MISB MOTION IMAGERY STANDARDS BOARD	MISB ST 1402.2
STANDARD	
MPEG-2 Transport Stream for Class 1/Class 2 Motion Imagery, Audio and Metadata	**27 October 2016**

1 Scope

This standard provides guidance and specifies requirements for encapsulating Class 1/Class 2 Motion Imagery (compressed), audio (compressed) and metadata within a MPEG-2 Transport Stream container.

2 References

[1] SMPTE RP 217:2001 Nonsynchronzied Mapping of KLV Packets into MPEG-2 System Streams.

[2] ISO/IEC 13818-1:2015 Information technology - Generic coding of moving pictures and associated audio information: Systems.

[3] MISB MISP-2017.1 Motion Imagery Standards Profile, Oct 2016.

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

[5] MISB ST 0603.4 MISP Time System and Timestamps, Feb 2016.

[6] MISB ST 0604.5 Timestamps for Class 1/Class 2 Motion Imagery, Oct 2016.

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

3 Revision History

Revision	Date	Summary of Changes
ST 1402.2	10/27/2016	• Deprecated REQ's -05, -6, -19, -23; these are required by ISO/IEC 13818-1 and should not be re-specified here. • Updated References

4 Definitions

Asynchronous Metadata Multiplex Method: Asynchronous metadata multiplexing into a MPEG-2 Transport Stream in accordance with SMPTE RP-217 [1].

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

5 Acronyms

DTS	Decode Time tamp
ES	Elementary Stream
IP	Internet Protocol
KLV	Key-Length-Value
MPTS	Multi-Program Transport Stream
MTU	Maximum Transmission Unit
PAT	Program Association Table
PES	Packetized Elementary Stream
PID	Packet ID
PMT	Program Map Table
PCR	Program Clock Reference
PS	Program Stream
PSI	Program Specific Information
PTS	Presentation Time Stamp
SPTS	Single Program Transport Stream
TS	Transport Stream
UDP	User Datagram Protocol

6 Introduction

A guiding principle set forth in the Motion Imagery Standards Profile (MISP) [3] is that Motion Imagery streams and files be exchangeable among systems; that is, Motion Imagery systems need to be interoperable. 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 outlines the use of MPEG-2 Transport Stream specified by ISO/IEC 13818-1 [2] as a container for Motion Imagery data, which promotes system interoperability.

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

MPEG-2 Transport Stream (TS), as specified by ISO/IEC 13818-1, is a container format for encapsulating media data, such as Motion Imagery, audio and other data (see Figure 1). TS is widely used in the commercial/broadcast communities for media distribution. Coded media data within a TS – called an elementary stream (ES) – is formatted as a packetized elementary streams (PES). The individual PES are then multiplexed together into one unified data stream with error correction and stream synchronization information to facilitate transmission integrity.

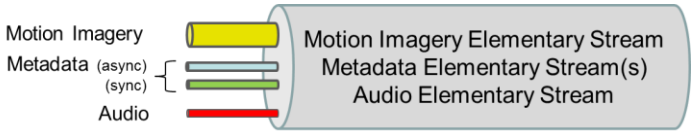

Figure 1: Example MPEG-2 Transport Stream

TS data is composed of 188 byte packets, where a packet consists of a 4-byte Header, an optional Adaptation Field, and Payload as shown in Figure 2. The header includes a Packet Sync byte, which indicates the beginning of a packet, and a Packet ID (PID), which uniquely identifies the packet. The optional adaptation field conveys information such as the Program Clock Reference (PCR). The Payload contains the data for an elementary stream.

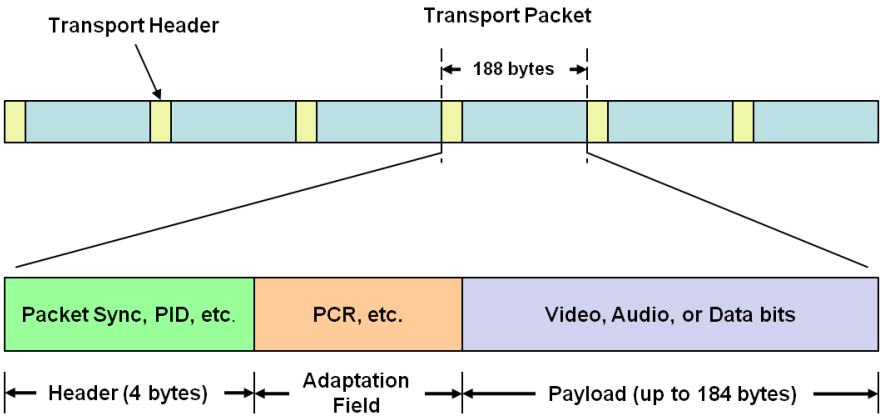

Figure 2: MPEG-2 Transport Stream Packet

MPEG-2 TS is a preferred container for several reasons (Appendix A outlines the limitations of the MPEG-2 Program Stream):

- MPEG-2 TS is designed for low latency, real-time delivery
- MPEG-2 TS is designed for delivery in error-prone transmission environments
- A ubiquitous number of receivers can decode and play a MPEG-2 TS
- A single MPEG-2 TS can carry numerous media streams of different types

A Single Program Transport Stream (SPTS) MPEG-2 TS carries one program, while a Multiple Program Transport Stream (MPTS) can carry multiple programs. Each program is composed of one or more PES containing data such as a Motion Imagery, Audio and Metadata. Each PES is assigned a unique Packet IDentifier (PID).

Several tables, collectively called the Program Specific Information (PSI) within the TS, guide a receiver in decoding the individual PES within a program.

7.1 Program Specific Information - PSI

The PSI is a collection of four tables, which provide information about the contents and use of the transport stream. Only two tables – the Program Association Table (PAT) and the Program Map Table (PMT) – are required by ISO/IEC 13818-1 in a TS stream; they are described next.

Table 1: PSI Tables

Program Specific Information (PSI)	Required per ISO/IEC 13181-1
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 PAT identifies all programs within a transport stream; it functions as the "program guide." The PAT is always assigned a PID of zero (PID = 0). Each program in the PAT is assigned a unique 16-bit program number. For each program number, there is an associated Program Map Table (PMT), which is also assigned a unique PID. As an example, shown in Figure 3, this PAT shows three programs, where the PMT for Program Number 1 is assigned the PMT PID Number = 20, Program Number 2 is assigned a PMT PID Number = 50, and Program Number 3 a PMT PID Number = 60. These PIDs are generally assigned by the TS multiplexer, and, their values may be re-assigned through subsequent stages of processing i.e. not persistent.

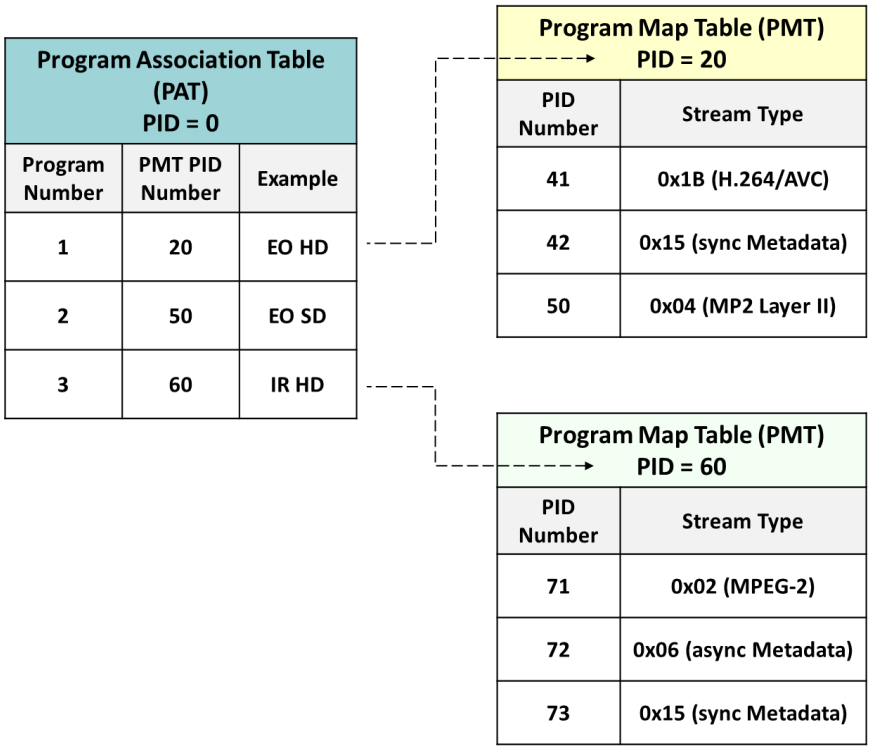

Figure 3: PAT and PMT Relationship (informative example)

7.1.2 Program Map Table - PMT

For each program in the PAT, there is a corresponding PMT. The PMT contains information about the media elements within a program, listing all elementary streams that comprise the program. Each elementary stream is labeled with a Stream Type as defined in ISO/IEC 13818-1.

Referring to Figure 3, the PMT with PID =20 (upper right) lists elementary streams for H.264/AVC (PID = 41), sync metadata (PID = 42) and MPEG-2 Layer II audio (PID = 50). Each media element identifies with a Stream Type i.e. 0x1B for H.264/AVC, 0x15 for synchronous metadata and 0x04 for audio.

The following requirement is supplied 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 facilitate rapid program acquisition.

7.2 Program Clock Reference - PCR

To synchronize content multiplexed into the transport stream a Program Clock Reference (PCR) is inserted at least once every 100 milliseconds (as required by ISO/IEC 13818-1) in the adaptation field of a TS packet. The PCR is a sample of the 27-MHz System Time Clock (STC) generated at the encoder used to regenerate the STC in the decoder.

7.3 PTS Frequency

Decoding Time Stamps (DTS) and Presentation Time Stamps (PTS) carried in a packetized elementary stream must meet accuracy requirements as defined in ISO/IEC 13181-1. The PTS is a sample of the STC coincident with a frame of Motion Imagery (or sampling of metadata), and provides timing information for synchronizing media elements at a receiver for display.

ISO/IEC 13181-1 specifies that successive PTS values must not differ by more than 0.7 seconds for each packetized elementary stream. At very low frame rates for H.264 [4], however, other Video Usability Information (VUI) parameters in the compressed stream are required by ISO/IEC 138181-1. The following MISB recommendation is intended to provide a tighter coupling between a Motion Imagery frame and metadata collected about that frame.

MISB Usability Recommendations

- That a PTS be inserted for every Motion Imagery frame, so there is a one-to-one association of a PTS to each Motion Imagery frame. This will ensure more accurate synchronization between Motion Imagery and metadata transported by the synchronous metadata multiplex method (see Section 9.3).
- At frame rates below 10 FPS a PTS should be present for every frame.

7.4 PTS to PCR Relationship

ISO/IEC 13181-1 specifies that all bytes for a given access unit (an access unit is 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 time indicated by the PTS. This ensures that all bytes of a Motion Imagery frame are available for decoding before the 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 [2] Section 2.4.2.3. Conversely, each picture should not occur too far ahead of its display time, as this will contribute to increased latency.

In ISO/IEC 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)." ISO/IEC 13818-1 further states that "For ISO/IEC 14496 [4] 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 Motion Imagery each picture must occur less than 1 second ahead of its display time; for H.264 each picture must occur less than 10 seconds ahead of its display time.

The above requirements help 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 Motion Imagery playback may stutter, freeze, or result in poor media synchronization, and can result in an inability to re-stream the content.

Usability Recommendations

- A TS packet at the start of a H.264/AVC RAP (Random Access Point: IDR picture or I picture with recovery point SEI, SPS and PPS from which Motion Imagery 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 TS 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 default PID of 0x1FFF, are used to form a constant bit rate stream. A Null packet is ignored by a receiver.

8 Delivery over Internet Protocol

A MPEG-2 TS can be transmitted in numerous ways. One common method is to encapsulate the TS into a datagram of the User Datagram Protocol (UDP). It is 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 is 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.

9 Multiplexing Motion Imagery, Audio and Metadata

9.1 Time Information

At least two types of timestamps are present in a Motion Imagery stream: the ISO/IEC 13818-1 Presentation Time Stamp (PTS) and the MISP Time System Precision Time Stamp [5]. The function of the PTS is to provide synchronization of inter-stream media elements, such as Motion Imagery, audio and metadata for display. Because these different media elements typically operate on different time scales, synchronization is maintained by each coding samples from the TS STC and into its data. This ensures that the decoder has the necessary timing information in each media stream for presentation at the intended time. The PTS has no meaning outside of the TS, and it has no inherent relation to an absolute time system (although it can derive its value from an absolute time reference, see MISB ST 0604 for more information). Once a TS is decoded, the PTS values are discarded.

The Precision Time Stamp, on the other hand, represents the sampling of an absolute time reference, and is totally independent of the PTS. In this way, the Precision Time Stamp is the only "true" indicator of wall clock time, and is persistent once a TS is de-multiplexed. This makes the Precision Time Stamp a necessary piece of information to include in Motion Imagery and metadata for analysis and exploitation. A Precision Time Stamp is mandated in Motion Imagery (see MISB ST 0604 [6] for guidance); likewise, a Precision Time Stamp is mandatory in MISP metadata.

A typical decoder does not use the Precision Time Stamp for synchronizing media for display. Thus, both types of timestamps are needed; the PTS for display/presentation purposes, and the Precision Time Stamp for persistent, absolute timestamping of events for post analysis.

9.2 Carriage of Class 1/Class 2 Motion Imagery

Class 1/Class 2 Motion Imagery represent compressed forms of Motion Imagery approved under the MISP [3]. ISO/IEC 13818-1 provides the guidance for multiplexing Class 1/Class 2 Motion Imagery. In addition to following the requirements in ISO/IEC 13818-1, the MISP mandates a Precision Time Stamp be inserted into Class 1/Class Motion Imagery; guidance for this is found in MISB ST 0604 [6].

9.3 Carriage of Audio

MISB ST 1001 [7] identifies MPEG 1 Layer II, and MPEG 2 Layer II and MPEG 2 AAC-LC as the types of audio compression approved for use by the MISP. ISO/IEC 13818-1 provides the guidance for multiplexing audio. Like the guidance for Motion Imagery in Section 7.3, ISO/IEC 13181-1 specifies that successive PTS values must not differ by more than 0.7 seconds for each packetized elementary stream. The frame size for MPEG 1 Layer II and MPEG 2 Layer II is 1152 samples, while that for MPEG 2 AAC-LC is 1024 samples.

9.4 Carriage of Metadata

Metadata is encapsulated in a MPEG-2 TS as private data. Two methods for encapsulating metadata within an MPEG-2 Transport Stream are defined by ISO/IEC 13818-1 and SMPTE RP 217, denoted here as the Synchronous Metadata Multiplex Method and the Asynchronous Metadata Multiplex Method, respectively.

9.4.1 Synchronous Metadata Multiplex Method

The **Synchronous Metadata Multiplex Method** provides a PTS for the metadata. Like the function of the PTS for Motion Imagery, the metadata PTS aides synchronizing private data (i.e. metadata) with other timed media for presentation on a display. This method is particularly useful where collected metadata describes events about a specific frame of Motion Imagery, and it is important to analysis that the metadata be displayed co-incident with that frame.

ISO/IEC 13818-1 defines a metadata header structure as shown in Figure 4, which contains the PTS and other information.

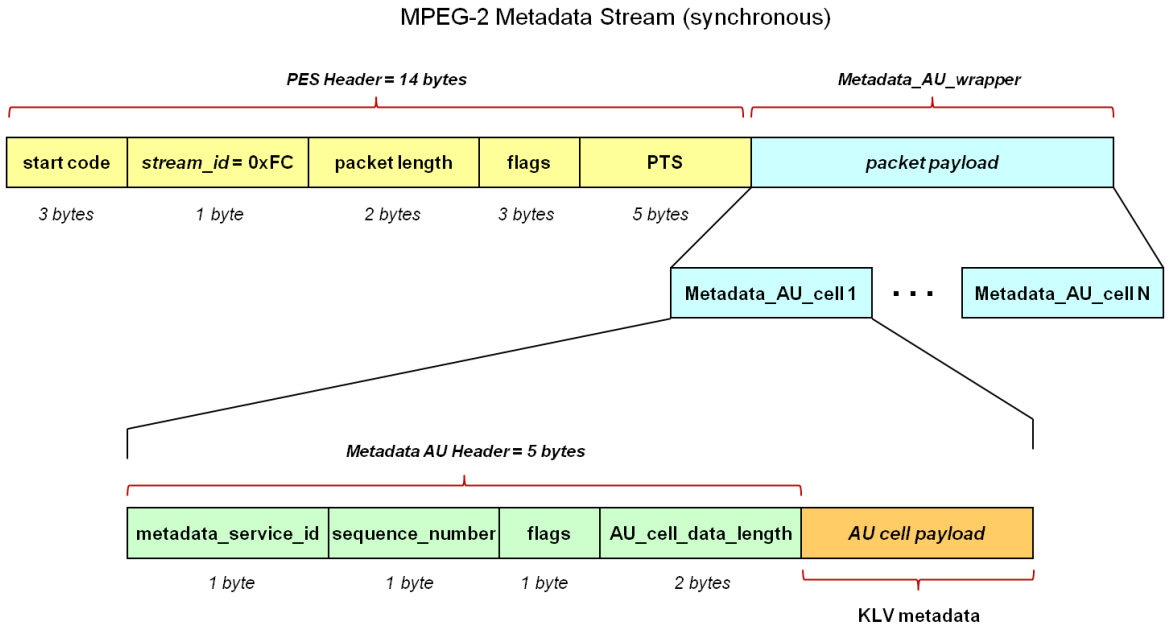

Figure 4: Synchronous Metadata Multiplex Method

Use of the Synchronous Metadata Multiplex Method is outlined in ISO/IEC 13818-1 Section 2.12.4 "Use of PES packets to transport metadata." If the PTS of the metadata is sampled at the same time as the PTS of a Motion Imagery frame, the metadata and Motion Imagery frame will have the same PTS; otherwise, they will differ by some time offset. System implementers need to account for this offset in displaying the data, otherwise errors can be made in interpreting metadata for a Motion Imagery frame. It is assumed that metadata is decoded instantaneously (i.e., there is no decoding time stamp (DTS) in this method).

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 metadata (i.e. the AU cell payload). Appendix B shows the field allocation of the Metadata_AU_cell.

A metadata service is defined in ISO/IEC 13818-1 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." When transporting metadata using this service, a unique metadata_service_id is assigned to each service.

The metadata_descriptor contains the metadata_service_id for the service it describes. The PMT may contain other descriptors. Appendix B shows sample values for the metadata_descriptor, metadata_std_descriptor and metadata_AU_cell header fields.

ISO/IEC 13818-1 Table-34 defines a stream_type = 0x15 for "Metadata carried in PES packets," and Table 2-22 defines a stream_id = 0xFC for "metadata stream."

The following requirements apply to the Synchronous Metadata Multiplex Method:

Requirement(s)	
ST 1402-04	When inserting metadata using the Synchronous Metadata Multiplex Method, it shall be implemented in accordance with ISO13818-1.
ST 1402-07	When implementing 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	When implementing the Synchronous Metadata Multiplex Method, the PTS_DTS_flags shall be set to '10'.
ST 1402-09	When implementing 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	When implementing 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	When implementing the Synchronous Metadata Multiplex Method, the PTS shall signal the time that the metadata Access Unit becomes relevant.
ST 1402-12	When implementing 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 elementary stream.
ST 1402-14	When implementing the Synchronous Metadata Multiplex Method, the metadata elementary stream shall be defined in the PMT as a separate stream within the same program as the Motion Imagery elementary stream.

ST 1402-15	When implementing the Synchronous Metadata Multiplex Method, the PMT shall contain a metadata_descriptor for each metadata service within the metadata stream.
ST 1402-16	When implementing the Synchronous Metadata Multiplex Method, the metadata_descriptor shall be within the descriptor loop for the metadata elementary stream.
ST 1402-17	When implementing the Synchronous Metadata Multiplex Method, the PMT shall contain a single metadata_std_descriptor for the metadata stream.
ST 1402.1-26	When implementing the Synchronous Metadata Multiplex Method, KLV metadata shall be identified by the registered metadata_format_identifier 0x4B4C5641 ("KLVA").

9.4.2 Asynchronous Metadata Multiplex Method

In the **Asynchronous Metadata Multiplex Method** (reference SMPTE RP 217) the metadata PES packets do not carry a Presentation Time Stamp (PTS), nor is there a Metadata Access Unit construct (see Figure 5). The time relationship between a Motion Imagery frame and metadata is established by proximity to when metadata is multiplexed into a TS, although this relationship cannot be guaranteed throughout decoding. Some receivers may decode/play asynchronous data as received, others may delay its output, and the time between decoded data may vary. In other words, there is no assurance that the metadata will be decoded/presented as inserted into the TS.

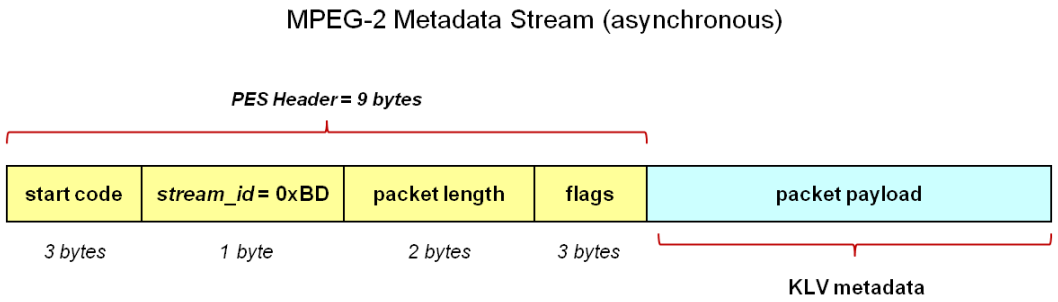

Figure 5: Asynchronous Metadata Multiplex Method

In ISO/IEC 13818-1, Table-34 defines a stream_type = 0x06 for "PES packets containing private data," and Table 2-22 defines a stream_id = 0xBD for "private_stream_1."

The following requirements apply to the Asynchronous Metadata Multiplex Method:

Requirement(s)	
ST 1402-18	When inserting metadata using the Asynchronous Metadata Multiplex Method (MISB ST 1402), it shall be implemented in accordance with SMPTE RP 217.
ST 1402-20	When implementing 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	When implementing 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	When implementing the Asynchronous Metadata Multiplex Method, the PTS_DTS_flags shall be set to 00 (no PTS or DTS present in PES packet header).
ST 1402-24	When implementing the Asynchronous Metadata Multiplex Method, the metadata elementary stream shall be defined in the PMT as a separate stream within the same program as the Motion Imagery elementary stream.
ST 1402-25	When implementing the Asynchronous Metadata Multiplex Method, the program element loop in the PMT shall contain a registration_descriptor as defined in ISO/IEC 13818-1 for legacy compliance with SMPTE RP 217.
ST 1402-03	When implementing the Asynchronous Metadata Multiplex Method, KLV metadata shall be identified by the registered format_identifier 0x4B4C5641 ("KLVA").

Appendix B provides sample values for the registration_descriptor.

Requirement(s)	
ST 1402-05 (Deprecated)	For the Synchronous Metadata Multiplex Method, the stream_id shall be 0xFC, indicating “metadata stream.”
ST 1402-06 (Deprecated)	For the Synchronous Metadata Multiplex Method, the stream_type shall be 0x15, indicating “Metadata carried in PES packets”.
ST 1402-19 (Deprecated)	For the Asynchronous Metadata Multiplex Method, the stream_id shall be 0xBD, indicating “private_stream_1.”
ST 1402-23 (Deprecated)	For the Asynchronous Metadata Multiplex Method, the stream_type shall be 0x06, indicating “PES packets containing private data.”

10 Appendix A: MPEG-2 Program Stream (PS) – Informative

The MPEG-2 Program Stream (PS) was developed to store/play MPEG-2 compressed files in an error-free environment. Program Stream is less versatile than Transport Stream. For example, PS lacks the error recovery and redundancy facilities of the TS, and it cannot carry more than one program. Program Stream is often not supported by decoders.

Program Stream is discouraged from use, but may legacy systems may have produced PS files. If a decoder is not able to play a PS but can handle the Transport Stream, then the PS will need to be converted to a TS. This is not a complex or computationally intensive operation.

11 Appendix B: Metadata Descriptors – Informative

Table 2 shows sample values for the metadata_descriptor, metadata_std_descriptor and metadata_AU_cell header fields when constructing a synchronous metadata elementary 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 shows sample values for the registrtion_descriptor when constructing an asynchronous elementary 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 shows values (as defined by MISB) for the metadata_application_format.

Table 4: KLV Metadata Type

metadata_application_format (type of KLV metadata)	
0x0100 (Default value only allowed)	General
0x0101 (deprecated)	Geographic Metadata
0x0102 (deprecated)	Annotation Metadata
0x0103 (deprecated)	Still Image on Demand