MISB MOTION IMAGERY STANDARDS BOARD	MISB ST 1910.1
STANDARD **Adaptive Bitrate (ABR) Content Encoding**	**29 October 2020**

1 Scope

This standard defines a file format for adaptive bitrate (ABR) content encoding of Motion Imagery, audio, and Key-Length-Value (KLV) metadata for applications within the NSG. It mandates the Common Media Application Format (CMAF) [1] container, with constraints defined in this document, for encoding and packaging segmented media – Motion Imagery, audio, and metadata for delivery via server/client-based protocols.

This standard only supports the profiles and levels of H.264/AVC [2] and H.265/HEVC [3] compressed imagery indicated herein which conform to the Motion Imagery Standards Profile (MISP) [4].

This standard provides guidelines to map media content from an MPEG-2 Transport Stream (MPEG-2 TS) to the Common Media Application Format (CMAF) based on ISOBMFF, fMP4 container.

This standard does not address the media delivery method from a server to an end-user client. MPEG-DASH [5] and HLS [6] both support delivery of CMAF files but their design supports different client architectures with unique manifest syntax and semantics in how they request and manage a CMAF presentation. Choice of either method is specific to application and implementation requirements, and therefore, outside the scope of this document. When used according to guidance in this standard, CMAF packaging of Motion Imagery, audio, and metadata enables the time-aligned retrieval and display of media essence by CMAF conformant clients.

ABR technology ideally suits video-on-demand applications. Given its ability to operate in thin client infrastructures, ABR offers an attractive delivery method for second and third phase Motion Imagery exploitation. Currently, ABR is not well suited for time-critical, latency-sensitive real-time applications. Sub-second delivery requires optimized server/client architectures; however, this continues to be area of active industry development.

2 References

[1] ISO/IEC 23000-19:2020 Information technology - Multimedia application format (MPEG-A) - Part 19: Common media application format (CMAF) for segmented media.

[2] ISO/IEC 14496-10:2014 Information Technology - Coding of audio-visual objects - Part 10: Advanced Video Coding.

[3] ISO/IEC 23008-2:2017 Information Technology - High efficiency coding and media delivery in heterogeneous environments - Part 2: High efficiency video coding.

[4] MISB MISP-2021.1 Motion Imagery Standards Profile, Oct 2020.

[5] ISO/IEC 23009-1 :2019 Information technology - Dynamic adaptive streaming over HTTP (DASH) - Part 1: Media presentation description and segment formats.

[6] IETF RFC 8216 HTTP Live Streaming 2nd Edition, Apr 2020.

[7] MISB MISP-2021.1: Motion Imagery Handbook, Oct 2020.

[8] ISO/IEC 14496-12:2015 Information technology - Coding of audio-visual objects - Part 12: ISO base media file format.

[9] ITU-R BT.709-6 Parameter values for the HDTV standards for production and international programme exchange, 06 2015.

[10] MISB RP 0802.2 H.264/AVC Motion Imagery Coding, Feb 2014.

[11] MISB ST 0107.4 KLV Metadata in Motion Imagery, Feb 2019.

[12] MISB ST 0601.17 UAS Datalink Local Set, Oct 2020.

[13] MISB ST 1001.1 Audio Encoding, Feb 2014.

[14] MISB ST 1402.2 MPEG-2 Transport Stream for Class 1/Class 2 Motion Imagery, Audio and Metadata, Oct 2016.

[15] MISB ST 0604.6 Timestamps for Class 1/Class 2 Motion Imagery, Oct 2017.

[16] ISO/IEC 13818-1:2018 Information technology - Generic coding of moving pictures and associated audio information: Systems.

[17] MISB ST 0603.5 MISP Time System and Timestamps, Oct 2017.

3 Revision History

Revision	Date	Summary of Changes
ST 1910.1	10/29/2020	• Deprecated Reqs -05 through -08, -10 through -15 • Added Reqs -16 through -33 • Modified structure of emsg value and id fields • Modified the scheme_id_uri to include document number and its version • Revised definitions, text, and figures for improved clarity • Reorganized content for readability • Updated references

4 Acronyms and Definitions

AAC Advanced Audio Codec
AVC Advanced Video Coding
ABR Adaptive Bitrate
CMAF Common Media Application Format
CTE Chunked Transfer Encoding

DASH	Dynamic Adaptive Streaming over HTTP
GOP	Group-of-Pictures
HEVC	High Efficiency Video Coding
HLS	HTTP Live Streaming
HTTP	Hypertext Transfer Protocol
IBMF	ISO Base Media File format
IDR	Instantaneous Decoding Refresh
IETF	Internet Engineering Task Force
ISO/IEC	International Standards Organization/ International Electrotechnical Commission
ITU-R	International Telecommunication Union Radiocommunication Sector
KLV	Key Length Value
MISB	Motion Imagery Standards Board
MISP	Motion Imagery Standards Profile
MPD	Media Presentation Description
MPEG	Moving Picture Experts Group
NSG	National System for Geospatial-Intelligence
OTT	Over-The-Top
PES	Packetized Elementary Stream
PMT	Program Map Table
PTS	Presentation Time Stamp
RP	Recommended Practice
ST	Standard
tps	ticks per second
TS	Transport Stream
ULL	Ultra-Low Latency
URN	Uniform Resource Name

Encoding Ladder	Content encoded into a variety of spatial resolutions, temporal resolutions, and bitrates designed to facilitate adaptation to changing network conditions. Different applications typically will choose a different encoding ladder as a function of compression type, content, and client device (i.e., computer, smartphone, etc.).
KLV Packet	see Motion Imagery Handbook [7]
KLV metadata stream	see Motion Imagery Handbook
Over-the-Top	Streaming media service directly from web servers to web clients without intermediate control systems.

5 Introduction

Adaptive Bitrate (ABR) streaming is a media-streaming model for delivery of media content controlled by the client. Commercial services adopted ABR to satisfy the growing demand for Over-the-Top (OTT) delivery of content to consumers using various device types for content

playback (e.g., computers, mobile devices, televisions). The convergence in the industry to adopt CMAF as the container file for ABR content provides a unifying model to support delivery of multimedia presentations to a variety of devices, such as set-top boxes and web browsers, using various means for streaming delivery such as MPEG-DASH and HLS.

ABR can provide widespread delivery of content stored in a cloud with access to content by the browser. This standard defines how to package both Motion Imagery and KLV metadata in a unified CMAF file. Mapping existing audio within the source content into its own CMAF file along with metadata if available follows the same packaging methods. A client application requests from the origin server the available CMAF files for decoding and rendering. Although applicable to other sources of content, the initial application for ABR in the NSG community is to map MPEG-2 TS collected Motion Imagery into CMAF.

Figure 1 illustrates an example of Motion Imagery and metadata content mapped from a MPEG-2 TS into CMAF. Although not shown, audio (plus metadata) would likewise map into its own CMAF file in a similar fashion. Using the example of Figure 1, the Motion Imagery content maps into CMAF fragments within Media Data Boxes. Metadata maps into the same CMAF fragments as emessages (emsg). This binds the metadata to the Motion Imagery into one unified package. This document describes the methods for packaging such content into CMAF along with appropriate signaling for MISP-approved compressed Motion Imagery, audio, and KLV metadata.

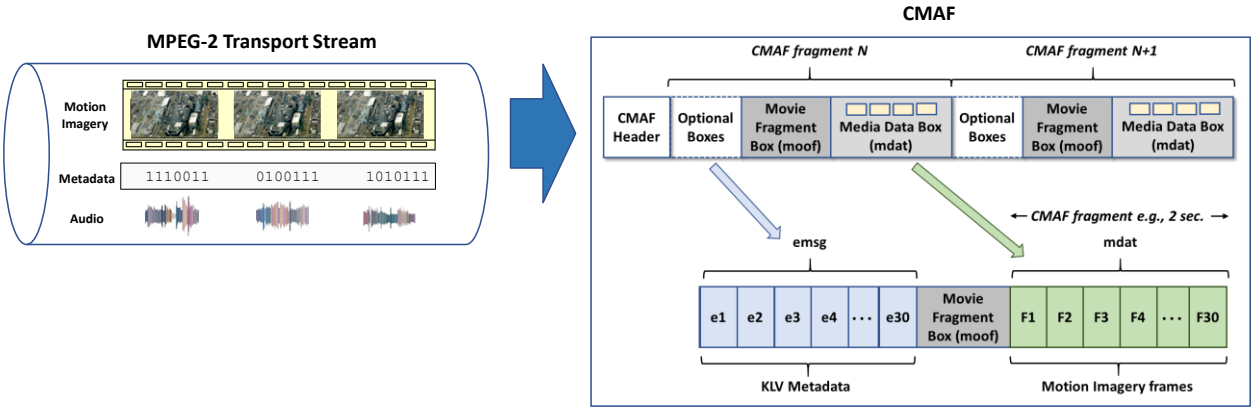

Figure 1: Example: MPEG-2 TS Content Mapped into CMAF

ABR operates within a server/client relationship. In Figure 2 an **ABR Server** hosts all CMAF resources composing a service, which are then accessible through multiple HTTP requests. Standard HTTP 1.1 (or later versions) compliant servers and caching proxies host and distribute media content to an **ABR Client**. The first stage, called CMAF Content Preparation, packages source media content into one or more CMAF files, which reside on a web server (i.e., ABR Server).

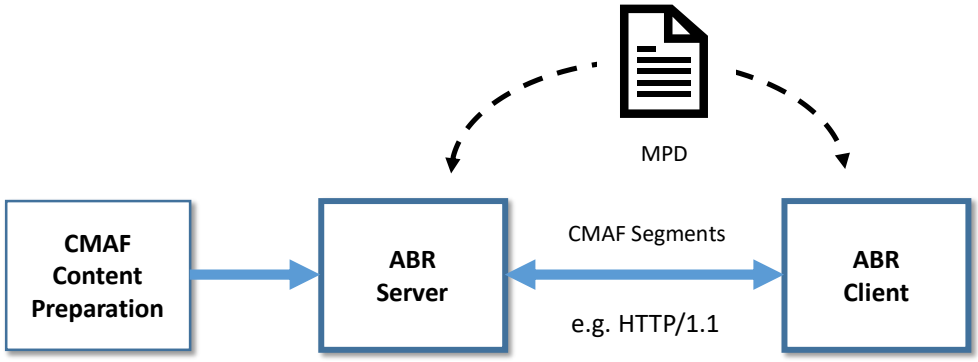

Figure 2: Architecture for ABR Delivery

A Media Presentation Description (**MPD**) manifest file, typically an XML formatted "instruction" file, provides the necessary information for an ABR Client to request media and prepare for receiving, processing, and displaying the content. The client parses the MPD and makes further requests based on its environment (network bandwidth, client rendering capabilities, etc.). As the network conditions between a server and client may change over time, the client adapts its requests to meet new conditions.

Evolved from the MPEG-DASH specification, CMAF constrains the IBMF container [8] further and provides for a unified MPD protocol to support both DASH and HLS applications. Developed principally for compressed video and audio, recent efforts have expanded its capabilities to carry additional types of media such as closed-captioning, subtitles, advertisements, etc. The NSG can leverage the CMAF technology to provide cloud-based processing and analytics to resource-constrained clients. This document leverages the baseline capabilities of CMAF with additional constraints to facilitate interoperability for NSG applications.

CMAF separates media types (e.g., video and audio) into their own independently managed CMAF tracks. Adapting CMAF for government application requires considering in-band security information for the encoded content. MISB requires tight indisputable binding of security information with Motion Imagery content, therefore, this document describes the packaging of metadata, to include security information, along with Motion Imagery unified within one CMAF track.

Figure 3 illustrates the source content-to-ABR delivery workflow and indicates the scope of the guidance provided by this standard. Source content includes Motion Imagery, metadata, and audio. This document has requirements for:

- The Motion Imagery, metadata, and audio encapsulated in CMAF.
- The allowed profiles and levels for H.264/AVC and H.265/HEVC compressed Motion Imagery types and MPEG-2 AAC LC compressed audio.
- The method to properly embed MISB Key Length Value (KLV) metadata.

This document has guidance for:

- CMAF event message (emsg) field values for KLV metadata.
- Mapping MPEG-2 TS streams (video, audio, metadata) into CMAF
- Recommendations on encoding parameters, such as encoding ladder, segment size, etc.

- Extending emsg signaling for different qualities of timing in MPEG-2 TS content

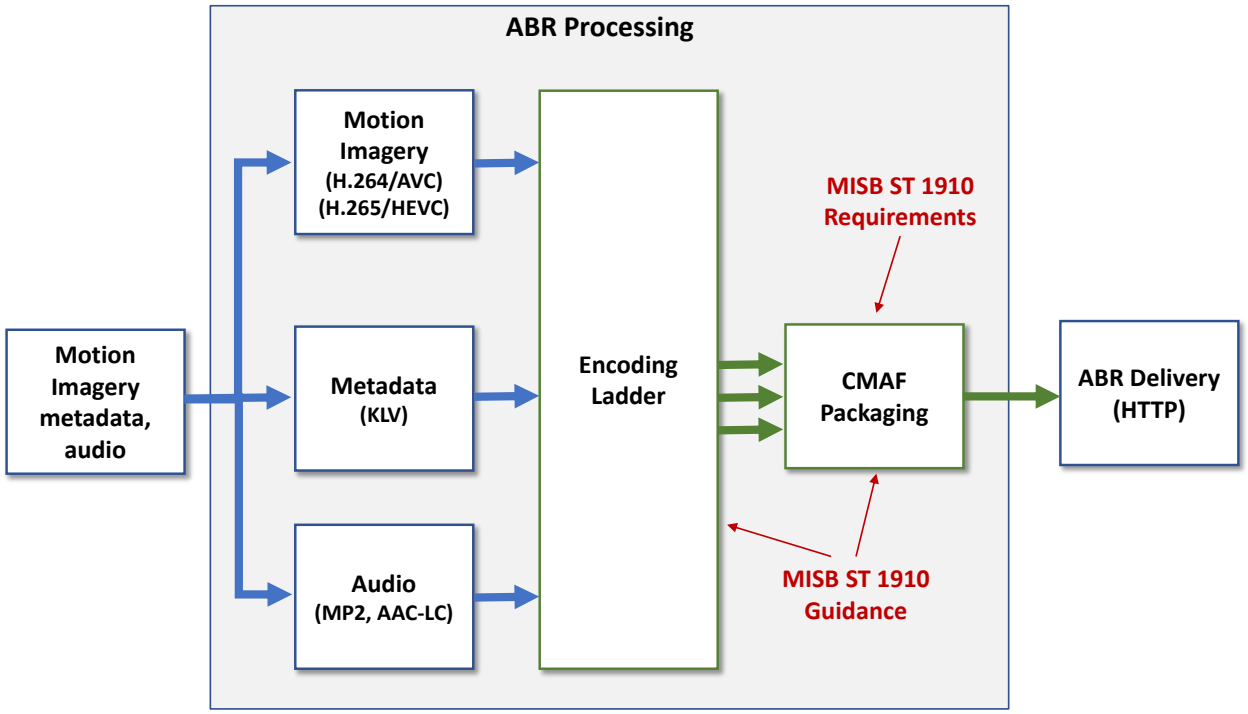

Figure 3: Source Content mapped to CMAF

6 CMAF Content Packaging Overview

The CMAF specification provides a complete overview using a hypothetical application model to describe several different ABR scenarios. The CMAF specification defines CMAF: Programs, Selection Sets, Switching Sets, Aligned Switching Sets, tracks, segments, and other constructs. This standard focuses only on the MISB requirements for packaging of Segments, use the other constructs in the CMAF specification as is.

6.1 Segments

Figure 4 shows a high-level structural view of a CMAF track, with one segment, and fragments. A CMAF **track** is a sequence of CMAF fragments that are consecutive in presentation time. CMAF requires a track to contain a header, optional index, and at least one CMAF segment. A CMAF **segment**, which consists of one or more consecutive CMAF fragments (in presentation order), is a compact self-contained set of media samples covering a short duration of media making it well suited to network transfer because it downloads quickly.

A CMAF **fragment** represents an encoded IBMF media segment conforming to CMAF constraints. A CMAF **chunk** contains a consecutive subset of the media samples of a CMAF fragment, where only the first CMAF chunk of a CMAF fragment is constrained to be an adaptive switching point. CMAF chunks are the smallest media object that can be encoded.

A CMAF segment can be delivered as a sequence of CMAF fragments (shaded blue) or as a sequence of CMAF chunks (shaded green). Both a fragment and a chunk have several box types: Optional boxes, which includes an Event Message Box, Movie Fragment Box and Media Data Box.

- **Optional Boxes:** Optional boxes provide for additional functionality, such as the Event Message Box. See CMAF specification for details.
- **Event Message (emsg) Box:** Event Message boxes provide inband media timed events such as closed captioning. Specific to this document, the MISB uses the emsg boxes for containerizing metadata (see Section 6.2).
- **Movie Fragment (moof) Box:** Movie Fragment boxes provide information for decoding the fragments. The moof box contains information about the fragment's or chunk's Media Data (mdat) boxes immediately following the moof. The CMAF specification requires a moof box per CMAF fragment or chunk.
- **Media Data (mdat) Box:** Media Data boxes contain the encoded video data in such forms as H.264/H.265.

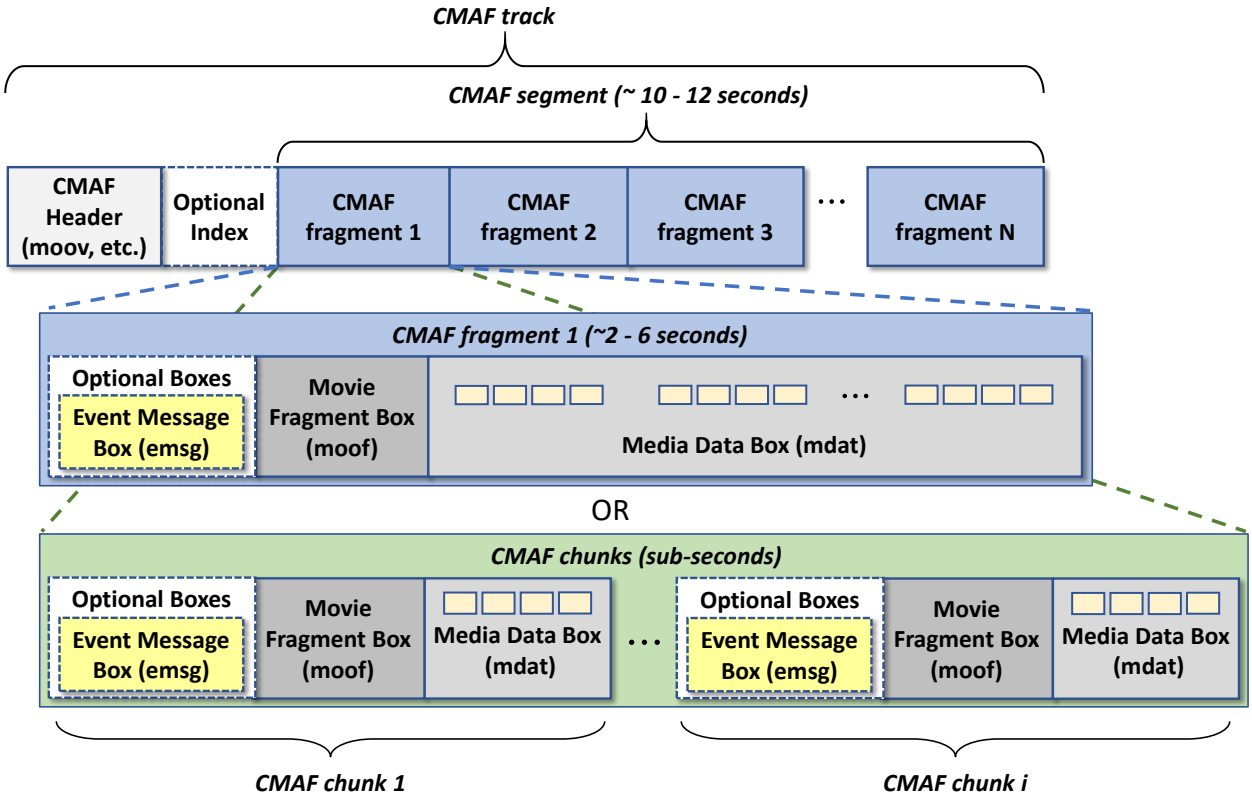

Figure 4: Structure of a CMAF Track

In typical applications, CMAF utilizes CMAF fragment delivery. In applications requiring much lower latency, such as broadcast's "live" streaming, CMAF utilizes Chunked Transfer Encoding (CTE) delivery in the ultra-low latency (ULL) mode. The ULL mode uses very

small chunk sizes (e.g., 200 milliseconds). Achieving such low latency through the transfer of small data units requires optimized infrastructure across a network to realize this benefit.

CMAF constrains a track to only contain one media essence type; thus, for example, one video encoding per track, one audio encoding per track, etc. Each CMAF track represents the encoding of one section in time of a media stream; thus, three different encodings (e.g., at different encoding rates) of the same content require three different CMAF tracks.

6.2 Event Message Box (emsg)

The inband Event Message, or emsg box, facilitates augmenting content with non-media information like closed captions or other types of metadata. An event message for non-media information specifies a data type, timing, identification, and data payload.

The MISB requires and uses emsgs to include KLV metadata with the Motion Imagery and uses the CMAF timing infrastructure to associate the KLV metadata to the Motion Imagery in the Media Data boxes (mdat).

Per the CMAF specification all emsg boxes pertaining to a fragment or chunk must precede the fragment or chunk (as shown in Figure 4); this minimizes the time for a client to detect and parse them. For example, if there is one emsg per image frame, such as illustrated in Figure 5, then an equivalent number of emsg boxes consistent with the number of frames present within that fragment or chunk need to precede the mdat box. In this 2-second segment there are 60 frames (i.e., F1, F2, ...F60) of Motion Imagery. Thus, 60 emsg boxes (i.e., e1, e2, ...e60) equating to the 60 image frames precede the mdat box.

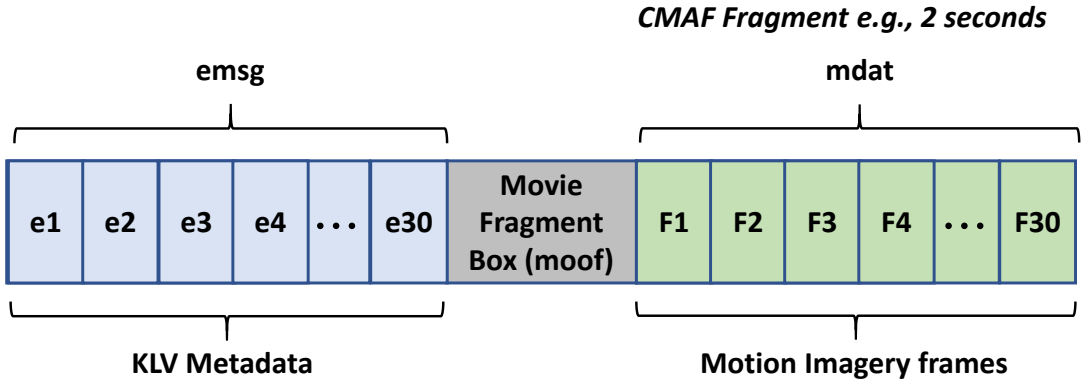

Figure 5: Example: KLV metadata as emsg and Motion Imagery in mdat box

Note that although this example shows one emsg for each media sample (i.e., frame), emsgs may be sporadic (i.e., not for every frame), or there may be multiple emsgs assigned to the same image frame.

6.2.1 Structure

The emsg box data structure, shown in Figure 6, provides signaling for generic events related to the media presentation time.

```
aligned(8) class DASHEventMessageBox extends FullBox (BoxType='emsg', version, flags = 0){
  if (version==0) {
    string    scheme_id_uri;
    string    value;
    unsigned int(32)    timescale;
    unsigned int(32)    presentation_time_delta;
    unsigned int(32)    event_duration;
    unsigned int(32)    id;
  } else if (version==1) {
    unsigned int(32)    timescale;
    unsigned int(64)    presentation_time;
    unsigned int(32)    event_duration;
    unsigned int(32)    id;
    string              scheme_id_uri;
    string              value;
  }
  unsigned int(8)    message_data[];
```

Version 1 used for CMAF 'emsg'

Figure 6: DASHEventMessageBox inband emsg data structure

The DASH specification defines the Event Message Box with two versions. This document does not use version 0. Version 1, shown in yellow shading, includes the field presentation_time. This makes emsg timing independent of box location within a CMAF track; more importantly it allows for aligning the message data within an emsg to the media in time.

Table 1 paraphrases from the MPEG-DASH and CMAF standards the version 1 descriptions for the DASHEventMessageBox fields. See the CMAF specification for additional constraints on the DASHEventMessageBox.

Table 1: DASHEventMessageBox Fields

Field	Description
Boxtype	'emsg' – fixed value denoting the box is an Event Message.
Version	Integer specifying the version of the box format. The CMAF specification recommends version 1.
flags	Additional parameters as needed. Set = 0 for this box type.
timescale	Equal to the timescale in the MovieHeaderBox[1], in ticks per second. Specifies the time scale for the entire presentation; this is the number of time units that pass in one second. For example, a time coordinate system that measures time in sixtieths of a second has a time scale of 60.

[1] The MovieHeaderBox is part of the CMAF specification.

presentation_time	The presentation time of the event measured on the CMAF track's presentation timeline (ticks), in the timescale (as declared in the MovieHeaderBox)
event_duration	The duration of the event measured on the CMAF track's presentation timeline (ticks), in the timescale (as declared in MovieHeaderBox). The value 0xFFFFFFFF indicates an unknown duration.
id	Identifies this instance of the message. Messages with equivalent semantics shall have the same value, i.e., processing of any one event message box with the same id is sufficient.
scheme_id_uri	Identifies the message scheme. The owner defines the semantics and syntax of the message_data[] for the scheme identified. The string may use URN or URL syntax.
value	Specifies the value for the event. The owner of the scheme must define the value space and semantics identified in the scheme_id_uri field.
message_data[]	The data for the event message. The owner of the scheme defines the syntax and semantics of this field identified in the scheme_id_uri field. Specific applications and users may define message schemes.

6.2.2 Frequency

A media sample can have one or more emsg boxes attached to it (note: a media sample defined in CMAF is media data associated with a single decode start time and duration such as a frame of video). Thus, a frame of Motion Imagery is a media sample and can have more than one associated emsg, thereby permitting grouping of metadata based on application needs. Section 7 describes the emsg value and id fields that facilitate this grouping.

6.2.3 Rate Partitioning

Rate partitioning is a method for separating multiple sources of sample data into streams of "like" sample rates. For instance, metadata sources sampled and captured at 10 Hz, 30 Hz, and 50 Hz can be "rate partitioned" into three separate streams with update rates of 10 Hz, 30 Hz and 50 Hz. In such cases, metadata sampling may be at a rate different than the Motion Imagery. Within CMAF the timescale property defines the timeline for the file which governs the timing for all information in the file. Choosing a suitable timescale which supports both the media and emsg data within the file allows rate partitioning of metadata.

CMAF defines the timescale in ticks per second. The choice in timescale needs to account for the highest expected frequency of the media or data; the timeline needs enough resolution to represent the smallest unit of time expected for a sample. Video typically runs at a near constant sample (i.e., frame) rate, whereas metadata may be aperiodic, more frequent than the frame rate, and need more precise timing than a frame sample.

As an example, a timescale of 60 ticks per second (tps) is enough to specify video at 60 frames per second (fps), where the duration for each media sample is one frame (i.e., 60/1 = 60 ticks per frame). Likewise, specifying a timeline of 96 000 tps but with a duration of 1600 ticks per frame produces the same result (i.e., 96 000/1600 = 60 frames/second). A greater timescale value enables sampling data to a finer resolution in time. Choosing a timescale to meet the resolution

requirements for the highest frequency signal in the file enables rate partitioning. The timescale does need however to be a multiple of the image frame rate to be evenly divisible.

A client can sort the different rates of metadata in successive emsg boxes based on its emsg presentation_time, the timescale, and the difference in presentation_time.

7 Source Content Mapping

The MISP mandates security information be present within any file containing Motion Imagery, metadata, or audio content. The security information represents the highest level of security for the combined content within a file. Security information when carried within metadata along with the Motion Imagery ensures Motion Imagery and metadata remain together in delivery to a client receiver.

Requirement	
ST 1910-02	A CMAF file shall contain security information.

In mapping source content containing metadata, the produced CMAF track Motion Imagery-to-metadata timing needs to reflect the Motion Imagery-to-metadata relationships within the source content.

Requirement	
ST 1910.1-16	The image frame-to-emsg association shall preserve the Motion Imagery-to-metadata temporal relationship in the source content.

7.1 Motion Imagery Essence

The CMAF specification limits the compression type, profiles, and levels for ABR video essence. Future versions of this specification will likely add additional formats. Table 2 lists the intersection of supported video profiles/levels by CMAF and those approved for use within the MISP. Updates to this document will reflect support of higher compression levels, such as H.265/HEVC Level 6.1 (also approved within the MISP), as ABR technology evolves. Note: the 5.1 Level for UHD/AVC is not an approved compression level in the MISP. This document supports this level to bridge the lack of browser compatibility for H.265.

Table 2: Motion Imagery Profiles

Motion Imagery Profile	Codec	Profile	Level	Color Primaries	Brand	Specification
HD	AVC	High	4.0	BT.709 [9]	‘cfhd’	CMAF Media Profile
HDHF	AVC	High	4.2	BT.709	‘chdf’	CMAF Media Profile
UHD	AVC	High	5.1	BT.709	‘avc1’	Advanced Video Coding extensions
HDD8	HEVC	Main10	5.0	BT.709	‘cud8’	CMAF Media Profile
UHD10	HEVC	Main10	5.1	BT.709	‘cud1’	CMAF Media Profile

Requirement	
ST 1910-03	A CMAF file shall conform to the Motion Imagery profiles listed in MISB ST 1910 Table 2: Motion Imagery Profiles.

7.1.1 Encoding Guidelines

The following recommendations for encoding Motion Imagery for a given source are to facilitate a consistent user experience.

- Fixed frame size, fixed aspect ratio, and fixed pixel density
- Constant temporal frame rate
- Key frame (I/IDR) present at least every GOP
- GOP of 2 seconds (equates to 60 frames at 30 frames per second)
- Closed-GOP structures for seamless switching between renditions
- One codec type per CMAF presentation (no codec changes)
- Coding structure (I-B-P) optimized for image quality and bandwidth based on application needs (see MISB RP 0802 [10] for additional guidelines).

7.1.2 Encoding Ladder Guidelines

Encoding parameters such as spatial/temporal resolutions and bitrates for the Encoding Ladder may vary based on application and workflow requirements. Table 3 is the exemplar set of bitrates in kilobits per second (kb/s) for encoding H.264/AVC and H.265/HEVC assuming 16:9 aspect ratio imagery and a temporal frame rate of 30 frames per second (fps).

Table 3: Exemplar Encoding Ladder

Spatial Density (samples)	Frame Rate (fps)	Aspect Ratio	H.264/AVC (kb/s)	H.265/HEVC (kb/s)
1920 x 1080	30	16:9	5500	3800
1280 x 720	30	16:9	3300	2300
1280 x 720	30	16:9	2000	1400
960 x 540	30	16:9	1200	850
640 x 360	30	16:9	750	500
640 x 360	30	16:9	450	300

Suggested guidelines for building an Encoding Ladder include:

- Sufficient maximum and minimum bitrates to meet anticipated network bandwidth fluctuations.
- Steps in the ladder given the selected range where the ratio between steps is 1.5 to 2. For example, the ratios of bitrates between the first two levels in Table 3 are:
  - 1080p AVC at 5500 kb/s and 720p at 3300 kb/s = 5500/3300 = 1.67
  - 1080p HEVC at 3800 kb/s and 720p at 2300 kb/s = 3800/2300 = 1.65

- Choose a reference frame size which optimizes image quality for its given bitrate; then from this bitrate determine the remaining ladder encodings.

7.1.3 CMAF Packaging Recommendations

A CMAF segment begins with an encoded imagery I/IDR frame to facilitate switching amongst representations (e.g., different bitrate encodings of the content). Although segment length may vary across applications, MISB recommends a segment length equal to two seconds, which provides a balance between access time and player requests. This means each GOP of encoded imagery will likewise be two seconds. Thus, at 30 frames per second one segment corresponds to 60 frames of imagery.

MISB recommends a CMAF fragment length equal to one second for delivery. Although this introduces additional overhead with an additional IDR frame, the shorter fragment affords finer control in stepping through content. For low latency applications, CMAF permits the use of CTE for ultra-low-latency delivery. This mode of stream delivery splits fragments into smaller addressable "chunks" ultimately reducing end-to-end streaming latency. A reasonable chunk length is 200 milliseconds; this equates to 6 frames of imagery at 30 fps. The values shown in Table 4 apply to all encoded spatial resolutions and bitrates within the encoding ladder.

Table 4: CMAF Packaging Guidelines: Imagery at 30 fps

GOP Type	GOP Size	Segment Length	Fragment Length	Chunk Length (opt. Low Latency)
Closed	2 sec (60 frames)	2 sec (60 frames)	1 sec (30 frames)	200 msec (6 frames)

Design choices in Encoding Ladder, encoding structure, GOP size, segment/fragment length, etc. depend on the requirements for its application. In lieu of such requirements, the parameters indicated in the above sections should produce a good solution.

7.2 KLV Metadata Event Message (emsg) Profile

As discussed in Section 6.2, the emsg box provides a structure to insert and carry metadata along with its corresponding media (e.g., Motion Imagery or audio). The following sections define the KLV metadata event message (emsg) profile of the CMAF event message box.

KLV metadata directly embeds into the emsg box. This standard requires emsg box version 1, which supports a presentation_time value to align an emsg in time with its corresponding media (e.g., Motion Imagery or audio) along a CMAF timeline.

Requirement	
ST 1910-01	The KLV metadata emsg profile shall use version 1 of the MPEG-DASH Event Message structure.

7.2.1 Time Scale Field (timescale)

The timescale defines the number of ticks per second as the measurement basis for the presentation_time and event_duration (both measured in ticks). Per the CMAF specification the timescale value is the same as the CMAF MovieHeaderBox timescale.

The KLV metadata emsg profile uses the timescale as defined in the CMAF specification.

7.2.2 Presentation Time Field (presentation_time)

The presentation_time field of the emsg defines the time, in ticks, of an emsg along the CMAF timeline. The presentation_time value in an emsg enables synchronization and binding of the KLV metadata to its associated frame of Motion Imagery.

The KLV metadata emsg profile uses the presentation_time as defined in the CMAF specification.

7.2.3 Event Duration Field (event_duration)

The event_duration specifies the duration of the event message, in ticks.

KLV metadata is a collection of individual metadata items each of which may or may not have a representative duration. Furthermore, MISB KLV metadata uses Report-On-Change (see MISB ST 0107 [11] and the Motion Imagery Handbook [7]) method which distorts the ability to define a duration. To properly understand the duration of metadata items, CMAF receivers must follow the rules in the underlying MISB KLV metadata standard(s). From the perspective of the emsg, the duration is unknown.

The KLV metadata emsg profile sets the event_duration to the "unknown duration" value of 0xFFFFFFFF – see DASH specification.

Requirement	
ST 1910.1-17	The KLV metadata emsg profile emsg event_duration shall be 0xFFFFFFFF.

7.2.4 Identifier Field (id)

The DASH event message id field provides each emsg a unique identity. The DASH specification envisions emsgs as signals to players to take actions (e.g., displaying advertisements, or sports scores). The emsg signals are repeatable to enable players just joining (or jumping into) a stream to receive the signal. The emsg id field identifies a particular signal. Once a player processes an emsg's signal action (identified with an emsg id), the player ignores all other emsgs with the same emsg id.

The KLV metadata emsg profile does not use repeatable emsgs; every emsg must have a unique id. To ensure uniqueness the KLV metadata emsg profile defines a method for the CMAF packager to create an emsg id comprising a count of segment number and a count of an emsg.

Requirement	
ST 1910.1-18	The KLV metadata emsg profile emsg id shall be the combination of the segment number in the most significant two bytes and the emsg count in the least significant two bytes. .

Figure 7 shows the KLV metadata emsg profile format of the for the 32-bit integer emsg id field.

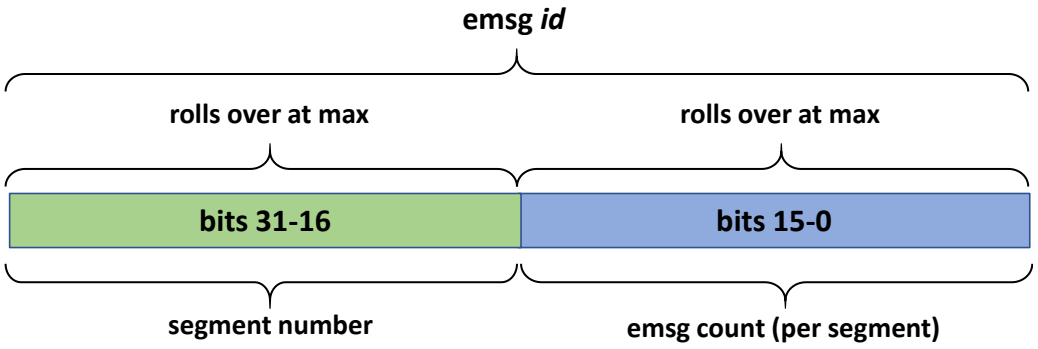

Figure 7: KLV metadata emsg profile id field structure

The KLV metadata emsg profile splits the id field into two counters: a 16 bit "segment number" counter to count CMAF segments (bits 31-16), and a 16-bit "emsg count" counter to count the number of emsgs per segment (bits 15-0). The segment number updates by one (1) for each new sequential CMAF segment. Per given segment number the emsg count increments by one (1) for each successive emsg. Thus, 16 bits or 65,536 emsgs are possible for each CMAF segment.

Requirement(s)	
ST 1910.1-19	The KLV metadata emsg profile emsg id segment number (bits 31-16) shall increment by one (1) for each successive CMAF segment.
ST 1910.1-20	The KLV metadata emsg profile emsg id emsg count (bits 15-0) shall increment by one (1) for each successive emsg within a CMAF segment.

This organization of the emsg id has several advantages: 1) metadata assigned to an emsg has a "signature" to aid in debugging and validating that metadata input from a source maps properly in both sequence and time to a corresponding emsg; and 2) government applications maintain a consistent id construction, which aids in minimizing test tool development.

Although the emsg id is a unique integer value, an emsg requires two additional pieces of information to assure an emsg is unique: 1) the emsg scheme_id_uri and 2) the emsg value. These three data must be present and as a group must be unique across emsgs. A client will ignore an emsg with the same three data (note: per the CMAF specification).

Note: from the client perspective it is only important that the emsg id is unique over the emsg time of use. Clients should utilize the full 32-bit field when evaluating uniqueness.

7.2.5 Scheme Identifier URI Field (scheme_id_uri)

The KLV metadata emsg profile for MISB KLV metadata defines the scheme_id_uri to be set to the combination of the utf8 strings “urn:misb:KLV:bin:” (with exact upper/lower case shown) and this document’s standard number, which for this version is ”1910.1”. Each emsg box carries this information in addition to the CMAF MPD.

Requirement	
ST 1910-04	The scheme_id_uri for emsg carriage of MISB KLV metadata shall be “urn:misb:KLV:bin:1910.1”.

7.2.6 Value Field (value)

The KLV metadata emsg profile for MISB KLV defines the emsg value field as a utf8 string composed of the two subfields, source-identifier and source-characteristic using the following syntax:

source-identifier : source-characteristic

Requirement	
ST 1910.1-21	The KLV metadata emsg profile value field shall be a utf8 string combining the source-identifier , a colon (“:”), and the source-characteristic.

The source-identifier subfield provides a unique identifier for a metadata stream. For example, a source-identifier = “KLV1” identifies all metadata as belonging to the KLV1 metadata stream; a source-identifier = “KLV12” identifies all metadata as belonging to the KLV12 metadata stream. An implementor may choose the nomenclature for s source-identifier. See section 8.1.1.2.3 for requirements on the value field in mapping MPEG-2 TS metadata.

The source-characteristic subfield signals the type of time relationship metadata has with respect to its imagery. For example, consider the source of content as an MPEG-2 TS. Metadata is either multiplexed as “synchronous” (i.e., assigned a timestamp called the presentation Time Stamp (PTS) in the packet header), or “asynchronous” (i.e., not having a PTS timestamp). The source-characteristic subfield signals this type of information. Different source data may have different source-characteristic values. MISB defines MPEG-2 TS source data source-characteristic values in section 8.1.1.2.3.

This document will add new source-characteristic values as new applications evolve.

Note: the designations synchronous and asynchronous as used here are consistent with terminology describing metadata carriage in MPEG-2 Transport Streams.

7.2.6.1 Value Field Signaling in MPD

The MPD lists the source-identifier:source-characteristic in the emsg value field as an inband event to signal the decoder how to interpret the data within the emsg boxes. In MPEG-DASH, for example with a source-identifier = “KLV2” and a source-characteristic = “01FC” (see section 8.1.1.2.3), the MPD signals the available inband event as:

InbandEventStream schemeIdUri="urn:misb:KLV:bin:1910.1" value="KLV2:01FC"/>

7.2.7 Message Data Field (message_data[ ])

The message_data[] is the data, or payload, information for the emsgs scheme_id_uri.

The KLV metadata emsg profile defines the message_data[] value to be KLV data in the form of one or more KLV Packets (as defined by the Motion Imagery Handbook) as a series of bytes. The message_data[] value is one or more complete KLV Packets.

Although the message_data[] field is agnostic to the data type stored, commercial/consumer decoder/players do not typically parse and interpret KLV. Thus, parsing and rendering this data requires special player decoding logic.

Requirement(s)	
ST 1910.1-22	An emsg shall only contain KLV metadata within the message_data[] field.
ST 1910.1-23	The emsg message_data[] field shall contain one or more complete KLV Packets (as defined in the Motion Imagery Handbook).

7.2.8 Example

Table 5 shows an example emsg box with parameters for scheme_id_uri and value defined by ST 1910 for MPEG-2 TS mapping. The boxtype is an emsg, the version is 1, and the flags are set to 0 as indicted in Table 1. The timescale of 25000 indicates 25000 ticks per second for the presentation timeline; thus, for example, this timescale together with the duration field in the MovieHeaderBox of 1000 ticks per second represents Motion Imagery at 25000/1000 = 25 frames per second. The emsg event_duration of 0xFFFFFFFF means each emsg has an unknown duration. The scheme_id_uri is the MISB-defined namespace of urn:misb:KLV:bin:1910.1. The Value field is one of the identifier strings defined in Section 7.2.6. Finally, metadata items from MISB ST 0601 [12] form the contents of the message_data[] field.

Table 5: Example ST 1910 emsg box

Field	Value (examples)	Comment
boxtype	emsg	
version	1	Required value
flags	0	Required no flags
timescale	25000	ticks per second (tps)
presentation_time	0	time (0 ticks /25000 tps = 0 seconds)
event_duration	0xFFFFFFFF	unknown
id	0x00050001	segment number = 0x0005XXXX (segment 5) emsg count = 0xXXXX0001 (emsg 1)
scheme_id_uri	urn:misb:KLV:bin:1910.1	Required value
value	KLV5:01BD	source-identifier = “KLV5”; source-characteristic = “01BD”. See Section 7.2.6
message_data[]	6, 14, 43, 52, 2, 11, 1, 1, 14, 1, 3, 1, 1, 0, 0, 0, 130, 1, 29, 2, 8, 0, 5, 33, …	Binary KLV data (hex): 060E 2B34 020B 0101 0E01 0301 0100 0000 8201 1D02 0800 0521…

Appendix A provides a sample of MISB KLV metadata mapped into an emsg box.

7.3 Audio Essence

If encoding audio, the CMAF specification allows a variety of compression types for ABR audio. Table 6 lists the intersection of supported profiles by CMAF and those approved by the MISP in MISB ST 1001 [13]. Refer to the CMAF specification for details on mapping audio to CMAF.

Table 6: Audio Media Profiles

Audio Profile	Codec	Profile	Level	CMAF Brand
AAC	AAC	AAC-LC	2	'caac'

Requirement	
ST 1910-09	A CMAF file shall conform to audio profiles listed in MISB ST 1910 Table 6: Audio Media Profiles.

When audio is present it accompanies Motion Imagery and metadata when collected, as all three relate to one another. This standard is Motion Imagery and metadata centric; however, this standard applies to packaging audio into CMAF as well. For example, metadata could accompany audio packaged into its own CMAF file using emsg boxes associated to specific portions of the audio.

8 Source-specific Packaging

Source content as identified within the MISP may come from Class 0 Motion Imagery, Class 1 Motion Imagery, Class 2 Motion Imagery or Class 3 Motion Imagery (properly converted to Class 1 or Class 2 Motion Imagery) acquisition systems. CMAF packaging applies to any type of content that meets both the CMAF specification and the approved methods for compression within the MISP as indicated in Table 2 and Table 6 above.

In this standard the packaging of content into CMAF and the signaling for metadata is the same regardless of the source content. This allows for extending the technology into new applications as they evolve. This document will update mappings for future sources in support of new applications as warranted. As the current demand is for mapping MPEG-2 TS content to CMAF, the following section pertains to this specific application.

8.1 MPEG-2 TS Content

Motion Imagery encapsulated within a MPEG-2 TS is either H.264/AVC or H.265/HEVC compressed data. MISB ST 1402 [14] provides guidelines for carriage of Motion Imagery, audio, and metadata in MPEG-2 TS. MISB ST 0604 [15] specifies the format and where to insert a

MISP timestamp into a Motion Imagery compressed stream. These two standards guide the requirements for mapping content from MPEG-2 TS to CMAF.

Figure 8 illustrates a conceptual example of an MPEG-2 TS with four essence streams: Compressed Motion Imagery, Synchronous KLV Metadata, Asynchronous KLV Metadata, and Audio. The Multiplexed View (yellow shaded area) illustrates the MPEG-2 TS multiplexing of compressed Motion Imagery (**I** and **P** frames in blue) frames, synchronous metadata (**M**$_S$ in dark green), asynchronous metadata (**M**$_A$ in light green), and audio (**A** in brown) data.

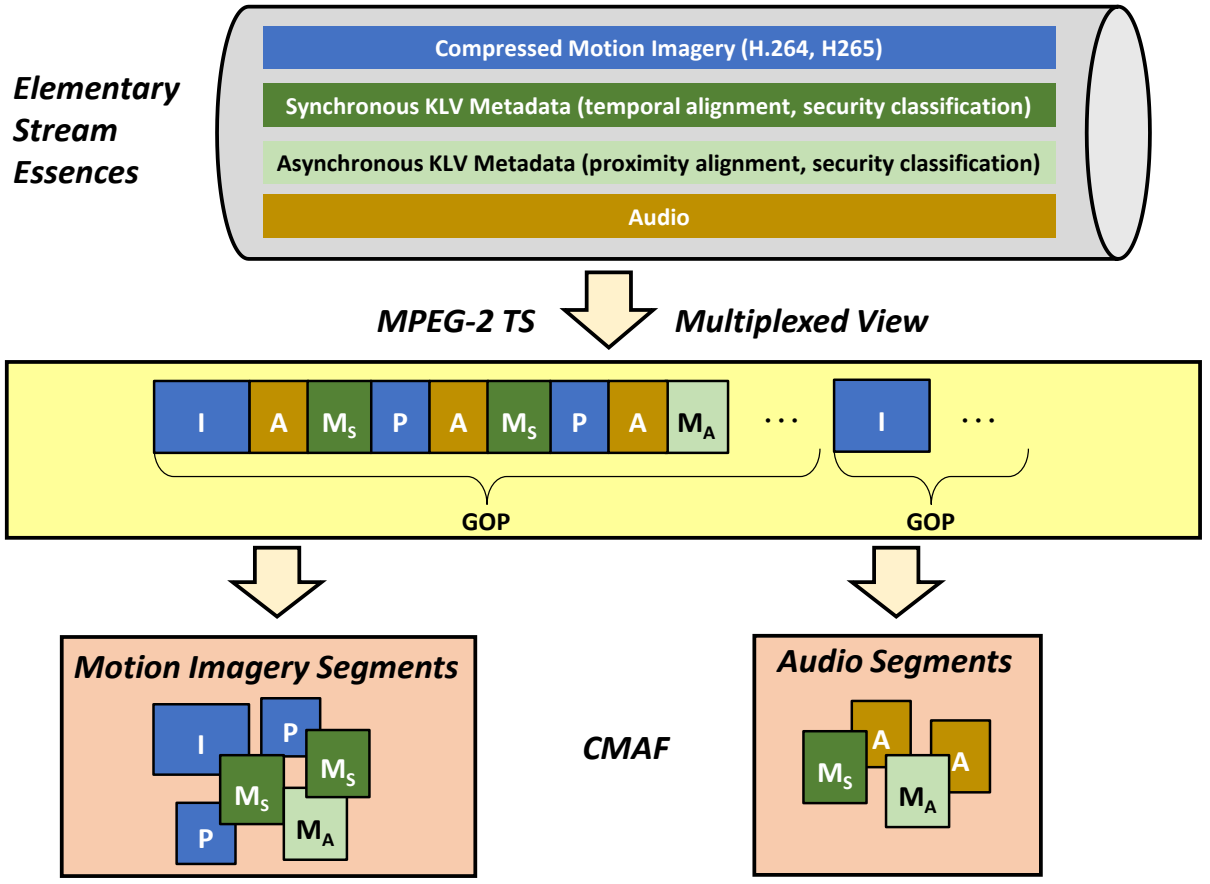

Figure 8: Example MPEG-2 TS essences mapped to CMAF file

As shown in Figure 8, when mapping the MPEG-2 TS essences to CMAF Motion Imagery segments the mapping separates the media essences into separate CMAF tracks; that is, the Motion Imagery and audio are each put into their own CMAF track. The mapping wraps the metadata into event messages and merges them with the Motion Imagery track. Likewise, the mapping could merge metadata with the audio track.

In this example of MPEG-2 TS multiplexing, a GOP is a sequence of one I frame, followed by a mix of multiple P frames, M_S, M_A, and A data, i.e., every I frame starts a new GOP. In general,

the CMAF mapping process maps GOPs to CMAF fragments while preserving the temporal synchronization between the frames and metadata. Alternatively, for low latency, the CMAF mapping process maps GOPs to CMAF chunks while preserving the temporal synchronization between the frames and metadata. The metadata (M_S and M_A) timing information does not need to align to the imagery (I and P) frames exactly; the metadata may be measurements recorded between frames.

MPEG-2 TS governs its time-alignment type between a Motion Imagery frame and metadata in two ways:

1) The Synchronous Metadata Multiplex Method (MISB ST 1402) prefixes the metadata with a PTS for multiplexing in similar fashion to the Motion Imagery to aid *temporal alignment* of metadata to imagery.
2) The Asynchronous Metadata Multiplex Method multiplexes the metadata in *proximity* alignment to the imagery, which could temporally displace the metadata away from the image frames by several image frames or more.

The CMAF mapping signals the time-alignment type along with a time-alignment "goodness" value. The mapping process maps the time-alignment type from the ISO/IEC 13818-1 [16] stream_id in the TS PES (packetized elementary stream) header. The PES indicates either synchronous data carriage (i.e., stream_id = 0xFC), or asynchronous data carriage (i.e., stream_id = 0xBD). In the case of synchronous carriage, the mapping process determines the time-alignment "goodness" value from the synchronous stream's metadata_application_format value. In asynchronous carriage, there is no such signaling provided in the TS. The time-alignment type information maps to CMAF emsg boxes to maintain the intended content integrity and acts as signaling for end user consumption.

The following sub-sections describe the transformation of MPEG-2 TS multiplexed data to CMAF segments.

8.1.1 Mapping MPEG-2 TS to CMAF Motion Imagery Segments

MPEG-2 TS Motion Imagery may map into multiple types of CMAF segments depending on the Encoding Ladder. Mapping requires a two-step process: transcoding and packaging. The transcoding step changes the Motion Imagery data into different compression profiles and levels as desired, see Section 7.1.1 for encoding guidelines. The packaging step builds the CMAF segment and its associated IBMF boxes in the correct order according to the CMAF specification.

This document requires the mapping process to include all the metadata from the MPEG-2 TS in the resulting CMAF Motion Imagery segments. Figure 9 illustrates an example of MPEG-2 TS multiplexed data repackaged into CMAF Motion Imagery segments. This figure illustrates a single fragment in a CMAF segment consisting of a series of emsg boxes containing the metadata (M_S and M_A), followed by a moof box that describes the mdat box, then a mdat box containing the Motion Imagery. The fragment building process repeats as needed to form a CMAF segment.

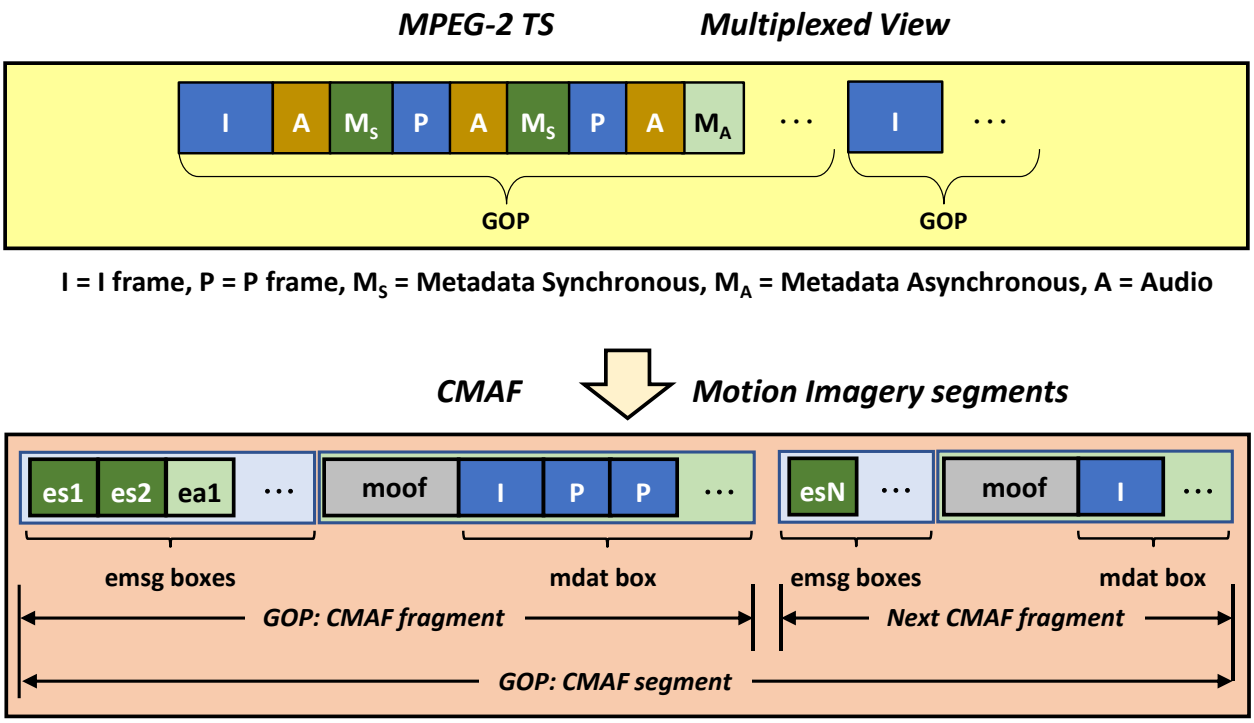

Figure 9: MPEG-2 TS to CMAF Motion Imagery Segments

	Requirement
ST 1910.1-24	The presentation_time of an emsg within a fragment shall be equal to or greater than the presentation_time of the fragment's first image frame and be less than the presentation time of the first frame of the next fragment's first image frame.

Multiple emsgs may have the same presentation_time, such as the case where multiple streams report their metadata at the same time, e.g., one synchronous and one or more asynchronous MPEG-2 TS essences.

8.1.1.1 Motion Imagery SEI Message: MISP Timestamp

Within the Supplemental Enhancement Information (SEI) message of the compressed Motion Imagery stream a MISP timestamp may be present (Note: current MISP guidance requires the SEI MISP timestamp; however, some legacy systems do not supply one). On packaging the Motion Imagery into CMAF this information is to remain intact. If the Motion Imagery is transcoded, the SEI messages are to remain in the transcoded result.

	Requirement
ST 1910.1-25	Motion Imagery extracted from a MPEG-2 TS and packaged into CMAF mdat boxes shall preserve all information in the SEI Message user_data_unregistered field.

8.1.1.2 Packaging KLV Metadata

Each MPEG-2 TS metadata stream contain a series of KLV Packets. The packaging process maps one or more KLV Packets, in their entirety, to emsgs. MISB KLV Packets potentially rely on previous packets, therefore KLV packets in the source MPEG-2 TS must map to emsgs corresponding to their image frames in their appropriate CMAF segment (refer to Requirement ST 1910-16). Packaging systems must not "drop" KLV Packets when packaging.

Requirement	
ST 1910.1-26	Every MPEG-2 TS metadata stream KLV Packet shall be present in an emsg.

The process of packaging MPEG-TS metadata is to set values in all the emsg fields based on the MISB emsg profile, see Section 7.2. Table 7 provides the instructions for packaging a single emsg.

Table 7: emsg Packaging Instructions

emsg Field	Instructions
timescale	Set to the same timescale value as the MovieHeaderBox timescale (this is a CMAF requirement)
presentation_time	See Section 8.1.1.2.1
event_duration	Set to 0xFFFFFFFF
id	See Section 8.1.1.2.2
scheme_id_uri	Set to "urn:misb:KLV:bin:1910.1"
value	See Section 8.1.1.2.3
message_data[]	See Section 8.1.1.2.4

8.1.1.2.1 Presentation Time Field (presentation_time)

Requirement ST 1910.1-16 dictates assigning a presentation_time to an emsg. The CMAF presentation_time measures an event along the CMAF presentation timeline; both the emsg box timescale value and the presentation timeline timescale are the same.

An emsg presentation_time depends on the type of MPEG-2 TS metadata multiplexing. With MPEG-2 TS synchronous metadata, the metadata's PTS values define the time of metadata along the TS timeline measured in 90,000 ticks per second. If for example, the CMAF timescale is 25,000 ticks per second, for a PTS value of 10 000 000 the conversion to CMAF presentation_time is (10 000 000/90 000) x 25 000 = 27 777 778 ticks. Table 8 provides additional examples of converting the TS timescale to the CMAF timescale.

Table 8: TS to CMAF Time Conversion - examples

MPEG-2 TS timescale (tps)	CMAF timescale (tps)	PTS (tps)	CMAF presentation_time (tps)
90,000	25,000	60000	16 667
90,000	60,000	60000	40 000
90,000	60,000	10 000 000	66 666 667

With asynchronous metadata the data locality within the stream provides an "approximate" timing of the metadata. There is no certainty of the exact time alignment of the metadata to a frame of imagery since there is no corresponding PTS. Methods may vary in how to choose an optimal presentation_time. One method might assign a presentation_time equal to the nearest image frame. Another might assign a presentation_time equal to the locality of the metadata along the MPEG-2 TS timeline.

8.1.1.2.2 Identifier Field (id)

The structure of the KLV metadata emsg profile emsg id field in section 7.2.4.

8.1.1.2.3 Value Field (value)

For MPEG-2 TS metadata the KLV metadata emsg profile value field contains a combination of both the metadata stream assigned by its PID in the MPEG-2 TS (defined by the source-identifier subfield), and the type of time-alignment of the metadata within the MPEG-2 TS (see section 7.2.6) defined by the source-characteristic subfield.

8.1.1.2.3.1 source-identifier for MPEG-2 TS

The "source-identifier" assigns a unique metadata stream of KLV metadata. For example, suppose an MPEG-2 TS contains one synchronous metadata stream on PID 52 and two asynchronous metadata streams on PIDs 55 and 56. The source-identifier for the synchronous stream might be "PID52", or "KLV1", while the source-identifiers for the two asynchronous streams might be "PID55" and "PID56", or "KLV10" and "KLV20" respectively. *Note: there is no guarantee that the CMAF packager will respect the MPEG-2 TS PID numbers and will likely assign new numbers.* This example is to illustrate the concepts only.

Requirement(s)	
ST 1910.1-27	Each MPEG-2 TS metadata stream shall have a unique emsg source-identifier in the emsg value field.
ST 1910.1-28	KLV metadata extracted from the same Packet Identifier (PID) metadata stream in a MPEG-2 TS shall use the same emsg source-identifier.

8.1.1.2.3.2 source-characteristic for MPEG-2 TS

For synchronous metadata, the source-characteristic derives from the MPEG-2 TS synchronous metadata's descriptor metadata_application_format value. For asynchronous metadata, the

corresponding registration_descriptor does not have such an indicator; in this case, there is only one source-characteristic value.

Table 9 lists the allowed source-characteristic for MPEG-2 TS metadata.

Table 9: Allowed MPEG-2 TS source-characteristic.

MPEG-2 TS		emsg value source-characteristic (utf8)	metadata type
stream_id	metadata_application_format		
0xFC	0x0100 – 0x0103	"01FC"	time-aligned (Level 2)
0xFC	0x11FC	"11FC"	time-aligned (Level 1*)
0xFC	0x12FC	"12FC"	time-aligned (Level 2*)
0xBD	none	"01BD"	time-proximity (Level 3)
*See Appendix B.1.3 for additional information on these codes.			

8.1.1.2.3.3 MPEG-2 TS Synchronous Metadata

MPEG-2 TS synchronous metadata with metadata_application_format values in the range 0x0100-0x0103 and stream_id = 0xFC map to the emsg value field source-characteristic as utf8 string "01FC". "FC" is readily identifiable as the MPEG-2 TS synchronous stream_id type.

Requirement(s)	
ST 1910.1-29	Where mapping MPEG-2 TS synchronous metadata with a stream_id = 0xFC and metadata_application_format in the range 0x0100-0x0103, the source-characteristic subfield of the emsg value field shall be equal to the utf8 string "01FC".
ST 1910.1-30	Where mapping MPEG-2 TS synchronous metadata with a stream_id = 0xFC and metadata_application_format = 0x11FC, the source-characteristic subfield of the emsg value field shall be equal to the utf8 string "11FC".
ST 1910.1-31	Where mapping MPEG-2 TS synchronous metadata with a stream_id = 0xFC and metadata_application_format = 0x12FC, the source-characteristic subfield of the emsg value field shall be equal to the utf8 string "12FC".

Note: use of the emsg source-identifiers "11FC" and "12FC" assume a remediated MPEG-2 TS. Appendix B discusses remediation, which is a post-collection process where analysis of the metadata-to-Motion Imagery timing may afford improved timing relationships. A remediated stream would leverage these additional emsg value codes. Ongoing community engagement continues towards best practices for remediation.

8.1.1.2.3.4 MPEG-2 TS Asynchronous Metadata

Unlike synchronous metadata, asynchronous metadata does not have a metadata_application_format descriptor. Signaling asynchronous metadata for CMAF packaging is through the MPEG-2 TS stream_id type (which is the value 0xBD) alone. MPEG-2 TS asynchronous metadata with stream_id = 0xBD maps to the emsg value field source-

characteristic as utf8 string “01BD”. “BD” is readily identifiable as the TS asynchronous stream_id type.

Requirement	
ST 1910.1-32	Where mapping MPEG-2 TS asynchronous metadata with a stream_id = 0xBD, the source-characteristic subfield of the emsg value field shall be equal to the utf8 string “01BD”.

8.1.1.2.3.5 Emsg example

In Figure 10 two MPEG-2 TS metadata PIDs maps to its own respective set of emsg boxes based on its source-characteristic, and the type of metadata carriage (i.e., synchronous, asynchronous signaled by the source-identifier). Synchronous metadata maps into emsg boxes denoted es1, es2, …es60; these emsgs carry the emsg value = “KLV1:01FC”. Similarly, asynchronous metadata maps to the two ea3 and ea15 emsg boxes; these emsgs carry the emsg value = “KLV2:01BD”. Thus, an emsg box has a type “personality” containing only that type of data.

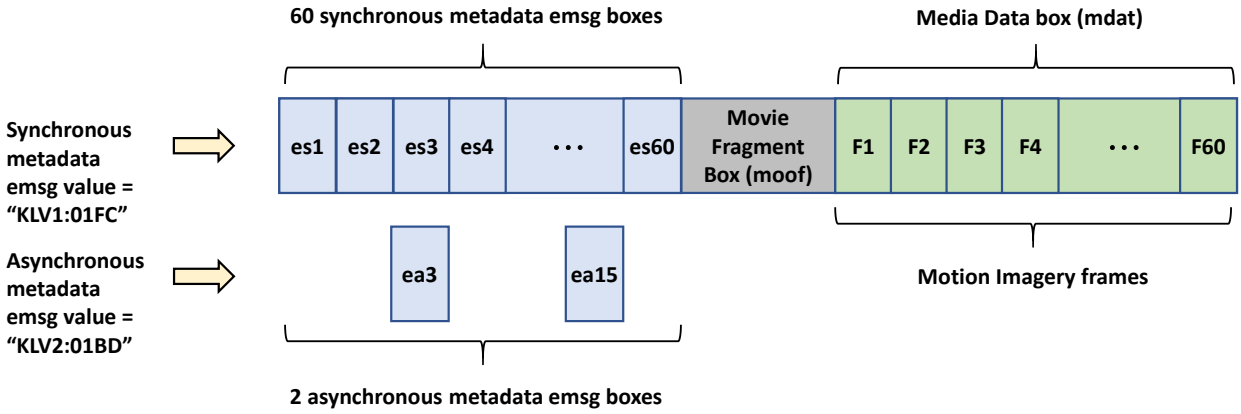

Figure 10: Example: synchronous (“es”) and asynchronous (“ea”) metadata mapped into respective value field of emsg boxes.

This example shows one emsg box per image frame (i.e., es1 contains metadata for F1, es2 contains metadata for F2, etc.) for the synchronous metadata, and two emsg boxes (i.e., ea3 and ea15) for the asynchronous metadata over the 60-frame image sequence. The incidence of ea3, ea15 falls along the media timeline only in proximity to a frame of imagery it may relate.

8.1.1.2.4 Message Data Field (message_data[])

The mapping process sets the message_data[] field with complete KLV Packets from the MPEG-2 TS metadata stream. The message_data[] field may contain multiple KLV Packets if they have equal KLV MISP Precision Time Stamps.

Requirement	
ST 1910.1-33	Where an emsg message_data[] contains more than one KLV Packet, the KLV MISP Precision Time Stamps in each KLV Packet shall be equal.

8.1.1.3 Client Compatibility

This standard defines the signaling given in Table 9 to indicate to a user the temporal alignment "quality" between the Motion Imagery and metadata. The word "quality" is meant to set an expectation and assurance that the metadata viewed or used in computation is an accurate depiction of data about the image displayed. Ideally, metadata about an image is in perfect time alignment, to better support accurate measurements and conclusions when exploiting the imagery; such a system produces "high quality" time alignment. However, systems vary in degrees of temporal alignment quality between the imagery and metadata. This standard defines the quality for three levels of time-alignment as indicated in Table 9. Appendix B offers more details on the origins of these levels of time alignment.

CMAF clients may elect to make a visual indication of the different qualities of time-alignment to a user. In an MPEG-2 TS, information provided by both the stream type and metadata descriptors serve as inputs to CMAF to set flags for the three levels of time alignment. The right-most two characters of an emsg value source-characteristic subfield indicates whether the metadata is of MPEG-2 TS synchronous or asynchronous origin. Parsing the left-most two characters in emsg source-characteristic field provides a further distinction in the level of time alignment quality. A client can render this information on the display so that a user understands the quality of the imagery/metadata time alignment when exploiting the data.

9 Player Functionality

9.1 Trick Play

The term "trick play" refers to playback other than forward playback at the recorded speed of the video/audio content ("1x"). Examples include fast forward, slow motion, reverse, single step, and random access.

These modes of playback fall to the responsibility of the player design. However, impacting the degree of control is the choice in CMAF segment length and download speed limitations. Encoded content I/IDR frame periodicity determines random access points. Player buffer size is a factor in limiting rewinding to earlier content. Decoding slower than normal and repeating frames can simulate slow motion.

Other options to implement trick play include converting I/IDR frames to thumbnail images which require far less storage at the player. Playing through the thumbnail images creates a "film strip" of the content for determining when to begin playback at a specific point. Another option is to request I/IDR frame only segments at a lower bitrate to speed delivery. Higher bitrates then resume once normal playback continues.

9.2 Browser Video / Metadata Synchronization

Although synchronization of Motion Imagery and metadata within a CMAF file is predictable and reliable, web browsers such as Chrome, Firefox, and Edge exhibit different behavior in rendering video with other timed media like metadata. As the use cases for these technologies are typically video with audio (and subtitles with loose timing with respect to a video frame), synchronizing metadata to be frame accurate continues to be works in progress. Thus,

applications should be cognizant of these differences. The MISB continues to evaluate delivery performance.

10 Deprecated Requirements

Requirement(s)	
ST 1910-05 (deprecated)	The emsg id field shall be composed of the KLV stream index in the least significant byte, and a KLV emsg count starting at zero and incrementing by 1, in the upper 3 bytes for every subsequent emsg box for that KLV stream index.
ST 1910-06 (deprecated)	The KLV track index shall be between 0x10-0xFF.
ST 1910-07 (deprecated)	The emsg presentation_time shall be within the duration of an image frame.
ST 1910-08 (deprecated)	Where the number of bytes for an emsg is unknown, the emsg event_duration shall be equal to the duration of one image frame.
ST 1910-10 (deprecated)	Where H.265/HEVC compressed Motion Imagery is transcoded, the information in the SEI Message user_data_unregistered field shall be preserved.
ST 1910-11 (deprecated)	Where H.264/AVC compressed Motion Imagery is transcoded, the information in the SEI Message user_data_unregistered field shall be preserved.
ST 1910-12 (deprecated)	Where MPEG-2 TS synchronous metadata with a stream_id = 0xFC and metadata_application_format in the range 0x0100-0x0103 is mapped to a CMAF emsg box, the emsg value field shall be set to the utf8 string “01FC”.
ST 1910-13 (deprecated)	Where MPEG-2 TS asynchronous metadata with a stream_id = 0xBD is mapped to a CMAF emsg box, the emsg value field shall be set to the utf8 string “01BD”.
ST 1910-14 (deprecated)	Where MPEG-2 TS synchronous metadata with a stream_id = 0xFC and metadata_application_format = 0x11FC is mapped to a CMAF emsg box, the emsg value field shall be set to the utf8 string “11FC”.
ST 1910-15 (deprecated)	Where MPEG-2 TS synchronous metadata with a stream_id = 0xFC and metadata_application_format = 0x12FC is mapped to a CMAF emsg box, the emsg value field shall be set to the utf8 string “12FC”.

Appendix A Sample of MISB KLV Mapped to an emsg

Figure 11 shows a sample of MISB KLV metadata mapped into a version 1 emsg box. The Motion Imagery framerate is 25 Hz. A timescale of 25000 ticks/sec with a duration of 1000 ticks provides the framerate timing (i.e., 25000 ticks/sec ÷ 1000 ticks = 25 Hz). This is frame 1 so the presentation_time is 0 with an id = 0x001A 0001 (emsg #1). Note that the next event message follows the first. In the second event message the presentation_time = 0x03e8 (i.e., 1000) indicating a time of 1/25 of 25000 and an id = 0x001A 0010 (emsg #2). The emsg value field of "KLV45:01FC" indicates both events come from the same MPEG-2 TS metadata essence (PID 45) signaled as 'KLV45' in the value field and are a Level 2 synchronization (value field includes '01FC'). The scheme_id_uri = urn:misb:KLV:bin:1910.1 which indicates a namespace for this dot 1 version of the standard.

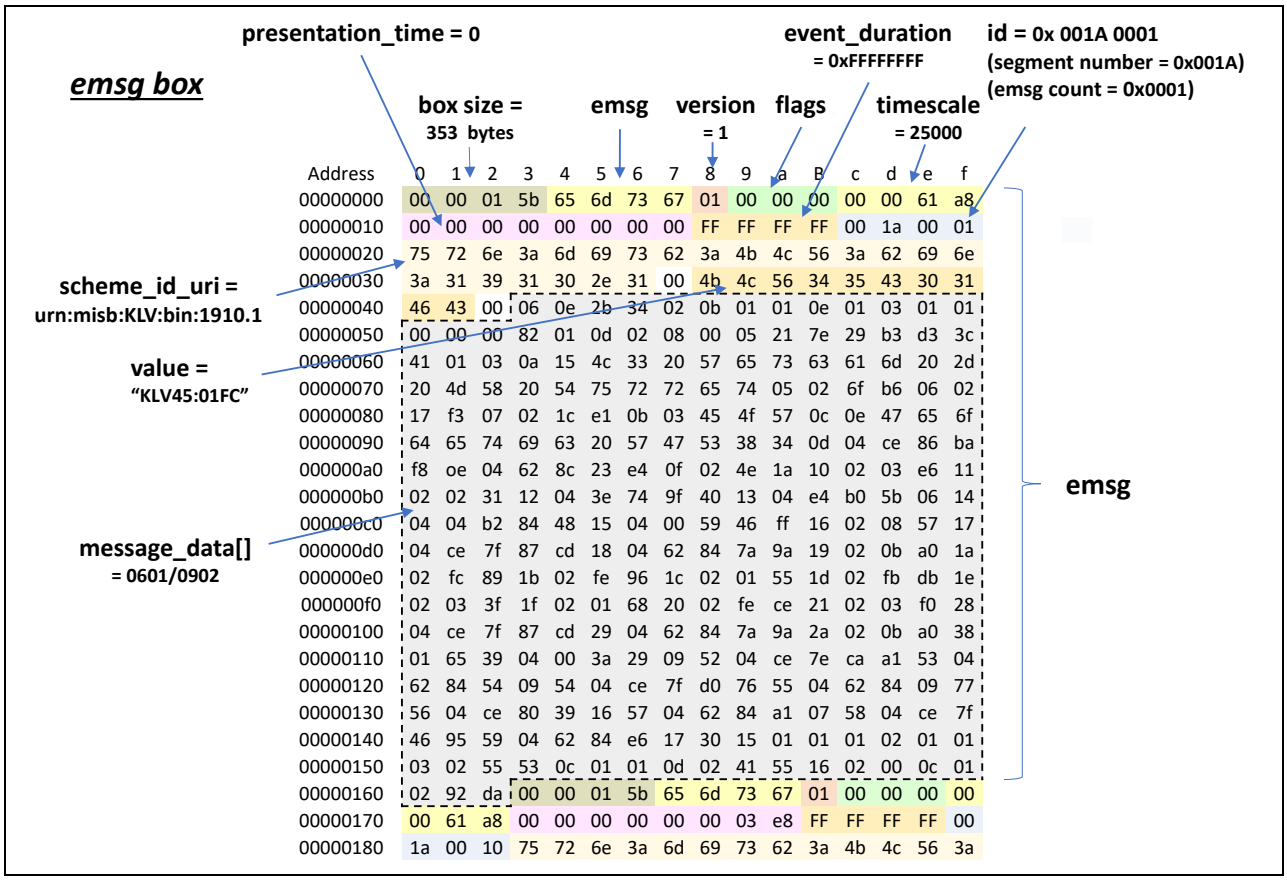

Figure 11: Sample of KLV metadata mapped to an emsg box

Appendix B MPEG-2 TS Motion Imagery/metadata Synchronization Informative

Class 1/Class 2 Motion Imagery content encapsulated within MPEG-2 Transport Steam typically carries additional content such as metadata and audio. The synchronization of metadata to a Motion Imagery frame at the source depends on how accurately the implementation inserts metadata with respect to the imagery. In an asynchronous metadata stream metadata may lose its timing with respect to the imagery because there is no facility to time the multiplexing of metadata into the transport stream – the only inherent timing is locality or proximity of the metadata to its corresponding related frame. In synchronous metadata streams the "goodness" or accuracy of the metadata to its respective image frame relies on the implementation supplying the metadata to the transport stream multiplexer at the correct time to its associated imagery.

In both cases improved accuracy of the synchronization is possible. This is the function of the post-collection remediation process described here. Note: the MISB is developing guidance for remediation so this is an informative overview only.

Figure 12 illustrates where in the workflow the remediation process occurs. A platform (e.g., airborne platform) delivers source content in a MPEG-2 TS with compressed Motion Imagery and metadata and audio to a receiver (e.g., ground station). The remediation process accepts this MPEG-2 TS and outputs a remediated MPEG-2 TS, which is an improved input to further downstream processes, such as Adaptive Bitrate (ABR) delivery or exploitation tools.

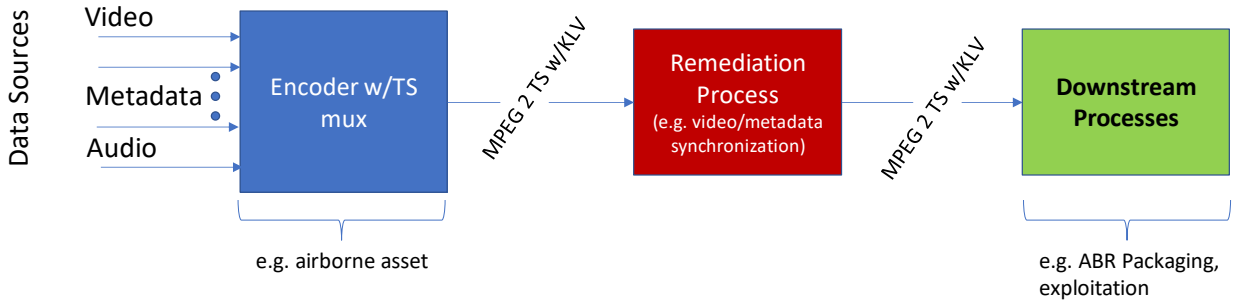

Figure 12: Remediation process of MPEG-2 TS content

B.1 Functions of the Remediation Process

Improving the time synchronization between Motion Imagery and metadata depends on whether a MISP timestamp is present in the Motion Imagery. MISB ST 0604 describes the format and the location for a MISP timestamp within the Supplemental Enhancement Information (SEI) message in compressed Motion Imagery when supplied. The first step in the remediation process is determining the presence of this timestamp; if present, accurate synchronization is possible. Leveraging the MISP timestamp in the Motion Imagery along with the MISP timestamp in the metadata enables accurate synchronization between the two.

In the event a timestamp is not present in either the Motion Imagery or the metadata, the MPEG-2 TS Presentation Time Stamp (PTS) in the header for synchronous metadata defaults as the most optimal timing available. For asynchronous metadata, the relative proximity of the

metadata to its image frame serves the synchronization function; this method produces the least accurate synchronization.

Levels of Synchronization – Timing "goodness"

The time synchronization between Motion Imagery and metadata depends on the method of metadata carriage selected and whether MISP timestamps are present in the Motion Imagery and the metadata. Three levels of synchronization assigned based on these conditions indicate to a client receiver the "goodness" of the synchronization for exploitation purposes. In legacy streams these assigned levels are absent; in these systems a user has little knowledge of the accuracy in the synchronization. A remediated stream introduces these levels as signaling carried forward for end-user awareness. The following sections describe the conditions which result in an assigned level of synchronization.

B.1.1 Motion Imagery with a MISP Timestamp

MISB mandates metadata incorporate a MISP timestamp as defined in MISB ST 0603 [17]. The MISP likewise mandates a MISP timestamp in the Motion Imagery. Because both timestamps derive from the same absolute time reference (see ST 0603), they provide a very accurate means for registering metadata to a frame of Motion Imagery when present. Unfortunately, systems built which precede these requirements do not include the timestamp in the Motion Imagery.

Case 1: Synchronous Metadata with MISP Timestamp

Synchronous metadata within a MPEG-2 TS is "synchronous" because the TS header for the metadata contains a Presentation Time Stamp, or PTS, assigned by the transport stream multiplexer much like that done for video and audio. Thus, synchronization of metadata to the imagery occurs at the input to the TS multiplexer. Assuming the implementation provides the data to the multiplexer consistent with the availability of its referenced image this is an accurate synchronization method. However, receivers of this data do not have enough information regarding the implementation to know the accuracy of the source multiplexed timing.

If the MISP timestamp is present in both the Motion Imagery and the metadata, synchronization of the two can be both guaranteed and known. With this information adjustments made to the PTS of the metadata with respect to the Motion Imagery form new inputs for re-multiplexing of the two. **This is Level 1 synchronization.** Again, this is only possible when MISP timestamps are present in the Motion Imagery.

Case 2: Asynchronous Metadata with MISP Timestamp

Asynchronous metadata carries no PTS information; multiplexing occurs in proximity to Motion Imagery when presented to the multiplexer. This produces uncertainty in the time synchronization between the imagery and the metadata. However, when the metadata and the Motion Imagery both contain a MISP timestamp, retiming is possible as in Case 1. Thus, asynchronous metadata post remediation becomes accurately synchronized metadata. **This is Level 1 synchronization**.

In both Case 1 and 2 the MISP timestamp guides increased accuracy in Motion Imagery / metadata synchronization. This is the optimal situation and one in which systems should adhere.

Case 3: Synchronous Metadata without MISP Timestamp

Assuming the metadata aligns coincident with its corresponding image frame at the multiplexer, this produces reasonably accurate synchronization. The PTS for both the Motion Imagery and the metadata provide the best timing information available. Unfortunately, without information on the implementation which constructed the stream timing, it is not possible to know the degree of accuracy. **This is Level 2 synchronization**.

Case 4: Asynchronous Metadata without MISP Timestamp

Where a MISP timestamp is not present in the metadata, remediation of the timing is not possible. **This is Level 3 synchronization**.

B.1.2 Motion Imagery without a MISP timestamp

Not all source content contains a MISP timestamp in the Motion Imagery. In these cases, remediation cannot improve the accuracy in synchronization of the two. Although remediation of the timing is not possible, the inherent timing provided by the MPEG-2 TS PTS can indicate that the Motion Imagery and metadata are in close, if not complete, alignment. In the following two cases, presence of a MISP timestamp in the metadata but not the Motion Imagery does not impact remediation, and therefore, this produces only these two cases.

Case 5: Synchronous Metadata

Assuming the metadata aligns coincident with its corresponding image frame at the multiplexer, this produces reasonably accurate synchronization. However, note this is an assumption. The PTS for both the Motion Imagery and the metadata provide the best timing information available. Unfortunately, without information on the implementation which constructed the stream timing, it is not possible to know the degree of accuracy. For this reason, the grading of the quality of synchronization is like Case 3 and is a **Level 2 synchronization**.

Case 6: Asynchronous Metadata

Without a MISP timestamp in the Motion Imagery and without the PTS synchronizing mechanism of the transport stream this situation provides the lowest level of synchronization timing – that is, a **Level 3 synchronization**.

B.1.3 Levels of timing synchronization "goodness"

Given the cases described the rating of the "goodness" or accuracy in the synchronization of metadata to Motion Imagery results in three levels: Level 1, Level 2, Level 3 listed in Table 10.

Table 10: Levels of KLV Metadata Synchronization

Case	Stream Metadata Type	MISP Timestamp		Level of Synchronization
		Motion Imagery	Metadata	
1	synchronous (PTS)	Yes	Yes	Level 1 (best)
2	asynchronous	Yes	Yes	
3	synchronous (PTS)	Yes	No	Level 2
4	asynchronous	Yes	No	Level 3

5	synchronous (PTS)	No	X	Level 2
6	asynchronous	No	X	Level 3

B.1.4 Conversion to synchronous elementary stream

In a remediated stream the Level of Synchronization in Table 10 is the value given in Table 11 into the metadata_application_format of the metadata descriptor in the Program Map Table (PMT) of the MPEG-2 TS. This signal provides an end user with information to improve their understanding of the Motion Imagery-to-metadata synchronization in the exploitation process. The levels correspond to identifiers for CMAF packaging according to Table 12.

Table 11: Signaling for "remediated" MPEG-2 TS CMAF packaging

Type of KLV Synchronization	Level of Synchronization*	metadata_application_format (MPEG-2 TS – MISB ST 1402)	stream_id (MPEG-2 TS)	emsg source-characteristic (utf8) CMAF
synchronous	Level 1	0x11FC (uint)	0xFC	11FC
synchronous	Level 2	0x12FC (uint)	0xFC	12FC
asynchronous	Level 3	N/A	0xBD	01BD
* defined in Appendix B Section B.1.3				

Table 12: Allowed KLV metadata emsg profile source-characteristic values

MISP timestamp in Motion Imagery SEI	MISP timestamp in KLV metadata	Remediated: Aligned by MISP timestamp's	Level of Synchronization	emsg source-characteristic (utf8)	Notes
yes	yes	yes	Level 1	11FC	Note 1
yes	yes	yes	Level 2	12FC	
yes	yes	no	Level 2	01FC	Note 2
no	yes	no effect	Level 3	01BD	Note 3
Note 1: Remediated MPEG-2 TS synchronous metadata Note 2: Legacy (non-remediated) MPEG-2 TS synchronous metadata Note 3: Legacy (non-remediated) MPEG-2 TS asynchronous metadata					