MISB MOTION IMAGERY STANDARDS BOARD	MISB ST 1402
STANDARD **MPEG-2 Transport of Compressed Motion Imagery and Metadata**	**27 February 2014**

1 Scope

This Standard defines requirements and provides guidance for multiplexing motion imagery and metadata within a MPEG-2 Transport Stream protocol [1].

2 References

2.1 Normative References

The following references and the references contained therein are normative.

[1] ISO/IEC 13818-1:2013, Information technology - Generic coding of moving pictures and associated audio information: Systems
[2] SMPTE RP 217:2001, Nonsynchronized Mapping of KLV Packets into MPEG-2 System Streams
[3] Motion Imagery Standards Profile 6.6, Motion Imagery Standards Board, Feb 2014
[4] ITU-T Rec. H.264 (04/2013), Advanced Video Coding for Generic Audiovisual Services
[5] MISB ST 0603.2 Common Time Reference for Digital Motion Imagery using Coordinated Universal Time (UTC), Feb 2014
[6] MISB ST 0604.3 Time Stamping Compressed Motion Imagery, Feb 2014
[7] ISO/IEC 13818-2:2000, Information technology - Generic coding of moving pictures and associated audio information: Video

2.2 Informative References

[8] SMPTE EG 40:2002, Conversion of Time Values Between SMPTE 12M Time Code, MPEG-2 PCR Time Base and Absolute Time

3 Revision History

Revision	Date	Summary of Changes
ST 1402	02/27/2014	• Initial draft; Includes only MPEG-2 transport guidance from ST 0604.2; for time stamping information see ST 0604.3 • Included material from RP 0101, ST 9701, RP 9717

4 Definitions

Asynchronous Metadata Multiplex Method: Metadata multiplexing into a MPEG-2 Transport Stream in accordance with SMPTE RP-217[2].

Synchronous Metadata Multiplex Method: Metadata multiplexing into a MPEG-2 Transport Stream in accordance with ISO/IEC 13818-1[1].

5 Acronyms

DTS Decode Time tamp
ES Elementary Stream
GOP Group of Pictures
KLV Key-Length-Value
MTU Maximum Transmission Unit
PAT Program Association Table
PES Packetized Elementary Stream
PID Program ID
PMT Program Map Table
PCR Program Clock Reference
PS Program Stream
PSI Program Specific Information
SPTS Single Program Transport Stream
TS Transport Stream

6 Introduction

The guiding principle set forth in the Motion Imagery Standards Profile (MISP) [3] is that motion imagery streams and files must be exchangeable among all systems. In order to achieve this, motion imagery must be interoperable at the file interchange level. Use of commercial standards and products affords a high degree of interoperability, where reliability and signal integrity for transmission and exchange are well understood. This Standard provides guidance to developers of digital motion imagery systems to ensure that their systems are interoperable. This document outlines the use of MPEG-2 Transport Stream at a high level; please refer to the governing ISO/IEC 13818-1 [1] for a much more complete discussion.

7 MPEG-2 Transport Stream (MPEG-2 TS)

MPEG-2 transport stream (TS) is a standard format for transmission and storage of audio, video, and data used prolifically in the commercial/broadcast communities.

MPEG-2 Transport Stream is specified by ISO/IEC 13818-1[1]. Transport stream specifies a container format for encapsulating packetized elementary streams [PES], with error correction and stream synchronization features for maintaining transmission integrity when the signal is degraded.

The MPEG-2 transport layer provides an infrastructure for the carriage of motion imagery, audio and metadata as a single motion imagery stream (see Figure 1).

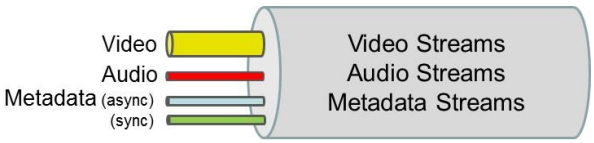

Figure 1: MPEG-2 Transport Stream

A transport stream packet consists of header fields, an optional adaptation field, and a payload field as shown in Figure 2. The header contains a sync byte that denotes the beginning of a TS packet and a Packet ID (PID) that uniquely identifies the packet. An optional adaptation field may be used to convey information such as the Program Clock Reference (PCR). The payload field contains the actual PES data.

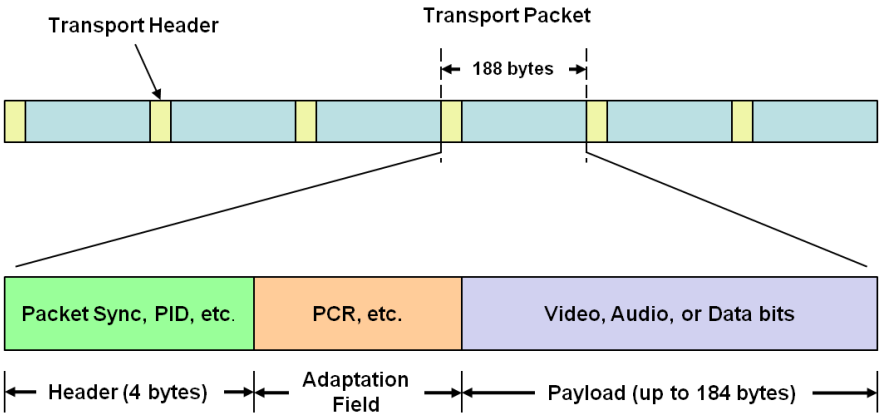

Figure 2: MPEG-2 Transport Packet

MPEG-2 TS is a preferred protocol for several reasons (Appendix A discusses Program Stream):

- The MPEG-2 TS protocol is intended for use in error-prone transmission environments for either file or streaming media applications. This does not mean they are not immune to errors, although most TS decoders are able to recover when frames are corrupted in transmission.

- Many codecs can decode the MPEG-2 TS protocol.
- An MPEG-2 TS may contain multiple video, audio, data, and metadata elementary streams together. This is a preferred method for transporting multimedia steams to ensure their time relationship to one another.
- KLV metadata is carried within a transport stream as its own elementary stream.
- KLV in a MPEG-2 TS can be ignored gracefully when a decoder cannot decode it. In other words, the presence of KLV does not break the decoding of the motion imagery.
- MPEG-2 TS is designed for real-time delivery. Developers should exercise care when editing or otherwise altering a MPEG-2 TS. "Slicing" or segmenting a TS into smaller TS streams need account for the GOP (Group of Pictures) structure.

The transport stream is composed of one or more Programs. A transport stream with a single program is a Single Program Transport Stream (SPTS). A transport Stream with more than one program is a Multiple Program Transport Stream (MPTS). Each Program is composed of one or more Elementary Streams, such as a Motion Imagery Elementary Stream, Audio Elementary Stream or Metadata Elementary Stream. Each elementary stream is assigned a unique Packet IDentifier (PID). All the PID's constituting a program are listed in a Program Map Table (PMT), which has its own assigned PID.

A receiver decoding a particular program merely locates the PMT for a program, and then uses the unique PIDs of the elementary streams within that program for decoding the payloads of each elementary stream.

It is highly recommended that the maximum integer number of TS packets encapsulated in a UDP datagram be selected to maximize throughput, minimize fragmentation, and minimize errors or losses. For example, in an IP/Ethernet network with a Maximum Transmission Unit (MTU) of 1500 bytes the recommended maximum number would be seven (7) TS packets.

Requirement	
ST 1402-01	All User Datagram Protocol (UDP) datagrams encapsulating MPEG-2 Transport Stream (TS) packets shall contain an integer number of TS packets. Each UDP datagram may contain a different integer number of TS packets.

7.1 Program Specific Information - PSI

The MPEG-2 Systems Program Specific Information (PSI) contains four tables useful for identifying where the various essence streams and timing information lie within the transport stream. Only two tables – the Program Association Table (PAT) and the Program Map Table (PMT) – are required in TS streams; the remaining tables are optional.

Table 1: PSI Tables

Program Specific Information (PSI)	Required per ISO/IEC 13181-1 [1]
Program Association Table (PAT)	Yes
Program Map Table (PMT)	Yes
Network Information Table (NIT)	No

Conditional Access Table (CAT)	No

7.1.1 Program Association Table - PAT

The Program Association Table (PAT) lists all programs available in the transport stream; it is the program guide. Each program is identified by a 16-bit value called a *program_number*. Each of the programs listed in the PAT has an associated PID for its Program Map Table (PMT) as shown in Figure 3.

7.1.2 Program Map Table - PMT

Program Map Tables (PMTs) contain information about programs. For each program, there is one PMT. The PMTs provide information on each program present in the transport stream, and lists the elementary streams that comprise the MPEG-2 program. Each elementary stream is labeled with a *stream_type* value as shown in Figure 3.

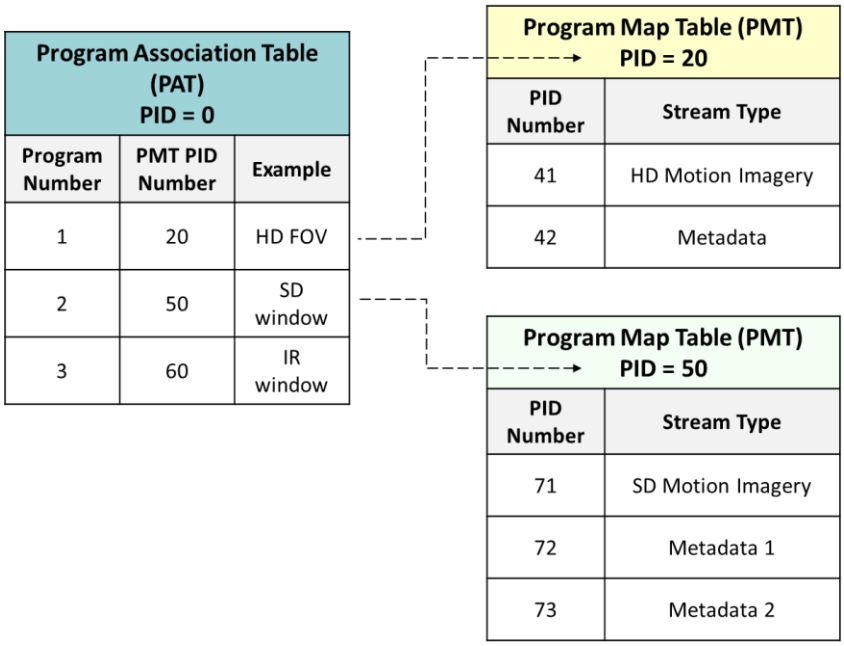

Figure 3: PAT and PMT Relationship (informative)

The Program Association Table (PAT) and the Program Map Table (PMT) are required in a MPEG-2 Transport Stream by ISO/IEC 13818-1 [1]. The following requirement was derived from community input for optimal stream usability.

Requirement	
ST 1402-02	Both a Program Association Table (PAT) and a Program Map Table (PMT) shall be inserted in the transport stream greater than 4 times per second (8 times per second recommended) throughout the program to allow rapid program acquisition.

7.2 Program Clock Reference - PCR

To synchronize content multiplexed into the transport stream a *Program Clock Reference* (PCR) is transmitted at least once every 100 milliseconds in the adaptation field of an MPEG-2 transport stream packet. The PCR is a sample of the 27-MHz System Time Clock (STC) generated at the encoder. The value of the PCR is used to regenerate the STC in the decoder.

7.3 PTS values for every video frame

Decoding Time Stamps (DTS) and Presentation Time Stamps (PTS) carried in the packetized elementary stream must be accurate as defined in the MPEG systems standard [1]. The PTS provides information necessary to synchronize the display of the packetized elementary streams (video, metadata, audio, etc.) in the transport stream.

The MPEG standard specifies that successive Presentation Time Stamp (PTS) values must not differ by more than 100 milliseconds for each Packetized Elementary Stream. Note: at frame rates below 10 FPS a PTS should be present for every frame.

Recommendation

- **That there be a one-to-one association of a PTS to each video frame. This will ensure tight synchronization between video and metadata transported in the synchronous metadata multiplex method (see Section 8.2).**

PTS values can also be used for file navigation, so access to a particular frame is possible if specified to frame increments.

7.4 PTS to PCR Relationship

The MPEG systems standard [1] specifies that all bytes for a given access unit (a coded representation of a presentation unit, such as data for a picture) must occur in the stream prior to the PCR time at which they must be decoded. Another way of saying this is that each PTS value must occur sufficiently ahead of the PCR so that all the bytes of the access unit will be available when the PCR clock reaches the PTS time. This ensures that all bytes of a video frame are available for decoding before the video frame is to be displayed. Note: when the encoder is operating in low delay mode, the buffer may underflow; this condition is further described in [1] Section 2.4.2.3.

Conversely, each video picture should not occur too far ahead of its display time, as this will contribute to increased latency.

In SMPTE 13818-1, it is required that "The delay of any data through the System Target Decoder buffers shall be less than or equal to one second except for still picture video data and ISO/IEC 14496 streams. Specifically: $tdn(j) - t(i) \leq 1$ second for all j, and all bytes i in access unit An(j)." SMPTE 13818-1 further states that "For ISO/IEC 14496 streams, the delay is constrained by $tdn(j) - t(i) \leq 10$ seconds for all j, and all bytes i in access unit An(j)." Thus, for MPEG-2 compressed motion imagery each picture must occur less than 1 second ahead of its display time, whereas for H.264 each picture must occur less than 10 seconds ahead of its display time.

The above SMPTE requirements help to prevent buffer overflow in decoders which may have limited memory resources. When the PTS value is too far head of the PCR or gets behind the PCR, the video playback may stutter, freeze, or result in poor video/audio synchronization, and can result in the inability to re-stream the content.

Recommendations

- A transport packet at the start of an H.264/AVC RAP (Random Access Point: IDR picture or I picture with recovery point SEI, SPS and PPS from which video decoding can begin successfully) should have *random_access_indicator* set to 1.
- Any error detecting devices in the transmission path set the *transport_error_indicator* bit in the transport packet when uncorrectable errors are detected. If this flag is set then the decoder can take a suitable concealment or error recovery measure.
- In support of future trick modes the *elementary_stream_priority_indicator* bit should be set whenever an access unit containing an I or IDR picture is present.

7.5 Null packets

Null packets, with a PID of 0x1FFF, are used to form a constant bit rate stream. A Null packet is ignored by a receiver.

8 Multiplexing Motion Imagery and Metadata in MPEG-2 Transport Stream

8.1 Carriage of H.264/AVC Compressed Motion Imagery

The H.264/AVC specification [4] distinguishes conceptually between a Video Coding Layer (VCL), and a Network Abstraction Layer (NAL). The VCL contains the "video" features of the codec (transform coefficients, quantization and motion information, loop filter parameters, etc.). The NAL layer formats the VCL data into Network Abstraction Layer units (NAL units) suitable for transport across the applied network or storage medium. A NAL unit consists of a one-byte header and a payload; the header indicates the type of the NAL unit and other information, such as the (potential) presence of bit errors or syntax violations in the NAL unit payload, and information regarding the relative importance of the NAL unit for the decoding process. When H.264/AVC is carried in MPEG-2 Transport Stream, NAL units are preceded by start code bytes.

8.2 Carriage of Metadata

Adding a Precision Time Stamp (see MISB ST 0603 [5] and MISB ST 0604 [6]) to motion imagery and metadata provides a way to correlate the two. The absolute time relationship between the motion imagery and the metadata is tied to the Precision Time Stamp, and is rather independent of the method selected for packaging the essence streams within an MPEG-2 transport stream. The Precision Time Stamp enables a method of time-stamping when a sensor imaged its image; the Precision Timestamp also enables a method of time-stamping when a metadata element was measured. The synchronous multiplex method only signifies when the two

different data streams were multiplexed together in the encoder. In Appendix B, the difference in synchronous and asynchronous data versus the methods to transport data synchronously and asynchronously is discussed.

Two methods for packaging metadata along with motion imagery for transport within an MPEG-2 Transport Stream are defined by ISO/IEC 13818-1[1] and SMPTE RP 217[2]; these are denoted here as the synchronous metadata multiplex method and the asynchronous metadata multiplex method, respectively.

8.2.1 Common Requirements

Requirement	
ST 1402-03	KLV metadata in a MPEG-2 Transport Stream shall be identified by the registered format_identifier 0x4B4C5641 ("KLVA").

8.2.2 Synchronous Metadata Multiplex Method

The **Synchronous Metadata Multiplex Method** ISO/IEC 13818-1[1] provides a Presentation Time Stamp (PTS) for metadata, similar to the PTS for the motion imagery. The PTS affords precise alignment of the motion imagery with the metadata at the presentation stage – such as viewing on a display.

The PTS is a relative timing reference used to provide stream-to-stream synchronization within an MPEG-2 transport stream. Its use is principally to guarantee that each stream with the same PTS is presented for display at the same time. It is not, however, intended to convey an absolute time reference; that is the function of the motion imagery and metadata Precision Time Stamp.

When implementing the **Synchronous Metadata Multiplex Method** this Standard requires the use of the method outlined in [1] Section 2.12.4 “Use of PES packets to transport metadata” for transporting metadata that is synchronized with the essence stream. This method provides a means to synchronize essence streams using the Presentation Time Stamp (PTS) located in the Packetized Elementary Stream (PES) header. The PTS is coded into the MPEG-2 Systems PES layer, and is relevant for both MPEG-2 and H.264.

When the PTS of the metadata is sampled at the same time as a PTS of the motion imagery frame the metadata and motion imagery frame will have the same PTS. If the metadata is not sampled at the same time as the motion imagery frame, it will have a different PTS, but will exist on the same timeline as the motion imagery frame—perhaps offset in time or at some non-periodic rate. It is assumed that metadata is decoded instantaneously (i.e., there is no DTS coded). If a motion imagery frame and a metadata Access Unit have the same PTS, then it is assumed that they are intended to be displayed at the same time.

ISO/IEC 13818-1 defines a metadata header structure which contains the PTS and other information. One important field is a service ID that affords the labeling of a particular piece of metadata. As additional metadata is added downstream, it can be assigned its own service ID, thereby adding flexibility for segmenting metadata based on what elements are added, what corresponding motion imagery elementary stream the metadata applies, priority of the metadata in decoding, etc.

Figure 4 shows the structure of a PES packet in the metadata bit stream. In the most common implementations, the packet payload will consist of a single metadata Access Unit or *Metadata_AU_cell*, which includes a five-byte header followed by the KLV metadata. Each metadata Access Unit may be carried in one or more Access Unit Cells.

A metadata service is defined in [1] as "a coherent set of metadata of the same format delivered to a receiver for a specific purpose… each metadata service is assumed to represent a concatenation (or collection) of metadata Access Units." Each metadata service is represented by a collection of metadata Access Units which are transported in PES packets. A metadata Access Unit Cell contains a 5-byte header (structured as in Appendix C). When transporting metadata using this service, a unique *metadata_service_id* is assigned to each service.

Multi-program Transport Streams and methods for associating metadata in one program to motion imagery in another are allowed [1]. Multi-program Transport Streams are not covered within the scope of this Standard.

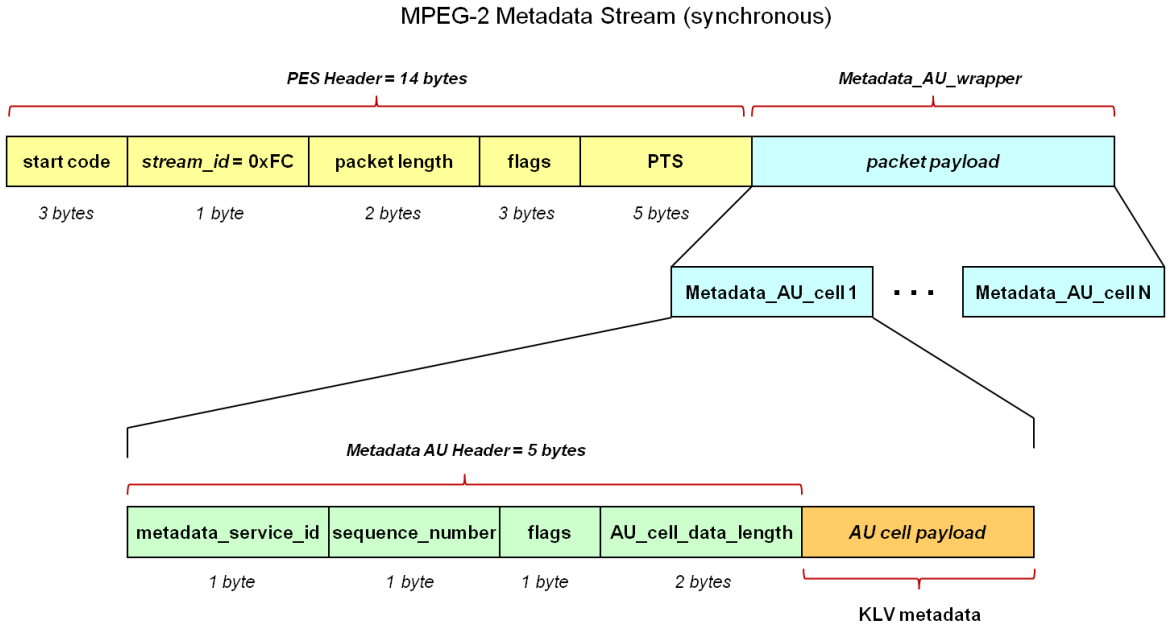

Figure 4: Synchronous Metadata Multiplex Method

Requirement	
ST 1402-04	When inserting metadata using the Synchronous Metadata Multiplex Method (MISB ST 1402), it shall be implemented in accordance with ISO13818-1 Section 2.12.4[1].
ST 1402-05	For the Synchronous Metadata Multiplex Method the stream_id shall be 0xFC, indicating "metadata stream."
ST 1402-06	For the Synchronous Metadata Multiplex Method the stream_type shall be 0x15, indicating "Metadata carried in PES packets".

ST 1402-07	For the Synchronous Metadata Multiplex Method each PES packet shall have a PTS to be used to synchronize the metadata with the motion imagery.
ST 1402-08	For the Synchronous Metadata Multiplex Method the PTS_DTS_flags shall be set to '10'.
ST 1402-09	For the Synchronous Metadata Multiplex Method the first PES packet data byte shall be the first byte of a Metadata Access Unit Cell.
ST 1402-10	For the Synchronous Metadata Multiplex Method the PTS in the PES header shall apply to each Access Unit contained in the PES packet.
ST 1402-11	For the Synchronous Metadata Multiplex Method the PTS shall signal the time that the metadata Access Unit becomes relevant.
ST 1402-12	For the Synchronous Metadata Multiplex Method the delay of any data through the System Target Decoder buffers shall be less than or equal to one second.
ST 1402-13	When inserting synchronous metadata into a transport stream which already carries synchronous metadata, new metadata shall be added to the existing synchronous metadata stream.
ST 1402-14	For the Synchronous Metadata Multiplex Method the metadata stream shall be defined in the PMT as a separate stream within the same Program as the motion imagery elementary stream.
ST 1402-15	For the Synchronous Metadata Multiplex Method the PMT shall contain a metadata_descriptor for each metadata service within the metadata stream.
ST 1402-16	For the Synchronous Metadata Multiplex Method the metadata_descriptor shall be within the descriptor loop for the metadata stream.
ST 1402-17	For the Synchronous Metadata Multiplex Method the PMT shall contain a single metadata_std_descriptor for the metadata stream.

The *metadata_descriptor* contains the *metadata_service_id* for the service it describes. The PMT may contain other descriptors such as the *content_labeling_descriptor* and the *metadata_pointer_descriptor*. Appendix B shows sample values for the *metadata_descriptor*, *metadata_std_descriptor* and *metadata_AU_cell* header fields when using constructing a synchronous metadata stream.

8.2.3 Asynchronous Metadata Multiplex Method

In the **Asynchronous Metadata Multiplex Method (**SMPTE RP 217 [2]) the metadata PES packets do not have a Presentation Time Stamps (PTS) or Metadata Access Unit construct, as shown in Figure 5. The time relationship between the metadata and a motion imagery frame is established by proximity.

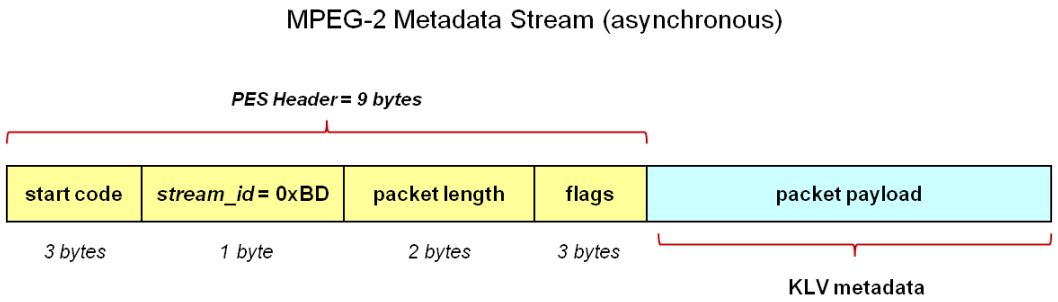

Figure 5: Asynchronous Metadata Multiplex Method

Requirement	
ST 1402-18	The transport of KLV metadata over an MPEG-2 transport stream in an asynchronous manner shall be confined to the method defined in SMPTE RP 217 Section 4.1.1 or Section 4.1.2[2].
ST 1402-19	For the Asynchronous Metadata Multiplex Method the stream_id shall be 0xBD, indicating "private_stream_1."
ST 1402-20	For the Asynchronous Metadata Multiplex Method the data_alignment_indicator shall be set to one when the PES packet contains the beginning of a KLV item.
ST 1402-21	For the Asynchronous Metadata Multiplex Method the data_alignment_indicator shall be set to zero when the PES packet does not contain the beginning of a KLV item.
ST 1402-22	For the Asynchronous Metadata Multiplex Method the PTS_DTS_flags shall be set to 00 (no PTS or DTS present in PES packet header).
ST 1402-23	For the Asynchronous Metadata Multiplex Method the stream_type shall be 0x06, indicating "PES packets containing private data."
ST 1402-24	For the Asynchronous Metadata Multiplex Method the metadata stream shall be defined in the PMT as a separate Stream within the same Program as the motion imagery elementary stream.
ST 1402-25	For the Asynchronous Metadata Multiplex Method the program element loop in the PMT shall contain a registration_descriptor as defined in ISO/IEC 13818-1[1] for legacy compliance with SMPTE RP 217[2].

Appendix B provides sample values for the *registration_descriptor* when constructing an asynchronous metadata stream.

9 Appendix A: MPEG-2 Stream Types – *Informative*

9.1 MPEG-2 Elementary Streams (ES)

An MPEG-2 Motion Imagery Elementary Stream [7] is a compressed essence bit stream. Its use in digital motion imagery applications as a stand-alone file or stream is generally discouraged for the following reasons:

- By definition, an ES is capable of carrying only one essence type (motion imagery, audio, or data). The nature of motion imagery collection, exploitation, archive, and distribution functions requires that complex multimedia contents (motion imagery, audio, data, and metadata) be carried together in one stream.
- No standards exist for the inclusion of KLV metadata into ES formats. Some systems insert KLV metadata into MPEG-2 ES private user packets, but these are not implemented uniformly and are not interoperable.
- Not all MPEG-2 decoders accept an MPEG-2 ES.
- The interchange of MPEG-2 ES between systems has proven to be problematic, especially with MPEG-2 software decoders. Because of apparent differences in codec implementations users have experienced difficulty decoding an ES created on encoders from different manufacturers. Interoperability criteria require that MPEG-2 streams created on one system shall be decodable by an MPEG-2 decoder from other manufacturers.

Developers may need to make use of MPEG-2 ES in processes that are internal to closed applications. Developers are discouraged from using the ES format for any exchange between different applications within a single system or between identical systems.

9.2 MPEG-2 Program Streams (PS)

The MPEG-2 Program Stream (PS) was originally developed to convey MPEG-2 in an error-free environment, such as a DVD, or for file storage. Program Stream is less versatile than TS. For example, it lacks the error recovery and redundancy facilities of the Transport Stream, and it cannot carry multiple programs. Program Stream is often not supported by decoders.

Program Stream is discouraged from use by MISB standards, but it may be required for STANAG compliance. If a decoder is not able to play a PS but can handle the Transport Stream, then the PS must first be "split" into TS components before decoding and playing. This is not a complex or computationally intensive operation, but must be taken into account if a system must deal with both PS and TS formats.

10 Appendix B: Metadata Descriptors – Informative

Table 2 contains sample values for the *metadata_descriptor*, *metadata_std_descriptor* and *metadata_AU_cell* header fields when using constructing a synchronous metadata stream using the **Synchronous Metadata Multiplex Method**.

Table 2: Sample Metadata Descriptors for Synchronous Metadata

	Value	No. of bits
metadata_descriptor		
descriptor_tag	0x26 (38)	8
descriptor_length	0x09 (9)	8
metadata_application_format	0x0100 - 0x0103 (see Table 4)	16
metadata_format	0xFF	8

metadata_format_identifier	0x4B4C5641 = "KLVA"	32
metadata_service_id	0x00	8
decoder_config_flags	'000'	3
DSM-CC_flag	'0'	1
reserved	'1111'	4
metadata_std_descriptor		
descriptor_tag	0x27 (39)	8
descriptor_length	0x09 (9)	8
reserved	'11'	2
metadata_input_leak_rate	(determined by encoder)	
reserved	'11'	2
metadata_buffer_size	(determined by encoder)	
reserved	'11'	2
*metadata_output_leak_rate**	(unspecified; recommend setting to 0)	
Metadata_AU_cell (5-byte header)		
metadata_service_id	0x00	8
sequence_number	(supplied by encoder; increments each cell)	8
cell_fragmentation_indication	'11', '10', '01' or '00'	2
decoder_config_flag	'0'	1
random_access_indicator	'0' or '1'	1
reserved	'1111'	4
AU_cell_data_length	(supplied by encoder)	16

*NOTE: the *metadata_output_leak_rate* is unspecified for synchronous metadata. The recommended value is 0.

Table 3 contains sample values for the *registrtion_descriptor* when using constructing an asynchronous metadata stream using the **Asynchronous Metadata Multiplex Method**.

Table 3: Sample Descriptors for an Asynchronous Metadata Stream

	Value	No. of bits
registration_descriptor		
descriptor_tag	0x05 (5)	8
descriptor_length	0x04 (4)	8
format_identifier	0x4B4C5641 = "KLVA"	32

Table 4: KLV Metadata Type

metadata_application_format (type of KLV metadata)	
0x0100 (default)	General
0x0101	Geographic Metadata
0x0102	Annotation Metadata
0x0103	Still Image on Demand