



The attached SMPTE Engineering Document has been declared “Stable” by the controlling Technology Committee.

The SMPTE Operations Manual for Standards states:

A document should be stabilized if it is believed to be substantially correct, does not contain harmful or misleading recommendations, may still be relevant to equipment or practices in use, is stable, but does not represent current technology, and need not be subject to future reviews.

A Stable document shall still be made available and offered for sale by the Society, but it shall be prefaced by a cover page explaining its current status.

At any time, a Technology Committee may revise, amend, or otherwise initiate a new Project on a Stable document.

A Stable document is “In Force”, and not deprecated or withdrawn.

*** * * * ***

Note:

SMPTE “Stable” documents were previously described as “Archived” and the attached document may be marked as “Archived”. The status of a SMPTE document described as “Archived” is exactly as described above for a “Stable” document.

Stable documents may not adhere to the latest style and format of SMPTE documents, or to current usage of normative language. Suitable care should be taken in interpretation.

SMPTE STANDARD

SMPTE 381-1-2005

Previously numbered SMPTE 381M-2005

for Television —

Material Exchange Format (MXF) — Mapping MPEG Streams into the MXF Generic Container



Page 1 of 33 pages

Table of contents

1	Scope
2	Normative references
3	Glossary of acronyms, terms and data types
4	Introduction
5	Mapping MPEG into the generic container
6	KLV coding of MPEG elements
7	SMPTE label for essence container identification
8	Essence descriptors for MPEG mappings
Annex A	Index tables for MPEG mappings
Annex B	Identifying MPEG picture types
Annex C	Requirements for mapping MPEG into MXF
Annex D	Guidance for mapping new or non-ISO defined essence in MPEG streams into MXF
Annex E	Bibliography

1 Scope

This standard specifies the mapping of MPEG streams, as identified by an ISO 13818-1 stream_id value, into the MXF generic container. This includes, but is not limited to, ISO 13818-2 MPEG video stream, ISO 11172-2 MPEG video streams, ISO 13818-3 MPEG audio streams, and ISO 11172-3 MPEG audio streams. Provision is made for mapping the content with or without the ISO 13818-1 PES layer. Provision is also made for single stream mapping and synchronized multi-stream mapping. The document outlines synchronization requirements for the MXF encapsulation of co-timed MPEG essence streams. This standard defines SMPTE universal labels to be used to uniquely identify specific MPEG implementations.

This standard also provides for clip wrapping a single program MPEG program stream or transport stream multiplex. Support for multi-program MPEG multiplexes is not provided in this standard. Support for transport streams where some of the transport stream packets have been removed is not provided in this standard. This standard does not map MPEG PSI or DVB-SI or ATSC-PSIP to any metadata scheme.

In order to achieve interoperability within any given operational pattern, restrictions may be placed on the way in which this essence container can be implemented. The reader is advised to carefully study the appropriate operational pattern document before implementation.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this standard. At the time of publication, the editions indicated were valid. All standards are subject to revision and

parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below.

SMPTE 336M-2001, Television — Data Encoding Protocol using Key-Length-Value

SMPTE 377M-2004, Television — Material Exchange Format (MXF) — File Format Specification

SMPTE 379M-2004, Television — Material Exchange Format (MXF) — Generic Container

SMPTE 385M-2004, Television — Material Exchange Format (MXF) — Mapping SDTI-CP Essence and Metadata into the MXF Generic Container

SMPTE RP 210, Metadata Dictionary Registry of Metadata Element Descriptions

SMPTE RP 224, SMPTE Labels Registry

ISO 13818 – Parts 1, 2, 3, 7, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems, Video, Audio, Advanced Audio Coding (AAC)

3 Glossary of acronyms, terms and data types

The full glossary of terms and acronyms used in the MXF specification is given in the MXF file format specification. Additional terms are given in the glossaries in SMPTE 379M and ISO 13818 (the MPEG2 specification). Terms are not repeated here to avoid any divergence of meaning.

AFD: Active format descriptor, defined by ETSI ETR154 – a parameter which defines the aspect ratio of the active content within the MPEG coded frame; e.g., 4:3 content within a coded 16:9 rectangle where black bars left and right are part of the MPEG coded signal.

BWF: Broadcast wave audio format (see SMPTE 382M in annex E)

ES: Elementary stream

GOP: Group of pictures

SPTS: Single program transport stream

PES: Packetized elementary stream

PMT: Program map table

PSM: Program stream map

VCD: Video CD

VOB: Video object (DVD term)

4 Introduction

In this introduction, it is assumed that a general understanding of MPEG video compression and the meaning of acronyms such as GOP, B-frame, and PES are understood. These acronyms are well defined in the normative references and will not be repeated here.

4.1 Basics of wrapping: Frames and access units (informative)

The MPEG video specifications enable the compression of picture-based video. These pictures may be frames or fields. The MPEG specification provides signaling for the number of pixels in the picture, the field / frame rate and other parameters including the aspect ratio of the pixels. Many of these parameters will be

copied into the MXF essence descriptor. In order to ensure interoperability, it is important to ensure that MPEG header information can be found when decoding a KLV wrapped MPEG file. For this reason, the text in section 5 refers to access units (AUs) rather than pictures. The full definition of an access unit (AU) is given in ISO 13818-1 and is summarized below for MPEG video and MPEG audio:

Audio: 1 AU = The coded representation of an audio frame

Video: 1 AU = The coded data for a picture and any stuffing that follows it, up to but not including the start of the next access unit. If a picture is *not* preceded by a `group_start_code` or a `sequence_header_code`, the access unit begins with a `picture_start_code`. If a picture is preceded by a `group_start_code` or a `sequence_header_code`, the access unit begins with the first byte of these start codes. If it is the last picture preceding a `sequence_end_code` in the bit stream all bytes between the last bytes of the coded picture and the `sequence_end_code` (including the `sequence_end_code`) belong to the access unit.

4.2 Timeline of the MXF tracks representing MPEG content (informative)

Many MXF files containing MPEG content will contain a picture track for the MPEG video data and sound tracks for any audio data. Many professional applications will use uncompressed audio which will have a simple timing relationship with the MPEG video. When an MPEG transmission or distribution feed is captured, the MXF representation may be more complicated. There is no intention within this mapping document to be able to exactly recreate a program stream or transport stream from its demultiplexed MXF components.

4.2.1 Creating an MXF file from an MPEG transport stream or an MPEG program stream

Consider an MPEG transport stream which contains compressed audio tracks and compressed video with other PES streams such as subtitles. It is likely that these PES streams may be bursty in nature; i.e., there may be many minutes of picture and sound with no other PES content present in the stream. These same issues also apply to program streams.

Capture option 1: Capture the entire transport stream or program stream and wrap in a single KLV packet (as outlined in section 6.3.5). This may be optimal when the stream is likely to be processed or used as a transport stream or program stream, respectively.

Capture option 2: Capture a partial transport stream (i.e., a stream where elements within the selected duration of the original stream have been removed) or a partial stream program stream (i.e., where elements within the selected duration of the original stream have been removed). These scenarios are outside the scope of this standard.

Capture option 3: Demultiplex the components of the captured transport stream or program stream into its elementary streams or PES streams and maintain the relative timing information by creating a valid MXF header on the fly. There are a few issues here:

1. In general the desired start frame is not the first presentation frame in a closed GOP sequence. The first "I" frame or sufficient frames before the desired frame (in the case where there are no I frames) must be captured to pre-charge the decoder.
2. The compressed audio AU duration is rarely the same as the video AU duration. To capture without recoding the audio requires calculating the temporal audio-video offset by inspection of PTS values and using these to create the MXF tracks with appropriate origin values.
3. PES data elements may have AU durations unrelated to the audio or video AUs. For example data such as ETSI EN300374 subtitle bitmaps may be delivered in the stream well before they are required to be displayed. A capture device may not know that the data is subtitles, only that the PTS is some time in the distant future. The MXF track which defines the subtitle shall be able to offset the origin of the track by a sufficiently large value to correct the timing.

4. Any PES data elements must be preserved with their PES headers which may contain extra information which is unknown to the capture device.
5. When demultiplexing takes place, it is assumed that preservation of the MPEG timing is sufficient and that preservation of the data in a form which allows streaming with the small MPEG buffers is not required. There is no buffer model for MXF streaming and, as a result, the streaming of the data in MXF form may have a higher latency than the streaming as a transport stream or program stream. It also means that re-creation of the transport stream or program stream will require an MPEG multiplexer with knowledge of the buffer models of all the tracks (possibly even the private PES ones) in the stream. The exact relative timing relationship between the ES and PES components of the program stream or transport stream depends on the original time base of the video stream, the sampling rate of the audio, the compression scheme used by the audio, and the PES strategy employed by the original multiplexer. These factors allow the relative timing relationships between the MXF tracks to be defined.

4.3 MXF tracks (normative)

In order to be able to represent the captured MPEG data in an MXF file, it must be possible to describe the start point of the track data, as well as the "in" and "out" points of the desired picture, sound, and PES data. This is shown in figure 1.

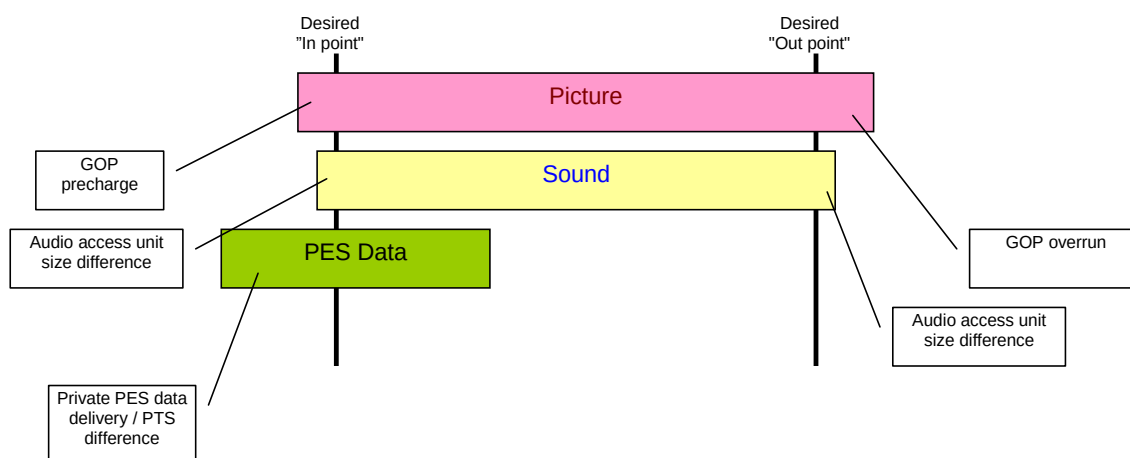


Figure 1 – MXF track timing for MPEG generic containers

The example shown in figure 1 shows the relationship between the desired start and end points of a stored generalized MPEG segment and the actual data start and end points which would be found in the stored essence. The track and sequence items within the header must describe accurately the synchronization scenario above.

The picture track shows that the desired start frame does not coincide with the first frame of an MPEG random access point. There is picture data between the first byte of stored data and the first byte of the desired picture information. In the MXF picture track, there shall be a positive **origin** value which shall correspond to the difference in **position** of the first picture stored in the generic container and the **position** of the desired start picture of the essence.

Caution is advised, however, when higher operational patterns are used with the MPEG generic container mapping. In an OP1a file, the material package shall have the same start position and duration as the top-level file package. Unwanted frames which were captured to allow decoding of the desired frames are, therefore, hidden using the mechanism described above. This is still a valid OP1a file.

In an operational pattern 3x file, editing is allowed; i.e., the material package may reference any portion of a top-level file package track. In this case, the desired start point of the file package track may be the very first image in the container, even when it is not decodable. The desired start position of the material package track may then be offset relative to the start of the top-level file package track so that a decodable portion of the essence is referenced.

If an MXF encoder captures an MPEG long GOP stream and is *unable* to determine the correct value of **origin**, then a decoder shall assume that the desired start point of the track is the first frame of the track. The first frame of the track may not be decodable and/or displayable. An MXF decoder should exhibit some consistent behavior under these circumstances; e.g., by displaying black frames until a decodable frame is reached.

The audio track shows that the decoded audio sample which should start the captured clip does not coincide with the first sample of the MPEG coded audio frame in which it lies. Compressed audio frames shall be decoded in order to obtain the desired audio sample within them. In the MXF sound track, there shall be a positive **origin** value which shall correspond to the difference in **position** of the first stored audio sample in the compressed audio frame and the **position** of the desired audio start sample. To achieve this, the **edit rate** of the sound track (e.g., 1/48 kHz) may be different to the **edit rate** of the picture track (e.g., 1/29.97 Hz). It is important to note that in order to construct an index table, all the indexed tracks shall have the same edit rate as described in the MXF format document.

The PES data track shows that there may be PES data in the file which is delivered a long time before it needs to be applied. The origin and edit rate of the data track shall be set to ensure that the data is delivered synchronously with the picture and sound information which had the same PTS values in the original MPEG stream.

It is the intention of this standard that the above scenarios can be calculated by using the PTS/DTS values in each captured stream.

4.4 External audio (informative)

In many applications involving MPEG video MXF files, the audio may be located in an external file as described in SMPTE EG 41. When the audio is organized in large chunks / frames, then the timing guidelines above must still be respected. Note that when the audio is external, there will be at least two file packages.

4.5 Wrapping options (normative)

4.5.1 Overview

The MXF generic container consists of one or more contiguous KLV wrapped content packages. The goal is that each content package should have the same duration and should contain the picture and sound data for that portion of the timeline. When long GOP MPEG video is used, this becomes difficult to achieve.

If we choose to wrap every access unit of long GOP MPEG video then we discover that the frames are being wrapped in transmission order rather than display order. If we interleave uncompressed audio sample in the same content package, we will not be re-ordering the audio samples to match the video. In addition, if the video is at a frame rate of 29.97 fps and the audio is sampled at 48 kHz then there will be a varying number of audio samples in the content packages as a result of the complex relationship of the clocks.

In general, wrapping modes are defined where one of the elements is deemed to be the "master" and all the other elements in the generic container are "slaves". The most common "master" will be the picture element.

NOTE – It is anticipated that frame wrapped long GOP MPEG ES with the picture element as "master" will be the most used wrapping mode.

4.5.2 Missing and variable elements (informative)

There will be occasions when a wrapping algorithm will not require data from a particular element. This may occur when, for example, there are no private PES packets for this content package. Under these circumstances, the element is missing from the content package.

SMPTE 379M states that each content package in an essence container should have a constant number of elements.

Elements within a content package may vary in size from one content package to another. These variations may be quite large and in the limit, the data for an element may not be present at all — the element may have zero length.

4.5.3 Converting a TS / PS multiplex to an MXF KLV multiplex

MPEG TS and PS that contain audio packetized at a different rate from video have already made a determination of how to interleave audio and video ES packets. Altering this interleaving should only be done if the resulting stream still maintains the integrity of the original MPEG buffers.

When encoding an MPEG PS or other previously multiplexed stream of PES or ES packets in an MXF generic container, packets shall be assigned to generic container content packages in the order they are encountered. A new content package shall commence with each video picture. For each other stream type, all packets encountered in the MPEG stream before the next video picture shall be grouped into a single GC element of that type, whose duration shall be the sum of the durations of the individual packets.

If no packets of a given stream type are encountered, a zero duration GC element shall be encoded in order to meet the MXF encoder guidelines in the generic container document.

4.5.4 Finding the content (informative)

When an application needs to find MPEG content on the timeline, it needs to search through the various header metadata items to locate the essence. The normative relationships between the parameters are defined in the MXF format document and the text here is an informative algorithm to locate the content:

- The MXF application has a material package picture track with a sequence of SourceClips;
- The desired content occurs during one of these SourceClips;
- From the SourceClip, it knows which position on which track of which file package it is looking for;
- From the file package, it gets the PackageID;
- From the MultipleDescriptor in the FilePackage, it gets the EssenceContainerUL, which it finds to be the UL of a generic container;
- From the track, it selects the FileDescriptor within the MultipleDescriptor for the track of interest, and it gets the EssenceContainerUL — which will be used by the essence codec;
- The FileDescriptor may have TextLocators or NetworkLocators which may indicate external essence;
- It then searches the ContentStorage::EssenceContainerData property for an EssenceContainerData set with the correct PackageID;
- From the EssenceContainerData set, it gets the IndexSID and BodySID;
- It then scans the partition packs in the file, to select partitions with the correct IndexSID;
- It then searches the index table for the desired position, then the slice and delta for the desired track, and obtains a byte offset into the essence data (in partitions of the correct BodySID);
- It then scans the partition packs in the file, to select partitions with the correct BodySID;
- It uses the BodyOffset in the partition pack to find the correct body partition;
- Within these partitions, it looks only in KLV packets with the correct essence element key;
- Bytes 13,14,15 and 16 of the key will match the EssenceTrackNumber item in the track ID;
- The contents of the KLV packets are then parsed to find the right content start point (depending on the wrapping method used).

4.5.5 MPEG in the generic container (informative)

From the generic container document: *"The premise for the MXF generic container format is that of a general purpose essence data and metadata container for the containment of many different kinds of essence and metadata elements into a single entity by interleaving the data streams in a defined and time-synchronous manner (typically over a 1-frame duration)."*

The figure below shows the use of a generic container to contain an MPEG stream. In general, the generic container was designed to create a streamable format where each repetition of the interleaved elements represents a single field or frame. In many MPEG applications this works well. These are usually the ones where there is information in each item every frame.

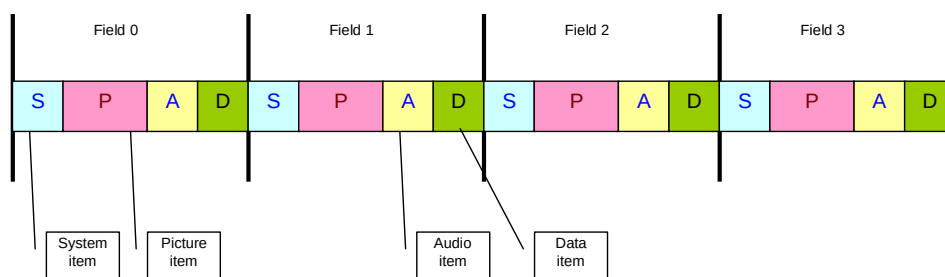


Figure 2 – MPEG in a generic container

Figure 2 shows an example of the interleave obtained by using the generic container. The picture implies that there is one field of picture in every content package, and that is associated with a system item, some audio information and some data. This follows the generic container policy that the audio and data are co-timed with the video information in the same content package.

Guidelines are required to ensure interoperability when the audio AU duration is not the same as the video AU duration. For example, when 48-kHz audio is coded with the MPEG-1 layer II algorithm, the AU size is 24 ms. The audio and video AUs will be synchronized every:

25 Hz:	every 3 picture frames (5 audio AU)
30 Hz:	every 18 picture frames (25 audio AU)
29.97 Hz:	every 32 seconds approx.

This gives rise to the situation where different content packages will contain different durations or different numbers of sample of the audio and data information. It is assumed that the video information is used as the "master" for creating content packages. This will be true for most cases, but it is important that MXF decoders follow MXF and KLV rules rather than make assumptions on how an MXF encoder might have behaved.

4.5.6 Interleaving guidelines (informative)

In order to give MXF encoders guidance for creating consistent generic containers between applications, interleaving guidelines are defined in section 5.1.2. The principle of creating the interleaving guidelines is to allow the creation of MXF files containing a mixture of unspecified components. For example:

- Long GOP MPEG ES + BWF sound;
- Long GOP MPEG PES + BWF sound;
- Uncompressed pictures + long GOP MPEG proxy + AES sound.

Different mapping mechanisms are listed in this and other mapping documents. Specific interleaving guidelines for implementing those mapping schemes are provided to improve interoperability. It should be

remembered that in addition to these low level interleaving guidelines, there may be higher level partition multiplexing rules operating which may be determined by the operational pattern of the file.

4.5.7 Multi-channel audio

There will frequently be multi-channel audio associated with long GOP MPEG pictures. This multi-channel audio may be coded as a single stream (e.g., 5.1 AC-3 encoding) or may be coded as 2 or more separate streams.

It is recommended that each separable audio stream be mapped into its own GC sound element, each of which follows the appropriate mapping and interleaving rules given elsewhere in this standard.

4.5.8 Picture identification

The MXF index table mechanism provides a deterministic method of identifying the key frames and dependencies of each indexed frame within the file. The user recommendations for the design of the MXF specification stated that there should always be index tables present in the file. The index tables are the primary mechanism for determining key frames and dependencies. There have been additional requests to identify the picture type within the KLV wrapping. This would allow easy building of index tables on the fly, all capture devices shall be able to parse the video stream to create this signaling. An MPEG video descriptor is defined in this document which gives parameters which allow fixed picture patterns to be identified.

If frame wrapping is used, then it is a simple procedure to determine these values. The value of the KLV triplet starts with the MPEG start_code of the video access unit. These are rapidly parsed to determine the picture type, coded_frame and AFD.

4.5.9 MPEG metadata

The MPEG descriptor detailed in section 8.1 provides a mechanism for mapping some MPEG descriptors to MXF structural metadata.

It is recommended that if time code is present in the MPEG GOP headers or has been inserted in accordance with SMPTE 328M then a time code track should be present in the MXF header metadata which accurately reflects the time code information found in the stream.

5 Mapping MPEG into the generic container

Three types of KLV wrapping are specified for the MPEG body structure:

- Frame wrapping;
- Clip wrapping;
- Custom wrapping.

A sequence of images shall be KLV coded as defined in SMPTE 336M. If this mapping specification is to be used with an MPEG multiplex (e.g., program stream or transport stream), then the multiplex shall be split into its component parts prior to wrapping. The splitting options are as follows:

- Picture Data → Demultiplex to ISO 13818-1 PES stream or demultiplex and remove PES headers to obtain elementary stream as appropriate;
- Sound Data → Demultiplex to ISO 13818-1 PES stream, demultiplex and remove PES headers to obtain elementary stream, B-Wav compatible stream, or other generic container compatible mapping as appropriate;
- Other Data → Demultiplex to ISO 13818-1 PES stream, demultiplex and remove PES headers to obtain elementary stream, or other generic container compatible mapping;
- Once demultiplexing has taken place, one of the KLV wrapping options below may be chosen.

5.1 Frame (picture access unit) wrapping

The "frame wrapping" methods for MPEG are shown in figures 3 and 4.

Figure 3 shows the whole of the image wrapped in a number of content packages with no other generic container elements in the container. Each content package has the duration of one MPEG video access unit. Note that this determines where the MPEG headers will be found (see section 4).

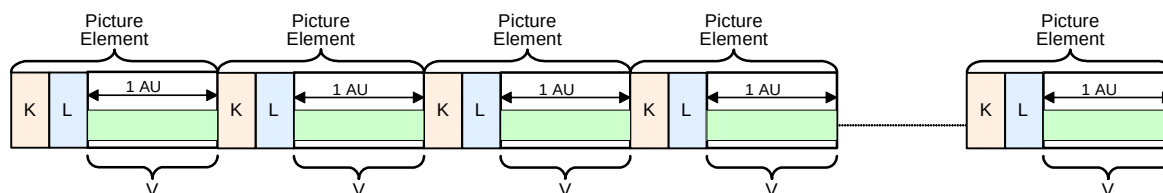


Figure 3 – Simple representation of frame wrapping

The frame wrapping method is intended to enable frame-by-frame access by MXF applications which process at the KLV level. This can be particularly useful for applications which support multiple generic container mapping types. Sufficient information is provided to allow individual frames to be identified at the KLV level without an MXF decoder having to parse or decode the essence data. Each frame of MPEG image data is KLV wrapped using a GC picture element key.

In some applications, the frame wrapped MPEG picture data will exist in content packages with other elements such as sound and data elements. An example generic container is shown in figure 4.

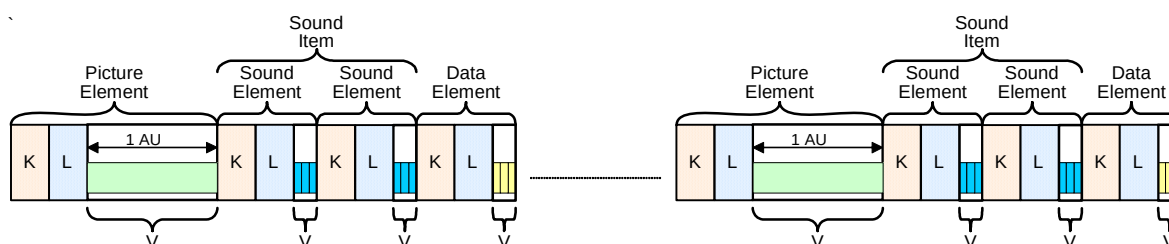


Figure 4 – Frame wrapping with other GC elements

The generic container mapping specifications for the sound and data elements will detail the key values and format of the data within the elements. Note that in this wrapping mode, the sound and data elements shall follow the guidance given in SMPTE 379M, section 5.4, concerning placement of the synchronized samples in essence containers. There are many situations where each wrapped frame of MPEG picture data will not exactly coincide with the sound/data elements within the same content package (e.g., a subtitle PES stream may have a duration of many seconds, but the bytes for the subtitle are included in a single data element in a single content package).

The sound/data elements may be PES wrapped, which will guarantee the synchronization. The MXF picture sound and data descriptors, along with the track definitions shall provide information to ascertain whether the picture/sound/data information is locked together. This specification allows a content package to be constructed from any MXF defined elements providing the generic container rules are respected.

5.1.1 Interleaving guidelines

An MXF encoder has many choices as to where data should be placed in a frame wrapped MPEG multiplex. Consistent rules will improve interoperability, especially when the relationship between the different elements is complex. Section 4.5.3 recommends that when wrapping an existing MPEG multiplex, the existing arrangements should be maintained where possible. When creating an MXF file from raw elementary stream components, the MXF encoder has greater flexibility. There are two different cases which will be considered:

- Creating an MXF file for interchange – easy access to all the components;
- Optimizing an MXF file for streaming – reducing buffer sizes and adding a small increase to the latency.

5.1.2 Interchange MXF file

In this case, it is desirable to group together elements which are synchronized on the timeline. The overall goal is to keep the synchronized video, audio and associated data in the same content package. This leads to the following rules.

In each content package:

- There shall be one video access unit.
- The synchronized sound sample should be in the first sound element in the same content package in accordance with SMPTE 379M, section 5.4. This implies that the start position of the video access unit should be equal to the start position of the sound element or fall within the duration of the first sound element. The unit of audio is defined in the appropriate GC sound element mapping document. Any valid GC sound element may be used. Examples might be an AES sample, a broadcast wave RIFF chunk, audio PES packet or MPEG-2 layer II audio frame.
- Sound elements should be placed in the content package until a sound element is found which may start a later content package. (Note that when the sound element duration is greater than the video access unit, this results in content packages with no or zero length sound elements).
- Any data element should start with the first indivisible unit of data where the start position of the video access unit is equal to the start position of the data element or falls within the duration of the first data element. The unit of data is defined in the appropriate GC data element mapping document. Any valid GC data element may be used.
- Any data element shall end with the unit of data whose position on the timeline is not later than the position of the next video access unit.

An implication of the above guidelines is that for each picture there will always be approximately the same time duration of data in the content package. Sometimes there will be empty elements and sometimes there will be more samples than average. This copes with the relationships between 29.97 frames/s video and its associated audio and data. Note that in many cases the sound and picture elements within the content package will have exactly the same duration.

Note that sound elements can be distinguished by the last 4 bytes of their key (see section 6.2). All sound element relating to a given track shall have the same last 4 bytes of the sound element key. This allows simple demultiplexing at the KLV level without having to parse the header metadata in advance.

5.1.3 Wrapping 3:2 sequences

The MPEG coding of 3:2 sequences and the definition of the AU is defined in the MPEG specifications (and explained in ISO13818-2, annex D). KLV wrapping of the 3:2 access units for interchange is, therefore, defined in section 5.1.2. Ensuring that an MPEG video access unit is wrapped defines the KLV wrapping point regardless of the regularity of the 3:2 sequence.

Construction of index tables for 3:2 sequences requires the index table construction application to calculate the correct byte offset for a given temporal offset regardless of the MPEG encoding used. This may be done