International Association of Sound and Audiovisual Archives
Internationale Vereinigung der Schall- und audiovisuellen Archive
Association Internationale d'Archives Sonores et Audiovisuelles
Asociación Internacional de Archivos Sonoros y Audiovisuales

iasa

Technical Committee
Standards, Recommended Practices, and Strategies

Guidelines for the Preservation of Video Recordings

IASA-TC 06

Part B. Video Signal, Preservation Concepts, and Target Formats

From IASA-TC 06, Edition 1
Revised version, 2019

B.1 THE VIDEO SIGNAL AND BITSTREAMS: FORMAT AND FEATURES

B.1.1 Conventional video carriers and formatting

B.1.1.1 Conventional video carriers and the video signal

A number of important and commonly encountered *video carrier formats* are the subject of part C, presented later in IASA-TC 06. Those sections explain how the formatting of both the carrier and the video signal that it carries are entwined and interdependent. Nevertheless, it is possible to consider the video signal separately, and such a consideration is the subject of this section.

This discussion of signal is intended, first, to provide an introductory answer to the question *What is video?* And, in this initial edition of the guideline, this means emphasizing analogue video. Second, and more important for IASA-TC 06, this section is drafted with preservation digitisation in mind, i.e., to call out the technical features of source recordings that must be considered when making copies, and to identify the features (like captions) that many archives will wish to retain in order to ensure that their preservation copies are complete and authentic. In addition, section B.1.4.2 discusses three "added" entities that are *not* part of the video signal as found on conventional carriers, described here because they often have a preservation value similar to that derived from elements captured and retained from the source recording.

Sidebar: the noun *video*

- In ordinary language, the word *video* is used in various ways; there is often ambiguity about the referent.
- Sometimes *video* is used in a broad way to name an entire entity or package.
- Sometimes *video* is used more narrowly to name one or more selected elements within the entity, e.g., the picture or the picture-and-sound.
- Since the *video signal* may include a number of components beyond picture and sound, e.g., captions (subtitles) and time code, this document occasionally uses the term *video payload* to remind readers about the important added data that may be part of a video recording.
- For specialists in the field, the nouns *video* and *signal* are understood to be the names of classes of entities, each with several members.

B.1.1.2 Conventional carriers compared to file-based video

This initial release of IASA-TC 06 concerns the preservation of video on *conventional carriers* (generally videotapes), and it discusses the main types of video signals encountered during the time period in which videotape prevailed. The heyday for videotape began in the early 1950s and continued to the mid- to late-1990s, although there were earlier glimmerings, and, to a degree, videotape continues to be used at the time of this writing. In the 1990s, file-based video systems began to come to the fore.

The distinction between *videotape* carriers and *file-based digital video* is tricky. Conventional videotapes may carry either analogue or digital signals. Recordings in these formats are *media-dependent*, i.e., the formatting of the carrier and the signal are interdependent. In contrast, file-based video, which only exists in digital form, contains signal—or perhaps more accurately *bitstreams*—formatted independently of the storage media. (See also *File-based digital video recordings*, section A.1.2.2.2 above).

What about the formatting of file-based digital video? Although not a topic for this edition of IASA-TC 06, it is worth noting that, compared to videotape formats, file-based video includes new factors that preservation-minded archives must consider:

- First, some components have been added, e.g., embedded fixity data (often frame by frame) to support tools that maintain content integrity.
- Second, the arrival of file-based digital video has expanded the range and diversity of picture and sound elements, including options such as Ultra High Definition (UHD) resolution, High Dynamic Range (HDR) tonal representation, and "immersive" sound.
- Third, the expansion noted in the previous bullet has, in turn, motivated an extension of embedded technical metadata.[1] In the past, with media-dependent videotapes, the data needed for proper playback and some technical metadata was embedded in the signal, generally as *ancillary data* carried in the brief intervals between fields (see section B.1.3.2 below). With file-based video, some signal-based (or bitstream-based) metadata carriage continues, albeit employing different structures and encodings. (This topic receives some discussion in sections B.3, in connection with file-based digital target formats for preservation.) At the same time, digital files also carry technical metadata in the file wrapper, often embedded as a file header.

Meanwhile, the digital era has also brought computer-generated imagery (CGI) to prominence. When this type of imagery is integrated into video productions destined for broadcast or theatrical projection, CGI technical characteristics are adjusted to match those of live-action video production created with broadcast or theatre in mind. In other applications—for example, some video games—the CGI technical characteristics may not be constrained in that way. In cases like these, moving image CGI material employs raster sizes, frame rates, brightness, and colour ranges that go beyond the limits associated with "normal" video. This topic is not discussed in IASA-TC 06.

B.1.1.3 Broadcast standards and the formatting of video recordings

The descriptions of common features in section B.1.2 and B.1.3 highlight the close relationship between *broadcast* rulemaking, especially in the United States and Europe, and its influence on the production and formatting of *video recordings*.

Rules promulgated by the U.S. Federal Communications Commission (FCC) are supported by a variety of standards from the Society of Motion Picture and Television Engineers (SMPTE) and made manifest in the design and development of video recording devices and signal/payload formatting. In the U.S., many important technical details were given shape by the National Television System Committee (NTSC), established by the FCC in 1940 to resolve the conflicts that emerged when analogue television systems became a national phenomenon. Subsequent NTSC specifications were central to the development of colour television in the 1950s.

In the United Kingdom, broadcast rulemaking is one role for the Office of Communications ("OfCom"). In Europe and in many other regions that do not employ NTSC specifications, regulations have been promulgated by the *Comité Consultatif International pour la Radio* (or *Consultative Committee on International Radio*, abbreviated as CCIR) or, as it has been officially named since 1992, the International Telecommunication Union

1 The term *technical metadata* can be used to name a wide range of types of information. In this context, the term refers to the "core" information found in a file header or its equivalent. This core information provides video players with facts needed for proper playback, e.g., information about picture resolution, scanning type (interlaced or progressive), picture aspect ratio, and the presence and types of soundtracks.

Radiocommunication Sector (ITU-R). CCIR System B was the broadcast television system first implemented in the 1960s and, during the four decades that followed but prior to the switchover to digital broadcasting, this system was used in many countries.[2] Meanwhile, just as SMPTE provides supporting engineering standards in the U.S., the European Broadcasting Union (EBU) provides engineering standards that support ITU-R regulations.

The broadcast-transmission-related technical rules from the FCC and CCIR did not specify how video is to be recorded but they influenced the development of videotape recorders and signal/payload formatting. The members of standards committees in SMPTE and EBU include specialists from hardware and systems manufacturers; these members and their parent companies thereby help shape the standards, and the overall process increases buy-in and adoption within the industry. Although never as universal as one might hope, these relationships also increase the level of standardization in video recordings.

Standards and specification from other branches of the industry have also influenced video formatting in our period of interest. One of the most important is *RS-170*, which spells out many of the intricacies of the synchronizing and timing of NTSC analogue composite picture data (see section B.1.2.6). This standard began its life under the auspices of the Electronic Industries Association (later renamed the Electronic Industries Alliance; EIA), a U.S. trade group for the manufacturers of electronic equipment, including television sets. As the standard took shape in the mid-1950s, it was also central to the NTSC specifications for television broadcasting in the United States, and it influenced parallel developments in other nations to fit the needs of the PAL and SECAM systems (see section B.1.2.1 below). In later years, the RS-170 standard was updated and republished by SMPTE.[3]

B.1.2 Analogue video unpacked, part one: key features and variants

Video may be a singular noun but it names a plural and varied set of entities: *types of video*. At a high level, these types have some features in common but even these common features may splinter into subtypes when closely examined.

The sections below (B.1.2.1 through B.1.2.7) and those in the following section (sections B.1.3.1 and B.1.3.2) describe the most important common features for the video types that are the subject of this initial version of IASA-TC 06, i.e., those on conventional carriers rather than in file-based form. These nine sections include high-level information about the feature and offer a sketch of how that feature varies from one video format type to another. Complete technical information about these features is beyond the scope of this guideline and often moves into advanced engineering areas. However, each of the nine common-feature sections includes a list of Wikipedia articles that provide significant amounts of added (and often excellent) technical information. Readers are also encouraged to consult the IASA-TC 06 bibliography (Part E) for additional references.

2 CCIR also specified systems A, G, H, I, and M, each used in selected nations or regions. System M is the ITU-R expression of NTSC.

3 RS-170 was first standardized in 1957 by EIA, an organization whose forebears include a trade group launched in the 1920s when radio broadcasting first came on the scene. The EIA continued until 2011, when the diversity of member activities led several subgroups to split off to form trade groups of their own. In 1994, the RS-170 specification was refined and published as SMPTE standard 170M (new nomenclature: ST 170), revised in 1999 and 2004 (SMPTE ST 170:2004 - SMPTE Standard - For Television — Composite Analog Video Signal — NTSC for Studio Applications). The standard's implementation is supported by the publication of SMPTE Engineering Guideline EG 27, most recently published in 2004.

B.1.2.1 Illusion of motion from a stream of still images

- **Common feature:** Picture data consists of a stream of still-image frames that, like movie film, create the illusion of motion.

- **Variation:** Frame rates differ from video system to system. In the analogue era, frame rates were—to simplify just a bit—30 frames per second in the United States and Japan (NTSC system) and 25 frames per second in Europe and many other regions (PAL and SECAM systems).[4]

  When colour came to television broadcasting in the 1950s, the NTSC system moved to *fractional* frame rates. (See also section B.1.3.1 below.) This frame rate adjustment was motivated by the need to continue to support the millions of black-and-white television sets already in homes. NTSC engineers played a complex game of mathematics in order to minimize the interference that resulted from mixing the colour subcarrier frequency with the sound intercarrier frequency. In terms of frame rate, the outcome was to divide the old rate of 30 frames per second by 1.001 (the *fraction* is "30 over 1.001"), yielding a new frame rate of 29.97 frames per second. Today, after the arrival of file-based digital video, a wide array of additional frame rates has come into use, and many specialists hope that fractional frame rates will slowly be phased out.

- Relevant Wikipedia articles:
  - https://en.wikipedia.org/wiki/Field_(video)
  - https://en.wikipedia.org/wiki/Flicker_(screen)
  - https://en.wikipedia.org/wiki/Film_frame
  - https://en.wikipedia.org/wiki/Frame_rate
  - https://en.wikipedia.org/wiki/NTSC
  - https://en.wikipedia.org/wiki/NTSC-J
  - https://en.wikipedia.org/wiki/PAL
  - https://en.wikipedia.org/wiki/PAL-M
  - https://en.wikipedia.org/wiki/Persistence_of_vision
  - https://en.wikipedia.org/wiki/SECAM
  - https://en.wikipedia.org/wiki/Television
  - https://en.wikipedia.org/wiki/Video

B.1.2.2 Sound data is carried in parallel with picture data

- **Common feature:** Most videotapes carry audio in a separate *longitudinal* track (or tracks) that runs parallel to the recorded picture information. At first, audio was limited to monaural sound. Stereo was added in the mid-1970s.

  By the early 1980s, broadcasters sought to transmit additional audio channels and Multichannel Television Sound (MTS) was added to the NTSC broadcast specifications in the United States in 1984 and was added to some PAL broadcast systems (in Europe and other regions) at about the same time. The additional tracks may support surround sound, soundtracks in which the spoken content is in "other" languages, or special features like Descriptive Video Service (DVS).

4 PAL is an acronym for Phase Alternating Line, while SECAM stands for Sequential Coleur avec Memoire (Sequential Colour with Memory). These two systems arose in order to support colour television (like the second round for NTSC), and they receive additional discussion in sections B.1.2.3, B.1.2.6, B.1.2.7, and B.1.3.1.1.

- **Variation:** The broadcast MTS requirement was reflected in the capabilities of tape formats. On some videotapes, added channels for audio may be recorded as additional longitudinal tracks. On others, the added sound data is modulated into the stream of picture information. For example, Betacam SP offered Audio Frequency Modulation (AFM) to provide four tracks. Meanwhile, the VHS and Hi8 tape formats offered HiFi audio. The added tracks in the HiFi system sometimes carried added sound information and sometimes simply provided higher-fidelity versions of the same sound data as the "normal" tracks. The number of audio tracks varies from one system to another; as noted, some are longitudinal and some are modulated into the picture data. In addition, some recordings employ Dolby or other noise reduction systems. In the digital realm, this variation increases and the digital encoding of the sound varies from instance to instance.

- Relevant Wikipedia articles:
  - https://en.wikipedia.org/wiki/Descriptive_Video_Service
  - https://en.wikipedia.org/wiki/Digital_audio
  - https://en.wikipedia.org/wiki/Second_audio_program
  - https://en.wikipedia.org/wiki/Multichannel_television_sound
  - https://en.wikipedia.org/wiki/VHS

B.1.2.3 Picture data consists of sets of horizontal scan lines

- **Common feature:** Video pictures are presented on a display monitor ("television set" or computer screen) as a series of horizontal lines that make up a rectangle, similar but not identical to the grid of pixels that comprise the rectangle in a digital still image. Both the video line-based image and the still image pixel set are referred to as a *raster* (more or less, a grid). During most of the period when conventional carrier formats prevailed, the picture presentation was *interlaced* (see section B.1.2.4), and the full set of scan lines consisted of two *fields*. The scan lines that include the actual image in a pictorial sense are referred to as active video. Other lines carry what is called *ancillary* data; see B.1.3.2.

- **Variation:** The quantities of lines differ from system to system. The NTSC format includes 525 lines per frame, with active video consisting of 486 lines (some authorities state 483). PAL and SECAM have 635 lines per frame, of which 576 are active video.

  These variations have increased dramatically with the arrival of digital video. The digital signal data also varies in how horizontal scan lines are encoded: the sequence of pixels for a given line may have different shapes (square, non-square).[5] In digital formats, the number and aspect ratio of the pixels, the number of pixels per line, and the number of lines, govern the aspect ratio of the picture as a whole. In the digital broadcast specification promulgated by the Advanced Television Systems Committee (ATSC), for example, the standard definition variant employs scan lines with progressive (non-interlaced) scan, usually abbreviated 480p, and this picture type may have either square or non-square pixels.

5 Katherine Frances Nagels provides an excellent explanation of pixel and picture aspect ratios in *PAR, SAR, and DAR: Making Sense of Standard Definition (SD) video pixels* (Nagels: 2016). The Wikipedia article "Pixel Aspect Ratio" also offers a good introduction and links to other sources of information, https://en.wikipedia.org/wiki/Pixel_aspect_ratio, accessed 24 November 2017.

- Relevant Wikipedia articles:
  - https://en.wikipedia.org/wiki/480p
  - https://en.wikipedia.org/wiki/Aspect_ratio_(image)
  - https://en.wikipedia.org/wiki/High-definition_television
  - https://en.wikipedia.org/wiki/NTSC
  - https://en.wikipedia.org/wiki/PAL
  - https://en.wikipedia.org/wiki/Pixel_aspect_ratio
  - https://en.wikipedia.org/wiki/SECAM
  - https://en.wikipedia.org/wiki/Raster_scan
  - https://en.wikipedia.org/wiki/Scan_line
  - https://en.wikipedia.org/wiki/Standard-definition_television
  - https://en.wikipedia.org/wiki/Television

B.1.2.4 Horizontal lines of picture data may be interlaced

- **Common feature:** For many years, limits on transmission bandwidth together with an interest in the reduction of flicker, led to the practice of dividing frames into fields, with each field carrying half of the lines in the frame, which are then interlaced on the display screen to recreate the original frame image. Interlacing is part of all analogue systems and for certain types of digital video.

- **Variation:** Since the number of lines per field is a function of the number of lines per frame, field sizes vary in parallel with the variation in frame size.

  For a certain period, successful video editing required careful determination and tracking of the "dominant field" (the first to be transmitted, which may consist of the odd-numbered or even-numbered lines) but advances in transfer-management technology have significantly reduced the risk of errors.

- Relevant Wikipedia articles:
  - https://en.wikipedia.org/wiki/Flicker_(screen)
  - https://en.wikipedia.org/wiki/Interlaced_video
  - https://en.wikipedia.org/wiki/Progressive_scan
  - https://en.wikipedia.org/wiki/Progressive_segmented_frame

B.1.2.5 Movies on film can be recorded as video

- **Common feature:** The images on motion picture film can be transferred to video using special processes. In a theatre, film from the sound era is projected at 24 frames per second (fps). With video standards differing (e.g., PAL at 25 fps and NTSC at 29.97 fps), the technology to transfer film to video varies. Audiences have long since accommodated the resulting anomalies.

- **Variation:** With PAL, the transfer was carried out on a frame-for-frame basis: 24 fps film to 25 fps video, speeded up about 4 percent. One outcome is that the soundtrack audio is about one-half semitone higher in pitch. Recently, the advent of digital tools has supported the adjustment of elapsed time, leaving the audio pitch unchanged, for PAL broadcast.

  In the United States and Japan, the use of a higher frame rate for video (nominally 30 fps, actually 29.97 fps) meant that speeding up a film would yield bothersome distortions in motion and sound fidelity. Thus, special approaches were developed for film transfer, notably what is called three-two pulldown (or 3:2 pulldown). One second of video contains (nominally) 30 frames; with interlacing, this means that 60 fields are in play. (See section B.1.2.3 and B.1.2.4 on picture lines, frames and fields, and interlacing.) With three-two pulldown, the 24 frames of film (one second's worth) are divided among the 60 fields. The resulting flow of imagery is thus a bit uneven, but the loss of smoothness is so subtle as to be virtually invisible.

More recently, when shooting film for television from approximately the 1970s forward, many producers shot footage at video frame rates to permit a frame-by-frame transfer, i.e., at 25 fps in Europe and other PAL/SECAM regions, and at 30 fps in North America and Japan, where NTSC prevailed.

- Relevant Wikipedia articles:
  - https://en.wikipedia.org/wiki/24p
  - https://en.wikipedia.org/wiki/NTSC
  - https://en.wikipedia.org/wiki/Telecine
  - https://en.wikipedia.org/wiki/Three-two_pull_down

B.1.2.6 Timing: video signal elements must be synchronized (RS-170)

- **Common feature:** The description that follows applies to analogue broadcasting and, to a degree, to digital video recordings in media-dependent formats. In contrast, digital file-based video is timed and synchronized via a different set of structures, albeit structures that have been carefully designed to accommodate elements inherited from earlier formats.

  The synchronization of the elements that comprise the video picture stream, together with sound and other ancillary data, employs a multipart technology that emerged over time. The most intricate nuances of synchronization pertain to the picture-data stream itself, where they concern the sequence, timing, and flow of scan lines, fields, and frames. Playback devices synchronize the elements in the picture-data stream by responding to embedded changes in electrical voltage often referred to as *pulses* and, in one case, *colour burst*. Some examples occur with each video scan line, e.g., the horizontal blanking pulse which includes the horizontal synchronizing pulse and colour burst (once per scan line). These elements occur during what is called the *horizontal blanking interval*. Other synchronizing elements are associated with each field e.g., the vertical synchronizing pulse and pre- and post-equalizing pulses. These elements occur during the *vertical blanking interval*.

  Why blanking? Cathode-ray tubes (CRTs) provided the picture display for television sets from the mid-1930s until the first decade of the twenty-first century. In a CRT, an electron gun emits a stream of electrons that sweep across a phosphorescent screen, creating the glowing lines that comprise the picture. In order to prevent unwanted "extra" lines from appearing on the screen, the electron gun was turned off (blanked) at the end of each line and field. (See also B.1.2.7 below.)

  This is an immensely complex subject that is often given central (and lengthy) treatment in books that describe video technology. The successful presentation of video content—to say nothing of success in digitization—depends upon proper management of video synchronization and timing.

- **Variation:** In the United States and Japan, where the NTSC system prevailed, synchronization and timing were based on the RS-170 standard and its (very similar) successors. Strictly speaking, RS-170 (and successors) specifies only the monochrome picture component although it is extensively used with the NTSC colour encoding specification. A version that applies to PAL colour encoding also exists.

  In the United States, the FCC adopted the RS-170 specification associated with the implementation of NTSC colour (referred to as *RS-170a*) for broadcast use in 1953. (This requirement was made obsolete by the switch from analogue to digital broadcasting.) Thus, for broadcast professionals, RS-170 carried the force of law and was precisely adhered to. Meanwhile, in non-broadcast settings, the specification was treated only as a recommendation, and many non-broadcast recordings do not meet RS-170 specifications. Nevertheless, when non-broadcast tapes are digitised for preservation, it is a good practice to apply technologies that bring the signal in line with RS-170 to the degree possible. For more information, consult Conrac's *Raster Graphics Handbook, Chapter 8* (Conrac: n.d.), Tomi Engdahl's *RS-170 video signal* (Engdahl: 2009), and Ray Dall's *NTSC Composite Video Signals, and the RS - 170A Standards* (Dall: 2006).

  In Europe and other regions that did not employ NTSC specifications, the colour standards called PAL and SECAM included rules for timing and synchronization that are comparable to RS-170.[6] Although comparable, additional intricacies come into play. For example, there is proper phase relationship between the leading edge of horizontal sync and what is called the zero crossings of the colour burst. This phase relationship is referred to as *SCH* (or *Sc/H*, for Subcarrier to Horizontal). SCH phase is important when merging two or more video signals. If the video signals do not have the same horizontal, vertical, and subcarrier timing and closely matched phases, there is a risk of unwanted colour shifts. This phase relationship in PAL is more complex than for NTSC due to the way that PAL's sync and subcarrier frequencies relate to one another.

  Similar standards pertain to certain types of closed circuit and military video signals, rarely encountered in memory institution archives and not described in IASA-TC 06.[7]

6 PAL and SECAM were designed to serve the European picture frequency of 50 fields per second. Both were developed during the 1950s and the early 1960s and implemented in the mid-1960s. PAL was developed in Germany and patented by Telefunken in 1962. The French electronics manufacturer Thomson later bought Telefunken, as well as Compagnie Générale de Télévision that had developed SECAM in the late 1950s. Since they post-date NTSC by a few years, PAL and SECAM include some improvements over RS-170.

7 The standards alluded to here include EIA-343 (formerly RS-343), a signal standard for non-broadcast high resolution monochrome video and EIA-343A (formerly RS-343A), a video signal standard for high resolution monochrome CCTV that is based on EIA-343. There seems also to have been an RS-343 RGB (525, 625 or 875 lines). Some information is available from the ePanorama.net page titled "RS-170 video signal," including the following, "RS-343 specifies a 60 Hz non-interlaced scan with a composite sync signal with timings that produce a non-interlace (progressive) scan at 675 to 1023 lines. This standard is used by some computer systems and high resolution video cameras. Precision imaging systems, infrared targeting, low-light TV, night-vision and special military display systems, usually operate to high-resolution, RS-343 standards (875-line, 30-frame scan). They require specialized and costly recording and display equipment." (ePanorama.net, n.d., http://www.epanorama.net/documents/video/rs170.html, accessed 13 November 2017).

Some older videotape formats predate or do not adhere to the NTSC, PAL, or SECAM specifications. The signal on these videotapes may have a poor "native" ability to present synchronizing elements when played back. In order to successfully digitise some formats, the transfer system must include such devices as a time base corrector, processing amp, and/or frame synchronizer. (See section D, on workflow and metrics.)

- Relevant Wikipedia articles:
  - https://en.wikipedia.org/wiki/Blanking_level
  - https://en.wikipedia.org/wiki/CCIR_System_A
  - https://en.wikipedia.org/wiki/CCIR_System_B
  - https://en.wikipedia.org/wiki/CCIR_System_G
  - https://en.wikipedia.org/wiki/CCIR_System_H
  - https://en.wikipedia.org/wiki/CCIR_System_I
  - https://en.wikipedia.org/wiki/CCIR_System_M
  - https://en.wikipedia.org/wiki/Colour_broadcast_of_television_systems
  - https://en.wikipedia.org/wiki/Color_framing
  - https://en.wikipedia.org/wiki/Component_video
  - https://en.wikipedia.org/wiki/Composite_video
  - https://en.wikipedia.org/wiki/Horizontal_blanking_interval
  - https://en.wikipedia.org/wiki/NTSC
  - https://en.wikipedia.org/wiki/PAL
  - https://en.wikipedia.org/wiki/SCH_Phase_Display
  - https://en.wikipedia.org/wiki/SECAM
  - https://en.wikipedia.org/wiki/Subcarrier
  - https://en.wikipedia.org/wiki/Vertical_blanking_interval

B.1.2.7 Range of picture brightnesses and blanking "brightness"

- **Common feature:** Broadcast authorities established the bandwidth for analogue broadcasting as 6 MHz (megahertz) in the United States and ranging from 6 to 8 MHz in Europe. These limits constrain the overall video signal: all parts, combined, must "fit" into the bandwidth. Although these rules pertain to over-the-air broadcasts, their requirements are inevitably reflected in the characteristics of the signal recorded on videotape.

  One key part of the video signal concerns the luma or brightness information, and it is constrained, in part, to help manage overall signal bandwidth.[8] Luma is important for two reasons. First, the human eye is exceptionally sensitive to differences in brightness and can easily discern subtleties in the picture related to the representation of light and dark areas. Second, when colour television emerged in the 1950s and 1960s, there were millions of black-and-white television sets that translated luma data into picture. Both

8 *Luminance* concerns what is reflected from objects in the world, i.e., it is an area-based photometric measure of luminous intensity for light travelling in a given direction. In the realm of video, *luma* represents the brightness in an image, i.e., the "black-and-white" or achromatic portion, distinct from the *chroma* or colour portion. This distinction is nuanced and common (even expert) usage is sometimes loose and inexact (a polite way of saying *wrong*). The colour expert Charles Poynton writes that in video "a nonlinear transfer function—*gamma correction*—is applied to each of the linear R, G and B. Then a weighted sum of the nonlinear components is computed to form a signal representative of luminance. The resulting component is related to brightness but is not CIE luminance. Many video engineers call it *luma* and give it the symbol Y'. It is often carelessly called *luminance* and given the symbol Y. You must be careful to determine whether a particular author assigns a linear or nonlinear interpretation to the term luminance and the symbol Y" (Poynton: 1997, pp. 6-7). See also https://en.wikipedia.org/wiki/Luma_(video) and https://en.wikipedia.org/wiki/Luminance.

broadcasters and regulatory authorities wanted to continue to serve this installed base: if luma could be separated from chroma (colour data), this would permit older television receivers to display programs in black and white, while newer sets could show the same broadcasts in colour.

- **Variation:** The Institute of Radio Engineers (founded in 1912, joined the Institute of Electrical and Electronics Engineers in 1963) established the IRE convention for measuring relative brightnesses when represented by electrical voltages, which are themselves relative in this context. For broadcast, the rules state that the brightest values ought not exceed 100 on the IRE scale (there are some exceptions) and black ought to have a very low value. In the NTSC system used in the United States, black in the picture includes what is called a *setup* (it is "set up" to a higher value) and is at 7.5 IRE. In contrast, for the PAL system in other nations, and for NTSC as implemented in Japan, "picture" black is specified to fall at 0 IRE. (There are other variations in different national implementations of PAL.)

  Section B.1.2.6 above mentioned the important role of *blanking* in video, roughly speaking the exceedingly short times needed for the electron beam (in analogue systems) to move from the end of one field or frame, or the end of one horizontal scan line, to the start of the next, often called the *retrace line*. (These matters of timing have been rearticulated in the digital realm). During these blanking intervals, the brightness value for the electron beam is set as black in many systems and "blacker than black" in others. The horizontal blanking interval also includes a horizontal sync pulse with a value of -40 IRE in the NTSC system and -43 in SECAM and some PAL systems.

  When digitising videotapes, it is important to know which luma specifications were employed when the tape was recorded in order to avoid incorrect tonal representations in the copy.

  The elements described in the preceding paragraphs pertain to composite video, the signal type that prevails for most of the media-dependent formats described in this edition of IASA-TC 06. However, some instances of conventional, media-dependent formats carry a signal that employs a different encoding: *colour-difference component video* (see section B.1.3.1). Although colour-difference component is most often encountered in file-based digital formats, its *analogue* expression is found in videotape formats like Betacam SP, a carrier that is described in section C.7. As colour-difference component recording moved into a digital mode, limits were established for broadcasters that are analogous to the IRE limits described above.[9]

- Relevant Wikipedia articles:
  - https://en.wikipedia.org/wiki/Color_television
  - https://en.wikipedia.org/wiki/Component_video
  - https://en.wikipedia.org/wiki/Composite_video
  - https://en.wikipedia.org/wiki/IRE_(unit)
  - https://en.wikipedia.org/wiki/Luma_(video)
  - https://en.wikipedia.org/wiki/Luminance
  - https://en.wikipedia.org/wiki/NTSC
  - https://en.wikipedia.org/wiki/NTSC-J
  - https://en.wikipedia.org/wiki/PAL
  - https://en.wikipedia.org/wiki/Rec._601
  - https://en.wikipedia.org/wiki/Rec._709
  - https://en.wikipedia.org/wiki/YCbCr

B.1.3 Analogue video unpacked, part two: key features and variants continued

B.1.3.1 Colour encoding for video on conventional carriers

- **Common feature:** Like timing and synchronization (section B.1.2.6 above), the encoding of colour is immensely complex, variable, and, as it happens, it is interrelated with signal timing and synchronization.[10] There are a number of ways to encode colour data in electronic formats. In the digital era, for example, the trio of red, green, and blue (RGB) colour components are frequently encountered in still images and are also used in certain types

9 This topic receives some elaboration in section B.2.3.1. In brief, the first of the three colour-difference components is *luma*, usually abbreviated as Y or, by careful writers, as Y' to distinguish it from *luminance*. (The word *luminance*, however, is widely used where even technical writers appear to be discussing *luma*.) The second and third components carry *chroma* or colour data, sometimes abbreviated as U and V. These abbreviations, however, are not defined in a precise way, and careful writers will instead refer to Pb and Pr for the chroma elements in analogue component signals and to Cb and Cr for the analogous elements in digital component signals.

For digital colour-difference component signals, the rules are spelled out in ITU-R recommendations BT.601 and BT.709, and this digital articulation provides the easiest way to illustrate how the limits work. The underlying idea—analogue or digital—is to provide a buffer or headroom at both ends of the possible ranges of luma and chroma colour-difference component values. In digital lingo: "avoid clipping". The effect may be compared to the way in which IRE limits control the range of brightnesses in a composite signal. The limits in BT.601 apply to the three signal components: for an encoding with 8-bits of data per sample, Y' has a permissible range from 16–235 levels (from a possible 0–255), while Cr and Cb are permitted to range across 16–240 levels (from a possible 0–255). For 10-bit recordings, there is a similar range of constraints against a "possible" range of 0–1023. Signals that adhere to this limit are often referred to as "video range" or "legal range". In contrast, in the realm of computer graphics, one may instead encounter "wide range" or "super white" values for Y' and Cr and Cb that run from 0 to 255 (with 8-bit sampling).

A further evolution as digital metrics come into play is seen in the recommendation from the broadcast standards body EBU in their 2016 document R 103, *Video Signal Tolerance in Digital Television Systems*, ver. 2.0, which associates luma levels with digital sample values (as seen on a histogram, for example) to take the place of traditional voltage measures (https://tech.ebu.ch/docs/r/r103.pdf).

10 Some writers limit their use of the term *encoding* to digital entities, and even to lossy types of data compression. IASA-TC 06, however, uses the term in a broader way, defining code as any set of rules that governs the conversion of any kind information into another form for communication or mediated storage, e.g., Morse code for the alphabet (some would say that the alphabet itself is an encoding) in a telegraphic system.

of moving images. RGB provides chroma (colour) and luma (brightness) information in the same units of data.

In contrast, the *video* formats described in this edition of IASA-TC 06 encode chroma data separately from luma data, or "separably" in the case of composite signal (see below). The separation of luma and chroma information opens the door for data reduction that usefully decreases the need for transmission bandwidth or space on storage media. (For still images, the immensely successful JPEG compression format demonstrates this: its encoding system depends upon separate luma and chroma data.) In the digital realm, this data reduction is referred to as *chroma subsampling*, images are encoded with less resolution applied to chroma than to luma. This approach succeeds because human visual acuity is lower for colour differences than for differences in brightness.

- **Variation:** There are three main colour encoding structures employed by the formats covered in this edition of IASA-TC 06: (1) composite (including "colour-under"), (2) S-video, and (3) colour-difference component. These encodings are described in the following paragraphs. At least two of the three may be divided into further subtypes.

B.1.3.1.1 Composite video

Composite video consists of a linear combination of the luma and a subcarrier frequency modulated by the chroma information; the phase and amplitude of this signal correspond approximately to the hue and saturation of the colour. Luma and chroma are "separable" when they are decoded from the *composite* signal stream. Details of the encoding process vary between the NTSC, PAL, and SECAM systems.

Composite was the first widely adopted formatting approach for colour television, implemented in the United States during the 1950s in a business-competitive environment. In order to promote standardization and interoperability, and to permit viewers at home to continue their use of black-and-white television sets, the FCC empowered the NTSC to define a best and compatible approach. NTSC colour came into use in the 1960s, paralleled by similar developments for PAL and SECAM in Europe. (See also sections B.1.1.3, B.1.2.3, B.1.2.6, and B.1.2.7.)

NTSC and PAL encode the chroma data in a subcarrier using *quadrature amplitude modulation* (QAM). The signal carries chroma data at the same time as luma data. One of the intricacies, however, concerns what is called *colour framing*, the term for the "cadences" used to apply the colour data. Colour framing is not paced in the same way in the NTSC and PAL systems.

Meanwhile, SECAM uses a different approach for the modulation of chroma data onto its subcarrier. Instead of QAM, SECAM modulates via *frequency modulation* (FM). In addition, while NTSC and PAL transmit the red and blue information together, SECAM sends one at a time, and uses the information about the other colour from the preceding line. Conforming television receivers store one line of colour information in memory, which accounts for the words "sequential, with memory" that underlie SECAM's acronym.

Composite reduces the size of the video signal data stream (always a plus when transmitting or recording electronic information) by taking advantage of the separation of luma and chroma: the decrease is accomplished by reducing the bandwidth of the modulated colour subcarrier.

An additional signal-size reduction was developed in the 1970s for tape formats like U-matic, VHS, and Betamax. The physical dimensions and transport speed of these tape formats limits bandwidths to less than 1 MHz. In order to record colour in this narrow band, the quadrature phase-encoded and amplitude-modulated sine waves from the broadcast frequencies are transformed to lower frequencies. These types of recording systems are referred to as *heterodyne systems* or *colour-under systems*, with slightly different implementations for NTSC, PAL, and other signal structures.[11] When played back, the recorded information is de-heterodyned back to the standard subcarrier frequencies in order to provide for colour display and/or for signal interchange with other video equipment.

B.1.3.1.2 S-video

The *S* in S-video stands for *separate*, and the format is sometimes referred to as *Y/C*. By separating the luma (usually stated as Y in this context, more correctly as Y') and colour (C) portions of the signal, S-video provides better image quality than composite video but does not match the quality of colour-difference component video. As with composite video, the luma portion carries brightness information and the various synchronizing pulses, while the chroma portion contains data that represents both the saturation and the hue of the video. The improvement in quality results from the separation of data streams, thus avoiding the composite-signal requirement to carry chroma via a subcarrier. The mixing of the main carrier frequency with a subcarrier (at a different frequency) inevitably causes interference.

B.1.3.1.3 Colour-difference component video

Like S-video, a colour-difference component signal carries the luma stream (Y') as a separate channel of data. Meanwhile, the chroma data is carried in two colour-difference component streams:

- U (termed *Pb* for analogue video, *Cb* for digital) = blue minus luma
- V (*Pr* for analogue video, *Cr* for digital) = red minus luma

The carriage of chroma data in two streams adds a greater degree of separation than for the single stream in the case of S-video, thereby further improving picture quality.

The Y'UV trio of signal components are typically created from a different trio of components: RGB (red, green, and blue), initially captured by an image source like a camera. The initial processing of the data from the camera sensor is generally carried out "under wraps" in the camera. The outcome is that weighted values of R, G, and B are summed to produce Y´, a measure of overall brightness or luma. U and V are computed as scaled differences between Y´ and the B and R values. In actual practice, this requires a more complex calculation than the simple "blue minus luma" statement above. Meanwhile, all of the data in play in the preceding mathematical calculations means that the "missing" information about the colour green can be calculated.

Data reduction from *chroma subsampling* is well implemented for colour-difference component encoding. This can be applied to analogue signals but most explanations, e.g., at Wikipedia, limit their explanations for chroma subsampling to the digital realm,

11 The Wikipedia article on Heterodyne (https://en.wikipedia.org/wiki/Heterodyne, accessed 22 December 2017) offers this added information: "For instance, for NTSC video systems, the VHS (and S-VHS) recording system converts the colour subcarrier from the NTSC standard 3.58 MHz to ~629 kHz. PAL VHS colour subcarrier is similarly down-converted (but from 4.43 MHz). The now-obsolete 3/4" U-matic systems use a heterodyned ~688 kHz subcarrier for NTSC recordings (as does Sony's Betamax, which is at its basis a 1/2" consumer version of U-matic), while PAL U-matic decks came in two mutually incompatible varieties, with different subcarrier frequencies, known as Hi-Band and Low-Band. Other videotape formats with heterodyne colour systems include Video-8 and Hi8."

sketching the meaning of the now-familiar expressions 4:2:2, 4:2:0, 4:1:1, etc.[12] *Chroma subsampling notation* by the colour expert Charles Poynton offers an excellent three-page discussion of this topic (Poynton: 2002).

- Relevant Wikipedia articles (pertaining to all forms of video colour technology):
  - https://en.wikipedia.org/wiki/Chroma_subsampling
  - https://en.wikipedia.org/wiki/Chrominance
  - https://en.wikipedia.org/wiki/Color_framing
  - https://en.wikipedia.org/wiki/Color_television
  - https://en.wikipedia.org/wiki/Colour_broadcast_of_television_systems
  - https://en.wikipedia.org/wiki/Component_video
  - https://en.wikipedia.org/wiki/Composite_video
  - https://en.wikipedia.org/wiki/Heterodyne
  - https://en.wikipedia.org/wiki/Luma_(video)
  - https://en.wikipedia.org/wiki/NTSC
  - https://en.wikipedia.org/wiki/PAL
  - https://en.wikipedia.org/wiki/Quadrature_amplitude_modulation
  - https://en.wikipedia.org/wiki/S-Video
  - https://en.wikipedia.org/wiki/SECAM
  - https://en.wikipedia.org/wiki/YCbCr
  - https://en.wikipedia.org/wiki/YPbPr
  - https://en.wikipedia.org/wiki/YUV

12 In 4:2:2 subsampling, 4 luma samples are coordinated with 2-plus-2 chroma samples. The 4:2:2 structure is widely used in the production of professional video footage. In 4:2:0 or 4:1:1 subsampling, 4 luma samples are coordinated with 2 chroma samples (in slightly different patterns), and the image quality is lower than that provided by 4:2:2 sampling.

Sidebar: colour and tonal specifications for digital video and related matters

Digital video bitstreams, colour and more. Although the focus of this edition of IASA-TC 06 is analogue video, the specifications for digital video *colour* and *tonal*[13] representation are relevant. These are "output referred" specifications in that they dictate the interpretation and presentation of picture data in display monitors and projectors: you should adjust your recording to cater to these output devices. Most of the standards for digital display are promulgated by the International Telecommunication Union, Radiocommunication Sector, ITU-R for short.

Since virtually all analogue video recordings are standard definition, when digitised, they would typically align with ITU-R Recommendation BT.601 (originally published 1982, latest update 2011), *Rec. 601* for short. The Wikipedia article on Rec. 601 reports that the standard governs the "encoding [of] interlaced analogue video signals in digital video form. It includes methods of encoding 525-line 60 Hz and 625-line 50 Hz signals, both with an active region covering 720 luminance [*luma* is meant] samples and 360 chrominance [chroma] samples per line. The colour encoding system is known as YCbCr 4:2:2." Indeed, many digitising systems use this as a default when producing files from analogue standard definition sources.

It is also the case that the digitisation sampling of standard definition picture results in a raster of pixels that are described as "non-square," meaning that display systems must process the picture data for the correct aspect ratio. For example, the horizontal lines in a digitised PAL image carry 720 luma samples, and there are 576 lines. If the picture data were treated as square pixels, the image would have an aspect ratio of 5:4 ("squished" a bit); the pre-display processing instead yields the desired 4:3 aspect ratio.[14]

In contrast, ITU-R Recommendation BT.709 (originally published 1990, latest update is BT.709-6, 2015), *Rec. 709* for short, is intended to support high definition video. It was first published ("part 1") for two video formats that were never fully implemented (1035i30 and 1152i25 HDTV) and then ("part 2") for the widely implemented 1080i and 1080p standards (1080 horizontal lines, interlaced scan or progressive scan). Generally speaking, picture that conforms to Rec. 709 employs square pixels. Its tonal representation also differs slightly from that of Rec. 601.[15]

A predicament arises when an archive makes materials available to users on, say, the web, via streaming services like Netflix, DVD disk players, or via other computer-based systems, which often feature monitors that are adjusted to present colour and tonal range as if the signal conformed to Rec. 709, and as if the scan were progressive with square pixels. There is additional discussion of this predicament (and what can be done about it) in a parallel sidebar that follows section D.1.4.2.5, titled *What about a version of the recording for digital dissemination today?*

13 Rec. 601 defines a nonlinear transfer function that is linear near 0 and then transfers to a gamma curve for the rest of the light intensity range.

14 See "PAR, SAR, and DAR: Making Sense of Standard Definition (SD) video pixels" (Nagels: 2016) and the Wikipedia article "Pixel Aspect Ratio" (https://en.wikipedia.org/wiki/Pixel_aspect_ratio, accessed 22 December 2017) both of which offer good introductions and links to other sources of information.

15 The Rec. 709 transfer function from linear to nonlinear is similar-but-not identical to that used in the sRGB colour model (a still-image standard intended to serve typical computer monitors). The Rec. 709 transfer function is linear in the bottom part and moves to a power function for the rest of the range.

New digital display standards emerge from time to time. Looking ahead to future digital moving image output-referred display standards, readers should be aware of the following standards, with more in the pipeline:

- ITU-R Recommendation BT.2020-2 (2015, Rec. 2020 for short)
  - Wikipedia reports that Rec. 2020 "defines various aspects of ultra-high-definition television (UHDTV) with standard dynamic range (SDR) and wide colour gamut (WCG), including picture resolutions, frame rates with progressive scan, bit depths, colour primaries, RGB and luma-chroma colour representations, chroma subsamplings, and an opto-electronic transfer function".
- ITU-R Recommendation BT.2100 (2016, Rec. 2100 for short)
  - Wikipedia reports that Rec. 2100 "defines various aspects of high dynamic range (HDR) video such as display resolution (HDTV and UHDTV), frame rate, chroma subsampling, bit depth, colour space, and optical transfer function...Rec. 2100 expands on several aspects of Rec. 2020".
- DCI-P3
  - Wikipedia reports "DCI-P3, or DCI/P3, is a common RGB colour space for digital movie projection from the US-American film industry...DCI-P3 was defined by the Digital Cinema Initiatives (DCI) organization and published by the Society of Motion Picture and Television Engineers (SMPTE) in SMPTE EG 432-1[16] and SMPTE RP 431-2.[17]As a step towards the implementation of the significantly wider Rec. 2020 it is expected to see adoption in television systems and in the home cinema domain".

Relevant Wikipedia articles:
https://en.wikipedia.org/wiki/Pixel_aspect_ratio
https://en.wikipedia.org/wiki/Rec._601
https://en.wikipedia.org/wiki/Rec._709
https://en.wikipedia.org/wiki/Rec._2020
https://en.wikipedia.org/wiki/Rec._2100
https://en.wikipedia.org/wiki/DCI-P3

16 Society of Motion Picture and Television Engineers (SMPTE), Engineering Guideline EG 432-1:2010, *Digital Source Processing — Color Processing for D-Cinema*.

17 Society of Motion Picture and Television Engineers (SMPTE), Recommended Practice *RP 431-2, D-Cinema Quality – Reference Projector and Environment for the Display of DCDM in Review Rooms and Theaters.*

B.1.3.2 Ancillary data

- **Common feature:** This section concerns the video signal's carriage of information that goes beyond picture and sound, i.e., the types of additional information that make a good answer to the question, "What's in the video payload?" IASA-TC 06 uses the term *ancillary data* to name a broad range of non-picture, non-sound data, as outlined in the sections that follow. Readers should note, however, that our usage is a bit looser than that of most video engineers, who limit the term to the information carried in the signal's Horizontal Blanking Interval (HBI) and the Vertical Blanking Interval (VBI); see sections B.1.2.6 and B.1.3.2.1. The term appears in a number of SMPTE and EBU standards, although with some variation in definition.[18]

B.1.3.2.1 Ancillary data in the vertical blanking interval

In the media-dependent, conventional-carrier types of video under discussion in this initial edition of IASA-TC 06, much (but not all) ancillary data is carried in the signal's vertical blanking interval (VBI). This interval is the split-second between the end of the final line of a field and the beginning of the first line of the next. During the VBI, no pictorial image information is transmitted. The original need for a VBI reflected the inertia of the magnetic coils that deflected the electron beam on a cathode ray tube: picture data had to wait while the beam in the tube caught up. Meanwhile, there is also a pause between successive scan lines in the picture, to allow the beam to retrace from right to left.

VBI can carry a range of types of data and there is great variation in what might be found in a given recording. When broadcast transmissions were still analogue, some VBI data was inserted at broadcast time and was never (or rarely) recorded onto videotapes. VBI data may include:

- Vertical Interval Test Signals (VITS)
- Vertical Interval Reference Signal (VIRS)
- Vertical Interval Time Code (VITC; see B.1.3.2.1.1 below)
- Closed Captioning and Teletext (see B.1.3.2.1.2 below)
- Copy Generation Management System - Analog (CGMS-A; copy-protection indicators)
- Extended Data Service (XDS) which repeatedly sends metadata like channel call letters, program title, time of day, etc.
- Commercial Insertion Data
- Satellite Data

Ancillary data can be a significant component in historical recordings, and archivists should take this into account when planning preservation projects. Some data, if present, will be important to future researchers and the production of complete and authentic copies depends upon its retention. Examples of components with potential long-term importance include time code, captioning, and teletext, and section B.1.4.1 below elaborates upon the value of these components to an archive. Meanwhile, other data, if present, can support quality assurance in the digitising process. Examples include vertical interval test signals and the vertical interval reference signal. This topic receives elaboration in part D, on workflow and performance metrics.

18 Regarding the signal associated with conventional carriers, video engineers generally use the term *ancillary data* to name the information carried in the Horizontal Blanking Interval (HBI) and the Vertical Blanking Interval (VBI); see sections 2.2.6 and 2.3.2.1. But not all engineers consider every type of VBI data to be ancillary. The EBU Core ontology page for *Ancillary Data*, for example, defines the ancillary data class with this comment, "Any ancillary data provided with the content *other than* captioning and subtitling" (EBU: n.d., *emphasis* by IASA-TC 06).

B.1.3.2.1.1 *Vertical interval time code*

Vertical Interval Time Code (VITC) is a form of SMPTE time code, typically inserted into the VBI. The VITC code is always repeated on two adjacent video lines, one in each field. This internal redundancy is exploited by VITC readers. There can be more than one VITC pair in a single frame of video, and this capability is sometimes used to encode extra data that will not fit in a standard time code frame. Meanwhile, different production units in an organization may sometimes wish to encode different sets of time code on the same tape.

Sidebar: drop-frame and non-drop-frame time code

In the United States, Japan, and other NTSC regions, the implementation of a colour specification led to a special requirement for time code. As noted in section B.1.2.1, the introduction of NTSC colour changed the frame rate from 30 to 29.97 frames per second. This change "lengthened" the nominal hour by 3.59 seconds. For broadcasters who require an orderly programming schedule, the resulting error of almost a minute and a half over the course of a day was not acceptable. To compensate, SMPTE developed *drop frame* time code. In drop-frame mode, no video frames are actually omitted, but some of the time code labels are dropped. Frame numbers 0 and 1 of the first second of every minute are omitted, except when the number of minutes is divisible by ten. When drop-frame time code is applied as video recordings are played for broadcast, the residual timing error is a fully acceptable 86.4 milliseconds per day.[19] NTSC video recordings being digitised for preservation may carry embedded time code in either the normal "non-drop-frame" mode or the drop-frame mode, and this must the taken into account when digitising.

B.1.3.2.1.2 *Closed captioning, subtitles, and teletext*[20]

In the NTSC signal, captions are encoded as binary data into line 21, which exists at the cusp between VBI and active picture. Many authorities state that this line is part of the vertical blanking interval, while others note that is also the first line of NTSC active video. Meanwhile, teletext is handled differently in the PAL and SECAM systems but the methods are similar. Some experts report that, in the European context, consumer videotape recorders (Betamax and VHS) were often not capable of recording or playing back teletext. Another European technology is the binary subtitle data often nicknamed "EBU STL," which was formerly exchanged between producers and broadcasters via floppy disk.[21]

19 Wikipedia, "Drop frame timecode" heading in the article SMPTE timecode, https://en.wikipedia.org/wiki/SMPTE_timecode#Drop-frame_timecode, accessed 16 December 2017.

20 This document uses the terms *captions*, *subtitles*, and *teletext* more or less interchangeably, to mean non-XML text intended for display over a timeline, in synchronization with image and sound essences. The term *Timed Text* carries the same meaning with the added constraint that such text is structured to comply with either the SMPTE or EBU Timed Text XML schema: SMPTE ST 2052-1:2013 *Timed Text Format (SMPTE-TT)* and EBU Tech 3350, EBU-TT Part 1: *Archive & Exchange*; EBU Tech 3360, EBU-TT Part 2: *STL Mapping*; EBU Tech 3370, EBU-TT Part 3: *Live Contribution*; EBU Tech 3380, EBU-TT-D, format for the distribution of subtitles over IP; EBU Tech 3390, EBU-TT Part M, EBU-TT metadata elements.

21 See the 1991 publication *Specification of the EBU Subtitling data exchange format*, *TECH. 3264-E* (European Broadcast Union: 1991), which states "The medium for exchange is a 3.5-inch high-density portable magnetic disk (microfloppy). The disk is formatted for 1.44 Mbytes (2 sides, 80 tracks, 18 sectors/track). . . . The datafile format is defined for use with an IBM PC/XT/AT or compatible computer. The format is based upon the operating system MS/PC-DOS, version 3.3. ((European Broadcast Union (EBU), 1991. *Specification of the EBU Subtitling data exchange format*, *TECH. 3264-E*, https://tech.ebu.ch/docs/tech/tech3264.pdf, accessed 24 November 2017.))

For digital recordings, a new set of "packet-based" specifications for captions and subtitles replaced the NTSC line-21 and EBU STL rules. In more recent years, however, professionals in the United States and Europe have begun to use an XML format for captions and subtitles. Called *timed text* and standardized by SMPTE and EBU, this approach was derived from a standard promulgated by the World Wide Web Consortium (W3C), the source for the formats Timed-Text Markup Language (*.ttml*), WebVTT, and DFXP (the distribution format exchange profile).[22]

Meanwhile, a number of other industry formats carry caption data, some of which are widely adopted and interoperable: Caption Center, Captions Inc., Cheetah, LRC, MPSub, NCI, Scenarist Closed Caption (*.scc*), SubRip (*.srt*), SubViewer, and Videotrol Lambda.[23]

B.1.3.2.2 Longitudinal time code

Longitudinal (or Linear) Time Code (LTC) is SMPTE time code carried as binary data on a videotape's audio track, time code channel, or address track, depending on the tape format and user preferences. For example, videotape formats such as D2 and Betacam SP have an address track upon which LTC is recorded. In contrast ½-inch videotape formats such as VHS do not have a separate address track and, if present, LTC is recorded onto an audio channel and can be seen as consistent tone on, say, an audio VU meter.

In NTSC regions, the time code carried as LTC may be formatted in either drop-frame or non-drop-frame modes; see the sidebar above.

Readers should note that other types of time code are occasionally encountered in older video recordings. These are relatively exceptional and are not discussed in IASA-TC 06.

B.1.4 Archival value of ancillary and associated data[24]

Ancillary data like captions and time code are embedded in, and can be extracted from, the source video for inclusion in digitised preservation copies. Several relevant types of ancillary data are described in the preceding sections. In contrast, *associated data* names entities that were not embedded in the source video, but may be created in the course of digitisation, e.g., PREMIS preservation metadata. Associated data could subsequently be made part of or bundled with the digitised preservation copies; it is discussed in section B.1.4.2 below.

Ancillary and associated data may have significant value for archives and their future users. For this reason, archivists should appraise the types of data that are embedded in the source video materials they plan to digitise, in order to determine if they wish to make an effort to retain it. A similar assessment should be made regarding potential associated data for a given category of source items. If the appraisal and assessment indicates that these components are valuable enough to retain or to create, this will influence the selection of tools and methods for digitisation as well as the selection of the target format for preservation. Tools and methods are discussed in part D, and target format options are discussed in section B.3.

22 See also section B.3.3.2.4.
23 See also section B.3.3.2.4.
24 This and the following sections have been derived from the FADGI paper titled "User Needs and MXF Options: Preservation Planning and the AS-07 Specification" (FADGI: 2015). The entities called *associated data* in IASA-TC 06 are termed *associated materials* in AS-07. In 2019, the AS-07 specification, with minor updates, was published as SMPTE RDD 48.

B.1.4.1 Value of ancillary data

B.1.4.1.1 Value of retained captions, subtitles, and teletext

Captions, subtitles, and teletext, once extracted and indexed, have clear value for archives, supporting search and retrieval as well as other outcomes. If part of a source video recording, their retention is also required in order to produce a complete and authentic copy. What form of captions or subtitles ought to be carried in a preservation file? The source video's binary formats will be awkward for future extraction, since their use will depend on the continued availability of decoding tools and may require real-time playback. (Nevertheless, although embracing the idea of having converted captions or subtitles, many archivists also want to retain them in their original forms in order to have an authentic copy.) A good case can be made to extract binary caption data and convert it to XML Timed Text for carriage in the preservation copy. Timed Text supports a variety of indexing approaches.

B.1.4.1.2 Value of retained time code

Source materials that are to be digitised or acquired for preservation may carry multiple time codes: vertical interval time code (VITC), longitudinal time code (LTC), and other specialized types not described in this document. Some are present on purpose, others by accident, some may have good integrity and continuity, while others may be discontinuous. The legacy time codes in videotapes and other sources may themselves be layered in ways that an archive wishes to track, e.g., a videotape may carry VITC and may additionally carry an earlier generation of LTC recorded, say, on an audio track. Any or all of these time codes may support forensic investigations by future researchers. A legacy or historical time code may be keyed to old documents like tape logs, may provide clues about the older source tapes that were assembled to create the video program you are now preserving, and may (as with footage of, say, space vehicle launches) represent elapsed time that can be correlated to other data streams. In many cases, this is data you do not want to lose. The European Broadcasting Union (EBU) calls these *Historical Source Timecodes*.[25]

B.1.4.2 Value of associated data

Associated data of the types described in the sections that follow can (a) document the preservation treatment applied to a recording when it is digitised and (b) provide enhanced description of the original object. Thus, associated data helps an archive track and monitor their preservation activities, and provides future researchers with greater insight into the original collection item. This is, however, a flexible area and the identification of relevant associated data is an "archivist's call" and should result from a process of appraisal and assessment.

B.1.4.2.1 Value of developing and storing supplementary metadata

The term *supplementary metadata* is intended to connote metadata beyond that required by the digital-file-based format selected for preservation. The writers assume that the digitisation process will create one or more digital preservation files that contain technical metadata about file characteristics (typically embedded by production equipment) sufficient to permit downstream applications to play the files correctly. This "expected" metadata is sometimes referred to as *parametric* or *core technical* metadata.

The supplementary metadata sketched in the sections that follow goes beyond the core technical metadata. However, the "core technical" and "supplementary" categories overlap, and the supplementary categories as outlined below also overlap with each other. Archivists are advised to consider the full range of possible elements when they identify the metadata structures that best fit their organization's preservation activities.

25 *EBU-Recommendation R 122 Material Exchange Format Timecode Implementation, version 2.0, November 2010*; https://tech.ebu.ch/docs/r/r122.pdf, accessed 25 November 2017.

An additional perspective on this topic will be provided in the IASA-TC 06 discussion of metadata, planned for the future expanded edition of this guideline.

Supplementary metadata is typically formatted as simple ASCII or marked-up as XML.

B.1.4.2.1.1 Supplementary metadata types, examples, and value

- Process metadata: additional technical metadata about the production or digitisation activity
  - Value: permits the archive to document the methods and equipment used to produce preservation copies; supports quality assurance by permitting audits back from problems discovered in preservation objects to the system or settings that had been used in production. To a degree, this category overlaps with another category in this list, "Information about validation, specification compliance, and/or quality review outcomes".
  - Example: reVTMD.
    - Reference: https://www.archives.gov/preservation/products/reVTMD.xsd.
  - Example: videoMD.
    - Reference: https://www.loc.gov/standards/amdvmd/

- Frame-by-frame (or time code second-by-second) documentation of signal anomalies, sometimes called "logging metadata".
  - Example: SAMMA metadata[26]

- *Information about validation, specification compliance,*[27] *and/or quality review outcomes*
  - Value: Permits the archive to document how the production of a preservation copy was monitored and approved. To a degree, this category overlaps with the preceding category in this list, "Process metadata".
  - Example: Reports produced by quality control and/or validation tools, such as JHOVE, Interra Systems Baton, MediaConch, or special routines offered by ffmpeg, all illustrative examples available at the time of writing (2018).

- *Information about the source item*
  - Value: Permits the archive to document the identity and, if implemented, various details about the original item's type and condition; provides staff or researchers with facts about a source item that help explain why the preservation copy has certain characteristics.

26 The sale of SAMMA systems ceased in about 2015; information about SAMMA features is provided because a number of archives continue to employ the system (or other systems that produce the same output). Regarding SAMMA process metadata, there is an XML structure for SAMMA process metadata but, at this writing, there is no schema, website, or formal description thereof. A more-or-less random example is embedded in AS-07 sample file 2 (along with essences and much other data) posted at the FADGI website in September 2016. Select the "AS-07 Golden File: JPEG2000 (zip file; 205859 kb)" from this page: http://www.digitizationguidelines.gov/guidelines/MXF_app_sampleFiles.html. Subsequent to the drafting of this section, as AS-07 is re-designated SMPTE RDD 48, the sample file identified here is being replaced by a similar RDD 48 file that also contains Timed Text.

27 The terms *validation* and *specification compliance* are defined in the documentation for the reVTMD schema (NARA: 2012). Validation is defined as, "The process of determining the level of conformance of a digital object to the normative syntactic and semantic rules defined by the authoritative specification of the object's format. Data could include output of tools such as JHOVE or Interra Systems Baton." Specification compliance is defined as, "An indication that item is compliant with pre-defined specifications [and the metadata ought to] name the specification to which the item is compliant." (NARA, *Schema for reVTMD v.1.0*, http://www.archives.gov/preservation/products/reVTMD.xsd, accessed 24 November 2017.)

- Example: See reVTMD and videoMD above.
- Example: Active Format Description (AFD) data, a set of codes specified in the SMPTE ST 2016 series of standards, and in European Telecommunications Standards Institute (ETSI) TS 101 154 v1.7.1, the digital video broadcast standard. Active Format Description (AFD) codes may be found in new digital recordings and, if not present in source recordings (analogue or digital), could be added to newly produced digital preservation copies. For example, embedding code 1001 when digitising most analogue video would provide the digital file with metadata that instructs standards-compliant digital players to display this recording full frame in a 4:3 window or pillarboxed in a 16:9 window.[28]

- Preservation metadata
  - Value: Supports long-term preservation management.
  - Example: PREMIS[29]
    - Reference: http://www.loc.gov/standards/premis/

- *Additional cataloguing or descriptive metadata*[30]
  - Value: If whole or partial catalogue-type data records are embedded, these will (a) provide direct access to descriptive metadata and (b) will provide backup data for disaster recovery if an archive's main collection database is lost or damaged. In contrast, a minimal approach that embeds, say, (i) a working title, (ii) the archive's name, and (iii) a content identifier, will provide information that an end-user can use to track back to additional metadata.
  - Example: Copy of the collection-management data record from a system like MAVIS,[31] used by multiple IASA member archives. Alternatively, an archive may choose to simply include the record number or identifier from MAVIS or some other collection-management or cataloguing system.

28 During playout, AFD codes are carried in the baseband SDI video signal and/or MPEG video stream and they provide playback devices (e.g., home television receivers) with information about the stream's aspect ratio and active picture characteristics. In MXF files, the embedded data should be embedded to conform with SMPTE ST 377-1:2011, with constant AFD values stored in the MXF Picture Descriptor and changeable values stored in a SMPTE ST 436:2006-compliant VBI/ANC GC Data Element formatted according to SMPTE ST 2016-3. See also Wikipedia, *Active Format Description*, https://en.wikipedia.org/wiki/Active_Format_Description, accessed 16 December 2017.

29 From *Understanding PREMIS* (Library of Congress: 2009), page 3: "The PREMIS Data Dictionary defines preservation metadata as 'the information a repository uses to support the digital preservation process.' Here are some examples of preservation activities and how metadata can support them . . . [1] Checksum information stored as metadata can be used to tell if a stored file has changed between two points in time. . . . [2] Metadata can support media management by recording the type and age of storage media and the dates that files were last refreshed. . . . [3] Both migration and emulation [preservation] strategies require metadata about the original file formats and the hardware and software environments supporting them. . . . [4] Metadata can help support authenticity by documenting the digital provenance of the resource . . . ."

30 Although most archives maintain such data in a collections management database, catalogue, or finding aids, some archivists feel that there is value in embedding some or all of this data in preservation digital objects. Others, however, prefer to limit what is embedded to a more skeletal set of information, e.g., ensuring the inclusion of the identity of the responsible archive, and one or more identifiers for the content. If well selected, the latter can be used to find additional descriptive metadata.

31 MAVIS is an Oracle-based application, developed in the 1980s and 1990s by Wizard Information Services in cooperation with Australia's National Film and Sound Archive (NFSA). See the MAVIS entry in the NFSA preservation glossary, https://www.nfsa.gov.au/preservation/preservation-glossary/mavis, accessed 7 December 2017. As of 2018, some archives continue to use the application, licensed and maintained at this writing (2018) by Feenyx Pty. Ltd. (http://www.feenyx.com.au/, accessed 20 January 2018).

- Example: PBCore.[32] A comprehensive and flexible cataloging standard for the description of audiovisual content and a data sharing tool. PBCore consists of Root Elements, Asset Elements, Instantiation Elements. Root elements pertain to PBCore's XML structure; the asset elements carry what librarians call *descriptive* and *administrative* metadata; the instantiation elements carry technical metadata about the physical or digital representation of the AV asset, including such features as format, media type, duration, file size, data rate, aspect ratio, and so on.

All of the preceding categories are textual and thus will benefit from being formatted as XML, with registered schemas preferred. XML schemas do currently exist for some examples, e.g., PREMIS, reVTMD, and videoMD. In the absence of an XML structure, text may be embedded as simple strings or blocks in ASCII, UTF-8, or UTF-16 formats.

B.1.4.2.2 Value of a digital object manifest

A manifest for a digital preservation object (file or bundle of files), especially one that may contain multiple elements, supports preservation and good housekeeping by offering an inventory of the object's parts and expresses the relationships between them. Through a mix of required and optional elements, a manifest provides a high-level inventory of the parts including their identifiers, data description, MIME type, size and location. This information can help the user to better understand the composition of the file and it will also provide machine-interpretable information for content processing in later phases of the life cycle. Manifests of one sort or another are included in several formats ranging from the digital library community's BagIt specification[33] to the Interoperable Master Format (IMF)[34] developed by the entertainment industry in Hollywood to the SMPTE RDD 48 (former AS-07) MXF application specification[35] developed the U.S. Federal Agencies Digital Guidelines Initiative (FADGI).

Just as there are schools of thought about the value of embedding supplementary metadata in preservation master files, there are similar schools of thought about the value of associating a manifest with a preservation master, either via embedding or by some form of multi-file packaging. Archives should consider this topic in terms of their own operations.

Digital object manifests are typically formatted as simple ASCII or marked-up as XML.

32 PBCore main webpage: https://pbcore.org. Readers with a special interest in source-item technical metadata should consult "instantiationPhysical Video Vocabulary" (https://pbcore.org/pbcorecontrolled-vocabularies/instantiationphysical-videovocabulary/). High-level information about the description of file-based digital video is a part of the "pbcoreInstantiationDocument" (https://pbcore.org/elements/pbcoreinstantiationdocument).

33 Wikipedia, *BagIt*, https://en.wikipedia.org/wiki/BagIt, accessed 7 December 2017.

34 Information about IMF is provided in Annie Chang's *SMPTE Standards PDA Webcast: IMF (Interoperable Master Format)*, available via YouTube, https://www.youtube.com/watch?v=bmhv36hmSP4, accessed 12 October 2020.

35 The FADGI MXF application specification and related materials are linked from http://www.digitizationguidelines.gov/guidelines/MXF_app_spec.html, accessed 1 January 2019.

B.1.4.2.3 Value of storing binary-form associated materials

This pertains to materials intimately associated with the file's primary essences[36] that exist in binary formats. One type is represented by scanned images of such entities as the marked-up videotape box or relevant documents found in the box.[37] The inclusion of such representations contributes to the completeness, comprehensibility, or usability of the video information object by the holding archive and by future researchers. These representations generally take the form of files in formats such as TIFF, JPEG, PDF, and the like.

Another binary-form associated material that some might categorize as supplementary metadata is caption data in the European Broadcast Union subtitling format (STL), standardized in EBU Tech 3264.[38]

36 Broadcast professionals and some archivists use the term *essence* in a slightly loose (and thus variable) manner. The terms in use here will be refined for eventual inclusion in an IASA-TC 06 glossary; this footnote is a placeholder. (An identical footnote is holding place in section A.1.5.3.) These definitions are based upon (and are revisions of) the glossary in Richard Hopper's *EBU Project Group P/META Metadata Exchange Standards* (Hopper: 2000). In Hopper's EBU document, *essence* is defined as "the audio, graphic or text itself—the physical output which can be heard or seen by the consumer"; *metadata* is "the information or data that identifies and describes associated essence"; *content* is essence plus metadata (p. 25). Other EBU P/META definitions include *media object* and *media asset*, which are combination entities (p. 25), in effect content-plus-wrapper, thus including metadata; the glossary highlights rights metadata, important to broadcasters; media objects or assets represent content that can be stored and/or played.

37 For recorded sound archivists, the archetypal example of a binary-form associated material is an image of the label for a commercial disc recording.

38 European Broadcast Union, *Specification of the EBU Subtitling data exchange format, TECH. 3264-E, 1991*, https://tech.ebu.ch/docs/tech/tech3264.pdf, accessed 24 November 2017.

B.2 PRESERVABLE OBJECTS AND THE SELECTION OF FORMATS FOR PRESERVATION

B.2.1 Preservable objects and sustainable formats

B.2.1.1 Media independence

The preservation of video content depends upon having the payload in a *preservable* digital form, i.e., a form that can be managed for the long term. The preservable form that we seek must be *media-independent* or, to use a current term in broadcasting, *file-based*.

As indicated in the introduction (section A.1), media independence is not absolute. The management of all forms of digital data, including video, is accomplished in systems that employ storage media. The preservation of digital content depends upon storage media, e.g., spinning disks, data tape, flash drives, or some combination thereof (or upon some new form of media). For preservation over the long term, the content must be "storage-system-migrated," i.e., the files must be copied from media-to-media and/or from device-to-device, when storage systems require refreshment due to obsolescence. This fact of digital system life underlies the requirement that the formatting of data, to the degree possible, be media and system independent. (Content will also require migration in terms of data formats--from format-to-format--but this variable is independent of storage media.)

B.2.1.2 Sustainable digital data

The goal for an archive is to *sustain content* for the very long and indefinite future: we might as well say *forever*. As suggested above, however, *format sustainability* is measured against an undefined *long term*, i.e., "a reasonable period," "as long as practicable," but not forever. More often than not, the content that an archive manages has been migrated from one format to another at some earlier point in its life and it will be migrated again as time passes. A sustainable format is one that, when it takes its turn in this succession of migrations, can support the three guiding principles identified in the introduction (A.1):

1. Sustain authentic and complete copies of the original recordings
2. Sustain very high levels of quality in terms of the reproduction of picture and sound
3. Sustain features or elements that support access by future users

For reference, the following sidebar reiterates seven sustainability attributes of formats using wording excerpted from the Library of Congress sustainable formats web site.[39] Other authorities have also developed similar attribute lists.[40] Readers should note that the selection of a format for a given application or project must also take functional and practical factors into consideration, beyond sustainability factors, as described in section B.3.

39 "Sustainability Factors" page on the Library of Congress *Sustainability of Digital Formats* website (Library of Congress: n.d.), https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml, accessed 16 December 2017.

40 A number of resources are listed on the "Related Resources for Digital Format Sustainability" page on the Library of Congress *Sustainability of Digital Formats* website (Library of Congress: n.d.), https://www.loc.gov/preservation/digital/formats/intro/resources.shtml, accessed 25 November 2017.

Sidebar: Library of Congress sustainability factors

Disclosure. Disclosure refers to the degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content. Preservation of content in a given digital format over the long term is not feasible without an understanding of how the information is represented (encoded) as bits and bytes in digital files.

Adoption. Adoption refers to the degree to which the format is already used by the primary creators, disseminators, or users of information resources. This includes use as a master format, for delivery to end users, and as a means of interchange between systems. If a format is widely adopted, it is less likely to become obsolete rapidly, and tools for migration and emulation are more likely to emerge from industry without specific investment by archival institutions.

Transparency. Transparency refers to the degree to which the digital representation is open to direct analysis with basic tools, including human readability using a text-only editor. Digital formats in which the underlying information is represented simply and directly will be easier to migrate to new formats, easier to manage via system emulation, and more susceptible to digital archaeology. Many digital formats used for disseminating content employ encryption or compression. Encryption is incompatible with transparency; compression inhibits transparency. However, for practical reasons, some digital audio, images, and video may never be stored in an uncompressed form, even when created. Archival repositories must certainly accept content compressed using publicly disclosed and widely adopted algorithms that are either lossless or have a degree of lossy compression that is acceptable to the creator, publisher, or primary user as a master version.

Self-documentation. Digital objects that are self-documenting are likely to be easier to sustain over the long term and less vulnerable to catastrophe than data objects that are stored separately from all the metadata needed to render the data as usable information or understand its context. A digital object that contains basic descriptive metadata (the analogue to the title page of a book) and incorporates technical and administrative metadata relating to its creation and early stages of its lifecycle will be easier to manage and monitor for integrity and usability and to transfer reliably from one archival system to its successor system. Such metadata will also allow scholars of the future to understand how what they observe relates to the object as seen and used in its original technical environment. The ability of a digital format to hold (in a transparent form) metadata beyond that needed for basic rendering of the content in today's technical environment is an advantage for purposes of preservation.

External dependencies External dependencies refers to the degree to which a particular format depends on particular hardware, operating system, or software for rendering or use and the predicted complexity of dealing with those dependencies in future technical environments. Some forms of interactive digital content, although not tied to particular physical media, are designed for use with specific hardware, such as a microphone or a joystick. Scientific datasets built from sensor data may be useless without specialized software for analysis and visualization, software that may itself be very difficult to sustain, even with source code available.

Impact of patents. Patents related to a digital format may inhibit the ability of archival institutions to sustain content in that format. Although the costs for licenses to decode current formats are often low or nil, the existence of patents may slow the development of open source encoders and decoders and prices for commercial software for transcoding content in obsolescent formats may incorporate high license fees. When license terms include royalties based on use (e.g., a royalty fee when a file is encoded or each time it is used), costs could be high and unpredictable. It is not the existence of patents that is a potential problem, but the terms that patent-holders might choose to apply.

Technical protection mechanisms. To preserve digital content and provide service to users and designated communities decades hence, custodians must be able to replicate the content on new media, migrate and normalize it in the face of changing technology, and disseminate it to users at a resolution consistent with network bandwidth constraints. Content for which a trusted repository takes long-term responsibility must not be protected by technical mechanisms such as encryption, implemented in ways that prevent custodians from taking appropriate steps to preserve the digital content and make it accessible to future generations.

B.2.2 Selected terms that pertain to digital formats and formatting

There is wide variation in the usage of these terms among specialists in the field. When these terms are encountered "in the wild," readers should attend to the context in order to assess their meaning. The following definitions represent usage in this guideline.

B.2.2.1 Terms that pertain to the components of digital formats: *wrapper* and *encoding*

Video data formatting in a digital file or files pertains to both the *file wrapper* and the *encoded bitstream* (or bitstreams) that represent the video payload. The definitions of terms that follow are derived from the FADGI glossary (FADGI: n.d.). Wrappers provide a way to store and, at a high level, structure the data represented in the encoded bitstream. In addition, the wrapper usually also provides a mechanism to store technical and descriptive information (metadata). An encoding, on the other hand, defines the way the picture and sound essence[41] data is structured at the lowest level. For example: Will the data be RGB or colour-difference component? (Colour-difference component is typically nicknamed *YUV*; many broadcast professionals prefer the more precise term *Y'PbPr* if analogue, *Y'CbCr* if digital.) If YUV, what is the chroma subsampling? The encoding also determines how much data will be captured: what is the sampling rate and how much information will be captured at each sample? Other encoding features include the frame rate and the bit depth at each pixel or macropixel.

- **File wrapper:** 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. Archetypal (non-video) examples include WAVE and TIFF. Files that are instances of these wrappers are distinguished in terms of their underlying bitstreams, e.g., WAVE files may contain (a) linear pulse code modulated (LPCM) audio, (b) highly compressed audio as used for digital telephony, or (c) other representations of sound. Meanwhile, the familiar TIFF header typifies the self-describing, content-declaring feature of a wrapper. Relatively more complex and facile wrappers like QuickTime may contain multiple objects, e.g., one or more video streams and separate audio streams.

41 The term *essence* is discussed in footnotes to sections A.1.5.3 and B.1.4.2.3.

Wrappers are often specific to a content category but they may be members of a class defined by a more generic specification. For example, the sound format WAVE is an instance of the Microsoft RIFF class, which also includes the video format AVI.

There are a number of formats that do not represent wrapper archetypes like the ones mentioned above. For example, AAF and MXF are two closely related formats employed in television and motion picture production. Video and motion picture creators often call these formats wrappers but they feature some of the characteristics associated with self-describing bundling formats. The metadata in an AAF or MXF file, for example, may reference video and audio content that exists as a separate entity as well as content that is encapsulated within the file.

- **Bitstream encoding:** The transformation of a signal or data into a code by means of a programmed algorithm. The code[42] may serve any of a number of purposes such as transforming analogue information into digital form, compressing information for transmission or storage, encrypting or adding redundancies to the input code, or translating from one code to another. In IASA-TC 06, the term is used broadly, generally as a way to refer to the particular structure of the stored digital bitstream that represents a still image, an audio or video waveform, or the way in which, say, XML-marked-up text is structured.

B.2.2.2 Terms that pertain to processes or actions: *migrating*, *digitising*, *transcoding*, and *rewrapping*

- **System Migration**. In the field of digital preservation, the term *migration* is used in two ways. The first—*system migration*, also called *physical migration*—is an element in the discussion in the preceding section. System migration refers to the movement of digital files from an obsolete data-management system to a new system, part of what is sometimes called *media refreshment*. In system migration "the bits don't change".

- **Format Migration**. *Format migration* also called *logical migration* refers the movement of a content item from one format to another: "the bits do change". The following three terms describe three types of format migration.

  - **Digitising.** IASA-TC 06 uses this term broadly, for an action that generally involves playing a video recording (usually but not always on videotape) in real time, and the processing of the video payload, also in real time, by a system that produces a digital file. In some cases, e.g., when a Digital Betacam videocassette is the source recording, the playback goes from digital-to-digital in real time, and some specialists will instead use the term *transcoding*, discussed below. Terminology notwithstanding, the process is the same.

    With an analogue source recording, e.g., a U-matic cassette, digitisation results from the combined action of (1) a video tape recorder (VTR) that plays back the recording (sometimes supported by other devices, as described in section D.1.3.1.4) and converts the payload from analogue to digital and (2)

42 As indicated in a footnote to section B.1.3.1, IASA-TC 06 uses the terms *code* and *encoding* in a broad way: a set of rules that governs the conversion of any kind information into another form for communication or mediated storage, including analogue examples like the Morse code formerly used in telegraphy.

a digitising system that encodes the payload elements and produces the file. In many cases the transmission will be from the playback device(s) to the digitising system via a Serial Digital Interface (SDI). In other cases, however, these devices and systems are bundled and integrated by manufacturers, and segments of the process happen "under the covers". (See section D.1.3.1.9 *Digitisation systems*.)

If the source recording is in a digital encoding that cannot be transferred "as data" (see *rewrapping*, discussed below), e.g., the Digital Betacam cassette alluded to above, the process does not include analogue-to-digital conversion. Digital Betacam cassettes—to elaborate on that example—carry video data compressed with a proprietary algorithm. The Digital Betacam VTR decodes (and transcodes) this data and transmits the signal via a Serial Digital Interface (SDI) just like the typical output from the analogue playback devices. In both cases, the digitisation process transforms important parts of the payload, e.g., the picture and sound essences,[43] and wraps the now-digital payload as a digital file or a bundle of related files.

- **Transcoding**. Transcoding generally entails the re-encoding of a video recording from one digital bitstream to another. In many cases, the input to the process is a digital file and the resulting output features both the transcoded bitstream and also a new file wrapper. One archetypal transcoding action is familiar to video creators who load newly shot footage into a non-linear editing system: a stream of video is transferred from a camcorder or other playback device into the editing workstation where the stream is transcoded and wrapped. For example, a camcorder's HDV footage is loaded into FinalCutPro and, in the course of this ingestion, the bitstream is transcoded to Apple's ProRes format and wrapped as a QuickTime *.mov* file.

- **Rewrapping**. Rewrapping entails the retention of a video recording: first, it is extracted from the file wrapper in which it was acquired and, second, it is placed in a new file wrapper without changing the payload. For example, an archive has an MPEG-2 recording in a file in the *.mpg* format and *rewraps* it in Matroska, MXF, or AVI. In another example, an archive has received an oral history recording in the form of a write-once DVD that, like most DVDs, carries underlying MPEG-2 video streams as *VOB* files. To prepare this content for long-term preservation, the archive may prepare a *disk image* and also *rewrap* the MPEG video signal from the VOB files as, say, Matroska, MXF, or AVI files.[44] As a third example, the video payloads on DV tapes can be copied "as data" and wrapped as a DV file or as Matroska or MXF.

B.2.3 Preservation format considerations

B.2.3.1 Factors that influence preservation format selection

Three considerations influence the selection of digital formats that support preservation:

- An assessment of the "life expectancies" of the formats under consideration (section B.2.3.2)
- The technical classification of the source video materials to be preserved (section B.2.4)
- The preservation strategies adopted by the archive (section B.2.4)

43 The term *essence* is discussed in footnotes to sections A.1.5.3 and B.1.4.2.3.
44 For more on DVD preservation, see the footnote to section B.2.4.5.

B.2.3.2 Format life expectancy and the inevitability of format migration

Preservation specialists speak of the *life expectancy* of physical media: polyester microfilm stored in an expert-recommended cool and dry environment has a life expectancy of 500 years. In contrast, some specialists estimate the life expectancy for videotape as ranging from 30 to 60 years, although the availability of playback equipment (or the lack thereof) may mean that the timeframe for recovering the video signal is shorter than the life expectancy of the tape as a medium.

Is there a life expectancy for digital file formats? Can this be taken into account when formats are selected as targets for preservation digitisation? There are no definitive answers to those questions. Specialists in the field have made varying attempts to assess digital-file-format life expectancies. At best, these assessments come to a range-of-years conclusion and, like the assessment of the life expectancy of videotape, are couched in terms of dependencies: will there be applications that can play the format, how widely adopted is the format, is it well documented in published specifications?

The safe conclusion is to assume that content in *any* digital file format will require migration to a new format at some point. The authors of IASA-TC 06 have made some educated guesses, and, in this guideline, we designate some formats as "for the long term," meaning that we anticipate life expectancies (time to a required format migration) on the order of twenty years or more. Other formats are designated as "for the medium term," meaning that anticipated life expectancies will be on the order of ten or twenty years. Some formats—often proprietary and not widely adopted—are considered to be "for the short term," i.e., ten years to a required format migration.

B.2.4 Preservation target formats, "if-then" strategies for 6 classes of video recordings

The term *target format* names the digital entity (a file or a bundle of files) that receives and carries the digitised content from a source video recording. The target formats recommended in this guideline are those that support the long-term preservation of the video content, and they are briefly introduced in the sections below. **Please note that a thorough discussion of the IASA-TC 06 first edition target format options is provided in section B.3, *Target Formats for Video Sources that Must Be Transferred "as Video" in Real-Time*.**

In order to frame its overall discussion of preservation strategies and methods, this guideline identifies six classes of source video recordings. These classes were introduced in section A.1.3.1. The sections that follow offer sketches of the six classes together with high-level notes on preservation strategies, methods, and target format options. In this "if-then" discussion, identical target formats are recommended for classes 1 and 2, and very similar recommendations are made for class 3. Regarding classes 4 and 5, the target format recommendations are similar to the preceding in terms of the wrapper, but they differ in terms of the encoded bitstream(s) to be carried within the wrapper. Class 6 is "retain as acquired".

The target format options that apply to classes 1 and 2 receive extensive coverage in section B.3 (*Target Formats for Video Sources that Must Be Transferred "as Video" in Real-Time*). Discussion of the formatting for classes 3, 4, and 5, and additional information on class 6, awaits the future expanded edition of IASA-TC 06.

B.2.4.1 Class 1: Analogue video recordings

- Characteristics:
  - The archetype for this class is the analogue videotape; laser videodiscs are also members of this class. The initial edition of IASA-TC 06 discusses the following class 1 source formats:
    C.2 Quadruplex 2-inch Videotapes
    C.3 EIAJ and Sony CV ½-inch Videotapes
    C.4 1-inch Helical-scan Open Reel Videotapes
    C.5 U-matic ¾-inch Videocassettes
    C.6 ½-inch Analogue Consumer and Semi-Professional Videocassettes
    C.7 Betacam ½-inch Professional Videocassette Family (analogue members)
  - Class 1 source recordings must be transferred by playing the tape in real time.
  - The real-time transfer of video requires significant resources: playback equipment, recording systems, systems to monitor metrics and support quality review, and skilled personnel. The transfer entails the conversion of the analogue signal to digital data.

- Target format strategy for class 1:
  - Since the resource requirements are significant, the recommended strategy for this category is to select a *target format* with a long-term "life expectancy".
    - *Option one:* "Marketplace" wrappers (AVI, QuickTime) that contain FFV1 or uncompressed v210; *detailed information in section B.3.1.2.1*.
    - *Option two:* Matroska/FFV1 (as specified via IETF in the MediaConch/CELLAR/Preforma project); *detailed information in section B.3.1.2.2.*
    - *Option three:* MXF that contains lossless JPEG 2000. (SMPTE RDD 48 and SAMMA[45]); *detailed information in section B.3.1.2.3.*
    - *Option four:* MXF that contains uncompressed v210. (Generic MXF/
    - v210 and SMPTE RDD 48); *detailed information in section B.3.1.2.3.*

B.2.4.2 Class 2: Digital videotapes with encodings that are "out of reach" or inappropriate for long-term retention

- Characteristics:
  - Members of this class are digital videotapes that carry proprietary digital bitstreams or employ composite-format digital encodings that are inappropriate for long-term file-based preservation, e.g., digital composite. The initial edition of IASA-TC 06 discusses the following class 2 source formats:
    C.7 Selected digital examples of Betacam ½-inch Professional Videocassette Family
  - Class 2 source recordings must be transferred by playing the tape in real time. DigiBeta content, for example, cannot be transferred "as data," and the transfer entails the transcoding of the proprietary data on the tape to standardized digital data.
  - The real-time transfer of video requires significant resources: playback equipment, recording systems, systems to monitor metrics and support quality review, and skilled personnel.

45 SAMMA was not marketed after 2015, but a number of archives continue to employ the system or systems that produce the same output and many hold files with the SAMMA MXF format profile.

- Target format strategy for class 2:
  - Since the resource requirements are significant, the recommended strategy for this category is to select a *target format* with a long-term "life expectancy".
    - *Option one:* "Marketplace" wrappers (AVI, QuickTime) that contain FFV1 or uncompressed v210; *detailed information in section B.3.1.2.1*.
    - *Option two:* Matroska/FFV1 (as specified via IETF in the MediaConch/CELLAR/Preforma project); *detailed information in section B.3.1.2.2.*
    - *Option three:* MXF that contains lossless JPEG 2000. (SMPTE RDD 48 and SAMMA); *detailed information in section B.3.1.2.3.*
    - *Option four:* MXF that contains uncompressed v210. (Generic MXF/v210 and SMPTE RDD 48); *detailed information in section B.3.1.2.3.*

B.2.4.3 Class 3: Digital videotapes with encodings that can be extracted "as data"

- Characteristics:
  - Members of this class are digital videotapes that carry digital bitstreams with open specifications and that are suitable for file-based, long-term preservation. The initial edition of IASA-TC 06 discusses the following class 3 source formats:
    C.7 Selected digital examples of Betacam ½-inch Professional Videocassette Family
  - Future editions of IASA-TC 06 will discuss other members of this class, e.g., the DV family and selected members of the "D" family (D-1, D-3, etc.).
  - Class 3 recordings must be played in real time in order to extract the content, which can be transferred "as data," e.g., Betacam IMX.
  - The real-time transfer of video requires significant resources: playback equipment, recording systems, systems to monitor metrics and support quality review, and skilled personnel.

- Strategies (A and B) and target format options for class 3:
  - Stratagem 3A: Since the source video digital formats are lossy compressed, they are smaller than the file sizes for long term formats (uncompressed or lossless compressed). However, their "life expectancy" is medium- or even short-term. Thus stratagem 3A consists of *transfer-and-wrap-without-transcoding*, together with a plan to migrate to a long-term format in ten or twenty years, when digital-data storage will be cheaper and easier to manage.
    - Example: If DV family, DV encodings wrapped as MXF, Matroska, or one of the marketplace wrappers (QuickTime, AVI).
    - Example: If IMX, MPEG-2 encoding (extracted from IMX) and other payload elements wrapped as MXF or Matroska
  - Stratagem 3B: Skip the medium-term "hold as is" step (i.e., stratagem 3A) and migrate to a target format with a long-term "life expectancy" upon acquisition or initial ingestion. This essence[46] migration from digital to digital entails *transcoding*. Stratagem 3B reduces future effort but requires the provision of more extensive digital-data storage in the near term since long-term formats are larger than short- or medium-term.
    - Example: Transcode any of the listed source encodings to one of the target formats recommended for classes 1 and 2.

- Discussion of target format types, standardized digital data, and related matters: future edition of IASA-TC 06.

B.2.4.4 Class 4: File-based digital video source materials that warrant (early) transcoding or rewrapping

- Characteristics:
  - Members of this class are digital files that carry digital bitstreams for which the "life expectancy" is moderate or poor. Examples include RealMedia video and Windows Video recordings. The "jury is still out" on some formats, e.g., ProRes encoding in QuickTime, although many specialists place this in class 6.
  - Class 4 recordings can be transferred and transcoded faster than real time.
  - This class will be discussed in future editions of IASA-TC 06.

- Strategy options for class 4:
  - Transcode the underlying bitstreams and use the target format options outlined for classes 1 and 2.

- Discussion of target format types, standardized digital data, and related matters: future edition of IASA-TC 06.

46 Discussion of the term *essence* is provided in footnotes to sections A.1.5.3 and B.1.4.2.3.

B.2.4.5 Class 5: Authored disc-based digital recordings

- Characteristics:
  - Write-once (non-protected) authored DVDs (i.e., not DVD-ROMs); in most cases, the underlying MPEG bitstreams within the DVD's VOB file structure can be reformatted without transcoding. IASA-TC 06 will not address the conversion of copyright protected and/or encrypted DVDs, although some of the same underlying principles and techniques apply.
  - Class 5 recordings can be transferred and transcoded faster than real time.
  - This class will be discussed in future editions of IASA-TC 06.

- Strategy options for class 5:
  - Typical method: disk image may be produced to serve as exact copy of the source item, often considered to be the *preservation master*; secondary copy generally produced by retaining the digital encoding and rewrapping as digital video files, often considered to be a *viewing copies* or a *production master*.[47]
  - Example: authored DVD, such as an unencrypted write-once disc from an amateur home recording.

- Discussion of target format types, standardized digital data, and related matters: future edition of IASA-TC 06.

B.2.4.6 Class 6: File-based digital video source materials that do not warrant transcoding or rewrapping

- Characteristics:
  - Members of this class are digital files that carry digital bitstreams for which the "life expectancy" is good. Examples include FFV1 encoding in AVI and MPEG in MXF. The "jury is still out" on some formats, e.g., ProRes encoding in QuickTime, although many specialists place this in class 6.

- Class 6 recordings do not require treatment and can be retained as acquired.

- This class will be discussed in future editions of IASA-TC 06.

47 Regarding DVDs, here are two key references: (1) Smithsonian Institution Archives, 2014. "Preserving Content from Authored Video DVDs," in *Creating and Archiving Born Digital Video Part II. Eight Federal Case Histories*, pp. 42-46, http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p2_20141202.pdf, accessed 11 November 2017. (2) George Blood Audio/Video, 2014. "Preserving Write-Once DVDs: Producing Disk Images, Extracting Content, and Addressing Flaws and Errors," a report for the Library of Congress, http://www.digitizationguidelines.gov/audio-visual/documents/Preserve_DVDs_BloodReport_20140901.pdf, accessed 11 November 2017. The terms used in the typical method statement for class 5 above are based on the following statement from the second reference: "[The production of] ISO disk image files from the optical discs [is] a process henceforth referred to as 'cloning.' Since these ISO files contain all of the data on the disc, and retain the logical structure of this data, the Library considers these ISO files as *archival master files*. The second action, henceforth referred as 'extracting,' is to extract the underlying digital video content from the ISO files. To the degree possible, the video and audio from the ISO files is left in the encoding 'as found' . . . . The Library considers this extracted content to represent *production master files*" (page 4).

B.3 TARGET FORMATS FOR VIDEO RECORDINGS TO BE DIGITISED "AS VIDEO" IN REAL TIME

B.3.1 Introduction to target formats

B.3.1.1 Evaluating and selecting target formats for digitisation projects

The term *target format* names the digital entity (a file or a bundle of files) that receive and carry the content transferred from a source video recording. This section discusses the target formats recommended when the transfer of video content occurs in real time, and the transfer is "as video" (not "as data"), categorized as *class 1* and *class 2* source materials in the preceding sections. Although the archetypal example is analogue videotape, this category also includes digital videotapes that must be played in order to acquire the content, e.g., SONY DigiBeta cassettes. (It also includes formats like laser videodiscs, not discussed in IASA-TC06.) Future expanded editions of IASA-TC 06 will discuss target format selection for recordings on conventional media that can be transferred "as data" or file-based recordings that can be *transcoded* and/or *rewrapped* in non-real time, usually faster than real time.

Section B.1 of this guideline, titled *The Video Signal: Format and Features*, describes the complex formatting of video. In an echo of that complexity, this section explores the formatting of the digital master files (or bundle of files) that carry and sustain video content to support preservation. Video data formatting pertains to both the *file wrapper* and the *encoded bitstream*[48] (or bitstreams) that represent the video payload.

Wrappers provide a way to store and, at a high level, structure the data represented in the encoded bitstream. In addition, the wrapper usually provides a mechanism to store technical and descriptive information (metadata), although some encodings (FFV1 is a notable example) also carry embedded metadata. For the most part, an encoding defines the way the picture and sound essence data is structured at the lowest level.[49] The encoding selected will determine, for example, the chroma subsampling represented in a given bitstream, the sampling rate and the extent of data per sample, the frame rate, and many other technical parameters.

The Video Signal: Format and Features describes *ancillary* and *associated data*, which includes a mix of essence (e.g., captions, subtitles, time code) and non-essence components (e.g., supplementary metadata) of a video payload. This topic receives elaboration in section B.3.3.2 titled "Capabilities regarding ancillary and associated data" below. Some target formats—usually the wrapper—can carry some or all of the ancillary and associated data that an archive determines to be of value; in other cases, however, additional "sidecar" files must be employed in order to retain this added data. Sidecar options are discussed in the sections that follow.

B.3.1.2 Three important format families

Preservation-oriented archives employ a number of target formats and the community is actively monitoring the development of several new implementations, i.e., refined profiles of wrappers and/or encodings that have been available for some time. At least three general types are of interest in the IASA-TC 06 context, each with its own applications and use cases, its own advocates, and a variety of pros and cons. The plural

48 Audiovisual specialists use the terms *encoding* and *codec* (from coding and decoding) in a loose and variable manner. Their use in IASA-TC 06 is broad, as suggested in the descriptions of analogue video in sections B.1.2 and B.1.3. Unlike some writers, we do not limit the terms to the realm of lossy compressed digital data. *A Primer on Codecs for Moving Image and Sound Archives* offers a helpful introduction to the terms and their application to audiovisual content (Lacinak: 2010).

49 Discussion of the term *essence* is provided in footnotes to sections A.1.5.3 and B.1.4.2.3.

term *general types* indicates that these are format families, groups of related formats, each with a slightly different combination of a wrapper and an encoding.

The formats and some subtypes receive brief descriptions in the sections below, and section B.2.1.2.4 reports memory institutions' ever-growing preference for the combination of the Matroska wrapper and FFV1 encoding. Detailed tables that compare these formats are provided in section B.3.4 and its associated annexes.

For the first category below, the IASA-TC 06 authors invented the term *marketplace wrapper* as a lexical convenience. It is the case, of course, that all of the wrappers in the list may be selected from the same metaphorical marketplace.

B.3.1.2.1 "Marketplace wrappers" with picture as lossless compressed FFV1 or as 10-bit-deep uncompressed, 4:2:2 chroma subsampling

- Subtype: OpenDML AVI wrapper,[50] FFV1 encoded video.[51] May require "sidecar" files to carry captions or subtitles, if present, and supplementary metadata to document some payload elements as described below.
- Subtype: OpenDML AVI wrapper, v210 encoded video.[52] May require "sidecar" files to carry captions or subtitles, if present, and supplementary metadata to document some payload elements as described below.
  - Comment: The OpenDML extension is required in order to support files greater than 2GB; this extension also offers additional support for valuable embedded metadata.
- Subtype: QuickTime wrapper,[53] v210 encoded video.

50 The AVI (Audio Video Interleave) format is based on the RIFF (Resource Interchange File Format) specification published in the *Multimedia Programming Interface and Data Specifications 1.0* (IBM and Microsoft: 1991). The WAVE audio-wrapper format is also a subtype of RIFF. Detailed specifications for AVI are provided in chapter 4 of the 1993 publication *Video for Windows Programmer's Guide* (Microsoft: 1993). Additional useful information about AVI is provided by John F. McGowan's *AVI Overview* (McGowan, John F.: 2004). Documentation of the OpenDML extensions is found in *OpenDML AVI File Format Extensions, Version 1.02* (OpenDML AVI M-JPEG File Format Subcommittee: 1997).

51 *FFV1 Video Coding Format Version 4* [main specification], draft-ietf-cellar-ffv1-v4-03 (draft version 03, 18 October 2018, expires 21 April 2019) in various formats: https://tools.ietf.org/html/draft-ietf-cellar-ffv1-v4/; related to earlier version *FFV1 Video Coding Format Version 0, 1, and 3* (draft version 06, 18 October 2018, expires 21 April 2019) in various formats: https://tools.ietf.org/html/draft-ietf-cellar-ffv1/. All preceding URLs accessed 19 January 2019; updating of all specifications continues; to identify latest versions, consult https://datatracker.ietf.org/.

52 The most complete v210 specification is provided in Apple developer documentation under the heading "v210' 4:2:2 CompressionType," part of the 1999 Apple technical note sometimes referred to as "tn2162," https://developer.apple.com/library/mac/technotes/tn2162/_index.html#//apple_ref/doc/uid/DTS40013070-CH1-TNTAG8-V210__4_2_2_COMPRESSION_TYPE, accessed 7 December 2017. In terms of bona fide standardisation, v210 was subsequently memorialised in SMPTE standard *ST 377-1:2011, Material Exchange Format (MXF) — File Format Specification*. See also "V210Video Picture Encoding" on the Library of Congress Format Sustainability website, http://www.digitalpreservation.gov/formats/fdd/fdd000353.shtml, accessed 16 December 2017. Like other encodings, v210 may be incorrectly formed by some applications and, in addition, may not be strictly lossless. Mike Casey in his valuable *White Paper: Encoding and Wrapper Decisions and Implementation for Video Preservation Master Files* reports: "If implemented correctly, v210 reserves the very high and very low sample values for synchronisation purposes. Observed samples in the reserved ranges are set to the closest valid value. Not all encoders follow the v210 specification correctly—for example, [one system] will use the reserved values for accurate sample values, thereby providing a true lossless uncompressed file that is, nevertheless, not fully compliant with the specification" (Casey: 2017, p. 4).

53 Both the "classic" (2001) and updated-through-2012 versions of the QuickTime File Format specification link from this menu page: https://developer.apple.com/standards/classicquicktime.html, accessed 21 January 2019.

B.3.1.2.2 Matroska wrapper with picture as losslessly compressed FFV1

- Subtype: format version that employs Matroska[54] and FFV1[55] as standardised by IETF (in progress), under the auspices of CELLAR[56]

B.3.1.2.3 MXF wrapper with uncompressed picture or JPEG 2000 compressed picture

- Subtype: MXF wrapper with picture as 10-bit-deep uncompressed, 4:2:2 chroma subsampling, generic version, conformant to SMPTE standards and realised in various digitisation systems[57]
- Subtype: Same wrapper and encoding, as described in the SMPTE RDD 48 (former FADGI AS-07) specification[58]
- Subtype: Same wrapper and encoding, as described in the BBC White Paper 241 (Glanville and Heritage: 2013)
- Subtype: MXF wrapper with picture as losslessly compressed JPEG 2000, described in the SMPTE RDD 48 (former FADGI AS-07) specification
- Subtype: Same wrapper and encoding, as produced by the SAMMA devices, now discontinued but formerly available from Media Matters and subsequently Front Porch Digital[59]

54 *Matroska Specifications* [main specification], draft-ietf-cellar-matroska-02 (draft version 02, 9 January 2019, expires 13 July 2019) in various document formats: https://tools.ietf.org/html/draft-ietf-cellar-matroska-02. Supported by additional specifications: *Matroska Codec* [mapping for codecs in Matroska] draft-ietf-cellar-codec-01 (draft version 01, 9 January 2019, expires 13 July 2019) in various document formats: https://tools.ietf.org/html/draft-ietf-cellar-codec-01; and by *Matroska Tags* [about metadata], draft-ietf-cellar-tags-01 (draft version 01, 9 January 2019, expires 13 July 2019) in various document formats: https://tools.ietf.org/html/draft-ietf-cellar-tags-01. Also relevant is *Extensible Binary Meta Language* [the EBML specification] draft-ietf-cellar-ebml-08 (draft version 08, 27 November 2018, expires 31 May 2019) in various document formats: https://tools.ietf.org/html/draft-ietf-cellar-ebml-08. All preceding URLs accessed 19 January 2019; updating of all specifications continues, to identify latest versions, consult https://datatracker.ietf.org/.

55 See footnote 48.

56 CELLAR is an acronym for Codec Encoding for LossLess Archiving and Realtime transmission; see https://datatracker.ietf.org/wg/cellar/charter/, accessed 21 January 2019.

57 The fundamental model for uncompressed video (v210 if 10-bit sampling is used) in MXF depends upon adherence to *SMPTE ST 377-1:2011, Material Exchange Format (MXF) — File Format Specification* (including amendment Am1:2012), and *ST 384:2005, Mapping of Uncompressed Pictures into the MXF Generic Container.* The production of conformant files has been and continues to be offered by commercial vendors.

58 The SMPTE RDD 48 (former FADGI AS-07) application specification and related materials are linked from http://www.digitizationguidelines.gov/guidelines/MXF_app_spec.html, accessed 7 December 2017.

59 There is no published specification for the output of the SAMMA device. Following Oracle's 2014 purchase of Front Porch Digital, SAMMA production and marketing was halted. In July 2016, GreyMeta (https://www.graymeta.com) announced that they offered SAMMA maintenance and service. That announcement remained on the GreyMeta website in 2019 but was no longer accessible in mid-2020. See also Patent 7,853,766, *Method and system for automated migration of media archives*, https://patents.google.com/patent/US7853766B2/en, accessed 21 January 2019.

- Other subtypes: MXF wrapper with other picture encodings, and the missing option: FFV1 wrapped in MXF.[60]

B.3.1.2.4 Varying format recommendations, zone of consensus, and levels of implementation

Why are multiple options listed in the preceding sections? Specialists in the field of video preservation have not reached full consensus about preferred digital-file formats for preservation. This is especially the case for recently acquired file-based, born digital video, where preservation strategies are still evolving.

For the digitisation of conventional videotapes, however, consensus is growing among memory institution archives in favor of the combination of the Matroska wrapper and the FFV1 encoding (B.3.1.2.2 above). In the United States, for example, notable examples include Indiana University and New York Public Library. Many have greeted this format combination with great enthusiasm, and supporting tools are coming into play prior to the completion of the IETF standardisation process, which should reach the point of final-standard publication during 2019. The success and acceptance of Matroska/FFV1 reflects its technical merits and the enthusiasm of its developers, who have maximised the use of open source specifications and tools.[61]

Meanwhile, the year 2019 saw the publication of SMPTE RDD 48, an MXF application specification largely derived from the FADGI AS-07 specification (B.3.1.2.3 above). This

60 The SMPTE standards that define the MXF wrapper have multiple layers: the standard SMPTE ST 377-1:2011, *Material Exchange Format (MXF) - File Format Specification*, represents the starting point. Moving down a level, MXF wraps various "containers" and, for picture and sound essences, the next structural waystation is the "Generic Container," specified in SMPTE ST 379-1:2009, *Material Exchange Format (MXF) - MXF Generic Container*, and SMPTE ST 379-2:2010, *Material Exchange Format (MXF) - MXF Constrained Generic Container*. Down another level are the specific essence mappings to the Generic Container. Each mapping is specified in a SMPTE standard and examples include MPEG Streams, SMPTE ST 381-1:2005 and SMPTE ST 381-2:2011; DV-DIF Data, SMPTE ST 383:2008; uncompressed picture, ST 384:2005; SDTI-CP Essence and Metadata, SMPTE ST 385:2012; Type D-10 Essence Data, SMPTE ST 386:2004 (Archived 2010); Type D-11 Essence Data, SMPTE ST 387:2004 (Archived 2010); VC-3 Coding Units, SMPTE ST 2019-4:2009; VC-1, SMPTE ST 2037:2009; and AVC Streams, SMPTE ST 381-3:2013. Many specialists have asked, "Can we not combine the FFV1 lossless encoding with the MXF wrapper?" The answer is "yes" but a fully standardised solution depends upon the publication of a SMPTE standard for mapping FFV1 to the MXF Generic Container. Readers should note that in 2015, the Apple Corporation began to follow this path with their ProRes compression codec, to further support its widespread use in professional video production. SMPTE first specified the relevant aspects of the underlying ProRes codec in SMPTE RDD 36:2015 - *SMPTE Registered Disclosure Docs - Apple ProRes Bitstream Syntax and Decoding Process*. The drafting of a standard to map the codec to the MXF Generic Container is in process at this writing (January 2019).

61 Matroska's origins can be traced to 2002, when Steve Lhomme employed the Extensible Binary Meta Language (EBML) instead of a binary format to a digital-media wrapper; see the Wikipedia article *Matroska* (https://en.wikipedia.org/wiki/Matroska, accessed 21 January 2019). Meanwhile, the FFV1 encoding had an early milestone in 2003 when its code became part of the widely used FFmpeg tool; see the Wikipedia articles *FFV1* (https://en.wikipedia.org/wiki/FFV1, accessed 21 January 2019) and *FFmpeg* (https://en.wikipedia.org/wiki/FFmpeg, accessed 21 January 2019). Development has continued in recent years with contributions from Michael Niedermayer, Peter Bubestinger, Dave Rice, and others. Supporting tools, beyond FFmpeg, have emerged from a set of linked activities, including MediaArea.net (https://mediaarea.net/, accessed 15 October 2017), the parent organisation for the MediaConch project and product (https://mediaarea.net/MediaConch/documentation/FAQ.html, accessed 15 October 2017). MediaInfo is another MediaArea.net project (https://mediaarea.net/en/MediaInfo, accessed 15 October 2017), an application that provides technical underpinnings for MediaConch. The MediaConch project and product is also an activity of PREFORMA (Preservation Format for culture information/e-archives, http://www.preforma-project.eu/, accessed 15 October 2017), a project undertaken by a consortium of European organisations with funding from the European Union. The IETF standardisation effort for Matroska and FFV1 is under the rubric CELLAR (Codec Encoding for LossLess Archiving and Realtime transmission, https://datatracker.ietf.org/wg/cellar/charter/, accessed 21 January 2019).

specification was developed as a refinement and improvement on the somewhat ad hoc MXF specification used in the SAMMA mass-migration device first presented in the period 2003-2005.[62] Although usually discussed in terms of lossless JPEG-2000-based picture encoding,[63] the RDD 48 specification (and its SAMMA predecessor) also covers wrapping v210 uncompressed picture. SMPTE RDD 48 includes features that permit, for example, well-documented inclusion of legacy time code(s) and XML-based Timed Text representation of captions and subtitles. The specification, however, is complex and, to date, not well supported by tools. Thus far, SMPTE RDD 48 has not seen significant implementation, although its developers believe that it will find favor with archives committed to SMPTE standards like MXF, e.g., organisations with extensive holdings of professional broadcast materials and DCP digital cinema materials. These latter formats also employ MXF wrappers and JPEG 2000 picture encoding.

It is also the case that some archives continue to use the AVI and QuickTime "marketplace" wrappers (B.3.1.2.1 above), formats that date from the 1990s, sometimes wrapping uncompressed picture encoding, sometimes FFV1 lossless video. This selection is often a reflection of momentum from the past: an archive (or the contractor who regularly services an archive) may have an installed equipment base (and related experience) and this base supports these older formats. For some classes of content with relatively simple payloads, e.g., with neither captions nor legacy time code, these formats work reasonably well. Their limited ability to carry embedded metadata accounts for the recommendation in sections 3.7.1 through 3.7.3 (below) that archives using AVI or QuickTime maintain supporting metadata and fixity data in sidecar files or in the archive's collectionmanagement database.

The user communities associated with the formats described above represent different perspectives and practicalities and took form at slightly different times. In 2007, the Library of Congress and the national libraries of Norway and Australia embraced the SAMMA digitisation system and its use of MXF and JPEG 2000. At that time, few memory institutions were familiar with Matroska and FFV1, formats that had not yet reached a mature state of development. The named institutions saw their selection of MXF/JPEG 2000 as consistent with their general stance favouring audiovisual standards from SMPTE, ISO, and IEC, and a good fit for their traditional acquisition of commercial digitisation systems. In 2017, plans were being made to implement AS-07 (now SMPTE RDD 48) at the Library of Congress, including the migration of the existing SAMMA files into the refined format (Murray: 2017).

Regarding Matroska and FFV1, a user perspective is presented in a very helpful paper published by Indiana University (Casey: 2017). The report includes notes about the value of the open-format approach and the risk-avoidance that the university's Media Digitization and Preservation Initiative (MDPI) saw in the adoption of this format combination. The Indiana University project postdated the MXF adoptions cited in the

62 JPEG 2000 is a picture codec developed by the Joint Photographic Experts Group and standardised by a joint committee of the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC). The standards form a series in which the central member is ISO/IEC 15444-1:2004, *Information technology - JPEG 2000 image coding system - Part 1: Core coding system*. In the course of the effort, a number of JPEG 2000 profiles have been standardised in support of moving image applications, including digital cinema and broadcasting, in order to further buttress the interoperability of JPEG 2000-encoded data. Meanwhile, to clarify the handling of picture data in the MXF wrapper, SMPTE published ST 422:2014, *Material Exchange Format - Mapping JPEG 2000 Codestreams into the MXF Generic Container*. Originally published in 2006 with a focus on digital cinema applications (all with progressive-scan picture), the 2014 update of ST 422 clarified the "rules" for mapping interlaced video, correcting a problem that emerged in some of the early implementations of MXF/JPEG 2000 for older digitised videotapes.

63 SAMMA's development began in about 2003, with a patent process launched a year or two later. United States Patent 7,853,766, *Method and system for automated migration of media archives*, was awarded in 2010 (https://patents.google.com/patent/US7853766B2/en, accessed 21 January 2019).

preceding paragraph, beginning with an early phase in 2011-12, using a version of the MPEG-2 lossy encoding format. The MDPI team revisited format selection in 2015 as the refinement and IETF standardising of Matroska and FFV1 was moving into high gear. The MDPI project has now embraced that format combination.

B.3.1.2.5 In addition: The Interoperable Master Format (IMF)

Some media-industry professionals see the Interoperable Master Format (IMF) as a very feasible fourth option in addition to the three families identified above.[64] IMF is standardized by SMPTE and at this writing has four proposed *applications* (that might be called *profiles* in other contexts), which themselves splinter into subtypes. It is likely that, for IASA-TC 06 readers, the best preservation fit for IMF will be for born-digital content. For example, the most fully developed relevant IMF application is Application 2 (SMPTE ST 2067-20), which supports High Definition (1920x1080 frame size) Standard Dynamic Range (SDR) picture and JPEG 2000 image coding. Application 2e (SMPTE ST 2067-21) extends Application 2 with support for High Dynamic Range (HDR) picture up to 4K frame size. (There is also an emerging Application 4 that pertains to film scanning.) As the IMF Forum web site reports, Applications 2 and 2e (often referred to collectively as *IMF Studio Profile*) have received support from studios, post production facilities as well as the manufacturing community and is being used as a delivery format to "over the top" (OTT) Internet based services like Netflix.

Meanwhile, attentive to the IMF effort, members of the North American Broadcasters Association (NABA) are developing MXF-based file format specifications to support distribution and archiving in a broadcast context.[65] At this writing, IASA-TC 06 takes no position regarding the advisability of using IMF or NABA specifications for preservation mastering.

B.3.2 Formats that employ lossy compression

B.3.2.1 The broadcasters' use case

Most broadcasting today depends upon file-based video, managed and played out for transmission (or handoff to the web) from video servers. Much of the new content—as shot, edited, and presented—is in high definition, with picture in a 16:9 aspect ratio and compressed using "lossy" algorithms, e.g., various forms of ProRes, DV, or MPEG compression. U.S. broadcasters refer to these as "acquisition formats," meaning what cameras shoot in the field or studio systems record at the production facility.

For news programs or documentaries, broadcasters frequently incorporate footage from older recordings, including segments taken from decades-old analogue and digital videotapes. In support of their ongoing production, broadcasters will create and maintain a *production archive* or *library* that contains their collected historical materials, old and new. In order to integrate the older footage, it is typically digitised to be in the same or compatible formats with the new footage. This may not entail cropping picture from a 4:3 aspect ratio to 16:9 (more often 4:3 footage will be intercut with pillarboxing), but the broadcasters' target formats will generally employ lossy compression.

64 Information about IMF is provided in Annie Chang's *SMPTE Standards PDA Webcast: IMF (Interoperable Master Format)*, available via YouTube, https://www.youtube.com/watch?v=bmhv36hmSP4, accessed 12 October 2020. Discussion of the term *essence* is provided in footnotes to sections A.1.5.3 and B.1.4.2.3.

65 The NABA efforts include implementation of a variant of the UK-centric Digital Production Partnership (DPP) efforts (based in MPEG-family essences) as well as the exploration of an IMF-inspired specification (JPEG 2000 essences). Both concepts are grounded in SMPTE standards, including MXF (NABA: 2017).

One easy-to-integrate standard definition format, for example, is 50 Mb/s MPEG-2 (SMPTE D-10), and this is sometimes selected as the target format for older SD videotapes, maintaining the 4:3 aspect ratio. If the historical holdings include 16:9 HD videotapes, these may be transferred to MPEG-family formats at data rates far higher than 50 Mb/s.

B.3.2.2 Lossy compression in other contexts

Various non-broadcast organizations employ lossy compression for their digital masters. Examples in Europe include the Nederlands Instituut voor Beeld en Geluid (Netherlands Institute for Sound and Vision); their format for standard definition is SMPTE D-10 at either 30 or 50 Mb/s (de Jong, Annemieke: 2016, p. 45).[66] The Beeld en Geluid specifications are more or less identical to those selected by broadcast archives, as described in the preceding section.

In the United States, two archives that collect content via off-air recording also use MPEG-2, although at data rates much lower, on the order of 5 and 6 Mb/s. The first is the Vanderbilt [University] Television News Archive,[67] with a collection of broadcasts from all major U.S. television networks dating back to 1968. Copyright restrictions generally prevent any reuse of this footage; the main use case for Vanderbilt's resource is academic and citizen research into network-news coverage of various topics, e.g., for teaching, an article, or a book.

The second example is SCOLA,[68] a small non-profit located in western Iowa, who capture, retransmit online, and record-and-archive broadcasts from one hundred or more nations around the world, in dozens of languages. One key use case for the material is in language classes for organizations like the U.S. Department of Defense language school and in a number of colleges and universities. As with the Vanderbilt example, there is little or no anticipated reuse of the footage in new productions. Organizations like these have relatively modest resources to apply to these activities, and three factors motivate the selection of low-bit-rate lossy compression for mastering: (1) modest organizational resources, (2) the inherent low quality of off-air recordings, and (3) customer use cases that do not feature reuse of footage in new productions.

B.3.2.3 IASA-TC 06 discourages lossy compression for preservation

The IASA Technical Committee and the IASA-TC 06 authors represent the voice of memory institution *archives* that seek to preserve content for the very long term, as distinct from broadcasters' typical *production archives*.

As will be discussed in the future expanded edition of IASA-TC 06, many specialists believe it is sensible to retain today's new video footage in its "acquisition" format, i.e., its native lossy format "as shot". Video can be held in this format for several years, after which the footage should be migrated into a lossless or uncompressed form for the very long term.

In contrast, when digitising older videotapes that must be played in real time for transfer, the IASA-TC 06 authors strongly prefer lossless compressed or uncompressed masters. The creation of *lossy* compressed copies of old videotapes, as the adjective im-

66 Section 9.3.1 in de Jong's document mentions two lossy formats, MPEG4 and SMPTE D10, a standard for an implementation of MPEG-2: "MXF is an open standard maintained by the AV standards organization Society of Moving Pictures Engineers (SMPTE). The format is intended for professional use and is supported by a large number of different transcoders and editing software packages. Of all ingested materials in this category, standard viewing versions are produced in MPEG4. Standard Definition (SD) Material must be encoded as MXF OP1a, D10-30 or D10-50, the standard for Digital Provision of the public broadcasters, on the basis of the SMPTE guidelines. High Definition (HD) material must be encoded as XDCAM HD422/50MBbps" (de Jong, Annemieke: 2016, p. 45).

67 Vanderbilt Television News Archive (website), http://tvnews.vanderbilt.edu/, accessed 7 December 2017.

68 SCOLA (website), http://www.scola.org/, accessed 7 December 2017.

plies, degrades quality (albeit only slightly). And it is often the first step in what is called the "cascading" (successive) migration of lossy compressed content over time. This further degrades quality: a web page from the Japanese equipment manufacturer NTT offers an excellent set of tabbed images that illustrate quality loss in a similar cascade.[69]

We recognize and respect the practical factors that motivate some organizations to select lossy compression, but we recommend against this practice.

B.3.3 Selecting target formats

B.3.3.1 Four principles that guide format selection[70]

B.3.3.1.1 Produce a complete and authentic copy

Four principles guided the IASA-TC 06 authors as they compared target formats for digitisation. The first concerns the production of an authentic and complete copy of the original recording. This principle led us to pay close attention to such payload elements as captioning and subtitles, multiple "legacy" time codes, the handling of soundtracks, and the metadata needed to support their presence. These payload elements are discussed in more detail in section B.3.3.2.

Although long-term data management is made easier if all of these payload components are embedded or packaged in one file or bundle, not all formats support such packaging. This lack of support impairs the format's ability to embody authentic copies of certain types of source materials. In some cases, an archive can compensate for a shortfall of this type by maintaining data in associated files or elements using metadata structures like reVTMD (NARA: 2012),[71] a schema developed by the U.S. National Archives and Records Administration (NARA) and tailored for reformatted video. Elements like captions or subtitles can be carried in separate files ranging from the W3C's Timed Text standard to the industry-specified SubRip file (srt). An organization's collection-management database may also carry important item-level technical metadata. For more on these subjects, see the discussion of video payload in section B.3.3.2.

B.3.3.1.2 Seek the highest possible reproduction quality

The second principle has to do with quality of reproduction. IASA-TC 06 emphatically favours formats that maximize quality in both picture and audio reproduction, i.e., uncompressed or losslessly compressed essences,[72] as indicated by the preceding section on lossy compression. Of course, we also recognize that no archive can be perfect in terms of retaining every nuance of the original picture essence. For example, many historical videotapes carry *composite* video recordings: PAL in much of Europe and many other parts of the world; NTSC in the United States, Japan, and elsewhere; and SECAM in France, the nations that formerly comprised the Soviet Union, and parts of Africa. Meanwhile, all of the digital encoding target formats that IASA-TC 06 compares employ *colour-difference component* (aka YUV) video. Thus when a PAL, NTSC, or SECAM videotape is played back, before it can be encoded and written to a file, the picture essence undergoes an irreversible transformation from the composite to the component colour model. And, for a variety of practical reasons, the new colour-difference encoding will employ 4:2:2 chroma subsampling, the standard and generally accepted data-reduction technique used by professionals.

69 *Differences in Image Quality between 4:2:2 and 4:2:0 Chroma Formats in Cascaded Codec Connections* (NTT: n.d.) An analogous effect will be seen if the cascading of lossy compressed picture is compared to a cascade based on lossless or uncompressed picture.

70 This section owes a significant debt to the FADGI report *Digital File Formats for Videotape Reformatting: Part 5. Narrative and Summary Tables* (FADGI: 2014b).

71 NARA, 2012. Schema for reVTMD v.1.0, http://www.archives.gov/preservation/products/reVTMD.xsd, accessed 24 November 2017. Note: During 2017–18, this metadata specification is undergoing additional development in another organization.

72 Discussion of the term *essence* is provided in footnotes to sections A.1.5.3 and B.1.4.2.3.

B.3.3.1.3 Produce masters that support the creation of access copies and related features

The third principle is the goal of producing archival masters that support the creation of access and access-support elements. When digitising a recording with closed captioning, selecting a format that permits the movement and storage of a copy of this textual captioning data as XML means that the archival master contains a resource that can be more easily extracted for indexing, just as having an OCR rendering of a book text means that the book can be indexed in order to be more accessible to researchers.

B.3.3.1.4 Produce masters that include fixity data

The fourth principle concerns the inclusion of fixity data, also known as content or media integrity data. The digital files of interest in IASA-TC 06 are generally destined for long-term archiving-and-preservation management, an objective that is supported by a number of actions, including the creation of fixity data, a broad term that includes both *cryptographic hashes* like MD5 and various SHA hashes, and *checksums* like CRCs. The preservation action consists of monitoring of those values for change over time, which would indicate a change to the data in the file.[73]

For digital library specialists, content or media integrity usually turns on whole-file fixity values, critical for a well-run asset management system. Whole-file fixity data is a critical part of storage and repository systems but are not relevant to this comparison of file formats. Incidentally, whole-file fixity data cannot be embedded in the file itself: that action would change the file, making the value different "next time," thus invalidating it for comparison and monitoring.

Most digital video specialists agree, however, that there is great value in the creation and carriage of fixity data on segments of the file, e.g., on a frame or some other small unit of video. One digital-video expert compares this concept to fire alarms. For a collection of documents in digital form (each file is small), a mismatch uncovered when monitoring fixity data (the alarm goes off) will successfully take you to the problem in a useful way. But due to the great extent of audiovisual files, in the words on one video expert, a mismatch in a whole-file data value (the alarm) only takes you "to the neighbourhood" without identifying the specific "house" (frame or group of frames) that is damaged (Rice: 2012, Murray and Fleischhauer: 2014). Fixity data on a more granular, per-frame level, facilitates an assessment of the extent or location of digital change when a file-monitoring check indicates a hash value mismatch. Frame-based fixity data also serves broadcasters, who often employ what is called *partial file restore* to pull a single segment from a longer archived recording that is needed for insertion in a new program. Frame-based data can be used to be sure that the partial restore was successful.

73 The U.S. National Digital Stewardship Alliance (NDSA) report *Checking Your Digital Content: What is Fixity, and When Should I be Checking It?* provides an excellent introduction to this topic (NDSA: 2014).

B.3.3.2 Capabilities regarding ancillary and associated data ("payload elements")

This section identifies the most important payload elements beyond picture and sound to consider when selecting a target format. These payload elements answer two needs. First, their retention supports the production of authentic and complete representations of the original source video. Second, some elements enhance the functionality of the file made for preservation, e.g., the inclusion of a *master time code* and/or the inclusion of frame-level *fixity data*.

To be sure, only some originals carry all of the legacy-payload elements listed, while many carry only some (or even none) of these elements. In addition, depending upon the particulars, some archives may not deem it necessary to retain every element that is present in the source video. Thus not every payload element will be retained in the digital file made for preservation. In general, however, IASA-TC 06 recommends the retention of the payload elements listed below. This recommendation is especially strong for time-based elements like captions and subtitles that ought to play out in sync with picture and sound essences.[74] It is the case that some target format wrappers, e.g., AVI and QuickTime, may not support embedded carriage of all of the desired elements. Frame-level fixity data produced via the ffmpeg tool generally exists as a separate file. Thus many archives will employ "sidecar files" to carry some of these elements and/or place some technical metadata in associated databases maintained by the archive. In addition, some archives may employ devices like added sound tracks that carry historical-legacy time codes.

B.3.3.2.1 Time code: retain legacy time code

The source materials that are to be digitised may carry multiple time codes: vertical interval time code (VITC), linear time code (LTC), and more. Adapting a term from the EBU, we call them *historical time codes*; they are also referred to as *legacy time codes*.[75] Some are present on purpose, others by accident; some may have good integrity and continuity (numbers that reliably increment for every frame), while others may be discontinuous (numbers that do not increment in a reliable way, and may be missing for sets of frames). The legacy time codes in videotapes and other sources may themselves be layered in ways that an archive wishes to track, e.g., a videotape may carry LTC and may additionally carry an earlier generation of time code recorded, for example, as audio track 3. Any or all of these time codes may provide forensic help for future researchers. A legacy or historical time code may be keyed to old documents like tape logs, may provide clues about the older source tapes that were assembled to create the video program you are now preserving, and may (as with footage of events like rocket launches by space agencies) represent elapsed time that can be correlated to other data streams. In many cases, this is data you do not want to lose. Legacy historical time codes need not be employed to control playback; that is the job for a *master time code*, described below. Archives should retain historical time code data to support future research and forensic activities.

74 Discussion of the term *essence* is provided in footnotes to sections A.1.5.3 and B.1.4.2.3.

75 The EBU term is *Historical Source Timecode*; see for example, see *EBU – Recommendation R 122 Material Exchange Format Timecode Implementation* (version 2.0, November 2010; https://tech.ebu.ch/docs/r/r122.pdf, accessed 22 December 2017).

B.3.3.2.2 Time code: provide coherent master time code

A high integrity, continuous time code may be present in the source video and thus will be inherited by the new digitised file. Often, however, this is not the case and a new *master time code* will be created when digitising. In some contexts, this is called *synthetic time code*. Systems that play back files (or recover video for new purposes) will benefit from, if not require, such a master time code.

B.3.3.2.3 Time code: label multiple time codes

Given the need to carry a master time code and (often) one or more historical source time codes, there is also a need to tag the time code so that users of the file can identify each one.

B.3.3.2.4 Captions and subtitles: retain and carry captions and subtitles

IASA-TC 06 uses the terms *captions* and *subtitles* more or less interchangeably, to mean non-XML text intended for display over a timeline, in synchronization with image and sound essence.[76] Captions and video subtitles are important features of broadcast collections although they are less frequently encountered in other categories of content. Over time, U.S. broadcast standards have required various flavours of binary-coded closed captioning (CC), beginning in the 1970s with carriage on line 21 of the analogue signal, at the boundary of active picture and the vertical interval. More recently, a number of SMPTE standards offer specifications for packet-based captions that may be part of essence streams or carried as a separate block of time-based data in the MXF file wrapper.[77] In Europe, there are some parallels, beginning with Teletext and later, EBU subtitles (STL), another binary system. Today, broadcast authorities on both sides of the Atlantic want to set these binary approaches aside and move toward XML-based Timed Text, derived from the W3C Timed Text standard and very applicable to the web dissemination of video.

Some target-format wrappers, notably MXF, can embed caption data (binary and/or XML Timed Text) in the same file as picture and sound. The binary forms, however, will be awkward for future extraction, since their use will depend on the continued availability of decoding tools and may require real-time playback. (Nevertheless, many archivists want to retain them in addition to a new XML version in order to have an authentic copy.) If the wrapper selected by an archive for digitisation is not capable of carrying embedded caption data—especially XML-formatted data—then a sidecar file can be made part of the bundle of preservation files. Caption and subtitle texts, once extracted and indexed, have clear value for archives, supporting search and retrieval as well as web presentations in which the video is synchronized with text. If part of a source video recording, their retention is also required when producing an authentic copy.

Three closely related standard formats for captions employ XML and, if stored as sidecar files, usually employ the filename extension *.ttml* or *.tt.xml*. The World Wide Web Consortium (W3C) is the source for the original standard.[78] Versions for professional broadcasting have been standardized in the U.S. and Europe by SMPTE[79] and EBU.[80] The IASA-TC 06 authors recommend the use of the preceding standards, if at all pos-

76 Discussion of the term *essence* is provided in footnotes to sections A.1.5.3 and B.1.4.2.3.

77 These standards have been revised from time to time; current versions include SMPTE ST 334-1:2015, *Vertical Ancillary Data Mapping of Caption Data and Other Related Data*; SMPTE ST 334-2:2015, *Caption Distribution Packet (CDP) Definition*; and SMPTE ST 436-1:2013, *MXF Mappings for VI Lines and Ancillary Data Packets*.

78 *Timed Text Markup Language 1* (TTML1) (Second Edition), W3C Recommendation 24 September 2013, https://www.w3.org/TR/ttaf1-dfxp/, accessed 7 December 2017.

79 SMPTE ST 2052-1:2010, *Timed Text Format (SMPTE-TT)*, https://www.smpte.org/sites/default/files/st2052-1-2010.pdf, accessed 7 December 2017.

80 EBU TECH 3350, *EBU-TT PART 1, Subtitling Format Definition, version 1.1* (2015), and other related documents cited therein, https://tech.ebu.ch/docs/tech/tech3350.pdf, accessed 7 December 2017.

sible. In addition, there are a number of industry and ad hoc structures for captions and subtitles. Although not specified by standard setting organizations, some of the following are widely adopted and reasonably interoperable: Caption Center, Captions Inc., Cheetah, DFXP, LRC, MPSub, NCI, Scenarist Closed Caption (*.scc*), SubRip (*.srt*), SubViewer, Timed-Text Markup Language (*.ttml*), Videotrol Lambda, and WebVTT.[81]

B.3.3.2.5 Audio track layout and labelling

Archives wish to digitise audiovisual content from a variety of source material with widely varying sound tracks. In terms of sound or aural field, examples range from silent research footage to monaural oral history recordings to performances with stereo, surround, or multichannel audio. In other cases, the tracks on a source item will include Descriptive Video Service (DVS), Second Audio Program (SAP), annotations (like a director's commentary for a dramatic program), as well as other types of multiple language content or other versioning elements. Sound tracks on certain videotape formats may also carry time code data, e.g., the carriage of LTC on track three of the 1-inch type C format. Archivists seeking to produce authentic copies wish to retain this source data and require metadata that labels or tags the tracks in a manner that will serve future users.

Source material audio tracks may or may not be labelled according to a standard or industry convention. When so labelled, the tagging may be in terms of such standards as SMPTE Multi-Channel Audio,[82] the EBU track allocation templates specified by EBU R 48[83] or EBU R 123,[84] or by an industry convention promulgated by a broadcast network, such as the PBS Audio Configuration specification cited in the MXF application specification AS-03.[85] Existing tagging should be retained in archive or preservation files.

B.3.3.2.6 Language tagging: provide a means to tag Timed Text languages

Archives may wish to tag primary and secondary languages in Timed Texts. In addition, primary language tagging is the subject of a SMPTE Recommended Practice specification.[86] General practices in broadcasting and archiving often rely upon IETF RFC 5646[87] and/or upon the coding approaches that underpin RFC 5646, especially ISO 639-2,[88] which is in turn an important basis for the MARC cataloguing standard and other library-oriented specifications.

Because collections are often relatively homogeneous in terms of language, many organizations will employ default language values ("boilerplate") in tags. In the U.S., for example, this will often be the code value for American English ("en-US").

81 DFXP (Distribution Format Exchange Profile), .ttml, and WebVTT are related formats associated with the important W3C Timed Text initiative; see https://www.w3.org/TR/ttml1/, https://www.w3.org/TR/webvtt1/, and https://w3c.github.io/webvtt/, all three accessed 7 December 2017. Additional caption-format information is provided by Gary McGath (McGath: 2016) and, for YouTube, on the Google support page *Supported subtitle and closed caption files* (Google YouTube: n.d.).

82 SMPTE ST 377-4:2012, *MXF Multichannel Audio Labeling Framework*.

83 EBU Technical Recommendation R48-2005, *Allocation of audio tracks on digital television recorders*, https://tech.ebu.ch/docs/r/r048.pdf, accessed 7 December 2017.

84 EBU-Recommendation R 123, *EBU Audio Track Allocation for File Exchange*, https://tech.ebu.ch/docs/r/r123.pdf, accessed 7 December 2017.

85 AS-03: MXF Program Delivery, http://www.amwa.tv/projects/AS-03.shtml, accessed 7 December 2017.

86 SMPTE RP 2057:2011, *Text-Based Metadata Carriage in MXF*, including Am1:2013.

87 IETF RFC 5646, *Tags for Identifying Languages*, https://tools.ietf.org/html/rfc5646, accessed 7 December 2017.

88 *Language codes-ISO 639*, http://www.iso.org/iso/home/standards/language_codes.htm, accessed 7 December 2017. In addition to a link to a list of codes, this introductory page explains that ISO 639 has been published in five parts (latest versions dating from 1998 to 2002) that offer lists of codes in 2- and 3-character forms.

B.3.3.2.7 Language tagging: retain language tagging associated with binary caption or subtitle data

Language tagging may be present in caption and subtitle source material, notably examples that employ the standardized approaches established for CEA-608, and -708 (called *caption service descriptors*), EBU STL, and SMPTE and EBU Timed Text. This language identification information should be retained in the output from a digitisation process.

B.3.3.2.8 Language tagging: provide a means to tag soundtrack languages

Archives may wish to tag primary and secondary languages in soundtracks. As noted above, general practices in broadcasting and archiving often rely upon the IETF RFC 5646 and/or upon the coding approaches that underpin it.

B.3.3.2.9 Embed text-based and binary data: provide carriage of supplementary metadata (text-based data)

The archivists' concern addressed in this section is not about metadata in general, i.e., the issues that every archive faces in terms of overall collection management and the provision of access to researchers. Nor does this concern pertain to basic information about file characteristics, the data that an application requires to play a file correctly, often referred to as parametric metadata. Most wrapper specifications require a certain level of parametric metadata, although as noted on the table in appendix A, some are deficient in this regard, especially regarding such arcana as aspect ratios, letter-boxing, pillar-boxing, and pan-and-scan.

This section concerns metadata that is supplementary in terms of the technical requirements for playback. Examples include additional technical metadata about the production or digitisation activity, sometimes called process-history metadata, information about the source item, about quality review outcomes, and preservation metadata, e.g., PREMIS. One example of process metadata is provided by the SAMMA digitisation system in use in several IASA member institutions: an XML-encoded, frame-by-frame record of the metrics associated with each tape transfer. Many organizations wish to maintain process metadata and some see value in embedding such data in files, as it the case with process history metadata in the EBU Broadcast WAVE audio format. And beyond this technical and administrative metadata, some archives plan to embed relatively complete descriptive (aka cataloguing) metadata in their files to provide a second copy in case of disaster. (This copy of the data will be date-stamped to make clear that it may not have been updated in step with the main institutional catalogue.)

As was the case for subtitles and captions, depending upon the target-format wrapper selected, supplementary metadata may be embedded in the main file or may instead be maintained as a sidecar file.

B.3.3.2.10 Embed text-based and binary data: provide carriage of a manifest (text-based data)

Although not a pre-existing element, this is a suitable place to highlight the importance of creating and carrying a manifest for a file, especially for one that may contain multiple elements. In such cases, a manifest supports preservation and good housekeeping by offering an inventory of the file's parts and expresses the relationships between them. Through a mix of required and optional elements, a manifest provides a high-level inventory of the parts including their identifiers, data description, MIME type, size and location. This information can help the user to better understand the composition of the file and it will also provide machine-interpretable information for content processing in later phases of the life cycle. Manifests of one sort or another are included in several formats ranging from the digital library community's BagIt specification to the IMF format, mentioned earlier in this chapter.

As was the case for supplementary metadata, depending upon the target-format wrapper selected, a manifest may be embedded in the main preservation file or may instead be maintained as a sidecar file.

B.3.3.2.11 Embed text-based and binary data: provide carriage of EBU STL, still images, documents, etc. (binary data)

Archives will also often also wish to retain and carry instances of the European Broadcast Union (EBU) binary subtitling format, called EBU STL and standardized in EBU Tech 3264.[89] Some archives will want to embed associated materials in the preservation file along with the main content item. In this context, *associated materials* name a class of binary representations of materials closely associated with the file's primary essences,[90] e.g., scanned images and documents, video trailers, etc. Associated materials contribute to the completeness, comprehensibility, or usability of the main information object in the file and take the form of files in formats such as TIFF, JPEG, MP4, PDF, and the like.

Some archives will prefer to maintain these types of materials as a part of preservation file bundle, rather than embedding them in main preservation file.

B.3.3.2.12 Frame-level fixity (content integrity) data

The value of frame-level fixity data is described in section B.3.3.1.4, concerning the principles that guide format selection. In some cases, frame-level fixity metadata is carried in a sidecar file that accompanies the main preservation file. This is generally the case with the hash values produced by the ffmpeg software's *framecrc* and *framemd5* capability.

In other cases, the integrity information is embedded in a file, generally in the wrapper but occasionally in the encoded essence bitstream. The BBC Archive Preservation File Format specified in BBC White Paper 241 (Glanville and Heritage: 2013) employs frame-level hash values, carried in the MXF wrapper. The digital cinema approach is standardized in SMPTE ST 429-6:2006, *D-Cinema Packaging – MXF Track File Essence Encryption.* In the latter example, fixity data is conjoined with data pertaining to encryption.

89 EBU, *Specification of the EBU Subtitling data exchange format, TECH. 3264-E, 1991*, https://tech.ebu.ch/docs/tech/tech3264.pdf, accessed 24 November 2017.

90 Discussion of the term *essence* is provided in footnotes to sections A.1.5.3 and B.1.4.2.3.

B.3.4 Format comparison tables

The compressed-summary table below outlines the most important features and capabilities of the four target-format families analyzed for IASA-TC 06. Detailed comparisons of the formats are provided in the appendix to this section. In the appendix, the same information is presented on three different tables. The first is a summarizing table. The second and third tables present rich detail. If printed, the second table requires large paper sheets (11x17-inches US; A3 European).

Content list for the B.3 Appendix

- Part 1. Summary Target Format Comparison Table (typing paper sheets)
- Part 2. Full Detail Target Format Comparison Table (large sheet version)
  - Footnotes to the Full Detail Target Format Comparison Table (typing paper sheets)
- Part 3. Full Detail Target Format Comparison Table (subdivided version, typing paper sheets)

In both the compressed-summary table (below) and in the detailed tables in the appendix, the main column headings identify the four format families while subsidiary columns identify the subtypes of special interest for IASA-TC 06. The tables employ 24 factors and subfactors that compare the features and capabilities of the target formats. The factors pertain to picture, sound, and other payload entities that are of interest in preservation planning. Some factors are relatively abstract and consider, for example, the standardization or adoption of a format. Others address the entities or characteristics that may be present in or relevant to a given source video recording. Archivists should use their judgment to determine which characteristics or entities are considered to be *essential* features for a given set of collection items, and thus to be retained in the digitised preservation copies. Under the heading *functionality factors*, the table reports on the format's *capability* to carry the entity in question, and also on whether the format provides for embedded *metadata* about the carriage.

These factors and the assessments have a rough-and-ready quality. Precise metrics do not exist for many factors and, in any case, the comparison is intended as a guide, a starting point for an archive's planning. Many of the factors are self-explanatory or are sufficiently explained by the brief statement in the detailed table in the annex. Some additional explanatory information is warranted, however, and is presented in the section on ancillary and associated data (B.3.3.2 above) and in the section that offers additional explanatory information (B.3.5 below).

B.3 Table 1. Compressed summary target format comparison

Category	Includes these factors	Marketplace wrappers with FFV1 or uncompressed v210			Uncompressed v210 in MXF		Lossless JPEG 2000 in MXF		FFV1 in Matroska
		FFV1 in Open-DML AVI	v210 in Open-DML AVI	v210 in Quick-Time	SMPTE RDD 48 (former AS-07) Base-band Shim	Standards compliant v210 in MXF	SMPTE RDD 48 (former AS-07) Baseband Shim	SAMMA Profile	Active IETF Internet Drafts
Sustainability factors (individually assessed)									
	Disclosure	Acceptable	Acceptable	Acceptable	Good	Good	Good	Acceptable minus	Good
	Adoption	Wide	Wide	Wide	Limited/little implementa-tion	No survey for this guideline	Limited/ little imple-mentation	Moderate	Growing
	Trans-parency	Slightly less transparent	Slightly more trans-parent	Slightly more trans-parent	Medium trans-parency	Medium trans-parency	Slightly less trans-parent	Slightly less trans-parent	Slightly less trans-parent
	Self-docu-mentation	Mid-level	Minimal minus	Minimal plus	Extensive	Extensive	Extensive	Mid-level	Mid-level plus
Quality factor (individually assessed)									
	Picture and sound encoded without loss	Good	Good	Good	Good	Good	Good	Good	Good
Functionality factors (twelve factors, detail in appendix, assessed as a group)									
	Summary assessment	Acceptable minus	Poor	Accept-able minus	Good	Good	Good	Acceptable	Good
Production and QC factors (five factors, detail in appendix, assessed as a group)									
	Summary assessment	Relatively easier	Relatively easier	Relatively easier	Relatively more difficult	Relatively more difficult	Relatively more difficult	Relatively more difficult	Relatively more diffi-cult; strong community support
Data-management, data-preservation factors (two factors, detail in appendix, assessed as a group)									
	Summary assessment	Good	Acceptable	Acceptable	Acceptable	Acceptable	Good	Good	Good

B.3.5 Additional information about selected comparison factors[91]

B.3.5.1 Sustainability factors

- Disclosure. About the publication formality for the format being described, whether a standard or profile (application specification); identification and formality information about important subsidiary formats (dependencies); identification of standardization or publication auspices.
- Adoption. Degree to which the format is already used by the primary creators, disseminators, or users of information resources.
- Self-documentation. Degree to which the format supports embedded descriptive, administrative, and provenance metadata. Note: on this table, technical metadata about the file in hand is covered in the section on *functionality factors*.

B.3.5.2 Quality factor

Picture and sound encoded without loss. As indicated by the summary and detailed table in the annex, all of the target formats carry digital reproductions of picture and sound that succeed in reproducing all or virtually all of original signal. This statement is qualified in a very modest way for three reasons.

- First, as noted above, the conversion of composite analogue signals to colour-difference component (Y'PbPr, Y'CbCr, or "YUV") data requires an irreversible transform that changes the picture information in subtle ways. (The same might be said for the changed appearance of picture displayed, today, on non-cathode-ray-tube monitors, as compared to the display, yesterday, of the original analogue signal on CRT monitors.)
- Second, also noted above, a variety of practical reasons force the selection of 4:2:2 chroma sub-sampling when digitising from analogue composite signals and this does represent a modest amount of data reduction, an approach that is universally adopted in broadcast television, well established and well accepted.
- Third, in the case of v210-encoded uncompressed video in the QuickTime wrapper, in some circumstances picture-data values at the very high and very low end (outside of broadcast range) may be lost. This could have a slight impact on the ability to post-process or correct video recorded with luma clipping.[92]

B.3.5.3 Functionality factors

As noted in the compressed summary table above, the format comparison took twelve functionality factors into account. The format-by-format assessment against these twelve formats is provided in the detailed tables in the appendix to section B.3, cited above. The following list identifies the twelve factors and provides background information:

B.3.5.3.1 4:2:2 chroma subsampling

IASA-TC 06 echoes the general preference in professional broadcasting for picture data with 4:2:2 chroma subsampling, i.e., 4 samples of luma or "brightness" data, and 2

91 This section owes a significant debt to *Digital File Formats for Videotape Reformatting: Part 5. Narrative and Summary Tables* (FADGI: 2014b).

92 See the Apple developer documentation under the heading "Scheme B: 'Video-Range' Mapping with Unsigned Y´, Offset Binary Cb, Cr," part of the 1999 Apple technical note sometimes referred to as "tn2162," https://developer.apple.com/library/content/technotes/tn2162/_index.html#//apple_ref/doc/uid/DTS40013070-CH1-TNTAG7-SCHEME_B___VIDEO_RANGE__MAPPING_WITH_UNSIGNED_Y____OFFSET_BINARY_CB__CR, accessed 7 December 2017. (Thanks to video expert Dave Rice for this reference.)

plus 2 samples of chroma or colour data.[93] Subsampling at the 4:2:2 level provides higher image quality than other feasible options. The 4:4:4 ratio offers even higher quality and is often used in high-end digital motion picture production, but practical considerations in terms of available equipment and interfaces generally preclude its use in video digitisation. 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-digitised 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. Image examples on a web page from the Japanese equipment manufacturer NTT illustrate quality loss in such a cascade, comparing 4:2:2 to 4:2:0 subsampling (NTT: n.d.).

B.3.5.3.2 Broadcast and wide video range and ITU-R indication

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?

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 some similar formulation.[94] 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 (nonsquare 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 some cameras (e.g., some DSLRs) and by computer graphics systems.

The significance of the factor pertains to playback or future re-digitisation. 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 digitisation may guarantee the uniformity of a given collection in terms of range, and this general knowledge can be used to guide future activities.

B.3.5.3.3 Scan types and field cadences

It is important to state when the source recording is progressive or interlaced and, if interlaced, to indicate which field was transmitted first.

93 For a technical explanation of chroma subsampling and the meaning of ratio statements like 4:2:2, 4:2:0, 4:1:1, see appendix B of part 5, FADGI report *Digital File Formats for Videotape Reformatting: Part 5. Narrative and Summary Tables* (FADGI: 2014b), p. 19.

94 Additional discussion of Rec. 601 and Rec. 709 is provided in the sidebar that follows section B.1.3.1.3 above.

B.3.5.3.4 Various aspect ratios

Picture in analogue source materials will generally be converted to digital in non-square pixels; digital source materials that require real-time playback will generally be represented as non-square pixels; metadata is needed to characterize the correct display aspect ratio; broadcasters provide this information via the Active Format Description (AFD) codes standardized in SMPTE ST 2016-1:2009 and ST 2016-3:2009.

B.3.5.3.5 Different bit depths

The presence of this factor reflects a desire to determine if a proposed format supports the 10-bit sampling recommended in IASA-TC 06. The explanation of chroma subsampling cited above identifies the elements being sampled in the digital image: luma data and two types of chroma data. Many digitising systems offer the option of recording either 8 or 10 bits per sample. IASA-TC 06 encourages the use of 10-bit sampling for the sake of higher image quality.[95] 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 colour. 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 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.

B.3.5.3.6 Primary and secondary time codes

See the discussion in sections B.3.3.2.1, B.3.3.2.2, and B.3.3.2.3.

B.3.5.3.7 Closed captioning and subtitles

See the discussion in sections B.3.3.2.4 and B.3.3.2.7.

B.3.5.3.8 Multipart recordings

Some archives may wish to copy multi-segment recordings, like a television news crew's field footage, into a single file but also with a method to mark and identify the segments.

B.3.5.3.9 Carriage of associated components

See the discussion in sections B.3.3.2.9, B.3.3.2.10, and B.3.3.2.11.

B.3.5.3.10 Fixity data

See the discussion in sections B.3.3.2.12.

B.3.6 Format selection: the influence of practical or circumstantial matters

For some archives, target format selection may be influenced by circumstances:

- Installed equipment base. Does the archive (or its usual contractor) already have certain equipment in place that favours one format over another? Is this reflected in the costs for digitising? Or is this reflected in the time requirements for conversion? Some production systems will support more than one preferred format, but will encode one format with high efficiency, and another with low efficiency.
- Preference for certain community alliances. Does the archive already favour and use systems that include significant open source elements, and will open

95 Some specialists argue, however, that there is no benefit for certain classes of material. Video specialist Dave Rice has said that at the City University of New York, they digitise 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 Rice examined the SDI video data from an SX tape on a binary display he could see that the 9th and 10th bits were always zero. Thus by taking only the first 8 bits, he captured all meaningful data, saving a considerable amount of digital storage space.

source applications (and the formats they support) be easier to integrate and employ? Or does the archive have a history or connections with commercial video production, broadcast and non-broadcast? If so, that archive may prefer to work with formats standardized by SMPTE and/or ISO, which are well supported by commercial systems marketed to professional production organizations.

B.3.7 Format recommendations in terms of source material characteristics

The characteristics of the source video materials to be digitised will always be significant considerations when selecting target formats and require an archivist's thoughtful judgment. To illustrate this point, consider the source content categories described in the sections that follow. (Note: these categories are orthogonal to the six *classes* discussed in sections A.1.3 and B.2.4.)

B.3.7.1 Ethnographic footage and oral history recordings

In general, collections of ethnographic footage or oral history recordings are unlikely to include elements like multiple time codes, captions or subtitles, or complex audio track configurations. Therefore, almost any of the formats on the table is likely to reproduce such items with reasonable success, assuming the workflow is attentive to chroma subsampling, retaining the full raster and making some record of aspect ratio and, if from interlaced sources, the field cadence.

- Practical option:
  - FFV1 in Open DML AVI
    - With supporting metadata carried in sidecar file or in the archive's database
- Acceptable options:
  - Uncompressed v210 in QuickTime
  - Uncompressed v210 in Open DML AVI
    - For both: supporting metadata and fixity data carried in sidecar file or in the archive's database
- Preferred options:
  - FFV1 in Matroska
  - Lossless JPEG 2000 in MXF
  - Uncompressed v210 in MXF

B.3.7.2 Edited documentaries and modest independent productions

Documentaries and modest independent productions may, of course, vary considerably in their makeup and production values. However, many can also be reproduced with reasonable success in most or all of the formats on the table, with the same caveats as for the previous category. It is worth noting, however, that edited productions may carry multiple time codes that may be worth retaining for two reasons. First, such legacy time codes may provide clues about the way in which footage was edited; if the original shooting rolls are available this data may make it easier for a researcher to retrace the editor's steps. Second, the time codes in an edited production may be a jumble of discontinuous numerical segments, and there is considerable value in selecting a format that can carry both a fresh primary ("master") time code as well as retaining legacy time codes for future research.

- Practical option:
  - FFV1 in Open DML AVI
    - With supporting metadata carried in sidecar file or in the archive's database
- Acceptable option:
  - Uncompressed v210 in QuickTime
    - With implementation of some of the features listed in the notes to the detailed table
    - With some supporting metadata and fixity data carried in sidecar file or the archive's database
- Preferred options:
  - FFV1 in Matroska
  - Lossless JPEG 2000 in MXF
  - Uncompressed v210 in MXF

B.3.7.3 Broadcast and other professional productions

This class of material will very often include elements like captions or subtitles, multiple time codes, and multitrack audio. More than the two other illustrative examples above, this category will benefit from careful use of the preferred options listed below.

- Practical option:
  - FFV1 in Open DML AVI
    - With captions/subtitles and supporting metadata carried in sidecar file or the archive's database
- Acceptable option:
  - Uncompressed v210 in QuickTime
    - With implementation of some of the features listed in the notes to the detailed table, although for this class of source material this option may become more difficult to implement in a manner that retains all essential elements
    - With some supporting metadata and fixity data carried in sidecar file or the archive's database
- Preferred options:
  - FFV1 in Matroska
  - Lossless JPEG 2000 in MXF
  - Uncompressed v210 in MXF