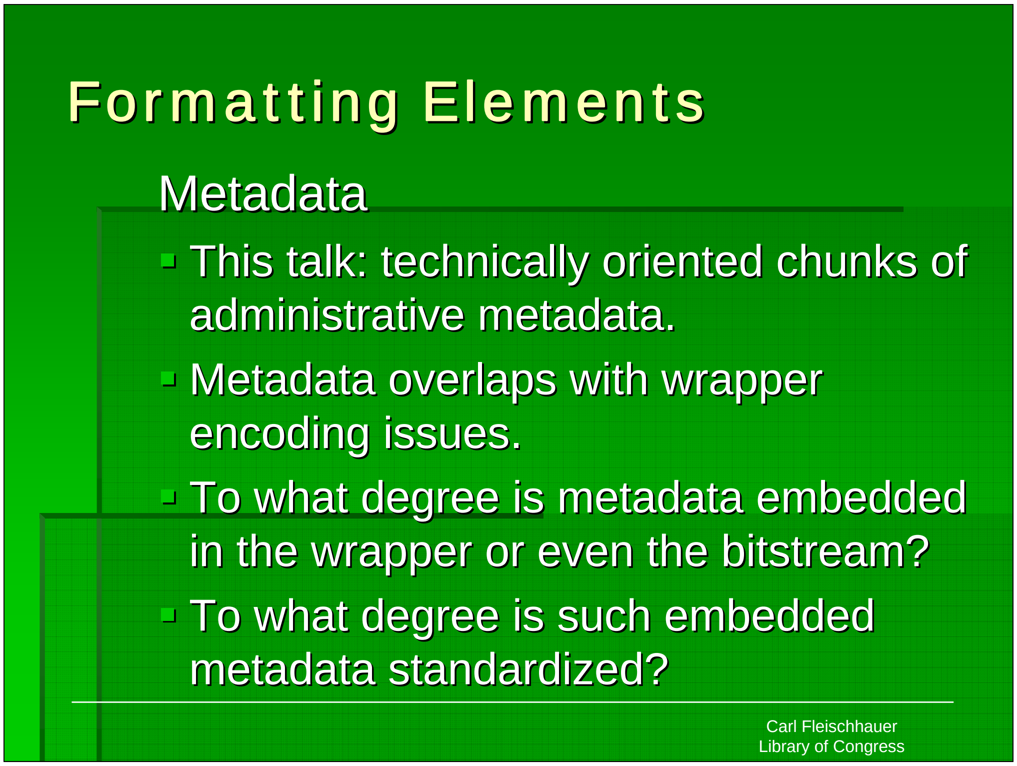
Formatting Elements

Wrapper

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

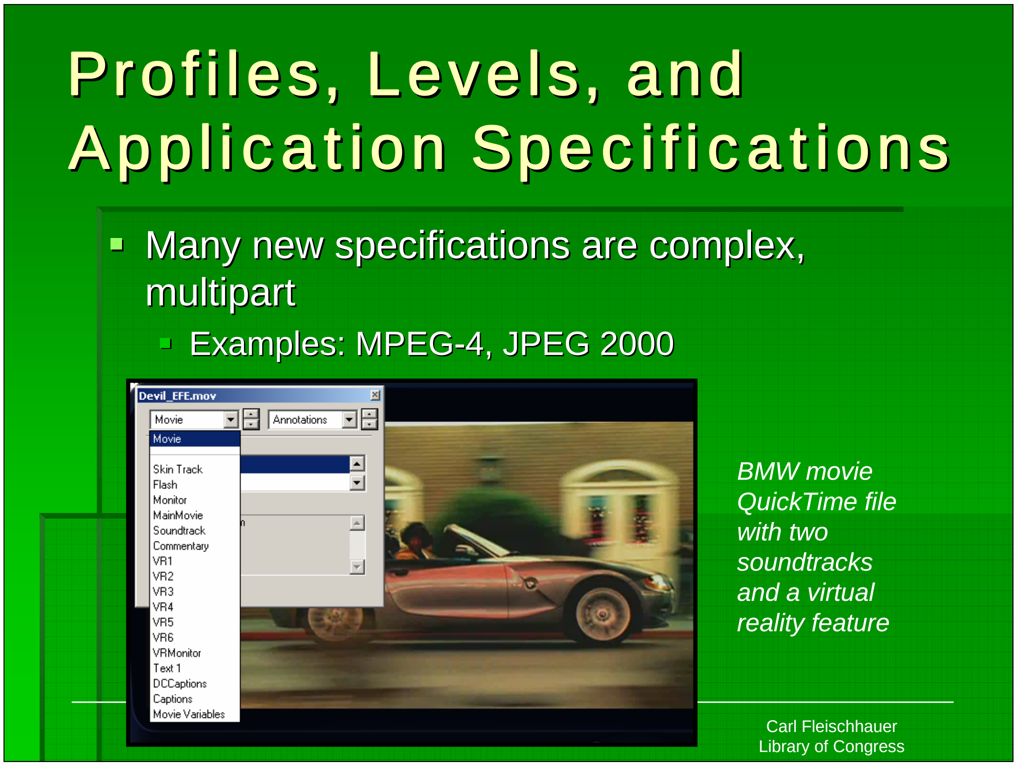
Carl Fleischhauer
Library of Congress

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.

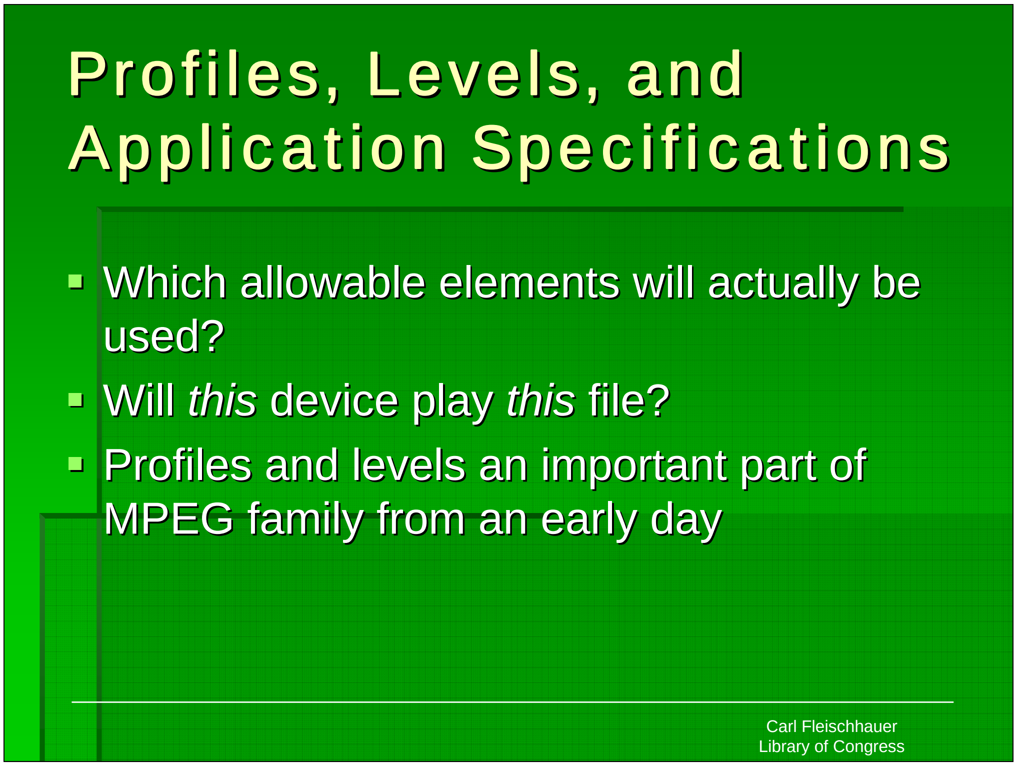


Regarding metadata, my talk will emphasize technically oriented chunks of administrative metadata. To what degree is metadata embedded in the wrapper or even the bitstream? To what degree is such embedded metadata standardized?

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



[Profiles 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.



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	low, normal and high quality decoding
HP	High Profile	P, I, B	4:2:2	1-3	

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.

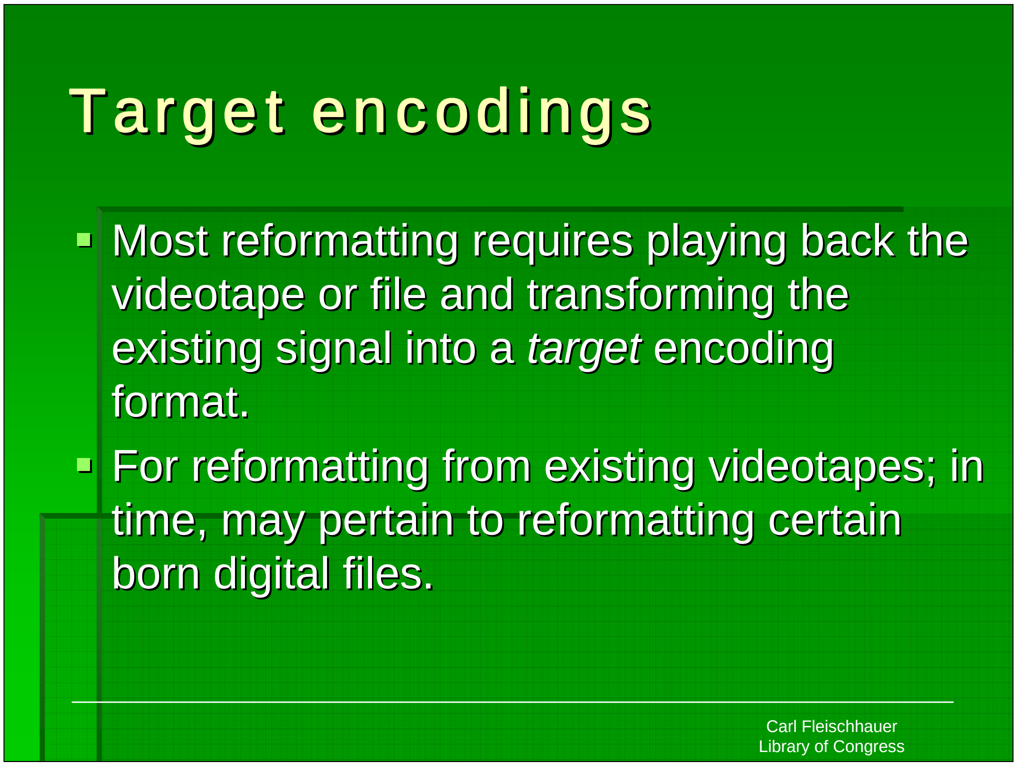
MPEG-2 Profiles and Levels

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			

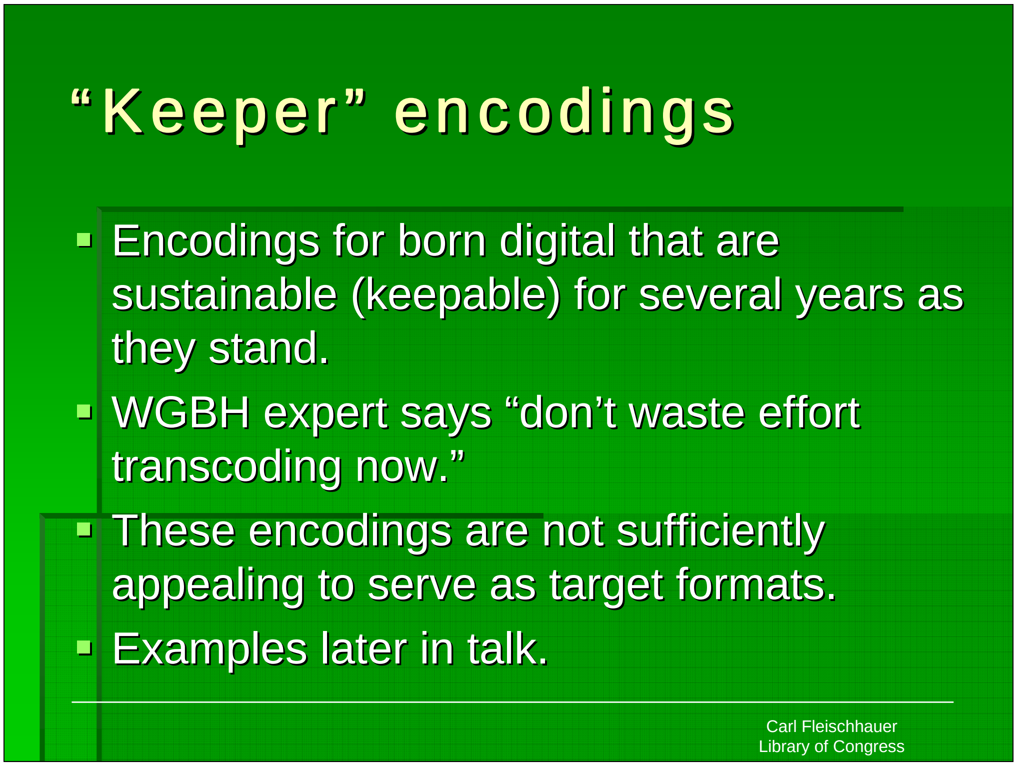
From Wikipedia

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.

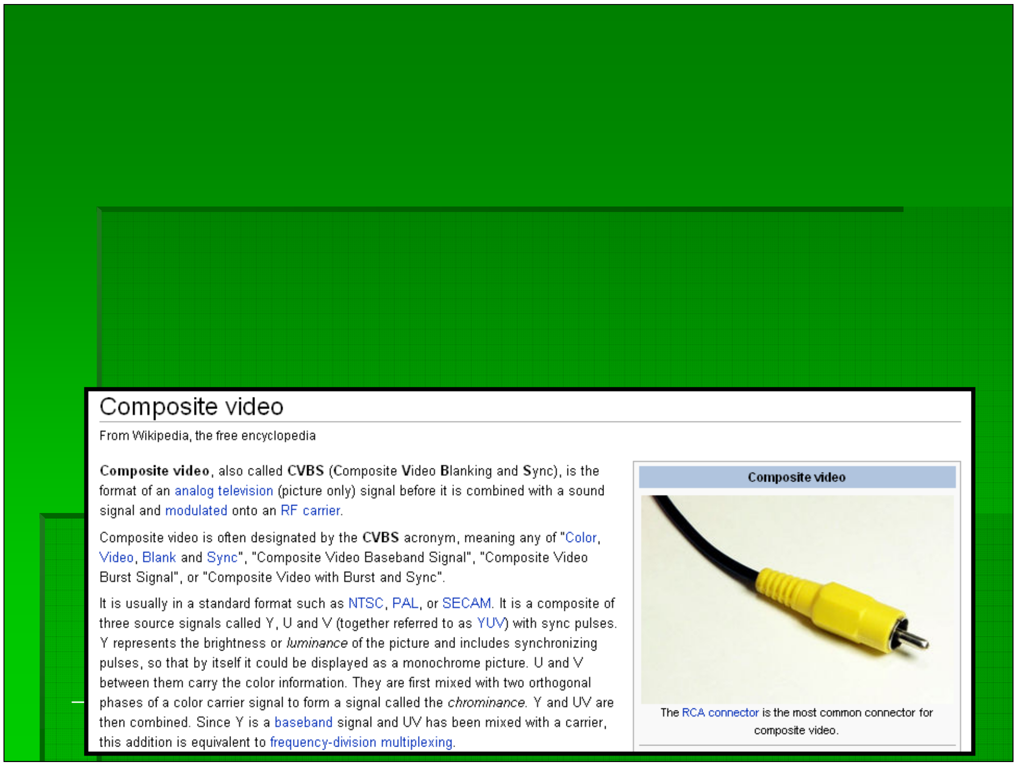


Now: about encodings: First, there are *target* encodings. Many organizations are reformatting from existing videotapes, mostly analog, and occasionally digital, like DV. In time, we may begin reformatting some of the digital files that we receive. Most reformatting requires playing back the videotape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



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.

[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 **V**ideo **B**lanking and **S**ync), is the format of an analog television (picture only) signal before it is combined with a sound signal and modulated onto an RF carrier.

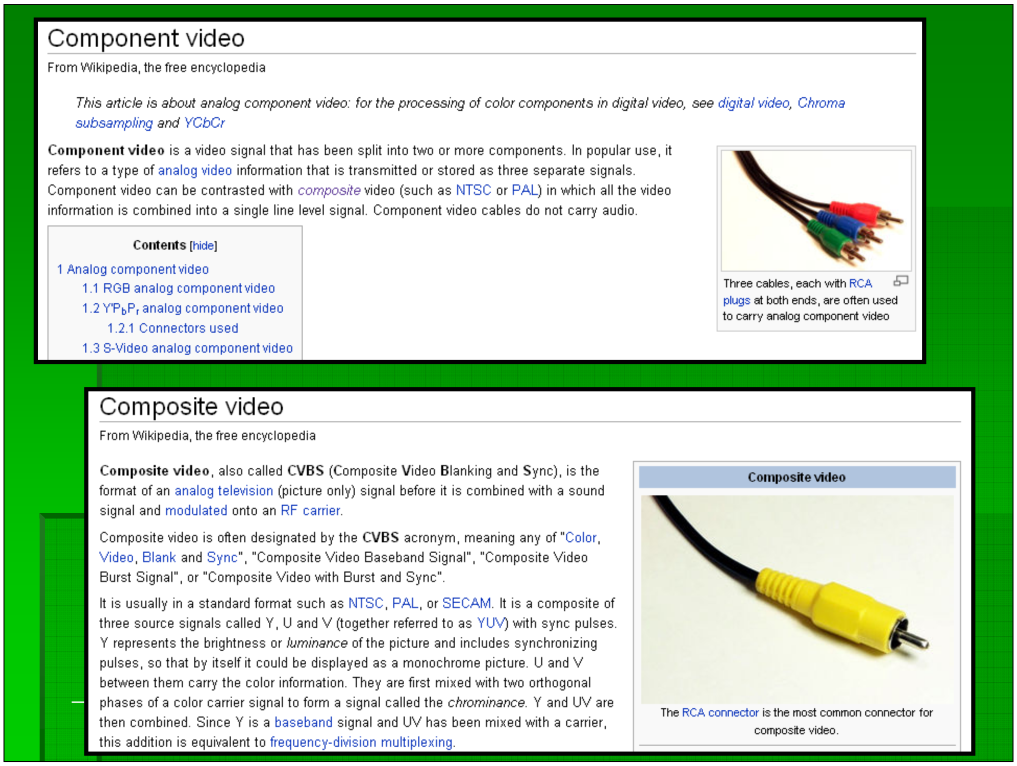
Composite video is often designated by the **CVBS** acronym, meaning any of "Color, Video, Blank and Sync", "Composite Video Baseband Signal", "Composite Video Burst Signal", or "Composite Video with Burst and Sync".

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.

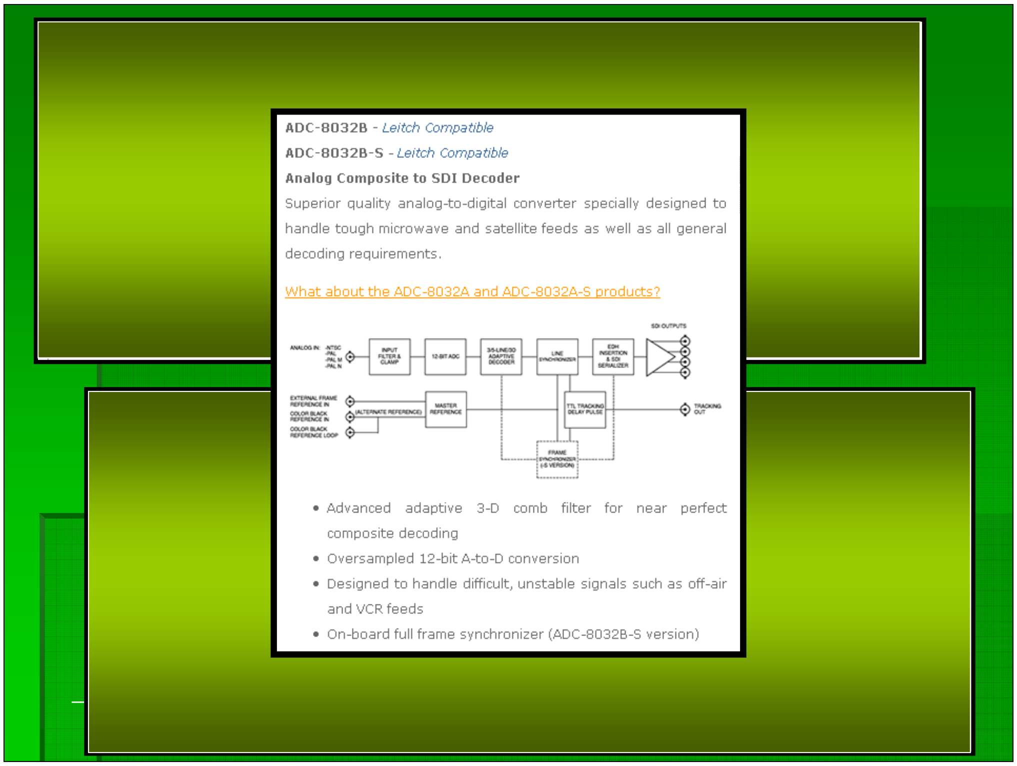
Composite video

The RCA connector is the most common connector for composite video.

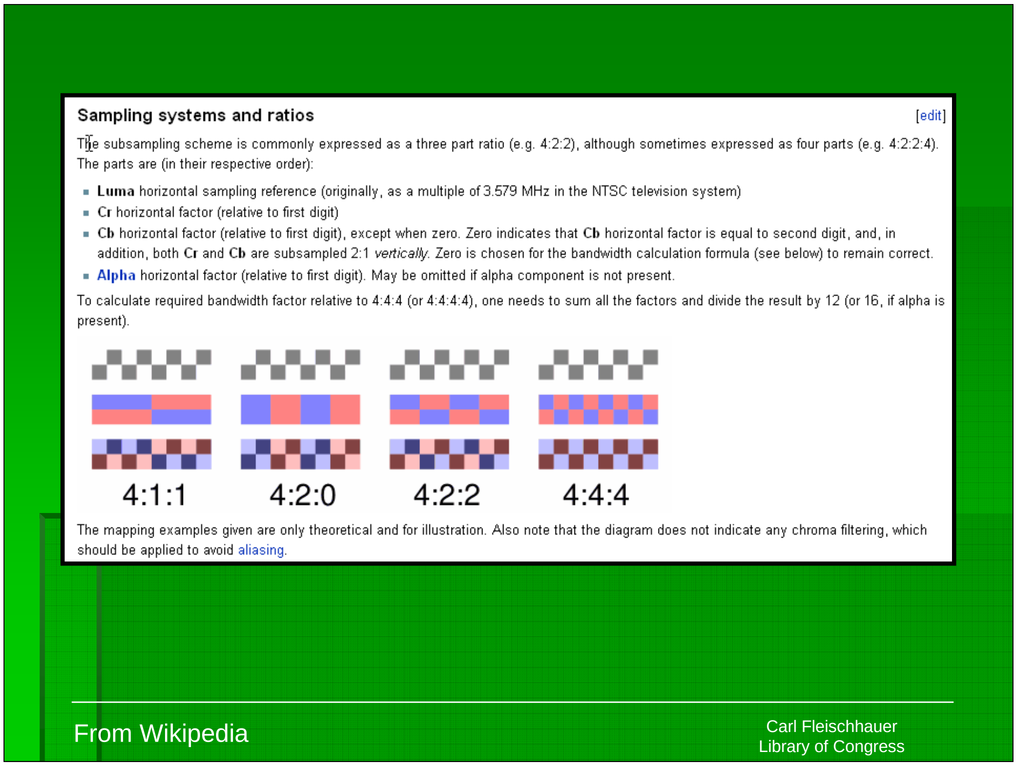
In contrast, *composite* video is represented by analog signals on older videotapes and the signals transmitted by broadcasters, until the great changeover to digital in February 2009. A composite signal blends the luminance and chrominance information.



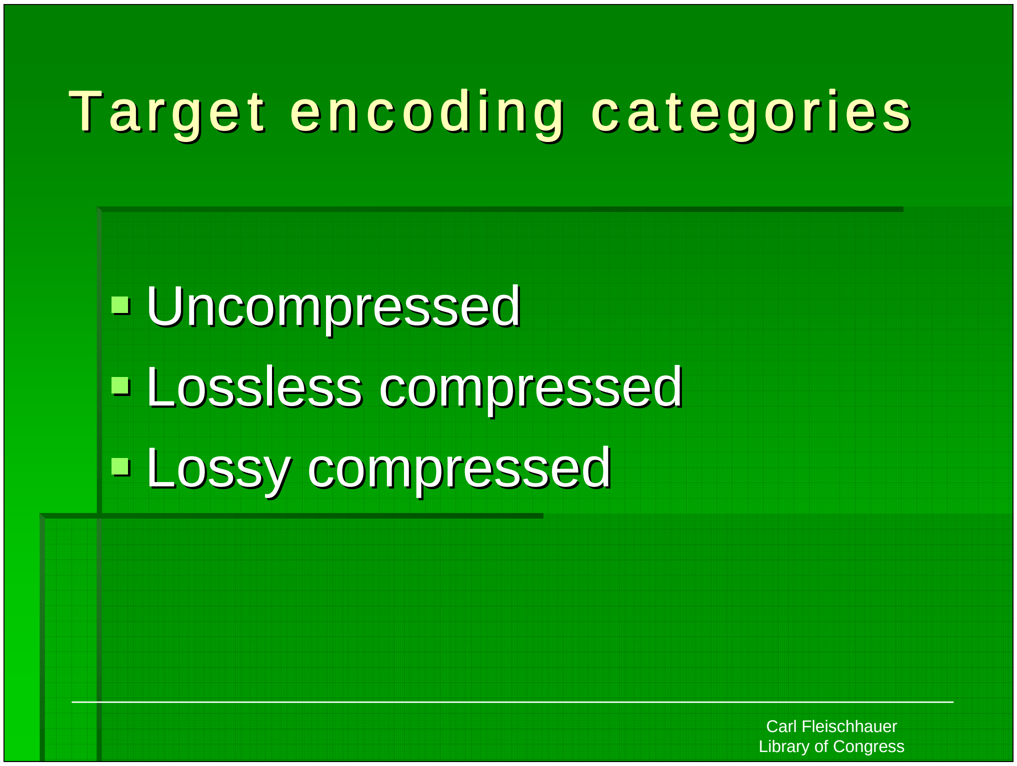
Old videotapes are composite and their output signal must be transformed into component in order to use any of the encoding options that I'll describe in a minute. In fact, the composite-to-component transform has been part of video reformatting for years--it happens when you record to a SONY BetaCam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



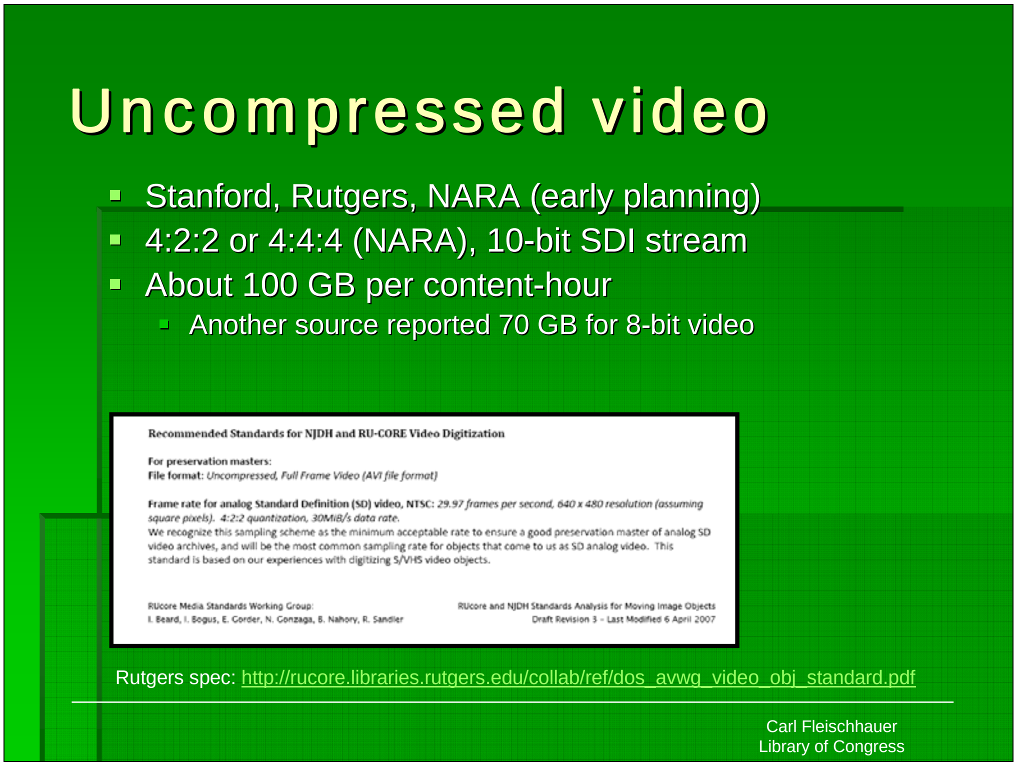
I wish I understood the degree to which different transform methods yield different results. I have been surprised at how little this part of the process has been discussed within our community.



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.

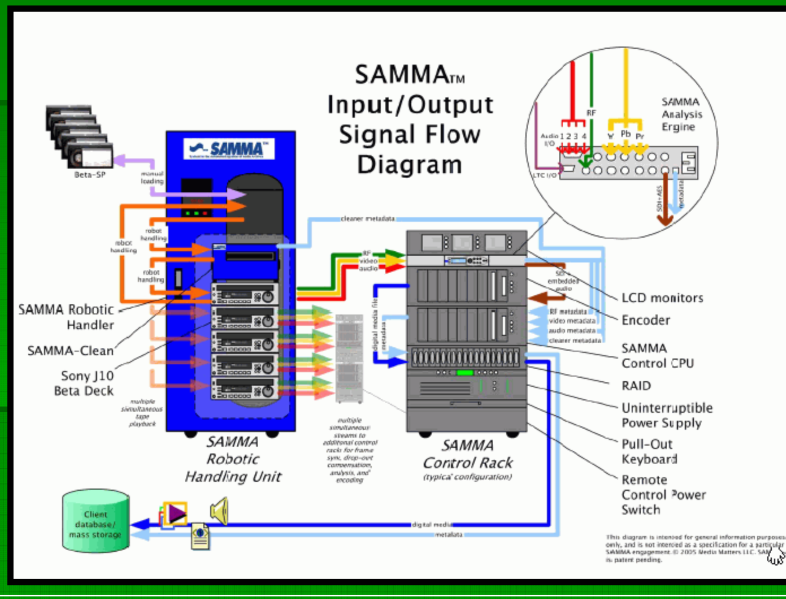


Generally speaking, existing preservation-oriented projects treat a 4:2:2 or 4:2:0 video signal, downstream from the composite-component transform, and their encoding falls into three broad categories: uncompressed, lossless compressed, and lossy compressed.



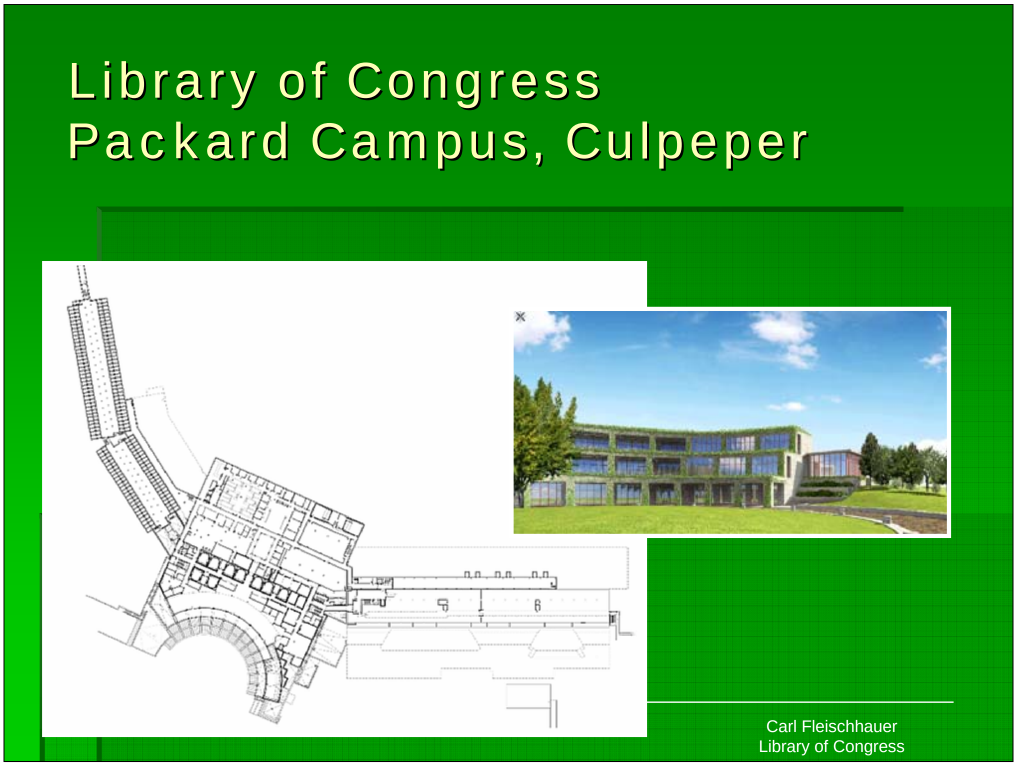
Projects at Stanford—overseen by Hannah Frost—and at Rutgers—overseen by Isaiah Beard—save the incoming signal without further compression. And I just learned that a planning group at the National Archives has made a very similar recommendation for their next phase of work. You could see this as the 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

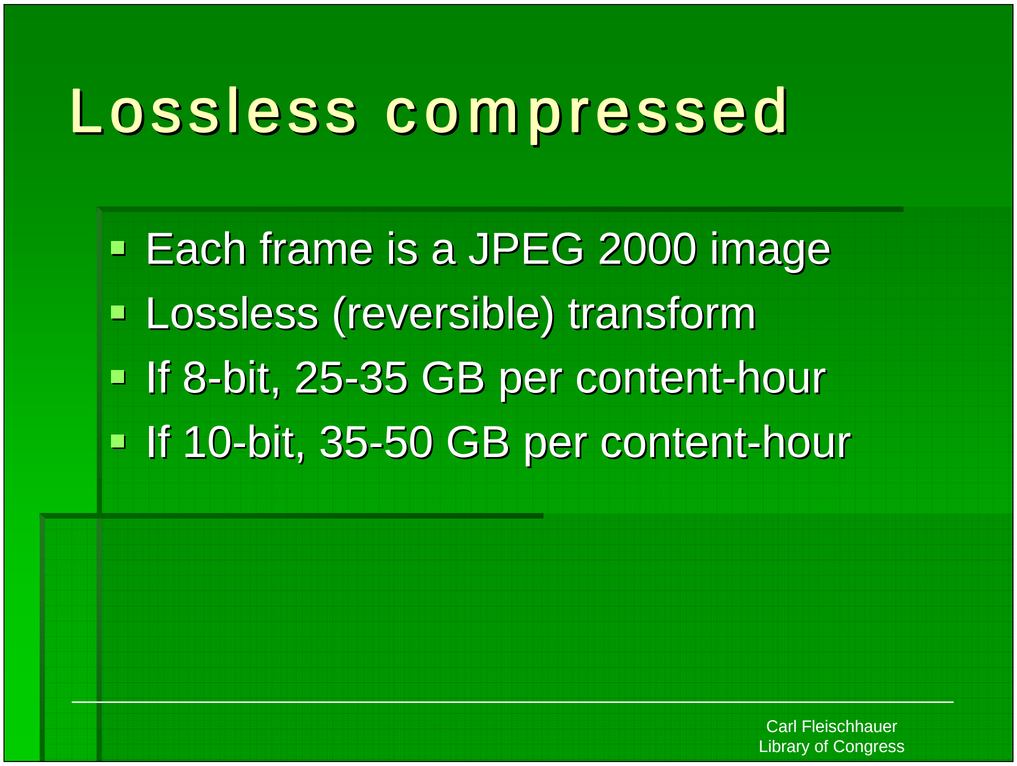


Carl Fleischhauer
Library of Congress

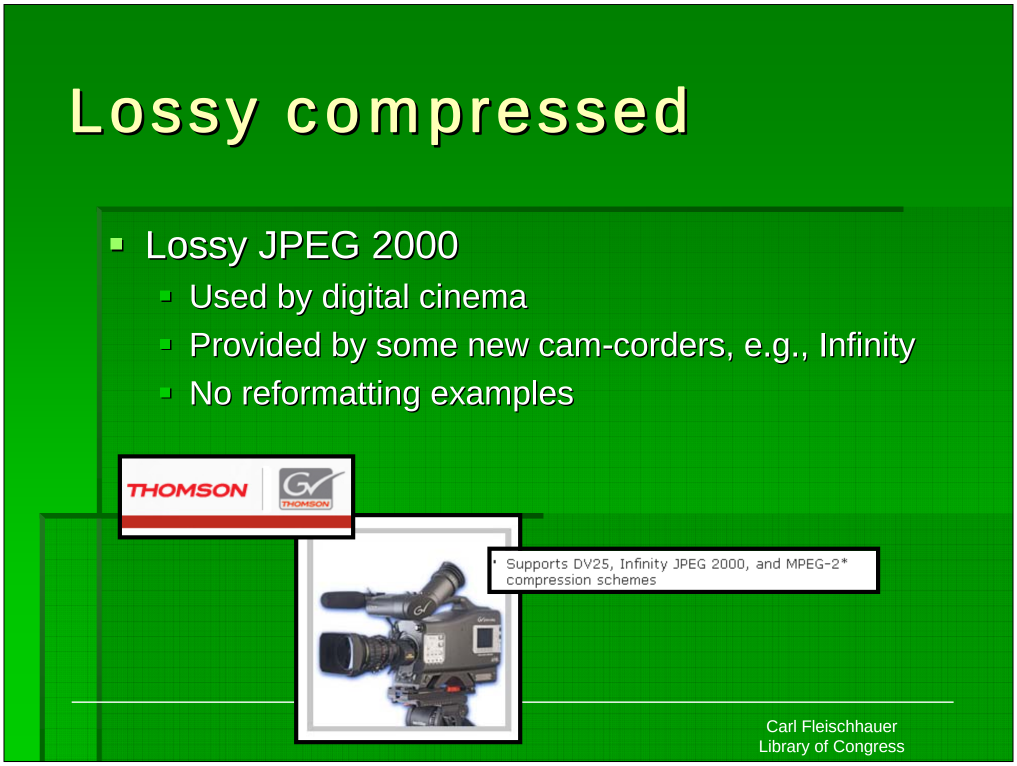
A second approach is to compress the picture using a lossless algorithm, the equivalent of saving your still image with LZW compression in a TIFF file. The leading proponent for this is Jim Lindner, whose company has developed an integrated system called SAMMA.



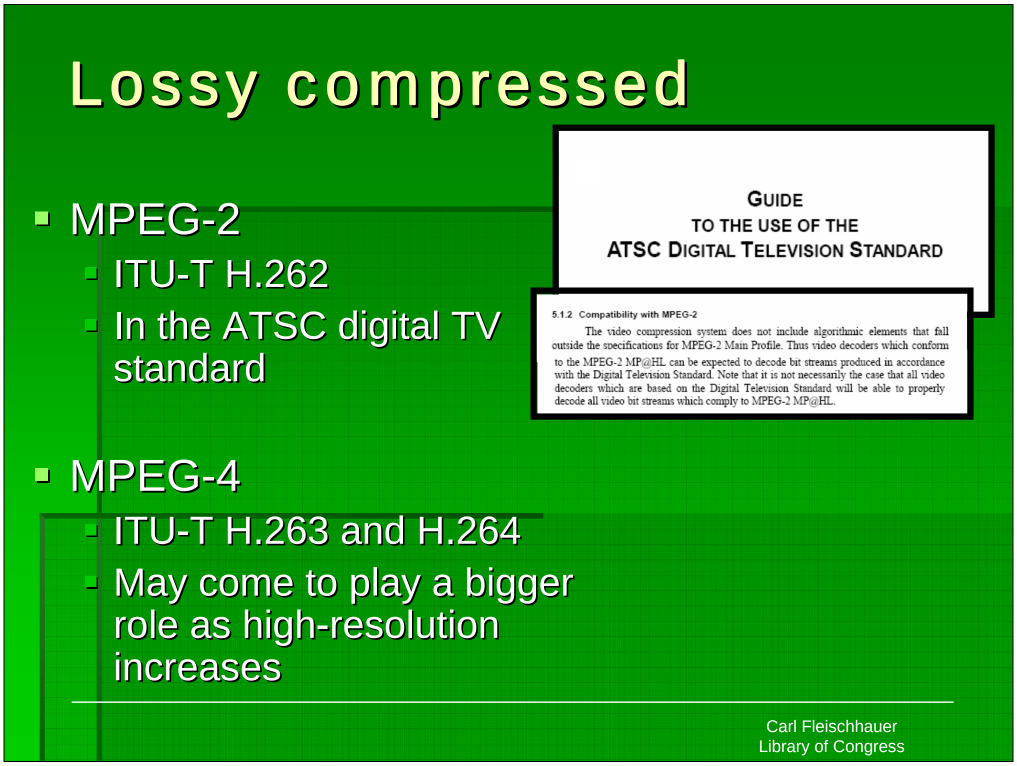
The Library of Congress Motion Picture, Broadcasting, and Recorded Sound Division is beginning to implement SAMMA in the new facility in Culpeper, Virginia.



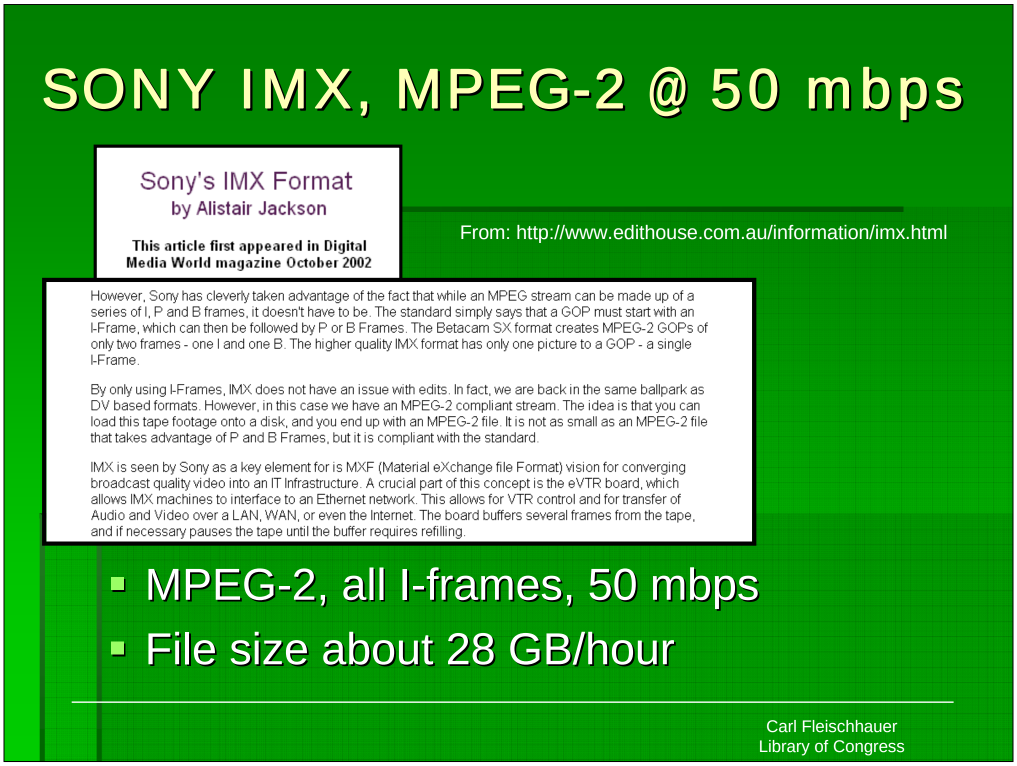
In this system, each video frame is compressed with the reversible (lossless) transform offered by the JPEG 2000 standard. Ian Gilmour, a member of the SAMMA team, reckons that 8-bit video will compress to something like 25-35 GB per hour; in one set of early tests, 10-bit came in at 35-50 GB per hour.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



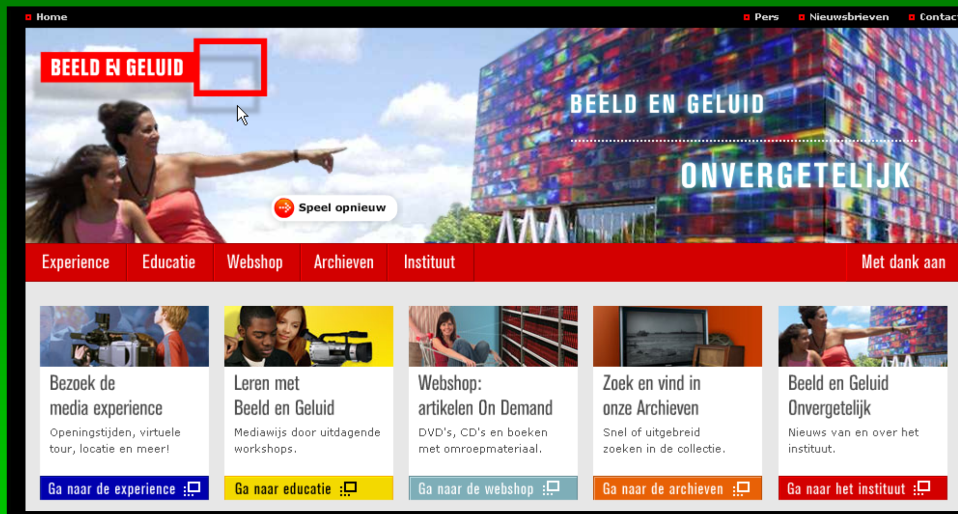
A third approach is to apply lossy compression to the picture information. You could do this using the irreversible transform in JPEG 2000, the approach used in the new digital cinema specification—movies for theaters. Although there is some uptake for lossy JPEG 2000 in born digital video in new cameras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



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.



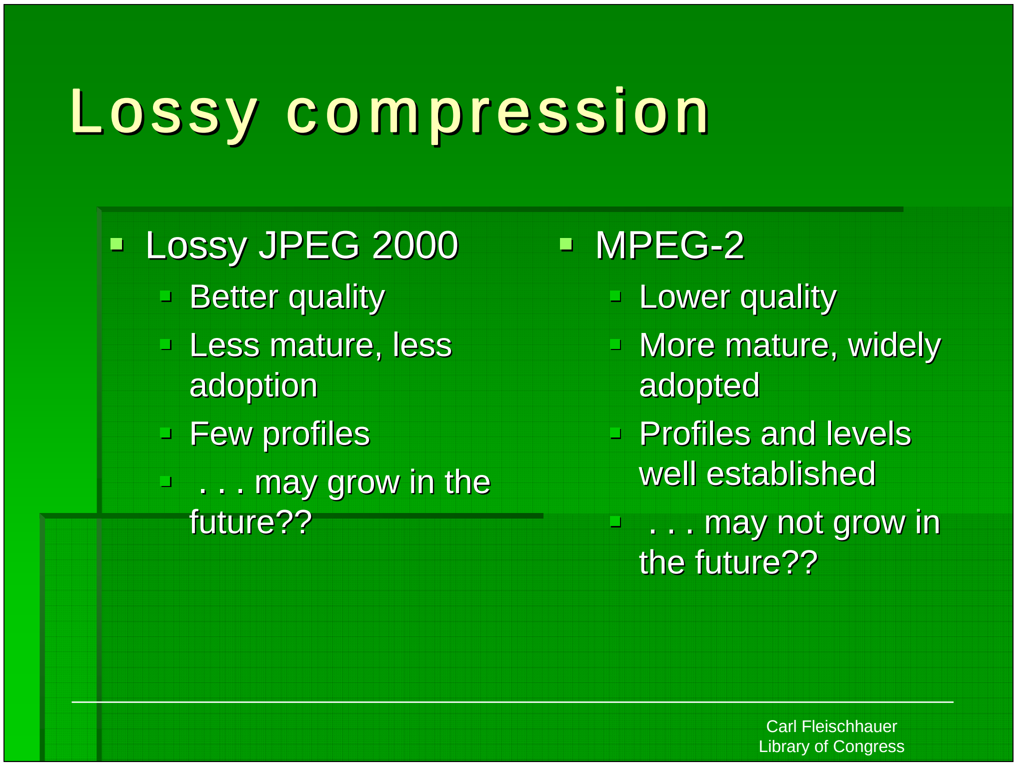
A number of broadcast organizations have selected MPEG-2 for archiving. SONY is a major player in professional circles and the widespread use of SONY IMX recorders has led to the acceptance of MPEG-2 with a data rate of 50 megabits per second as a benchmark. This encoding consists of all I-frames, meaning 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.



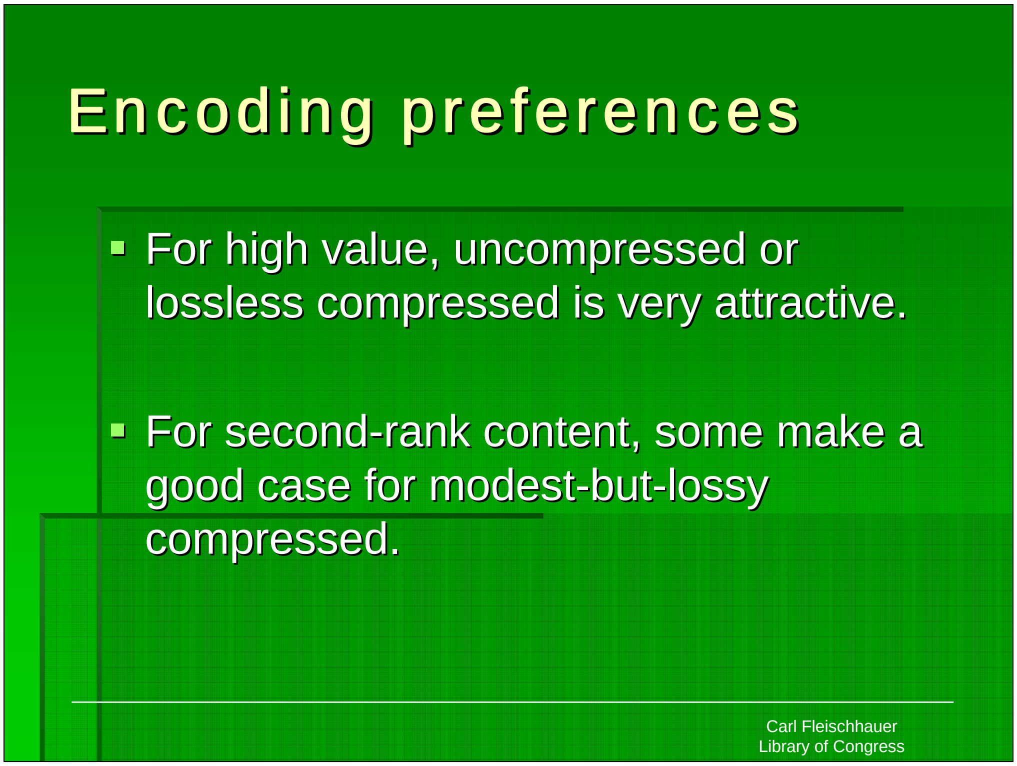
The Netherlands Institute for Sound and Vision
MPEG-2 @ 50 mbps and 30 mbps (news)

Carl Fleischhauer
Library of Congress

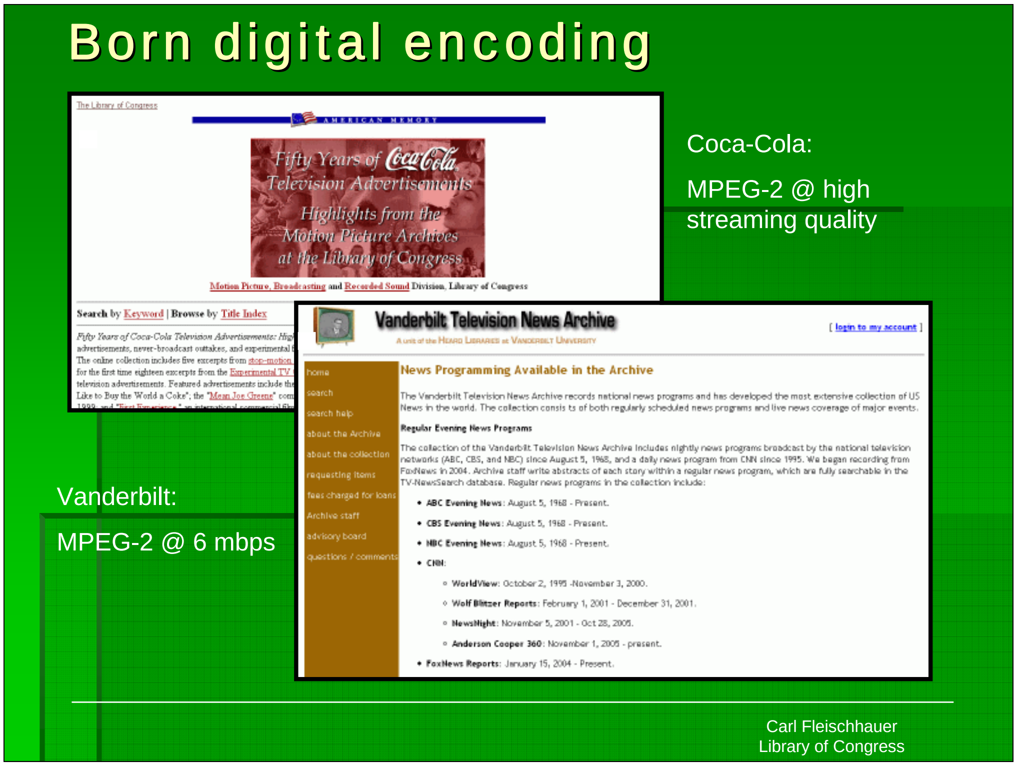
There is some use of MPEG-2 for archiving in Europe. For example, I just read an account of work at the Netherlands Institute for Sound and Vision. They have begun to digitize their extensive collection, with funding from the Dutch government, using MPEG-2 at 50 Mbps for much of their television material,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



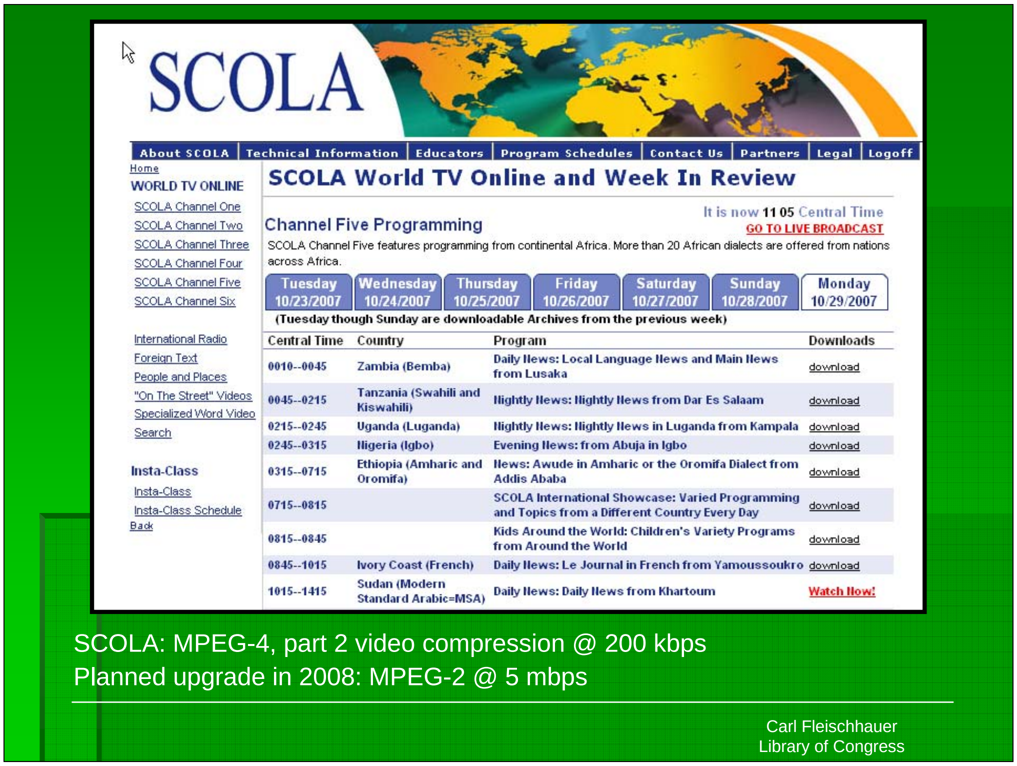
How do specialists in the field compare the two lossy contenders, JPEG 2000 and 50 mbps MPEG-2? My sense is that they feel that lossy JPEG 2000 compression can provide better quality, but that its implementation is less mature and, except for digital cinema, moving image profiles are lacking. In contrast, 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.



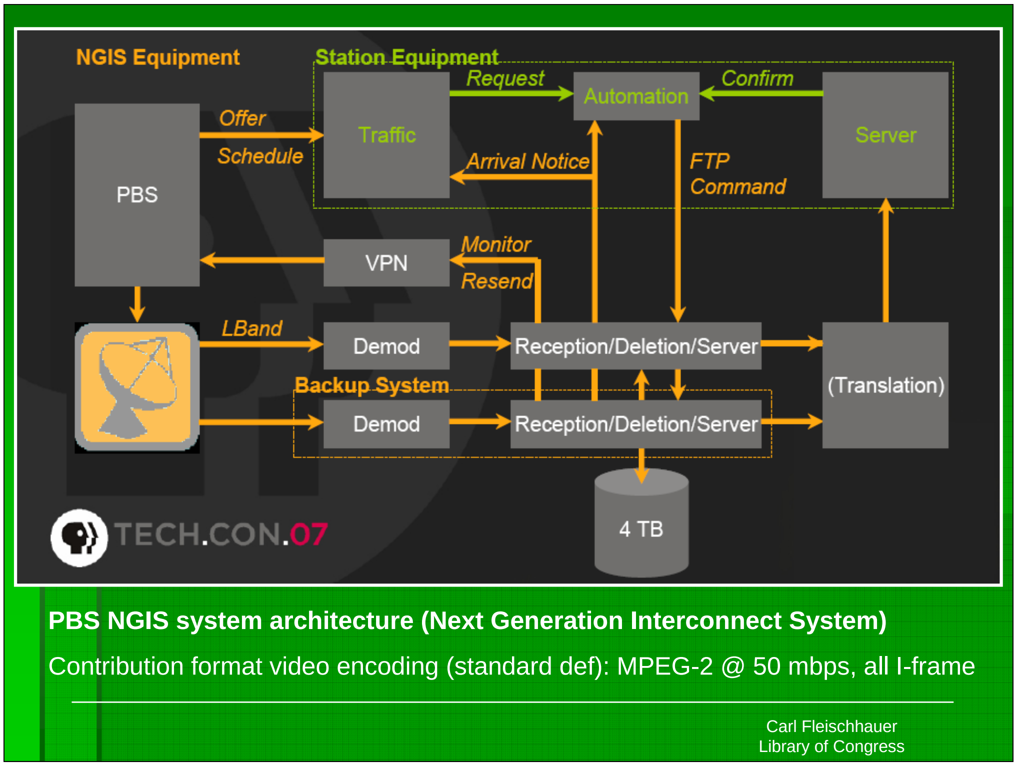
How shall we vote on reformatting target encodings overall? For high value content, it is hard not to be drawn to uncompressed or lossless encodings, the latter adding complexity to the bitstream but reducing storage requirements significantly. For second-rank content, some will make a case for modest-but-lossy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



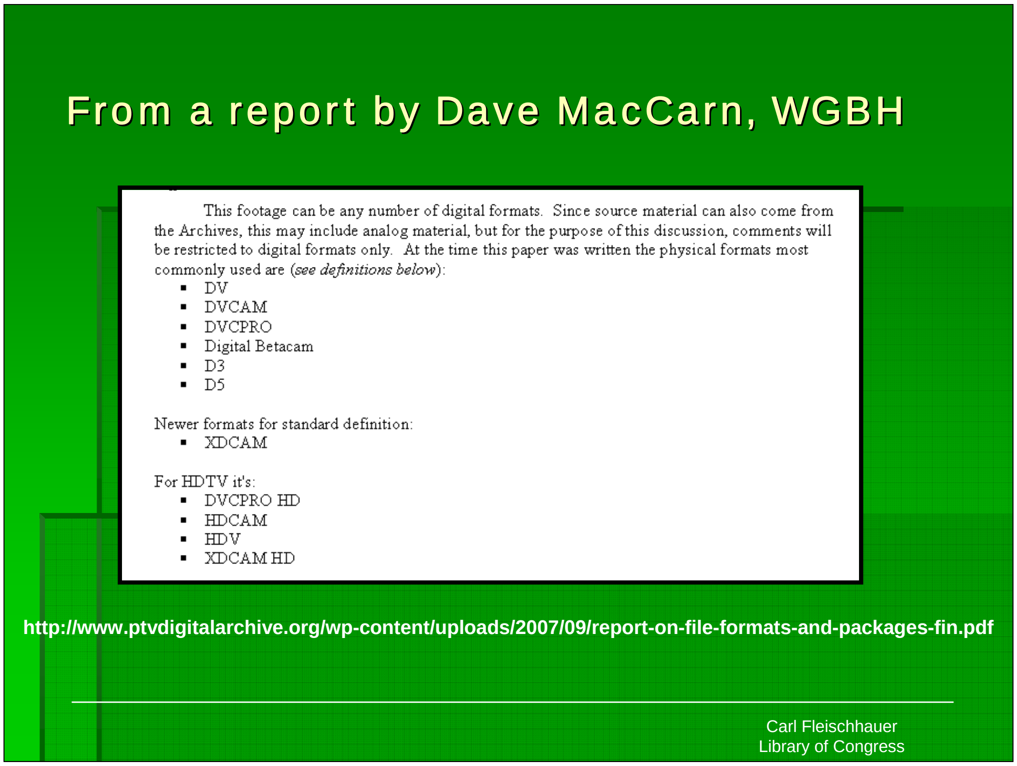
[Born digital encodings.] On the born digital side, when transcoding is necessary, the same target options recur: uncompressed, lossless compressed, and lossy compressed. But what are examples of the *keeper encodings* I mentioned earlier? In the Motion Picture, Broadcasting, and Recorded Sound Division, 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.



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.

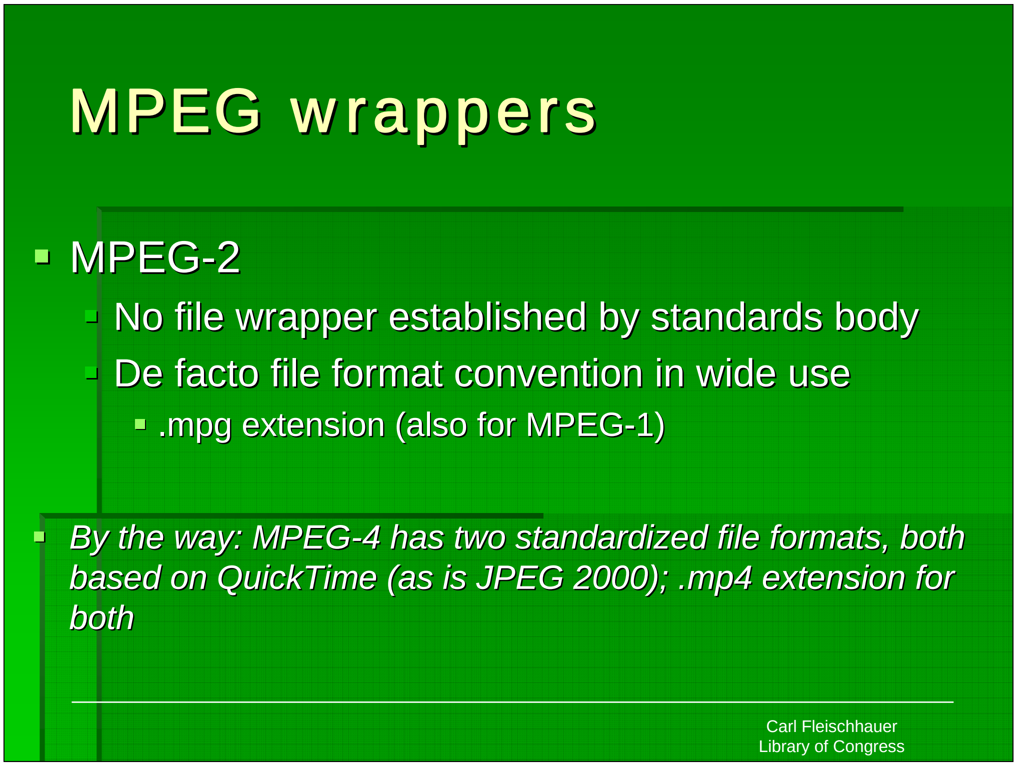


Another NDIIPP project connects us to a higher-end solution being devised in the public television community. Using PBS's new interconnect system, the producers of public TV content plan to contribute finished standard-definition programs as 50 mbps MPEG-2 files. The PBS team is developing an application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



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.



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

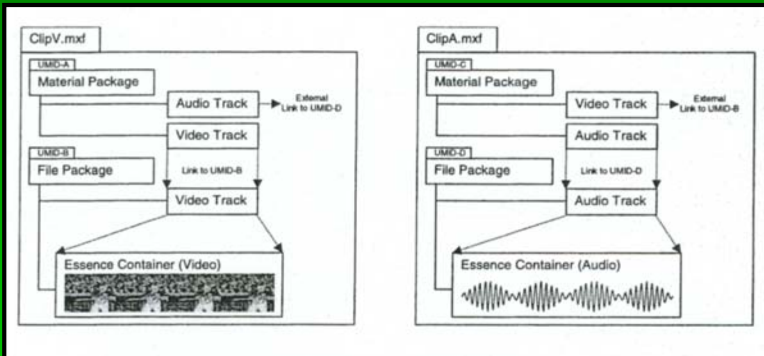


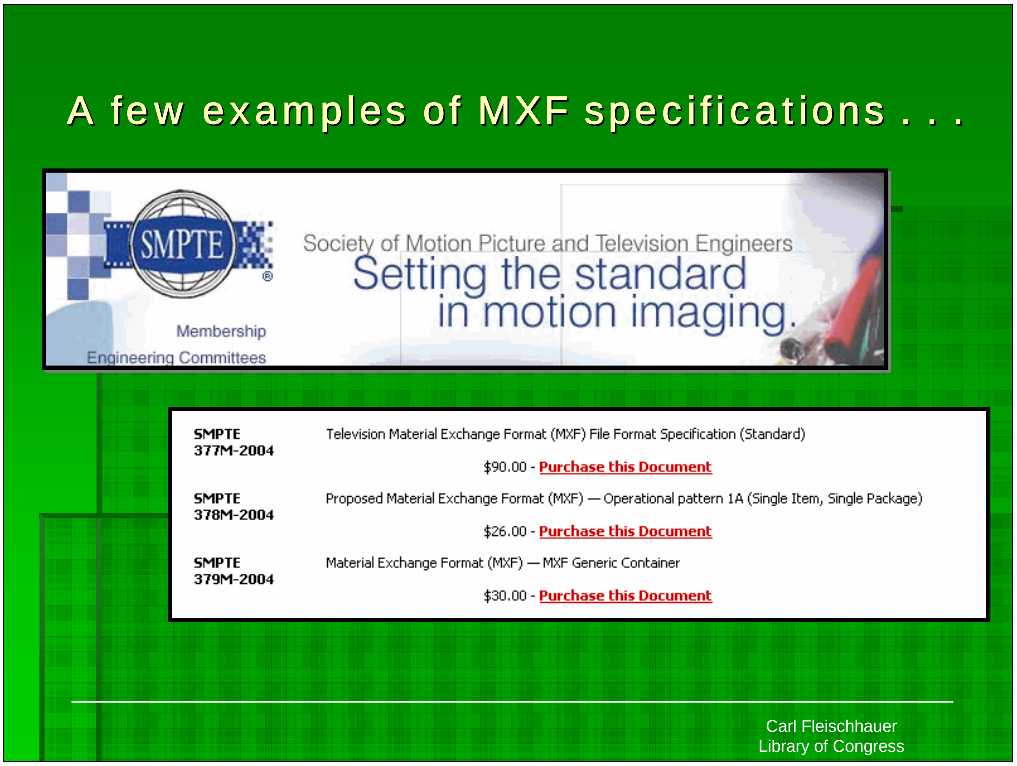
Figure A.1 – OP-Atom files showing a parallel relationship

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

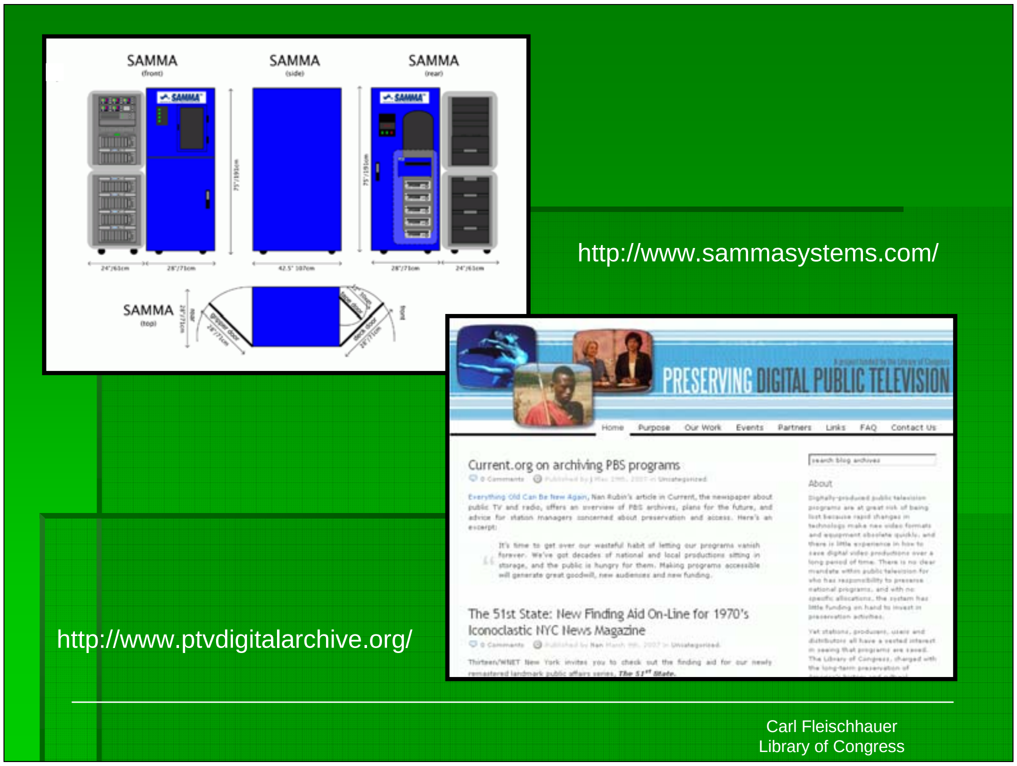
BTW: AAF (Advanced Authoring Format) is a close cousin to MXF.

Carl Fleischhauer
Library of Congress

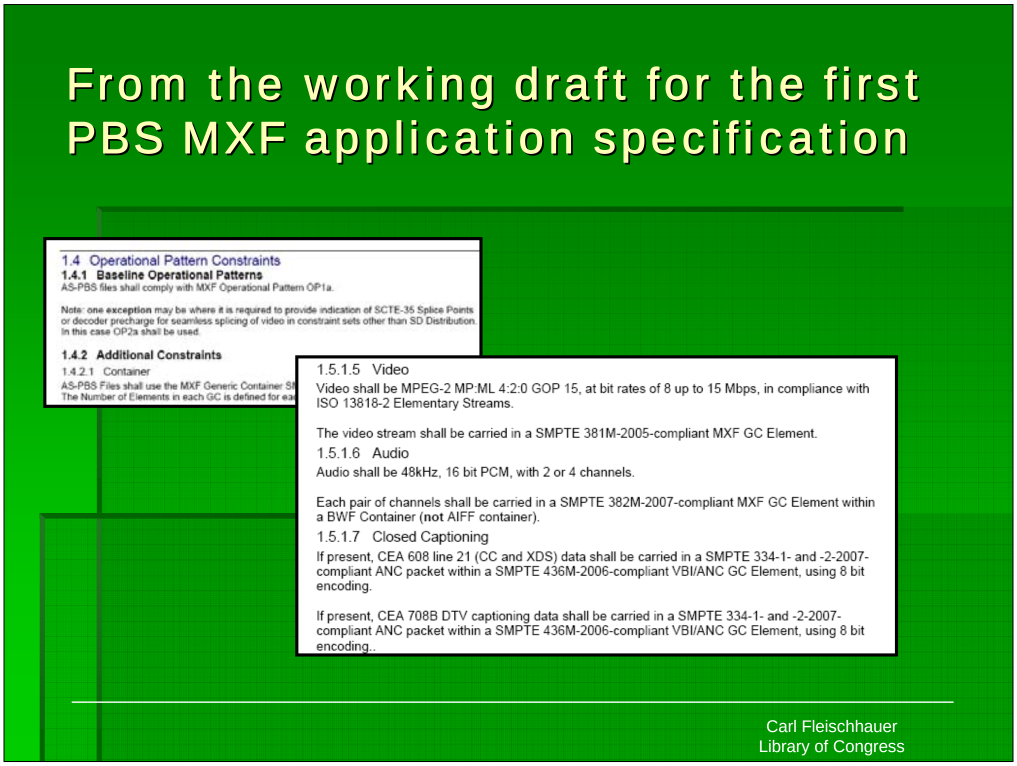
In professional circles, the wrapper buzz these days concerns the Material Exchange Format (MXF), standardized by SMPTE. MXF is an object-based file format that bundles video, audio, timecode information, closed captions, and what amounts to an "edit decision list." Complexity of structure is categorized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



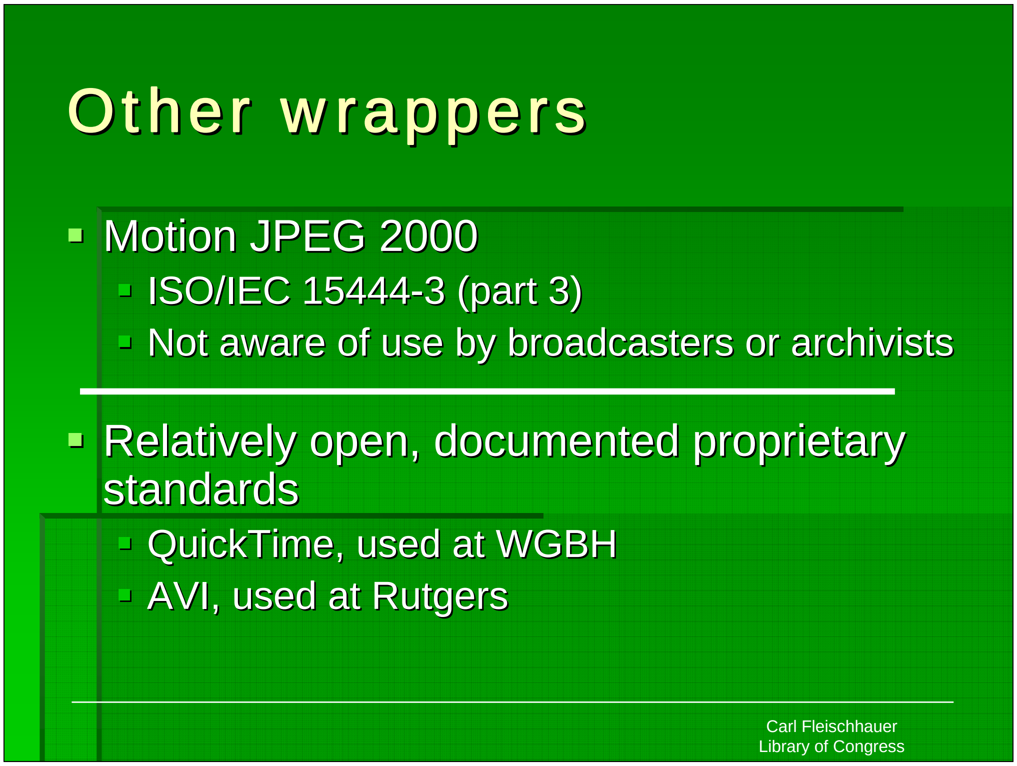
MXF is intended to support content interchange between creators and distributors, and to be implemented in cameras, recorders, and computer systems. It is used in the digital cinema specification. MXF is complicated and new: between 2004 and the present, SMPTE has published more than thirty specification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



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.

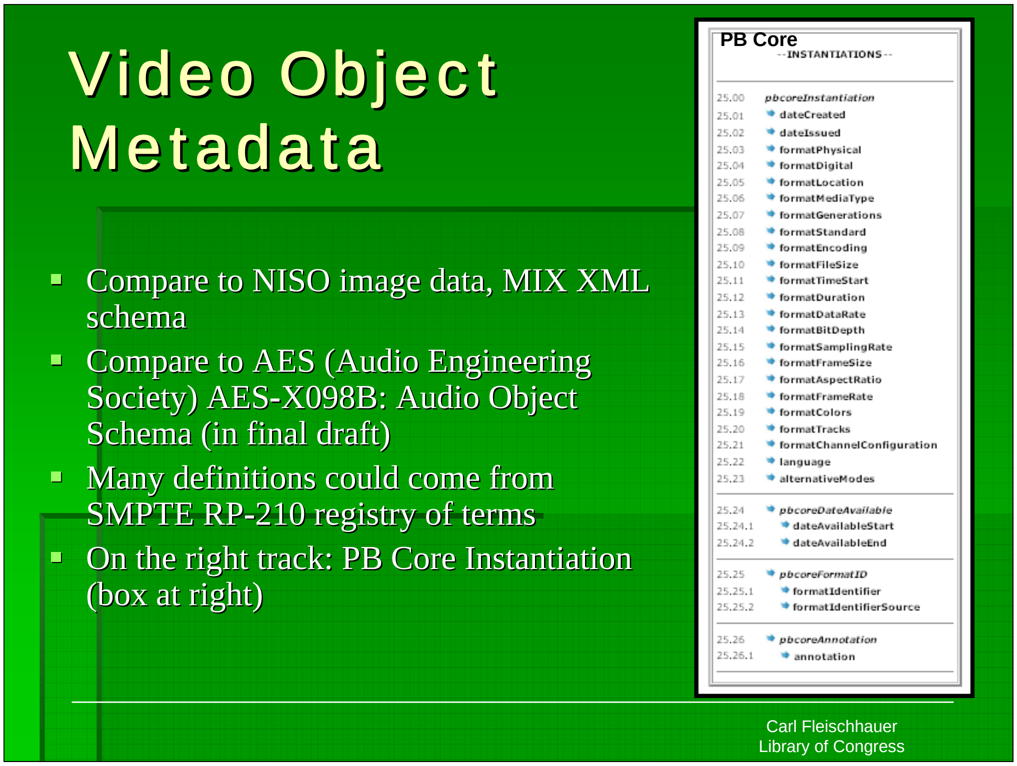


The public television folks are attending to profiling: they have drafted one MXF *application specification* (“AS”) for their moderate-resolution *distribution* files, and PBS is about to draft an AS for the high-res *contribution* files.

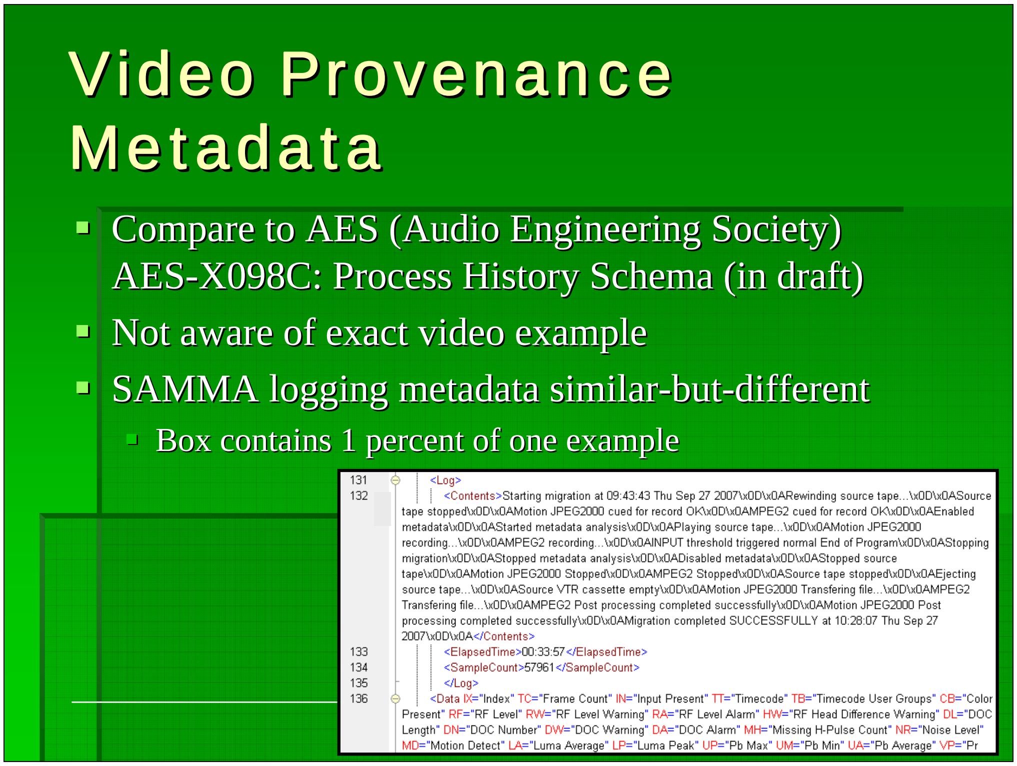


Meanwhile, there seems to be little or no uptake for ISO's Motion JPEG 2000, a wrapper designed for use with JPEG 2000 frame encoding. The folks I talk to use MXF to wrap JPEG 2000 frames instead.

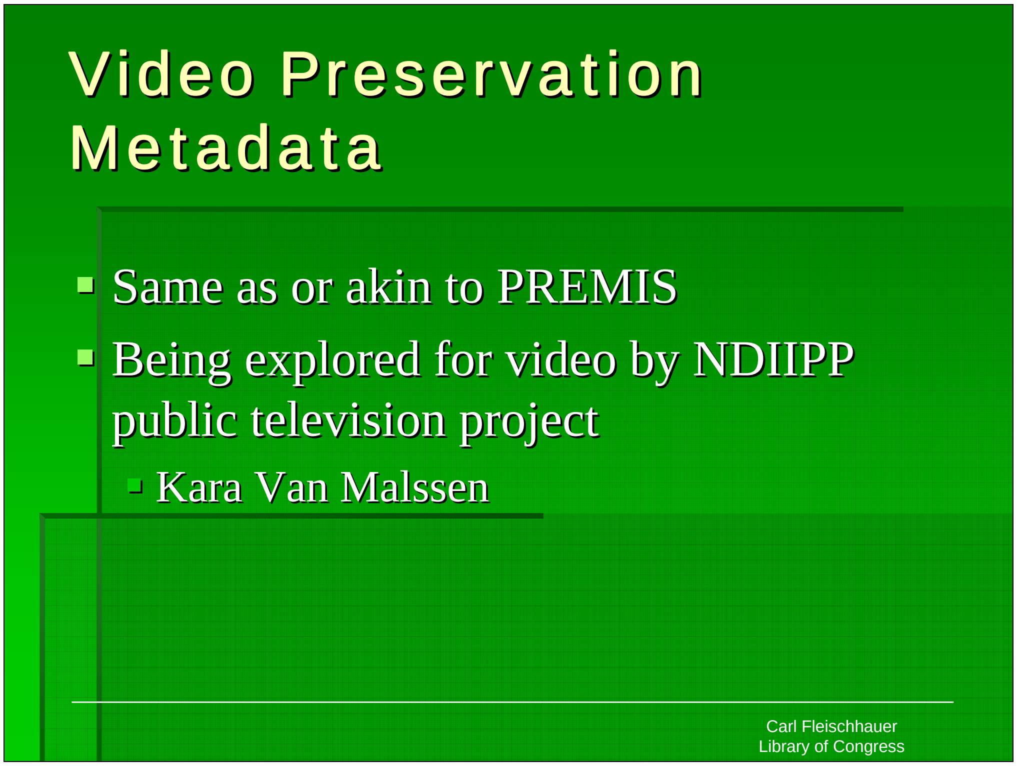
Another meanwhile: broadcasters and archivists sometimes employ proprietary wrappers, several of which have relatively open, mostly public specifications. WGBH uses the QuickTime wrapper while they wait for better tools to support MXF. For the time being, Isaiah Beard at Rutgers wraps his uncompressed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



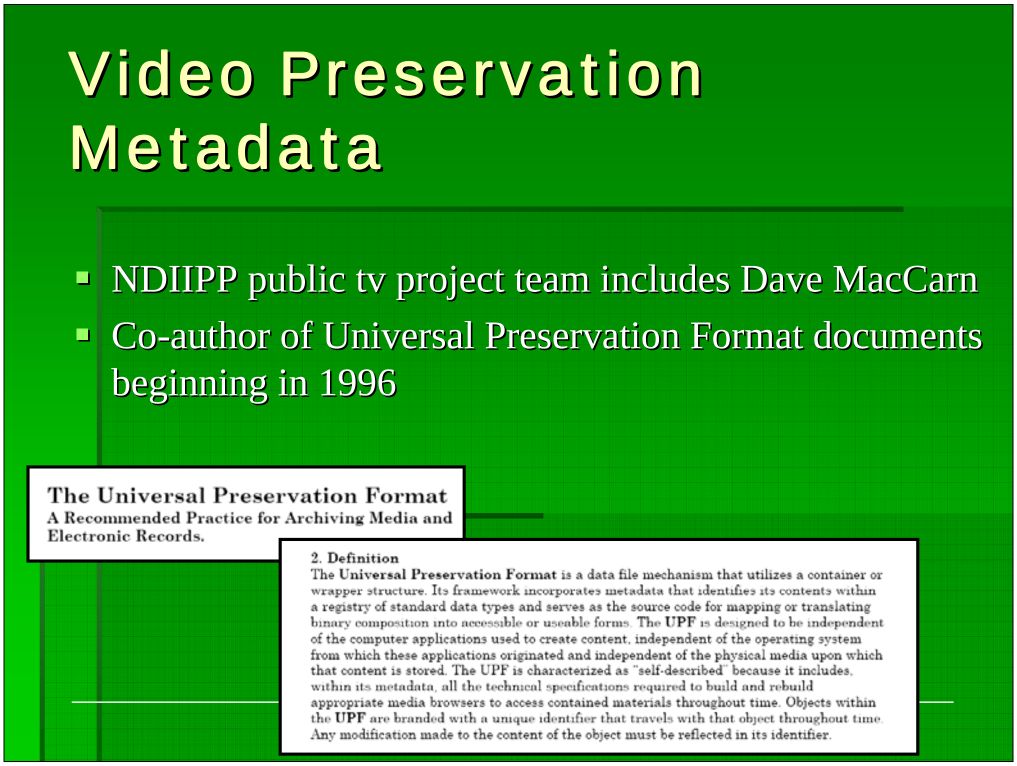
[Metadata.] Let me close with a snapshot of three metadata subcategories. The first concerns the technical characteristics of the video *object* at hand, comparable to the NISO data set for *still images*, aka MIX, and to the *audio object* metadata specification from the Audio Engineering Society (AES).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



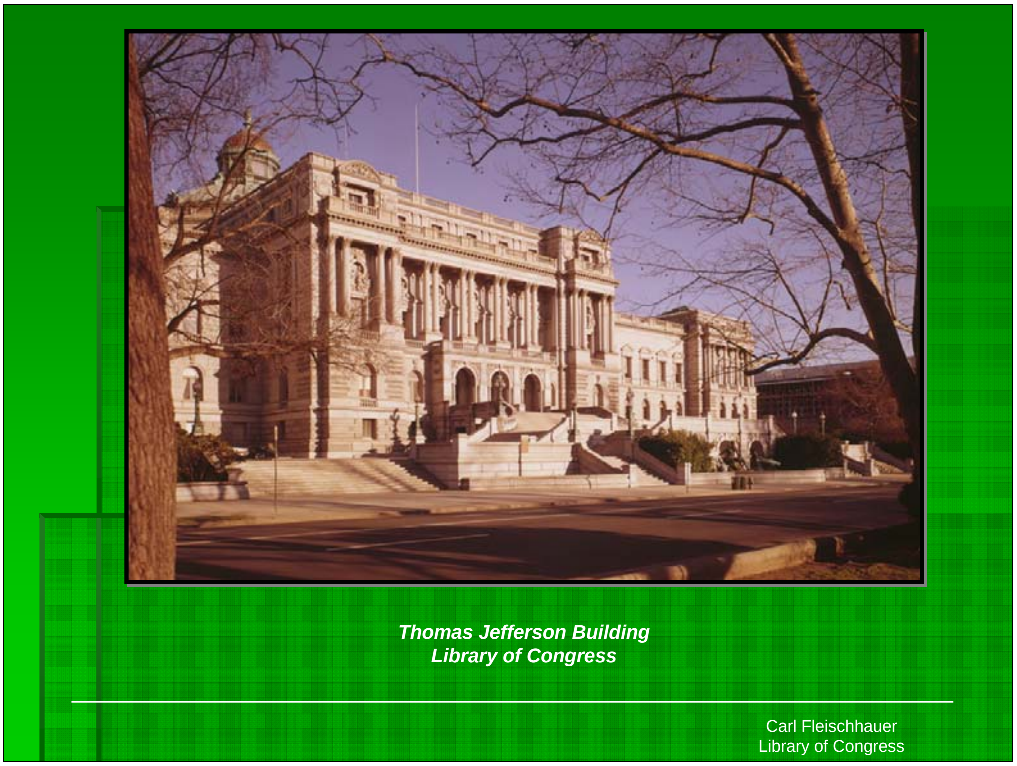
For audio, AES has been working on what we sometimes call “digital provenance” metadata, to document what we did to make the file we have. There’s nothing quite like this in the video arena, although the SAMMA system collects extensive logging data and encodes it as XML.



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 --



-- 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.



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.