

SMPTE STANDARD



D-Cinema Packaging - Immersive Audio Track File

Table of Contents		Page
	Foreword	3
	Intellectual Property	3
	Introduction	3
1	Scope	4
2	Conformance Notation	4
3	Normative References	4
4	Terms and Definitions	5
5	Carrying Immersive Audio Data in the MXF Generic Container	6
5.1	Frame Wrapping	6
5.2	Clip Wrapping	6
5.3	Custom Wrapping	6
5.4	Element and Item Constraints	6
6	KLV Coding of Immersive Audio Data Elements	6
6.1	Data Element Key	6
6.2	Length	6
6.3	Value	7
7	Constraints	7
7.1	General	7
7.2	Bitstream Constraints	7
8	Label for Immersive Audio Data Essence Container Identification	8
9	Immersive Audio Data Essence Descriptor	8
10	Version Number of the Data Essence Coding Item UL in the Primer Pack	9
11	SMPTE Label for Immersive Audio Data Essence Coding	9
12	Immersive Audio Data Essence SubDescriptor	9
12.1	Immersive Audio Version	10
12.2	Max Channel Count	10
12.3	Max Object Count	10
12.4	Immersive Audio ID	10

12.5	First Frame	11
12.6	IAB Sample Rate	11
13	Packaging Constraints	11
14	Composition Playlist Extensions	11
14.1	Extension Elements	11
14.2	Namespace	11
14.3	AuxData	12
14.4	DataType	12
15	CPL Extension Schema	12
Annex A	XML Schema (informative)	14

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 its Standards Operations Manual. This SMPTE Engineering Document was prepared by Technology Committee 21DC.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

This document specifies a method for carrying Immersive Audio data essence, as specified in SMPTE ST 2098-2, in an MXF file based on the MXF Generic Container for use in a D-Cinema Package. While this specification is written to allow synchronization of the Immersive Audio data track with picture or sound, the synchronization mechanism is outside the scope of this document.

1 Scope

This standard specifies the mapping of an ST 2098-2 Immersive Audio Bitstream into an MXF file for use within a Digital Cinema Package. The MXF Generic Container (GC), as described in SMPTE ST 379:2004, is used as a normative reference for this mapping, but ST 429-18 uses Class 14 ULs which prevents full compliance. The resultant file will be referred to as an Immersive Audio Track File.

This standard specifies the Key, the Length and the Value fields of the Immersive Audio Data Element. This standard also defines the Essence Container and the Essence Descriptors.

This standard defines the SMPTE ST 429-7 Composition Playlist asset type extension required for inclusion of an Immersive Audio track.

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword "reserved" indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword "forbidden" indicates "reserved" and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; then formal languages; then figures; and then any other language forms.

3 Normative References

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

to revision, and parties to agreements based on this engineering document are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

SMPTE ST 377:2004¹, Television — Material Exchange Format (MXF) — File Format Specification

SMPTE ST 379:2004, Television — Material Exchange Format (MXF) — MXF Generic Container

SMPTE ST 400:2012, SMPTE Labels Structure

SMPTE ST 429-3:2007, D-Cinema Packaging — Sound and Picture Track File

SMPTE ST 429-7:2006, D-Cinema Packaging — Composition Play List

SMPTE ST 2029:2009, Uniform Resource Names for SMPTE Resources

SMPTE 2098-2:2019, Immersive Audio Bitstream Specification

[XML] World Wide Web Consortium (W3C) (2004 February 4). Extensible Markup Language (XML) 1.0 (Third Edition)

[XML Schema Part 1: Structures] World Wide Web Consortium (W3C) (2004, October 28). XML Schema Part 1: Structures (Second Edition)

[XML Schema Part 2: Datatypes] World Wide Web Consortium (W3C) (2004, October 28). XML Schema Part 2: Datatypes (Second Edition)

[RFC 2396] Internet Engineering Task Force (IETF) (1996, November). RFC 2396 — Uniform Resource Identifiers (URI): Generic Syntax

4 Terms and Definitions

For the purposes of this document, the following terms and definitions apply.

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

The reader is also directed to the glossary terms defined in SMPTE ST 429-7.

¹ The reference to SMPTE ST 377:2004 and SMPTE ST 379:2004 are intentional. SMPTE ST 377-1:2011 and SMPTE ST 379-1:2009 or future versions are not appropriate for use with this document.

5 Carrying Immersive Audio Data in the MXF Generic Container

5.1 Frame Wrapping

The "Frame Wrapping" method for data essence is illustrated in Figure 1. Frame wrapping shall be used for the Immersive Audio track file.

Figure 1 shows a series of data elements, each wrapped in a single Content Package Data Element with no other Generic Container Elements in the Container. Each Content Package has the duration of one edit unit.

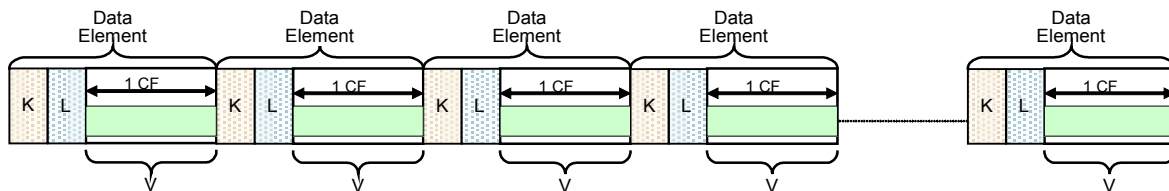


Figure 1: Simple Representation of Frame Wrapping (Informative)

The Frame Wrapping method enables frame by frame access by MXF applications which process at the KLV level. 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 data frame shall be KLV wrapped using a GC Data Element Key.

5.2 Clip Wrapping

Clip wrapping shall not be used in an Immersive Audio track file

5.3 Custom Wrapping

Custom wrapping shall not be used in an Immersive Audio track file.

5.4 Element and Item Constraints

Only Data Elements shall be carried in an Immersive Audio track file. Each Data Item shall contain only a single Data Element.

6 KLV Coding of Immersive Audio Data Elements

6.1 Data Element Key

The Immersive Audio Data Element key shall be set as specified below:

06.0E.2B.34.01.02.01.05.0E.09.06.01.00.00.00.01.

NOTE: The Immersive Audio Data Element key is not a valid SMPTE UL, is not listed in the SMPTE Metadata Registers, and does not conform to SMPTE ST 379.

6.2 Length

The length field and its application shall comply with SMPTE ST 377.

6.3 Value

6.3.1 Frame Wrapped

The value field shall comprise one IABitstream frame as defined in SMPTE ST 2098-2.

7 Constraints

7.1 General

The Immersive Audio Track file shall be constrained to comply with SMPTE ST 429-3, Sound and Picture Track file with respect to operational pattern, header metadata, and asset identity constraints.

7.2 Bitstream Constraints

7.2.1 General Bitstream Constraints

All IAFrames in the bitstream shall have the same value for FrameRate, BitDepth and SampleRate.

7.2.2 AudioData Element

Bitstreams shall not contain the AudioDataPCM element. Instead, AudioDataDLC shall be used for all audio essence.

7.2.3 Sample Bit Depth

The bit depth of audio samples shall be 24 bits.

7.2.4 Maximum Frame Size

Each individual frame of the Immersive Audio Bitstream shall be constrained to have a maximum size as listed in Table 1 below.

Table 1 - Maximum Size of Frame

Edit Rate	Maximum Size (bytes)
24	781,250
25	750,000
30	625,000
48	390,625
50	375,000
60	312,500
96	195,313
100	187,500
120	156,250

7.2.5 UserData Element

The UserData element as defined in ST 2098-2 shall not be used.

7.2.6 UseCase Codes

UseCase codes between the values of 0x30 and 0xFE inclusive shall not be used.

7.2.7 ChannelID Codes

ChannelID codes greater than 0x7F shall not be used.

8 Label for Immersive Audio Data Essence Container Identification

The Immersive Audio Data Essence Container UL is defined in Table 2:

Table 2 - Immersive Audio Data Essence Container UL

Kind	Leaf
Name	MXF-GC IADData Frame Wrapped
Symbol	MXF_GC_IADData_Frame_Wrapped
Description	Identifies Container for Frame Wrapped Immersive Audio Data
UL	urn:smpte:ul: 060E2B34.04010105.0E090605.00000000

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.

Note: This UL is from a Class 14 node so does not comply with the UL construction defined in SMPTE ST 379.

9 Immersive Audio Data Essence Descriptor

The File Descriptor sets are those structural metadata sets in the Header Metadata that describe the essence and metadata elements defined in this document. The Immersive Audio Data Essence Descriptor shall be a sub-class of the MXF Generic Data Essence Descriptor (SMPTE ST 377). File Descriptor sets shall be present in the Header Metadata for each Essence Element. Implementations that carry specific data types may extend the Immersive Audio Data Essence Descriptor using a SubDescriptor. Implementations complying with this specification shall ignore unrecognized SubDescriptors.

Table 3 - Immersive Audio Data Essence Descriptor

Item Name	Type	Len	Local Tag	UL Designator	Req ?	Meaning	Default
Immersive Audio Data Essence Descriptor	Set UL	16	dyn	06.0E.2B.34.02.53.01.05.0E.09.06.03.00.00.00.00	Req	Identifies the Immersive Audio Data Essence Descriptor Set (a collection of Parametric metadata)	
Length	BER Length	4			Req	Set length	
All items from the Data Essence Descriptor in SMPTE ST 377 Annex D.4 to be included							

The Data Essence Coding property shall be present in the Immersive Audio Data Essence Descriptor

10 Version Number of the Data Essence Coding Item UL in the Primer Pack

In the Primer Pack, the UID associated with Local Tag 0x3E01 shall be set to:

06.0E.2B.34 01.01.01.05.04.03.03.02 00.00.00.00

NOTE: The UL above corresponds to the Data Essence Coding Item UL with its Version Number (Byte 8) set to 0x05. The UL is not listed in the SMPTE Metadata Registers, and does not conform to SMPTE ST 377.

11 SMPTE Label for Immersive Audio Data Essence Coding

The Data Essence Coding UL is used in the Immersive Audio Data Essence Descriptor. The UL and associated values are defined in Table 4.

Table 4 - Label for Immersive Audio Coding

Kind	Leaf
Name	Immersive Audio Coding
Symbol	ImmersiveAudioCoding
Description	Identifies Immersive Audio Coding per ST 2098-2
UL	urn:smpte:ul: 060E2B34.04010105.0E090604.00000000

12 Immersive Audio Data Essence SubDescriptor

The File Descriptor sets are those structural metadata sets in the Header Metadata that describe the essence and metadata elements defined in this document. File Descriptor sets shall be present in the Header Metadata for each Essence Element. The Immersive Audio Data Essence SubDescriptor is a supplementary Essence Descriptor that can be strongly referenced by any Data Essence Descriptor and shall be an instance of MXF SubDescriptor. In order that the strong reference can be made, the MXF Generic Descriptor (as defined in SMPTE ST 377) has an additional optional property as defined in Table 17.

The Local Tag value associated with this additional optional property (called “SubDescriptors”) shall be dynamically allocated (dynamic) as defined in SMPTE ST 377. 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 ST 377.

Some of the values of the items defined in the Immersive Audio Data Essence SubDescriptor are derived from values used in the Immersive Audio frame. The specific items whose values are derived are as follows:

Table 5 Immersive Audio Data Essence SubDescriptor Values

Item Name	Type	Len	Local Tag	UL Designator	Req ?	Meaning
Immersive Audio Data Essence Sub Descriptor	Set UL	16	dyn	06.0E.2B.34.02.53.01.05.0E.09.06.06.00.00.00.00	Req	Defines the Immersive Audio Data Essence Descriptor Set (a collection of Parametric metadata)

↔ Length	BER Length	4			Req	Set length
Immersive Audio Version	Uint8	1	dyn	06.0E.2B.34.01.01.01.05.0E.09.05.06.00.00.00.00.	Opt	Immersive Audio Coder version used to create the source Bitstream
Max Channel Count	Uint16	2	dyn	06.0E.2B.34.01.01.01.05.0E.09.05.07.00.00.00.00.	Opt	Maximum number of channels in the bitstream
Max Object Count	Uint16	2	dyn	06.0E.2B.34.01.01.01.05.0E.09.05.08.00.00.00.00.	Opt	Maximum Number of objects in the bitstream
Immersive Audio ID	UUID	16	dyn	06.0E.2B.34.01.01.01.05.0E.09.05.09.00.00.00.00.	Opt	UUID of the Immersive Audio project
First Frame	UInt32	4	dyn	06.0E.2B.34.01.01.01.05.0E.09.05.0A.00.00.00.00.	Opt	Specifies an edit unit for alignment with the FFOA of the picture track
IAB Sample Rate	Rational	8	dyn	060E2B34.0101010E.04020301.0F000000	Opt	Sample Rate of the audio essence contained in the bitstream.

Some of the values stored in the Subdescriptor are derived from the Immersive Audio Bitstream as indicated below. However, the values for Immersive Audio ID and First Frame are set during initial bitstream coding and should not be changed. Implementations should copy the values for Immersive Audio ID and First Frame from the Immersive Audio Data Essence SubDescriptor of the unmodified file in the event that the file must be re-wrapped.

12.1 Immersive Audio Version

This integer is derived from the: Version field of a single IAFrame element. All IAFrame elements within a file shall have the same value.

12.2 Max Channel Count

Decoders shall ignore this item

12.3 Max Object Count

Decoders shall ignore this item.

12.4 Immersive Audio ID

This item is deprecated and should not be present.

12.5 First Frame

This parameter shall be an integer indicating the edit unit that is to be synchronized to the beginning of the associated image sequence, also known as the FFOA.

12.6 IAB Sample Rate

The Sample Rate parameter shall indicate the rational representation of the IAFrame SampleRate field.

13 Packaging Constraints

13.1 Edit Rate Value

The Edit Rate of the essence track shall be derived from the IAFrame:FrameRate code.

13.2 Sample Rate Value

The Sample Rate value for the MXF Generic File Descriptor shall be derived from the IAFrame:FrameRate code.

13.3 Encryption

Files created with this specification shall be encrypted using SMPTE 429-6, MXF Track File Encryption, if encryption is necessary.

14 Composition Playlist Extensions

14.1 Extension Elements

To reference the Immersive Audio Track in a Composition, the extension elements defined in this section shall be used to extend the Reel element of a Composition Playlist, as specified in SMPTE ST 429-7.

14.2 Namespace

The `AuxData` extension element defined in this specification shall be associated with a unique XML namespace name that shall be the string value `"http://www.dolby.com/schemas/2012/AD"`. This namespace name conveys both structural and semantic version information, and serves the purpose of a traditional version number field.

XML namespace names used in this standard are identified in Table 6. Namespace names are represented as Uniform Resource Identifier (URI) values per RFC 2396².

² Readers unfamiliar with URI values as XML namespace names should be aware that although a URI value begins with a "method" element ("http" in this case), the value is designed primarily to be a unique string and does not necessarily correspond to an actual on-line resource. Applications implementing this standard should not attempt to resolve URI values on-line.

Table 6 - XML Namespace

Qualifier	URI
cpl	http://www.smpte-ra.org/schemas/429-7/2006/CPL
ia	http://www.dolby.com/schemas/2012/AD

URIs listed in Table 6 are normative, whereas the namespace qualifier values themselves (used in Table 6 and elsewhere in this standard) are not normative. Thus, namespace qualifier values may be replaced in instance documents by any arbitrary XML-compliant namespace qualifier, meaning that conformant implementations shall expect any XML-compliant namespace qualifier value that is associated with a URI from Datatypes from other schemas that are used in this document will be prefixed with the appropriate namespace qualifier (e.g. `xs:dateTime`). See XMLSchema Part 2: Datatypes for further information about these types.

14.3 AuxData

The `AuxData` extension element defines the Immersive Audio asset intended for use with the composition. The actual data essence is contained in an external Immersive Audio Track File.

The `AuxData` element shall be an instance of the `DataTrackFileAssetType` element, which is derived from the `TrackFileAssetType` whose structure is defined in SMPTE ST 429-7.

The element defined below replicates values contained in the underlying track file and shall remain consistent with the content of the underlying track file at all times. It is included in the Composition Playlist to alleviate the need for theater management software to access and parse individual track files when scheduling content. In the event an inconsistency exists, the values contained in the underlying track file shall take precedence.

14.4 DataType

The `DataType` element is a UL that matches the value of the Data Essence Coding parameter of the Data Essence Descriptor in the Immersive Audio Data Track File. This allows identification of the type of data essence that is referenced by the Immersive Audio Data track. It shall be coded as type `urn:smpte:ul` as specified in SMPTE ST 2029. Only one value shall be valid, which is the UL specified for Immersive Audio essence in section 11.

15 CPL Extension Schema

The XML Schema document presented in this section normatively defines the structure of the SMPTE 429-7 Composition Playlist extensions previously described using a machine-readable language. While this schema is intended to faithfully represent the structure presented in the normative prose portions of this specification, conflicts in definition may occur. In the event of such a conflict, the normative prose shall be the authoritative expression of the standard.

```
<?xml version="1.0" encoding="UTF-8"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://www.dolby.com/schemas/2012/AD"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:cpl="http://www.smpte-ra.org/schemas/429-7/2006/CPL"
  xmlns:ia="http://www.dolby.com/schemas/2012/AD"

  elementFormDefault="qualified" attributeFormDefault="unqualified">
  <xs:element name="AuxData" type="ia:DataTrackFileAssetType" />
  <xs:complexType name="DataTrackFileAssetType">
    <xs:complexContent>
      <xs:extension base="cpl:TrackFileAssetType">
        <xs:sequence>
          <xs:element name="DataType" type="xs:anyURI" />
        </xs:sequence>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</schema>
```

Annex A XML Schema (informative)

This specification is accompanied by the following element, which is an XML schema document as specified in XML Schema Part 1: Structures.

st429-18a-2019.xsd

This element collects the XML schema definitions defined in this specification. It is informative and, in case of conflict, this specification takes precedence.