for Television — Material Exchange Format — Mapping JPEG 2000 Codestreams into the MXF Generic Container

Table of contents

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in Part XIII of its Administrative practices.

This SMPTE standard was prepared by Technology Committee W25 on Metadata and Wrappers.

Introduction

This standard maps the codestream for each JPEG 2000 (ISO-IEC15444-1) coded still picture into a MXF generic container. JPEG 2000 is a picture-by-picture coding scheme where each picture is entirely independent and can be extracted as an independent entity. However, the codestreams can be simply concatenated to form a sequence of compressed pictures.

This standard maps the JPEG 2000 codestream as either frame-wrapped where each JPEG 2000 codestream is individually mapped into a frame or clip-wrapped where a sequence of JPEG 2000 codestreams is mapped into a clip. This standard defines the KLV coding, the essence container and compression label values and the essence descriptor.

This standard specifies the mapping of ISO-IEC 15444-1 Annex A codestreams into the MXF generic container. This document does not specify mappings for ISO-IEC 15444-1 Annex I, nor does it specify mappings for any other parts of ISO-IEC 15444.

1 Scope

JPEG 2000 is a picture-by-picture compression coding defined by ISO-IEC 15444-1 and used for both individual pictures and picture sequences. This standard specifies the mapping of JPEG 2000 codestreams into a picture essence track of the MXF generic container in both frame-wrapped and clip-wrapped forms.

The MXF generic container is the native essence container of the material exchange format (MXF) file body. The MXF generic container is defined for the interchange of streamable audio-visual material.

This standard defines the data structure at the signal interfaces of networks or storage media. This standard does not define internal storage formats for MXF compliant devices.

2 Normative references

The following documents 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 edition of the standards indicated below.

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

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

SMPTE 400M-2004, Television — SMPTE Labels Structure

SMPTE RP 210, Metadata Dictionary Registry of Metadata Element Descriptions

ISO-IEC 15444-1 2004, Information Technology — JPEG 2000 Image Coding System.

3 Glossary of acronyms, terms and data types

The general glossary of acronyms, terms and data types used in the MXF specification is given in SMPTE 377M and is supplemented in SMPTE 379M. These glossaries are not repeated here to avoid any divergence of meaning.

Definitions of terms, abbreviations and symbols relating to JPEG 2000 are given in ISO-IEC 15444-1.

4 Introduction

The MXF generic container (GC) is fully described in SMPTE 379M. This standard specifies the mapping of JPEG 2000 codestreams as a picture element that may be used in the picture item of the MXF GC. The picture element may contain either individual JPEG 2000 codestreams using frame-wrapping or a sequence of JPEG 2000 codestreams using clip-wrapping.

This standard specifies the key, the length, and the value fields of the JPEG 2000 coded picture element. This standard also defines the essence container and compression label values and the essence descriptor.

4.1 JPEG 2000 coding summary (Informative)

JPEG 2000 is a picture-by-picture coding scheme, so each picture is independently coded and can be extracted as an independent entity. However, sequences of JPEG 2000 coded bitstreams can be simply concatenated to form a sequence of compressed images.

A JPEG 2000 coded bitstream for a single compressed image is defined as a codestream. This codestream is defined by a start codeword that identifies the start of the codestreams and an end codeword that identifies the end of the codestream. In between the start and end codewords are other codewords for identification of key parts of the codestream together with the raw compressed image data. The syntax of the codestream is fully defined in ISO-IEC15444-1.

This standard specifies only the mapping of ISO-IEC 15444-1 Annex A codestreams into an MXF file. JPEG 2000 codestreams encapsulated in the MPEG-4 base media file format as defined by ISO-IEC 15444-12 and ISO-IEC 15444-3 are known as an ISO-IEC Motion JPEG 2000. Any ISO-IEC Motion JPEG 2000 file that is converted to an MXF file should transfer appropriate file metadata to the MXF file.

4.2 Application in the MXF generic container

This mapping shall use the MXF generic container in either the frame-based wrapping or clip-based wrapping mode defined in SMPTE 379M.

4.3 Frame-based wrapping

An essence container that frame-wraps only JPEG 2000 compressed image data shall comprise one or more KLV triplets each of which shall contain a single JPEG 2000 compressed picture as illustrated in figure 1. A system item is optional in this essence container.

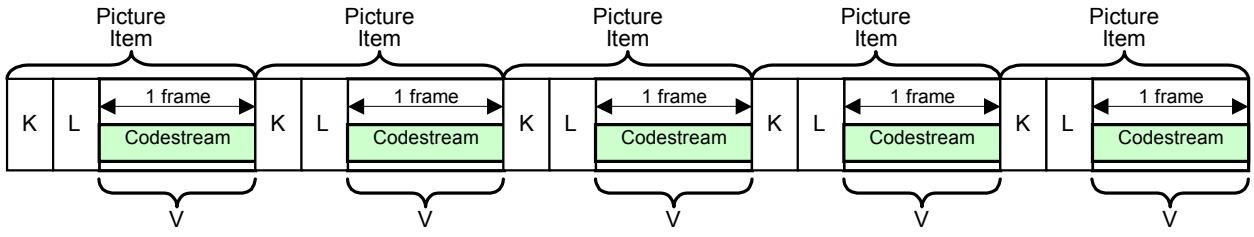


Figure 1 – Frame-based wrapping JPEG 2000 picture elements in the generic container

The JPEG 2000 compressed images may optionally be interleaved with other essence components in the frame-wrapped essence container as illustrated in figure 2.

These other essence components shall be as defined by this or other MXF mapping standards. All essence element types shall be frame wrapped. For simplicity of operation, each frame should contain essence data that is independent of adjacent frames. Interleaved essence elements that are inter-frame coded are not prohibited, but their inclusion may impact the performance of codecs. All essence elements in each interleaved frame should be time coincident within the limits of human recognition.

NOTE – The term 'frame-based wrapping' is defined by SMPTE 379M as the individual wrapping of one or more content packages each having a basic sample unit. This basic sample unit is defined by the JPEG2000 codestream and in television systems, may be the result of coding fields from an interlaced scanned picture or frames from a progressively scanned picture.

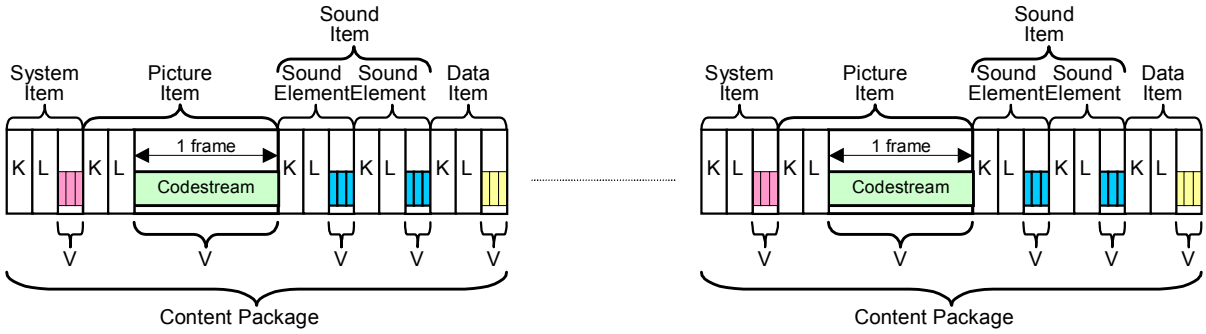


Figure 2 – Frame-based wrapping of interleaved items and elements in the generic container

NOTE – Interleaving essence elements may often involve timing tolerances whose specification is beyond the scope of this standard. However, the design of the frame-based interleaved MXF generic container is predicated on the concept of essentially time-aligned essence elements within each content package.

Individual applications may define the JPEG 2000 picture element as the only element present in each content package.

4.2.2 Clip-based wrapping

An essence container that clip-wraps only JPEG 2000 compressed image data shall comprise one KLV triplet containing a sequence of JPEG 2000 codestreams as illustrated in figure 3.

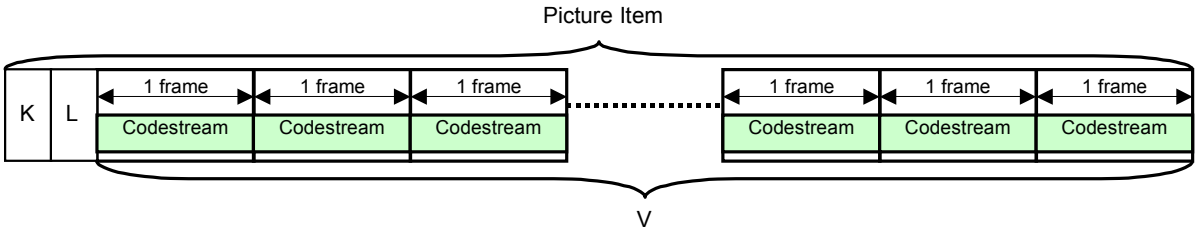


Figure 3 – Clip-based wrapping of a sequence of JPEG 2000 codestreams in the generic container

The clip-wrapped JPEG 2000 essence element may be the sole component in the MXF generic container content package.

The clip-wrapped JPEG 2000 essence element may also be used in the MXF generic container content package in sequence with other clip-wrapped essence elements. Note that each essence element should have the duration of the entire clip.

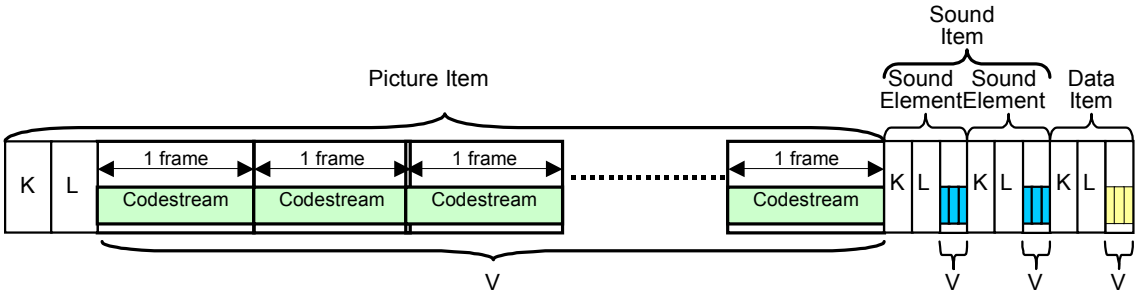


Figure 4 – Clip-based wrapping with other essence elements

5 Key-length-value coding

5.1 Essence element key

The values of the first 12 bytes of the essence element key are defined in SMPTE 379M. The values of the last four bytes of the picture element key are given in table 1.

Table 1 – Key value for the JPEG 2000 picture element

Byte No.	Description	Value (hex)	Meaning
1~12	Defined in SMPTE 379M		See SMPTE 379M
13	Item Type Identifier	15h	GC Picture Item (as defined in SMPTE 379M)
14	Essence Element Count	kkh	Count of Picture Elements in the Picture item
15	Essence Element Type	08h 09h	Frame-wrapped JPEG 2000 Picture Element Clip-wrapped JPEG 2000 Picture Element
16	Essence Element Number	nnh	The Number (used as an Index) of this Picture Element in the Picture Item

5.1.1 Essence element count — Byte 14

This is a count of the number of picture elements in the picture item of the generic container.

5.1.2 Essence element type — Byte 15

The value of 08h identifies that each JPEG 2000 codestream is frame-wrapped.

The value of 09h identifies that the sequence of JPEG 2000 codestreams are clip-wrapped.

5.1.3 Essence element number — Byte 16

This is a number used as an index to identify this instance of the element type within the picture item. Each element within an Item shall have a unique value between 00h and 7Fh, as defined by SMPTE 379M, which shall remain constant within the generic container.

5.2 Length

The length field shall comply with SMPTE 379M, section 5.5.1.

The length field should be 4 byte BER long-form encoded (i.e., 83h.xx.yy.zz) for frame-based wrapping and should be 8 byte BER long-form encoded (i.e., 87h.aa.bb.cc.dd.ee.ff.gg) for clip-based wrapping.

NOTE – Although the use of 8-byte BER long-form encoding for clip-based wrapping will be sufficient for all foreseen applications, the use of 4 byte BER long-form encoded length fields limits the size to 16 Mbytes. This may be insufficient for very high resolution pictures in which case longer length fields should be used.

5.3 Value

5.3.1 Frame-wrapped

The value field shall comprise a single JPEG 2000 codestream as defined in ISO-IEC 15444-1, Annex A.

NOTE – The JPEG 2000 codestream starts with the unique SOC (start of codestream) 2-byte marker and ends with the unique EOC (end of codestream) 2-byte marker.

5.3.2 Clip-wrapped

The value field shall comprise a sequence of one or more concatenated JPEG 2000 codestreams where each JPEG 2000 codestream is as defined in ISO-IEC 15444-1, Annex A.

NOTE – Each JPEG 2000 codestream starts with the unique SOC (start of codestream) 2-byte marker and ends with the unique EOC (end of codestream) 2-byte marker. Users should be cautioned that that the code values for SOC and EOC are not protected and may accidentally occur within the image size marker segment (SIZ), quantization marker segment (QCD), comment marker segment (COM) and other places. Thus, it is not safe to parse the concatenated JPEG 2000 codestreams by merely scanning for SOC and/or EOC values. The structure of a JPEG2000 codestream is essentially key-length-value, thus it is easy to read the lengths of the various codestream pieces and compute the length of the entire codestream. But applications which attempt to parse the bitstream at a randomly accessed point within the sequence of codestreams should be aware that SOC and EOC values are not guaranteed to be absent between the true SOC and EOC markers.

SMPTE ULs for JPEG 2000

5.4 Essence container UL

The values for the essence container UL are given in table 2.

Table 2 – Specification of the essence container label

Byte No.	Description	Value (hex)	Meaning
1-12	Defined by Generic Container		As defined in SMPTE 379M
13	Essence Container Kind	02h	MXF Generic Container
14	Mapping Kind	0Ch	JPEG 2000 Picture Element (as listed in SMPTE RP224)
15	Content Kind	01h 02h	Frame- wrapped Picture Element Clip- wrapped Picture Element
16	Reserved	00h	

The essence container UL is used within a batch of ULs in partition packs and the preface set and on its own in the essence descriptor. These UL values are listed in the SMPTE labels registry (SMPTE RP 224).

5.5 Picture essence compression UL

Values for the picture essence compression UL are given in table 3.

Table 3 – Specification of the picture essence compression label

Byte No.	Description	Value (hex)	Meaning
1-7	Registry Designator	See SMPTE 400M	Designator value is defined in SMPTE 400M
8	Registry Version Number	vvh	Version of the Registry
9	Parametric	04h	Node used to define parametric data
10	Picture Essence	01h	Identifies picture essence coding
11	Picture Coding Characteristics	02h	Identifies picture coding characteristics
12	Compressed Picture Coding	02h	Identifies compressed picture coding
13	Individual Picture Coding	03h	Identifies individual picture coding
14	JPEG 2000 Picture Coding	01h	Identifies JPEG 2000 picture coding
15	JPEG 2000 Picture Coding Variant	01h	Identifies JPEG 2000 coding according to ISO/IEC 15444-1
16	JPEG 2000 Picture Coding Constraints	xxh	Identifies coding constraints for the intended application. A value of '00h' indicates a generic application that has no coding constraints. Other specifications will define the meaning of non-zero values.

The picture essence compression UL is used in the generic picture essence descriptor. This UL is listed in SMPTE labels registry (SMPTE RP 224).

6 Application issues

6.1 Application of the KAG and the KLV fill item

There are no specific KAG requirements for the JPEG2000 mapping. MXF encoders and decoders shall comply with the KAG rules defined in SMPTE 377M, section 5.4.1.

The default value of the KAG is '1'. Other KAG values may be used within the range defined by SMPTE 377M, section 5.4.1.

The KLV fill item may be used to maintain a constant content package size so permitting the use of a single index table segment.

6.2 Index table usage

Since the coding is frame-based, the KLV fill item can provide for a constant edit unit size for all frames in many applications.

Where the application defines a constant edit unit size, an index table shall be used. This includes the cases where the JPEG 2000 essence element is the sole essence component and where it is interleaved with other essence components.

Where the application has a variable edit unit size an index table should be used wherever possible.

SMPTE EG41 gives examples of how index tables can be created for both mono and multi-essence mappings and for both constant and variable length edit unit sizes.

6.3 Operational pattern usage

This essence mapping may be used with any generalized operational pattern.

NOTE – This does not preclude the use of specialized operational patterns.

6.4 Mapping track numbers to generic container elements

Each track number value for an essence element defined in this standard shall be derived as described in the MXF generic container specification (SMPTE 379M).

6.5 Essence container partitions

Frame wrapping maintains each content package of the generic container as a separate editable unit with the contents of the system, picture, sound and data items in synchronism. If a frame-wrapped essence container is partitioned, then individual content packages should not be fragmented by the partitioning process.

If the essence container is clip wrapped it is recommended that each essence element be multiplexed in a sequence of partitions.

NOTE – SMPTE 377M, section 5.2.2 (Partition Rules Summary) summarizes the use of partitions in MXF files.

7 Essence descriptors

7.1 File descriptor sets

The file descriptor sets are those structural metadata sets in the header metadata that describe the essence and metadata elements defined in this standard. The structure of these sets is defined in the MXF file format specification (SMPTE 377M) and in some generic container mapping specifications.

The values of the metadata defined in the sub-descriptor defined below are copies of values used in the syntax of the JPEG 2000 codestream. If there is any discrepancy between values, those in the codestream shall take precedence and the values in the sub-descriptor should be updated.

With the exception of those properties that have been defined in SMPTE 377M, all local tag values in descriptors defined in this standard shall be dynamically allocated (Dyn) as defined in SMPTE 377M, section 8.2.2 (local tag values). The translation from each dynamically allocated local tag value to its full UL value can be found using the primer pack mechanism defined in SMPTE 377M, section 8.2 (primer pack). The full 16-byte UL values are defined in SMPTE RP 210.

7.2 JPEG 2000 picture sub-descriptor

Essence tracks that use the JPEG 2000 essence mapping may use the values of the JPEG 2000 picture sub-descriptor as defined in table 6. The JPEG 2000 picture sub-descriptor is coded as a local set using 2-byte tag values and 2-byte length values consistent with all MXF descriptors.

This sub-descriptor is a supplementary essence descriptor that can be strongly referenced by any file descriptor. It is intended that this JPEG 2000 sub-descriptor be referenced either by the CDCI picture essence descriptor or the RGBA picture essence descriptor both of which are defined by SMPTE 377M. In order that the strong reference can be made, the MXF generic descriptor (as defined in SMPTE 377M) has an additional optional property as defined in table 4.

Table 4 – Additional optional property for the MXF generic descriptor

Element Name	Type	Len	Local Tag	UL Designator	Req?	Element Description	Default
All elements from the Generic Descriptor defined in SMPTE 377M Table 17							
Sub Descriptors	StrongRefArray (Sub Descriptors)	8+16 n	Dyn	06.01.01.04.06 .10.00.00	Opt	Ordered array of strong references to sub descriptor sets	

NOTE – The JPEG 2000 picture sub-descriptor is a sub-class of the MXF header metadata abstract superclass and inherits only the InstanceUID and GenerationUID properties. In order to use this set, the new "sub-descriptors" property in the MXF generic descriptor allows both the CDCI and RGBA picture essence descriptors to inherit this property and thus either can make a strong reference to the JPEG 2000 picture sub-descriptor.

The JPEG 2000 picture sub-descriptor includes only those properties from the main header of the codestream that are required in ISO/IEC 15444-1, Annex A, table A.2.

Note that the JPEG 2000 picture sub-descriptor may only be used if the required properties are consistent for all JPEG 2000 codestreams in the essence container.

Annex B illustrates the chain of MXF descriptors and their relationships.

7.2.1 Key value

The set key of the JPEG 2000 picture sub-descriptor shall be as defined in table 5.

Table 5 – Key value for the JPEG 2000 picture sub-descriptor

Byte No.	Description	Value	Meaning
1~13	As defined in SMPTE 377M, table 13		Values for all MXF structural metadata sets
14~15	Set Kind	01.5Ah	Defines the Key value for the JPEG 2000 Picture Sub Descriptor
16	Reserved	00h	Reserved value

7.2.2 Length value

The set length shall be BER long form encoded. It is preferred that the length field uses 4 bytes.

7.2.3 Descriptor value

Table 6 – Specification of the values of the JPEG 2000 picture sub-descriptor

Element Name	Type	Len	Local Tag	UL Designator	Req?	Element Description	Default
Instance UID	UUID	16	3C.0A	01.01.15.02. 00.00.00.00	Req	Unique ID of this instance [**RP210** The ISO/IEC 11578 (Annex A) 16 byte Globally Unique Identifier]	
Generation UID	UUID	16	01.02	05.20.07.01. 08.00.00.00	Opt	Generation Identifier [**RP210** Specifies the reference to an overall modification]	

Rsiz	UInt16	2 bytes	Dyn	04.01.06.03.01.00.00.00	Req	An enumerated value that defines the decoder capabilities. Values are defined in ISO/IEC 15444-1 Annex A.5 Table A-10. Other values may be defined in amendments to ISO/IEC 15444-1 or in related international standards documents.	
Xsiz	UInt32	4 bytes	Dyn	04.01.06.03.02.00.00.00	Req	Width of the reference grid, as defined in ISO/IEC 15444-1 Annex A.5.1.	
Ysiz	UInt32	4 bytes	Dyn	04.01.06.03.03.00.00.00	Req	Height of the reference grid, as defined in ISO/IEC 15444-1 Annex A.5.1.	
XOsiz	UInt32	4 bytes	Dyn	04.01.06.03.04.00.00.00	Req	Horizontal offset from the origin of the reference grid to the left side of the image area, as defined in ISO/IEC 15444-1 Annex A.5.1.	
YOsiz	UInt32	4 bytes	Dyn	04.01.06.03.05.00.00.00	Req	Vertical offset from the origin of the reference grid to the top side of the image area, as defined in ISO/IEC 15444-1 Annex A.5.1.	
XTsiz	UInt32	4 bytes	Dyn	04.01.06.03.06.00.00.00	Req	Width of one reference tile with respect to the reference grid, as defined in ISO/IEC 15444-1 Annex A.5.1.	
YTsiz	UInt32	4 bytes	Dyn	04.01.06.03.07.00.00.00	Req	Height of one reference tile with respect to the reference grid, as defined in ISO/IEC 15444-1 Annex A.5.1.	
XTOsiz	UInt32	4 bytes	Dyn	04.01.06.03.08.00.00.00	Req	Horizontal offset from the origin of the reference grid to the left side of the first tile, as defined in ISO/IEC 15444-1 Annex A.5.1.	
YTOsiz	UInt32	4 bytes	Dyn	04.01.06.03.09.00.00.00	Req	Vertical offset from the origin of the reference grid to the top side of the first tile, as defined in ISO/IEC 15444-1 Annex A.5.1.	
Csiz	UInt16	2 bytes	Dyn	04.01.06.03.0A.00.00.00	Req	The number of components in the picture as defined in ISO/IEC 15444-1 Annex A.5.1. If this Sub Descriptor is referenced by the CDCI Descriptor, the order and kind of components shall be as defined by the Essence Container UL in the MXF File Descriptor. If this Sub Descriptor is referenced by the RGBA Descriptor, the order and kind of components shall be as defined by the Pixel Layout property of the RGBA Descriptor.	

Picture Component Sizing	J2K ComponentSizingArray	8+3n bytes	Dyn	04.01.06.03.0B.00.00.00	Req	Array of picture components where each component comprises 3 bytes named $Ssiz^i$, $XRSiz^i$, $YRSiz^i$ (as defined in ISO/IEC 15444-1 Annex A.5.1). The array of 3-byte groups is preceded by the array header comprising a 4-byte value of the number of components followed by a 4-byte value of '3'.	
Coding Style Default	J2K CodingStyleDefault	var	Dyn	04.01.06.03.0C.00.00.00	Opt	Default coding style for all components. Use this value only if static for all pictures in the Essence Container. The data format is as defined in ISO/IEC 15444-1, Annex A.6.1 and comprises the sequence of Scod (1 byte per table A-12), SGcod (4 bytes per table A.12) and Spcod (5 bytes plus 0 or more precinct size bytes per table A.12)	
Quantisation Default	J2K QuantisationDefault	var	Dyn	04.01.06.03.0D.00.00.00	Opt	Default quantisation style for all components. Use this value only if static for all pictures in the Essence Container. The data format is as defined in ISO/IEC 15444-1, Annex A.6.4 and comprises the sequence of Sqcd (1 byte per table A.27) followed by one or more $Sqcd^i$ bytes (for the i^{th} sub-band in the defined order per table A.27).	

Annex A (informative)
JPEG2000 coding of television signals

JPEG2000 may be used to compress any known television standard. With interlaced scanning, the duration of the JPEG2000 codestream may be 1 frame or 1 field. With progressive and segmented frame scanning, the duration of the JPEG2000 codestream will be 1 frame.

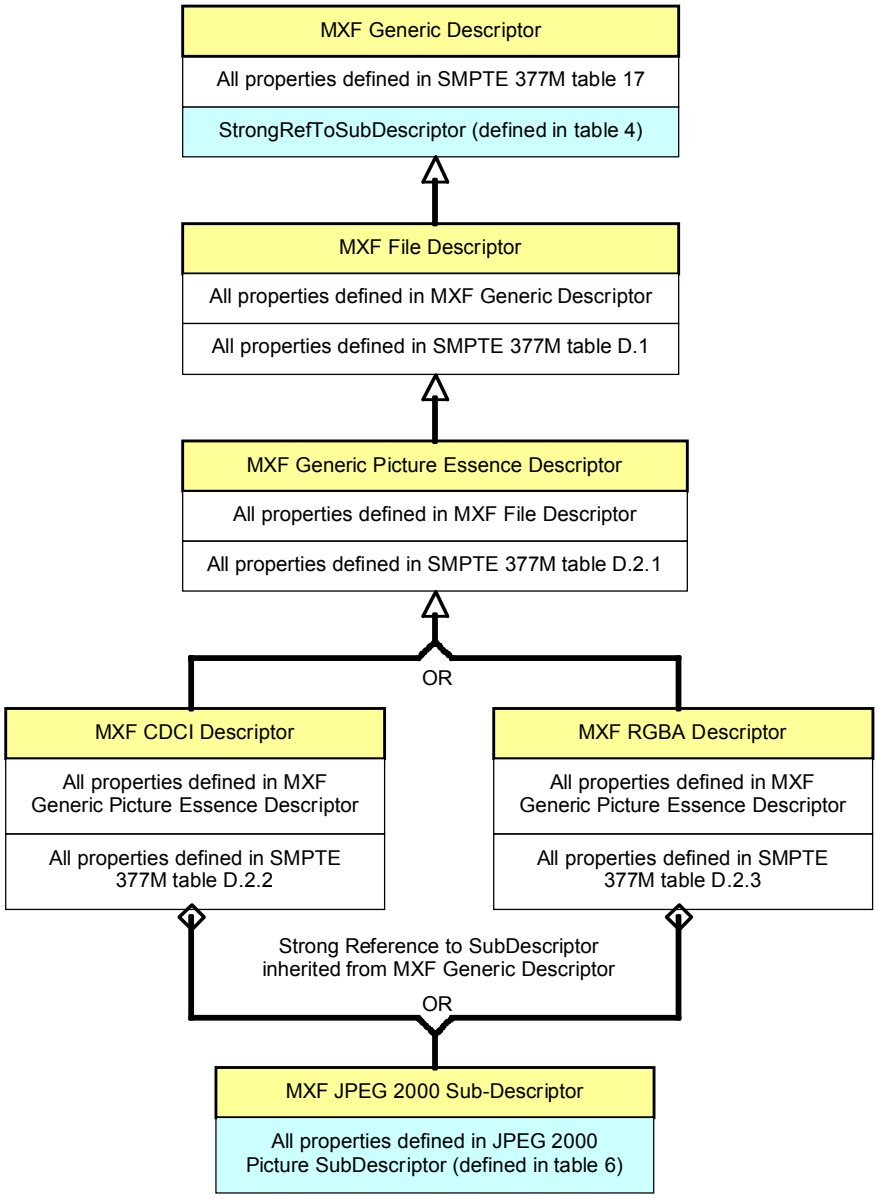
The format of the source coding can be found from the ‘frame layout’ property in the generic picture essence descriptor. This indicates whether the source picture was full frame, separate fields, mixed fields or any other permitted layout value.

When using separate fields, the optional ‘FieldDominance’ property, also in the generic picture essence descriptor, defines the field number (1 or 2) which is considered to be temporally the first field of an interlaced frame.

The main television standards are listed in annex C for information purposes. Other television standards not listed may also be used.

Annex B (informative)
Illustration of the JPEG 2000 picture sub-descriptor

The figure below illustrates how the MXF JPEG 2000 sub-descriptor can be the target of a strong reference from either the CDCI descriptor or the RGBA descriptor. The distinction between the arrow and diamond symbol symbols is that the arrow symbol represents inheritance (dependency) and the diamond symbol represents ownership (composition). The modelling method used in this illustration is based on the Unified Modelling Language (UML) and more on metadata modelling can be found in SMPTE EG 42 (annex C).



Annex C (informative)
Bibliography

ANSI/SMPTE 298M-1997, Television — Universal Labels for Unique Identification of Digital Data

SMPTE 274M-2005, Television — 1920 × 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

SMPTE 296M-2001, Television — 1280 × 720 Progressive Image Sample Structure — Analog and Digital Representation and Analog Interface

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

SMPTE RP 224, SMPTE Labels Registry

SMPTE EG 41-2004, Television — Material Exchange Format (MXF) — Engineering Guideline

SMPTE EG 42-2004, Television — Material Exchange Format (MXF) — MXF Descriptive Metadata

ISO/IEC 15444-3:2003, Information Technology — JPEG 2000 Image Coding System — Part 3: Motion JPEG 2000 Amendment 2: Motion JPEG 2000 Derived from ISO Base Media File Format

ISO/IEC 15444-12:2003, Information Technology — JPEG 2000 Image Coding System — Part 12: ISO Base Media File Format

ITU-R BT.601-5, Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios