SMPTE ST 429-2:2020
Revision of SMPTE ST 429-2:2019

SMPTE STANDARD

D-Cinema Packaging — DCP Operational Constraints

Table of Contents

Pages

Approved October 27, 2020

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 Standards Operations Manual.

SMPTE ST 429-2 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.

1 Scope

This document specifies a D-Cinema Package (DCP), a collection of files containing d-cinema essence and related metadata to be ingested and reproduced by a d-cinema playback system.

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.

3 Normative References

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 382:2007) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 382M-2007). Documents with the same root number (e.g. 382) and publication year (e.g. 2007) are functionally identical.

The following standards contain provisions which, through reference in this text, constitute provisions of this recommended practice. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this recommended practice are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

SMPTE ST 377:2004[1] Material Exchange Format (MXF) — File Format Specification

SMPTE ST 377-4:2012 MXF Multichannel Audio Labeling Framework

SMPTE ST 382:2007[2] Material Exchange Format — Mapping AES3 and Broadcast Wave Audio into the MXF Generic Container

SMPTE ST 400:2012 SMPTE Labels Structure

SMPTE ST 422[3]:2006 Material Exchange Format — Mapping JPEG 2000 Codestreams into the MXF Generic Container

[1] The reference to SMPTE ST 377:2004 is intentional. SMPTE ST 377-1:2011 or future versions are not appropriate for use with this document.
[2] The omission of the reference to Amendment 1:2012 to SMPTE ST 382:2007 is intentional. This Amendment is not appropriate for use with this document.

SMPTE ST 428-1:2019 D-Cinema Distribution Master — Image Characteristics

SMPTE ST 428-2:2006 D-Cinema Distribution Master — Audio Characteristics

SMPTE ST 428-7:2014 D-Cinema Distribution Master — Subtitle

SMPTE ST 428-10:2008 D-Cinema Distribution Master — Closed Caption and Closed Subtitle

SMPTE ST 428-12:2013 D-Cinema Distribution Master — Common Audio Channels and Soundfield Groups

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

SMPTE ST 429-4:2020 D-Cinema Packaging — MXF JPEG 2000 Application

SMPTE ST 429-5:2017 D-Cinema Packaging — Timed Text Track File

SMPTE ST 429-6:2006 D-Cinema Packaging — MXF Track File Essence Encryption [4]

SMPTE ST 429-7:2006 D-Cinema Packaging — Composition Playlist

SMPTE ST 429-8:2007 D-Cinema Packaging — Packing List

SMPTE ST 429-10:2008 D-Cinema Packaging — Stereoscopic Picture Track File

SMPTE ST 429-12:2008 D-Cinema Packaging — Caption and Closed Subtitle

SMPTE ST 429-16:2014 D-Cinema Packaging — Additional Composition Metadata and Guidelines

SMPTE ST 430-2:2017 D-Cinema Operations — Digital Certificate

SMPTE ST 2029:2009 Uniform Resource Names for SMPTE Resources

SMPTE ST 429-17:2017 XML Constraints

ISO/IEC 10646:2017 Information Technology — Universal Coded Character Set (UCS)

ISO/IEC 15444-1:2019 Information Technology — JPEG 2000 Image Coding System: Core Coding System

ISO/IEC 15948:2004 Information Technology — Computer Graphics and Image Processing — Portable Network Graphics (PNG): Functional Specification

IEC 61966-2-1:1999 Colour Measurement and Management in Multimedia Systems and Equipment — Part 2-1: Default RGB Colour Space - sRGB

Internet Engineering Task Force (IETF) (July 2005). RFC 4122 A Universally Unique Identifier (UUID) URN Namespace

Internet Engineering Task Force (IETF) (February 2006). RFC 4246 International Standard Audiovisual Number (ISAN) URN Definition

Internet Engineering Task Force (IETF) (Sept 2016). RFC 7972 Entertainment Identifier Registry (EIDR) URN Namespace Definition

---

[3] The reference to SMPTE ST 422:2006 is intentional. SMPTE ST 422:2013, Amendment 1:2014 to SMPTE ST 422:2013, SMPTE ST 422:2014 or future versions are not appropriate for use with this document.
[4] The reference to SMPTE ST 429-6:2006 is intentional. SMPTE ST 429-6:2006 Am1:2018 is not appropriate for use with this document.

4 Glossary and Acronyms

In addition to the glossary terms and acronyms presented here, the reader should also be aware of terms defined in SMPTE ST 377, SMPTE ST 429-7 and SMPTE ST 428-7.

d-cinema digital cinema

DCP Digital Cinema Package

ISAN International Standard Audiovisual Number

UMID Unique Material Identifier

UUID Universally Unique Identifier

XML eXtensible Markup Language

5 Overview (Informative)

D-cinema content is composed of a number of distinct elements such as Composition Playlists and Track Files (d-cinema *assets*). For delivery to d-cinema systems, assets are combined into a logical D-Cinema Package (DCP). The syntax and semantics of these assets and the DCP are described by the family of d-cinema specifications depicted in Figure 1. To promote modularity and layering, each document has a limited scope and often defines a single structure or format.

This specification describes operational constraints applicable to the complete DCP. While structure-specific constraints are addressed in the document that defines a particular structure, this document defines constraints that apply to the combined set of structures that comprise a DCP. For instance, constraints specific to the Composition Playlist, such as those related to content markers, must be defined in the Composition Playlist (CPL) specification, whereas constraints that apply to the DCP as a whole, such as composition edit rate, will be defined in this document.

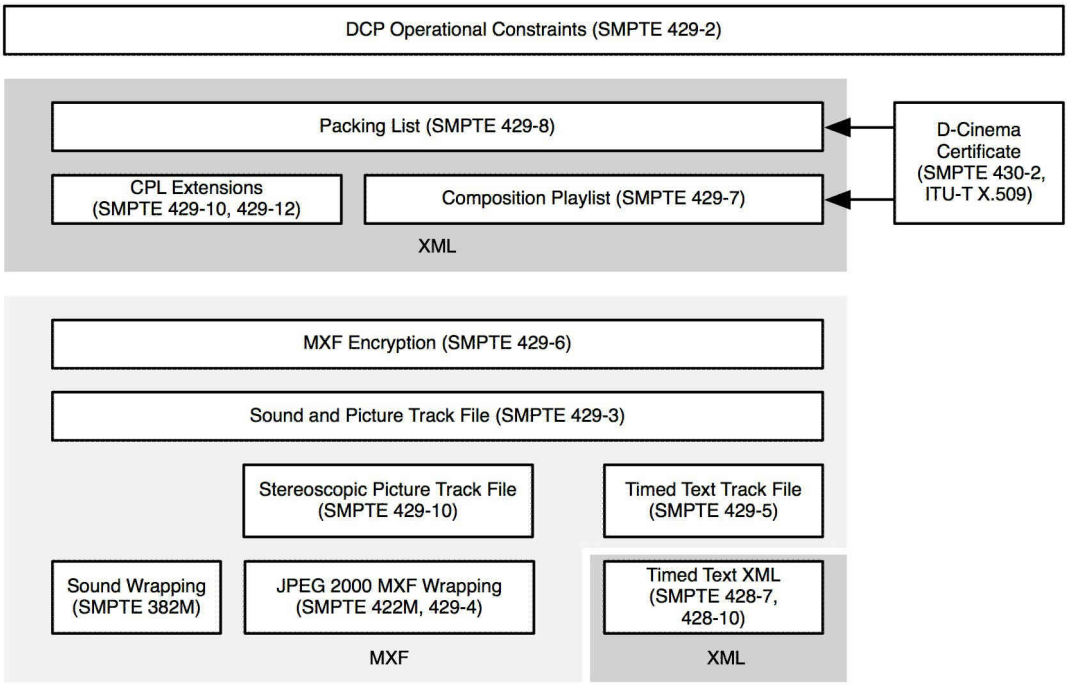

Figure 1 – DCP Family of Specifications

5.1 D-Cinema Package

A D-Cinema Package (DCP) is a set of files consisting of one (1) Packing List (SMPTE ST 429-8) and each of the files referenced by that Packing List. Figure 2 illustrates this structure. The figure shows a Packing List with ten asset references. Each asset reference points to one of the nine track files or the Composition Playlist. A Packing List may reference any combination of Track Files and Composition Playlists, however the set of referenced files must contain no duplicates.

A DCP may contain one or more complete Compositions, or it may contain components of compositions destined to complete, augment or replace previously distributed material.

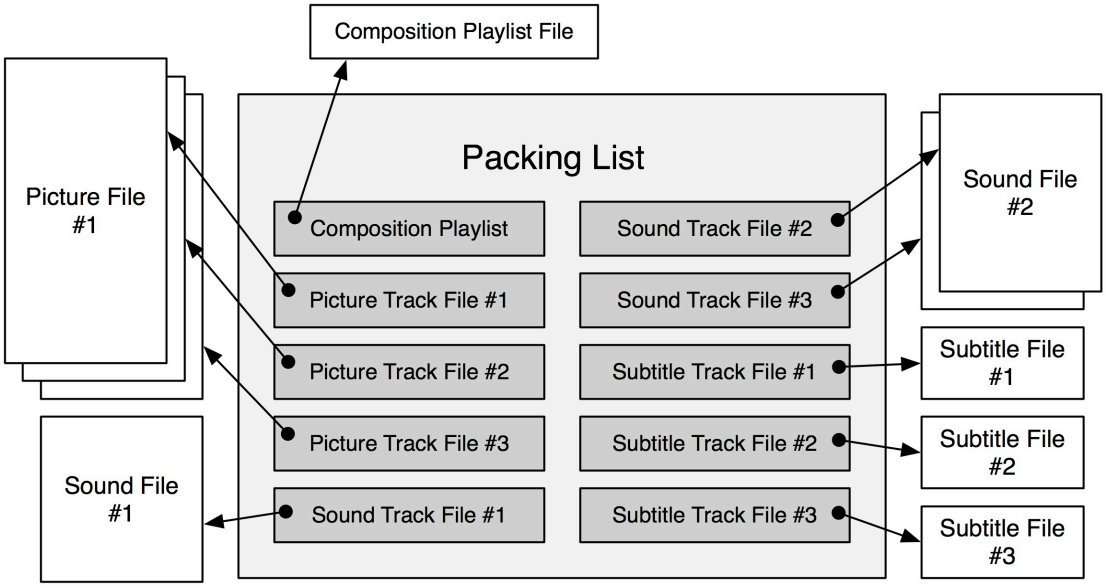

Figure 2 – A D-Cinema Package consists of a Packing List and the files to which it refers

5.2 D-Cinema Composition

A Composition is a set of files consisting of one (1) Composition Playlist document (SMPTE ST 429-7) and each of the Track Files (see Section 10 below) referred to from within that Composition Playlist. Figure 3 illustrates this structure for a composition having three reels of image, sound and subtitles.

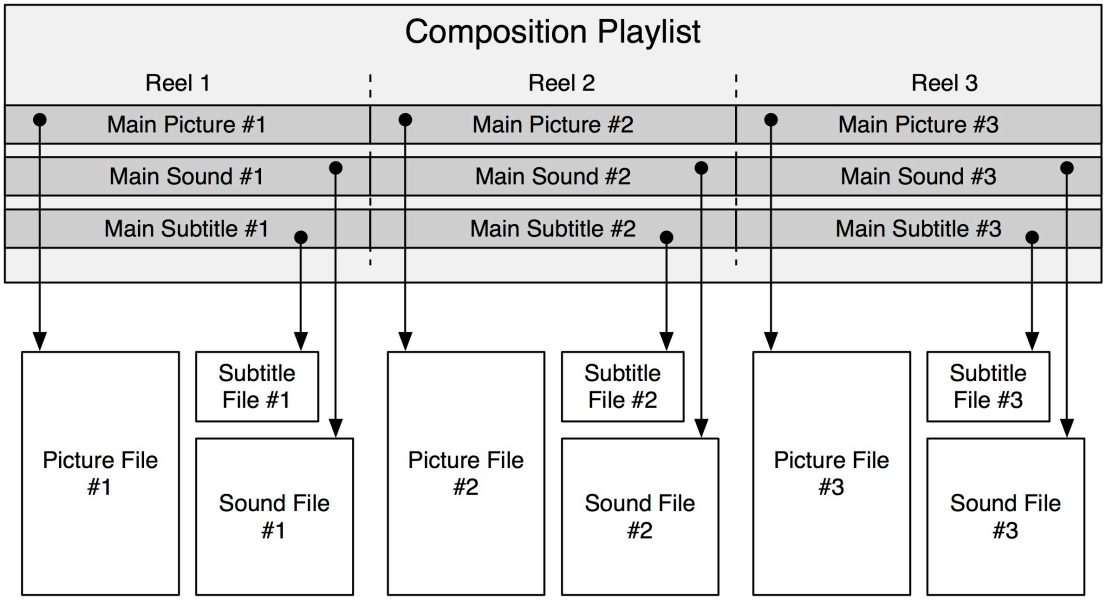

Figure 3 — A Composition consists of a Composition Playlist and the Track Files to which it refers

6 DCP Constraints

6.1 Minimum Contents

A DCP shall consist of one Packing List and one or more assets (i.e., Composition Playlists and/or Track Files), referenced by the Packing List.

6.2 UUID Generation

UUID values are used throughout the DCP to uniquely identify assets and data structures. All UUID values in a DCP shall be generated as specified in RFC 4122. UUID values which identify assets or encryption keys shall be generated using a truly-random or pseudo-random number source, and shall have a Version field value of '4' (0100b[5]) (RFC 4122).

6.3 XML Constraints

XML documents (SMPTE ST 428-7, SMPTE ST 429-7, SMPTE ST 429-8, SMPTE ST 429-10, SMPTE ST 429-12) in a DCP shall be encoded using the UTF-8 character encoding (ISO/IEC 10646-1) and shall comply with SMPTE ST 429-17).

7 Packing List Constraints

The Packing List document which defines the DCP contents shall be created as specified in SMPTE ST 429-8. Note that the specification requires that each Packing List document must have a unique UUID value in the top-level Id element. A Packing List may reference assets which are referenced by other Packing Lists.

7.1 Asset Identity

The value of the `Id` element within each `Asset` element shall be extracted from the referenced asset per the specification for the asset (see SMPTE ST 429-3 and SMPTE ST 429-7.)

7.2 Unique Set of Assets

Each `Asset` element shall contain an Id element value that is unique within the Packing List.

7.3 Digital Signature

When a Packing List document is digitally signed as specified in SMPTE ST 429-8, digital certificates in the signer's certificate chain shall conform to the provisions of SMPTE ST 430-2.

7.4 Group ID

7.4.1 Composition Packages

A *Composition Package* is a DCP containing only the complete set of assets comprising one or more compositions. The `GroupId` element shall not be present in the Packing List of a Composition Package.

[5] The 'b' suffix on this value indicates a binary encoding, most significant bit (MSB) first.

7.4.2 Asset Packages

An *Asset Package* is a DCP containing Track Files and/or Composition Playlists comprising one or more incomplete compositions (*i.e.*, some assets needed to complete the composition are not present in the package.) Asset Packages shall be identified by the presence of the `GroupId` element in the Packing List. An Asset Package should contain only related assets (*i.e.*, partial sets of assets from two unrelated compositions should be listed in separate Packing Lists using different `GroupId` values.) When two or more Asset Packages contain related assets, the Packing Lists should have the same `GroupId` value.

8 Composition Constraints

A Composition (i.e., a Composition Playlist and referenced Track Files) may be delivered in a single DCP or it may be spread across several DCPs. Regardless of the number of DCPs used to convey a Composition, a Composition shall conform to the following constraints.

8.1 Edit Rate

The composition shall have an Edit Rate of 24/1, 25/1, 30/1, 48/1, 50/1 or 60/1.

8.2 Picture Essence Encoding

Picture essence tracks shall be encoded as specified in SMPTE ST 428-1. The pixel array size and frame rate shall be one of the formats listed in Table 1. Monoscopic picture essence tracks shall have matching frame rate and edit rate. Stereoscopic picture essence tracks shall be limited to the 2K formats, and shall have a frame rate of 48/1 and an edit rate equal to half the frame rate ($r_e = r_f / 2$). (See SMPTE ST 429-10 for an explanation).

Source images having an aspect ratio not listed in Table 1 should be encoded so that the image fills either the horizontal or vertical dimension of the desired Full pixel array (2K or 4K). To fill the pixel array in the opposite dimension, the image should be padded with an equal number of black pixels on each side, *i.e.*, “letter-box” (top side, bottom side) or “pillar-box” (left side, right side).

Table 1 – Pixel Array Dimensions

Format	Horizontal Pixels	Vertical Pixels	Frame Rate
2K Scope (2.39:1)	2048	858	24/1, 25/1, 30/1, 48/1, 50/1 or 60/1
2K Flat (1.85:1)	1998	1080	24/1, 25/1, 30/1, 48/1, 50/1 or 60/1
2K Full (1.90:1)	2048	1080	24/1, 25/1, 30/1, 48/1, 50/1 or 60/1
4K Scope (2.39:1)	4096	1716	24/1, 25/1 or 30/1
4K Flat (1.85:1)	3996	2160	24/1, 25/1 or 30/1
4K Full (1.90:1)	4096	2160	24/1, 25/1 or 30/1

8.3 Sound Essence Encoding

Sound essence tracks shall be encoded as specified in SMPTE ST 428-2. Section 10.3.3 and Annex A specify means of identifying the content of these essence tracks.

8.4 Timed Text Essence Encoding

Timed Text essence shall be encoded as XML data as specified in SMPTE ST 428-7, and may be constrained per SMPTE ST 428-10. Sub-pictures shall be encoded as Portable Network Graphics (PNG) images as specified in ISO/IEC 15948.

8.4.1 Fonts for Timed Text

When `Text` elements are present in the Timed Text essence, one (1) `LoadFont` element shall be present. Timed Text essence shall not contain more than one (1) `LoadFont` element.

Within the scope of any given `Subtitle` element, all `Font` elements shall have the same `EffectSize` attribute value.

The font resource should not be larger than 10MB.

Note: Legacy implementations might not be able to support font resources larger than 640 KB.

Note: Operational testing has determined that a Font size smaller than 8pt might be difficult to read, and that, depending on the length of the subtitle, a very large Font size might take too long to appear and might go beyond the dimension of the Primary Picture.

8.4.2 Text Color Interpretation

Color values encoded in the Timed Text essence (in the `Color` and `EffectColor` attributes of the `Font` element) shall be encoded as sRGB values (IEC 61966-2-1).

8.4.3 Images for On-Screen Timed Text

PNG image resources used per SMPTE ST 428-7 shall have three (3) 8-bit color components (*R*, *G*, and *B*). An alpha channel may be present. If an alpha channel is present, the decoder shall use it when creating the composite image. PNG image resources shall contain the sRGB chunk per ISO/IEC 15948.

The width and height of a subpicture shall be equal to or less than the width and height, respectively, of the associated main picture.

8.4.4 Maximum Rate of Occurrence for On-Screen Timed Text

Up to two (2) subtitle instances may be visible on screen at any time. The visibility period of an instance shall include fade-in and fade-out times. A subtitle instance shall contain no more than six (6) `Text` elements or three (3) `Image` elements.

8.4.5 Constraints on Stereoscopic Control

All `Text` and `Image` elements to be displayed at the same time shall have the same depth information specified through `Zvalue` within `VariableZ` and/or `Zposition` attributes.

8.4.6 `IntrinsicPictureResolution` Attribute

When present, the value of the `IntrinsicPictureResolution` attribute of the `SubtitleReel` element (see SMPTE ST 428-7) shall be one of the values listed in Table 2 below.

Table 2 – `IntrinsicPictureResolution` Attribute Values

Attribute Value
2K Scope
2K Flat
2K Full
4K Scope
4K Flat
4K Full

Note: The `IntrinsicPictureResolution` attribute is intended to guide the mastering operator to select the appropriate subtitle resources for the Primary Picture content.

8.5 Sound and Picture Sample Rates

The sample rate of sound essence in a Composition shall be one of the combinations *a*-*l* listed in Table 3.

Table 3 – Sample Rate Constraints

	Sound Sample Rate	Composition Edit Rate	Samples per Edit Unit
a	48 kHz	24/1	2000
b	48 kHz	25/1	1920
c	48 kHz	30/1	1600
d	48 kHz	48/1	1000
e	48 kHz	50/1	960
f	48 kHz	60/1	800
g	96 kHz	24/1	4000
h	96 kHz	25/1	3840
i	96 kHz	30/1	3200
j	96 kHz	48/1	2000
k	96 kHz	50/1	1920
l	96 kHz	60/1	1600

8.6 Track File Edit Rates

All essence tracks in a Composition shall have an identical Edit Rate.

8.7 Homogenous Essence

Essence tracks in a Composition shall have homogenous encoding parameter values throughout the Composition. Picture essence shall have constant frame rate and pixel array size. Sound essence shall have constant sample rate, language, channel count, and channel assignment parameters.

9 Composition Playlist Constraints

9.1 Minimum Essence Requirement

A Composition Playlist shall have one picture essence track and one sound essence track in each `Reel` element.

9.2 Composition Playlist Uniqueness

Two Composition Playlist documents having different contents shall have different values in the top-level Id element.

9.3 ContentVersion Id

The `Id` element within the `ContentVersion` element shall contain a URI value conforming to one of the following types:

a Basic UMID (SMPTE ST 2029)
an ISAN (RFC 4246)
a UUID (RFC 4122)
an EIDR (RFC 7972)

Note: The `Id` element of the `ContentVersion` element is intended to remain constant across multiple Composition Playlist instances referencing the same underlying content. For instance, both a pre-release and a final version of a Composition Playlist associated with the same feature can have the same `ContentVersion` / `Id`, while their `Id` elements are different. In a typical application, `ContentVersion` / `Id` can be used as a reference to an internal booking system.

9.4 Reel Duration

The `Duration` element shall be present within every Asset element that refers to an external track file. The value of all Duration elements in a reel, with the exception of timed text elements, shall be equal. The Duration of the Reel shall be determined by the `MainPicture` element, per the provisions of SMPTE ST 429-7, or the `MainStereoscopicPicture` element, whichever is present.

9.5 Track Files

Track files referenced by a Composition Playlist shall conform to the provisions of Section 10 of this document.

9.6 Picture Tracks

Each `Reel` element in a Composition Playlist document shall contain one (1) `MainPicture` element (SMPTE ST 429-7) or one (1) `MainStereoscopicPicture` element (SMPTE ST 429-10). This element shall refer to a Picture Track File as defined by SMPTE ST 429-3. If the element name is `MainStereoscopicPicture`, the referenced Track File shall also conform to SMPTE ST 429-10.

9.6.1 Essence Characteristics

All picture assets in a Composition Playlist shall have identical values for the following metadata items:

element name (*i.e.*, `MainPicture` or `MainStereoscopicPicture`)

`EditRate` element

`FrameRate` element

`ScreenAspectRatio` element

9.7 Sound Tracks

This element shall refer to a Sound Track File as defined by SMPTE ST 429-3.

9.7.1 Essence Characteristics

All sound assets in a Composition Playlist shall have identical values for the following metadata items:

`EditRate` element

`Language` element

9.8 Timed Text Tracks

A timed text track is established by the presence of a timed text asset (e.g. `MainSubtitle`, `MainCaption`, `ClosedSubtitle`, or `ClosedCaption`) in at least one Reel of a Composition. Once a timed text asset appears in one Reel, the established track shall be assumed to exist for the entire Composition, even if related timed text `Asset` elements are not present in all Reels.

Each `Reel` element in a Composition Playlist document may contain one on-screen text track, either `MainSubtitle` as defined by SMPTE ST 429-7 or `MainCaption` as defined by SMPTE ST 429-12. When present, the `MainSubtitle` element shall refer to a Timed Text Track File as defined by SMPTE ST 429-5, containing an XML resource conforming to SMPTE ST 428-7. When present, the `MainCaption` element shall refer to a Timed Text Track File as defined by SMPTE ST 429-5, containing an XML resource conforming to SMPTE ST 428-10. A Composition Playlist shall contain no more than one on-screen text track type (`MainSubtitle` or `MainCaption`).

Each `Reel` element in a Composition Playlist document may contain up to six (6) off-screen (closed) text tracks, using any combination of `ClosedSubtitle` and `ClosedCaption` elements as defined by SMPTE ST 429-12. When present, an off-screen text element shall refer to a Timed Text Track File as defined by SMPTE ST 429-5, containing an XML resource conforming to SMPTE ST 428-10. When more than one off-screen text track asset of the same type (`ClosedSubtitle` or `ClosedCaption`) is present, the `Language` attribute shall be used. The `Language` attribute value of each off-screen text track shall be unique among the set of similarly-typed off-screen text tracks. The value of the `Language` attribute shall be used to identify material of the same off-screen text track from Reel to Reel for each `Asset` type instance.

The maximum number of timed text tracks in a Composition Playlist document is seven (7); one (1) on-screen text track plus six (6) off-screen text tracks. Each off-screen text track with a unique combination of element name and Language shall be considered a distinct off-screen text track.

In order to illustrate the concepts in this section, the example diagram in Figure 4 shows a collection of Composition assets on the left, and a Composition with tracks on the right. Each reel shown on the left contains a number of off-screen timed text assets that appears to be within the specified limit of this standard. However, in the example, the number of off-screen text tracks possible is seven, which is more than that allowed by this

standard. The Composition on the right is correctly constrained. Note that each timed text track exists for the duration of the Composition, even though it might not be represented by an asset in every reel.

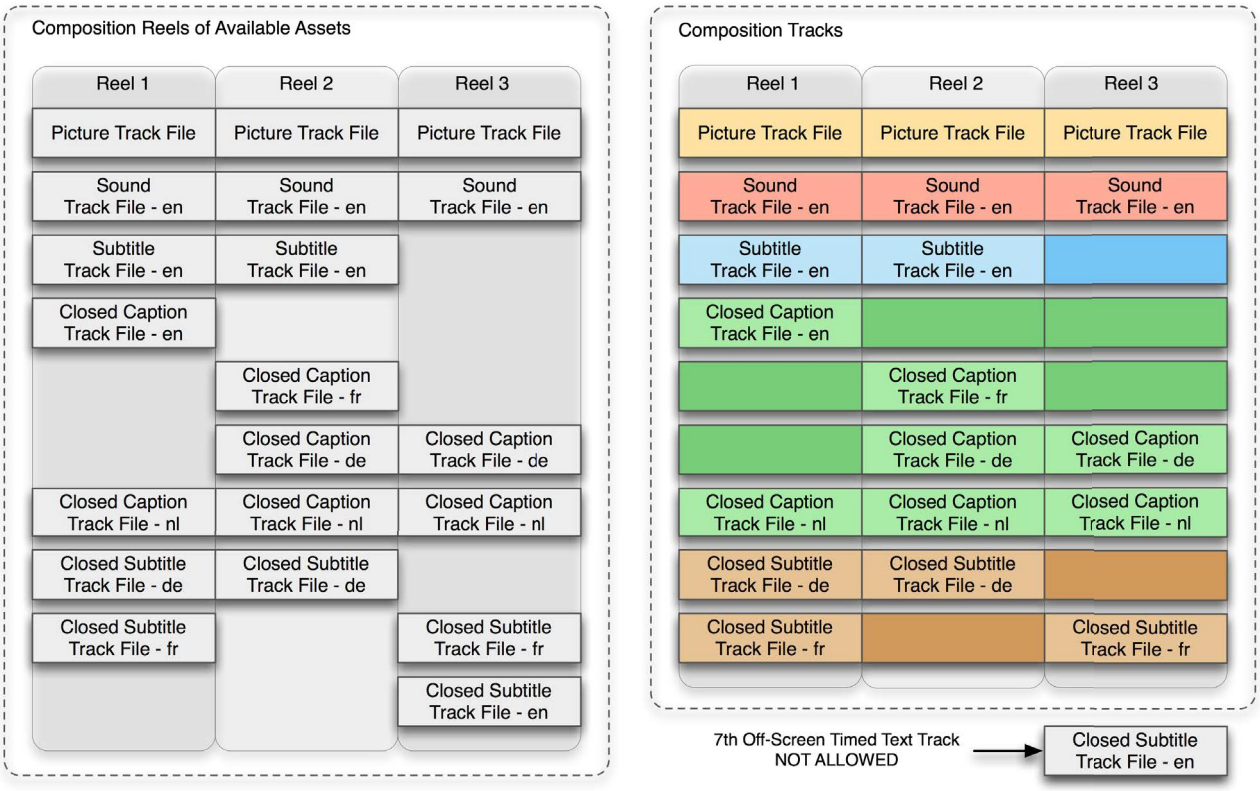

Figure 4 — Example of allocating timed text assets to timed text tracks

9.9 Marker Tracks

When present, a `MainMarkers` asset shall not contain any marker with an `Offset` value, minus the `EntryPoint` of the enclosing `MainMarkers` element, that exceeds the duration of the Reel.

9.10 Cryptographic Keys

No more than 256 distinct cryptographic keys, as uniquely identified by their Key ID, shall be used to encrypt the assets referenced by a Composition Playlist.

9.11 Hash Element

The `Hash` element shall be present in an asset when the `KeyId` element is present (*i.e.*, when the referenced Track File is encrypted).

9.12 Digital Signature

When a Composition Playlist document is digitally signed as specified in SMPTE ST 429-7, digital certificates in the signer's certificate chain shall conform to the provisions of SMPTE ST 430-2.

9.13 Composition Metadata

The `CompositionMetadataAsset` element defined in SMPTE ST 429-16 should be present.

10 Track File Constraints

Essence data shall be contained in MXF files (SMPTE ST 377).

10.1 Encryption

When cryptographic protection is required, Track Files shall use KLV encryption per SMPTE ST 429-6. Each encrypted Track File shall be encrypted with exactly one (1) 128-bit symmetric key.

10.2 Picture Track Files

In addition to the essence encoding constraints specified in Section 8 above, Picture Track Files shall have the following properties.

10.2.1 Operational Pattern

Picture Track Files shall conform to the provisions of SMPTE ST 429-3.

10.2.2 Compression

Picture essence shall be compressed using JPEG 2000 (ISO/IEC 15444-1).

There shall be 5 wavelet transform levels for 2K picture essence.

There shall be 6 wavelet transform levels for 4K picture essence.

10.2.3 Wrapping

Picture essence shall be frame wrapped according to SMPTE ST 422 and SMPTE ST 429-4. Stereoscopic picture essence shall also conform to SMPTE ST 429-10.

10.3 Sound Track Files

In addition to the essence encoding constraints specified in Section 8 above, Sound Track Files shall have the following properties.

10.3.1 Operational Pattern

Sound Track Files shall conform to the provisions of SMPTE ST 429-3.

10.3.2 Wrapping

Sound essence shall be frame wrapped per SMPTE ST 382. Sound essence shall be contained in KLV packets labeled with the Wave Frame Wrapped Element UL. A Wave Audio Essence Descriptor shall be present in the Top-Level File Package.

10.3.3 Channel Assignment

Channel assignment defines what reproduction channel is carried in each channel of the distributed track. Sound Track File channel assignment shall be indicated by a UL value in the Channel Assignment property of the Wave

Audio Essence Descriptor. The UL may indicate a fixed channel assignment. Annex A of this document defines a set of channel assignments and respective UL values based on this method. The UL may also indicate a channel assignment scheme defined in another specification. In this case, additional details regarding channel assignment shall be provided by the specification that defines the UL.

If the Channel Assignment property is not present, Channel Configuration 1 (Annex A, Table A.3) shall be assumed by the decoder. Routing of the container channel to the system audio output is not in the scope of this document.

10.4 Timed Text Track Files

In addition to the essence encoding constraints specified in Section 8 above, Timed Text Track Files shall have the following properties.

10.4.1 Timed Text Essence Format

Timed Text essence shall be encoded as XML data as specified in SMPTE ST 428-7, and may be constrained per SMPTE ST 428-10. See Sections 8.4 and 9.8 above.

10.4.2 Track File Format

Timed Text Track Files shall be created according to SMPTE ST 429-5.

10.4.3 Timed Text Essence Descriptor

If the DCDM Subtitle file contains the `IntrinsicPictureResolution` attribute (see SMPTE ST 428-7), then the “Intrinsic Picture Resolution” property of the Timed Text Essence Descriptor, defined in Annex D of this document, should be present in the Timed Text Track File and, when present, shall represent the same value.

If the DCDM Subtitle file contains the `DisplayType` element (see SMPTE ST 428-7), then the “Display Type” property of the Timed Text Essence Descriptor, defined in Annex D of this document, should be present in the Timed Text Track File and, when present, shall represent the same value.

If the Timed Text Essence Descriptor property “RFC 5646 Language Tag List” is present, it shall contain at least the language code specified in the DCDM Subtitle file.

If at least one subtitle instance of the DCDM Subtitle file contains a `Zposition` attribute (as defined in SMPTE ST 428-7), the “Z-Position In Use” property of the Timed Text Essence Descriptor shall be non-zero.

Annex A Audio Channel Assignment Label (Normative)

Note: Implementation behavior is undefined when a Sound Track File fails to adhere to the normative provisions specified herein.

SMPTE ST 382 carries multi-channel PCM sound samples by using sample interleave on a channel basis. Each sample position can be thought of as a channel within the SMPTE ST 382 container.

The number of channels within the Sound Track File shall be an even number. The inclusion of a channel of silence may be required to achieve this.

Annexes A.1 and A.2 each specifies a method for unambiguously identifying the channels present in Sound Track Files and indicating their intended reproduction location in the theater. Each method uses the ChannelAssignment property of the WaveAudioEssence Descriptor in a Sound Track File, as specified in Section 10.3.3 above.

Compliant playback devices shall use the ChannelAssignment property to identify the sound channels being used.

A.1 Static Container Channel Configurations

Each table in this Annex defines a container channel configuration that has a corresponding Universal Label (UL) for use as a value of the ChannelAssignment property. Container channels are numbered in sample packing order. The first sample is carried in container channel 1, the second in container channel 2 and so on.

The number of channels contained in a Sound Track file shall be less than or equal to the number of channels defined by the table associated with the ChannelAssignment property. However, if a given container channel is present, it shall be used according to the table. The WaveAudioEssence Descriptor ChannelCount property may be used in combination with the ChannelAssignment property to determine actual channel usage. For instance, a ChannelAssignment label indicating Channel Configuration 1 may accompany a container with a ChannelCount value of 6, indicating that channels 7 and 8 (Hearing Impaired and Visually Impaired-Narrative) are not present.

The special case of no specified channel configuration is also provided for. See Table A.6, Channel Configuration 4. The label associated with this table shall mean “no configuration specified”. This may be used for test or experimental purposes.

Note: For the purpose of setting appropriate transport flags, implementations should not assume that all audio channels in Channel Configuration 4 contain linear PCM audio samples suitable for direct conversion to an analog audio signal.

A.1.1 Channel Label Set ULs

Table A.1 – Specification of the Channel Assignment Label when Static Container Channel Configurations are used

Byte No.	Description	Value (hex)	Meaning
1-7	Registry Designator	See SMPTE ST 400	
8	Registry Version Number	0bh	Version of SMPTE RP 224 in which this label first appears
9	Parametric	04h	Node used to define parametric data
10	Sound Essence	02h	Identifies sound essence coding
11	Sound Coding Characteristics	02h	Identifies sound coding characteristics
12	Sound Channel Labeling	10h	Identifies sound channel labeling
13	Sound Channel Labeling SMPTE ST 429-2	03h	Identifies sound channel labeling as defined in this document (SMPTE ST 429-2)
14	Channel Label Sets	01h	Identifies Static Sound Channel Label Sets
15	Channel Configuration	See Table A.2	Identifies sound Channel Configuration
16	Reserved	00h	Reserved

Table A.2 – Values for Table A.1, Byte 15

Channel Configuration	Byte 15 Value
Channel Configuration 1 (Table A.3)	01h
Channel Configuration 2 (Table A.4)	02h
Channel Configuration 3 (Table A.5)	03h
Channel Configuration 4 (Table A.6)	04h
Channel Configuration 5 (Table A.7)	05h

A.1.2 Channel Configuration Tables

Table A.3 – Channel Configuration 1

Container Channel	SMPTE ST 428-12 Name
1	Left
2	Right
3	Center
4	LFE
5	Left Surround
6	Right Surround
7	Hearing Impaired
8	Visually Impaired-Narrative

Table A.4 – Channel Configuration 2

Container Channel	SMPTE ST 428-12 Name
1	Left
2	Right
3	Center
4	LFE
5	Left Surround
6	Right Surround
7	Center Surround
8	Not Used
9	Hearing Impaired
10	Visually Impaired-Narrative

Table A.5 – Channel Configuration 3

Container Channel	SMPTE ST 428-12 Name
1	Left
2	Right
3	Center
4	LFE
5	Left Surround
6	Right Surround
7	Left Center
8	Right Center
9	Hearing Impaired
10	Visually Impaired-Narrative

Table A.6 – Channel Configuration 4

Container Channel	Name
1	CH01
2	CH02
3	CH03
4	CH04
5	CH05
6	CH06
7	CH07
8	CH08
9	CH09
10	CH10
11	CH11
12	CH12
13	CH13
14	CH14
15	CH15
16	CH16

Table A.7 – Channel Configuration 5

Container Channel	SMPTE ST 428-12 Name
1	Left
2	Right
3	Center
4	LFE
5	Left Side Surround
6	Right Side Surround
7	Left Rear Surround
8	Right Rear Surround
9	Hearing Impaired
10	Visually Impaired-Narrative

Note: Earlier revisions of this specification used terminology from SMPTE ST 428-3, instead of SMPTE ST 428-12, to define the mappings from container channels to audio channels. Although the mappings remain unchanged, the terms used to refer to a few of the audio channels have changed. For instance, SMPTE ST 428-12 differentiates Side Surrounds (Lss/Rss) from Left and Right surrounds (Ls/Rs) and uses Lrs to refer to the Left Rear Surround channel, whereas SMPTE ST 428-3 uses Rls.

A.2 Configurations using MXF Multichannel Audio Framework

When the ChannelAssignment of the WaveAudioEssence Descriptor in a Sound Track File contains the UL defined in Table A.8, the framework specified in SMPTE ST 377-4 shall be used in conjunction with the constraints defined in Annexes A.2.2, A2.3 and C to unambiguously identify the audio channels and soundfield group carried in the Sound Track File.

Note: Items defined in SMPTE ST 377-4 that are not specified in this section can nevertheless be present in the Sound Track File and describe particular aspects of an audio channel or soundfield group. Implementations can safely ignore these items.

The MXF Multichannel Audio Framework (MCA Framework) associates audio channels and soundfield groups contained within a D-Cinema Sound Track File with an MXF SubDescriptor that contains metadata, including a unique identifier. This enables D-Cinema implementations to properly route and process audio channels, e.g. the Hearing Impaired and Left channels may be handled by different devices. It also enables straightforward extensibility for the purpose of both experimentation and widespread use: new standalone audio channels can be defined without impacting existing soundfield groups and new soundfield groups can be introduced with minimal effort.

Figure 5 illustrates the use of the audio channel and soundfield group information contained in a Sound Track File, as specified here.

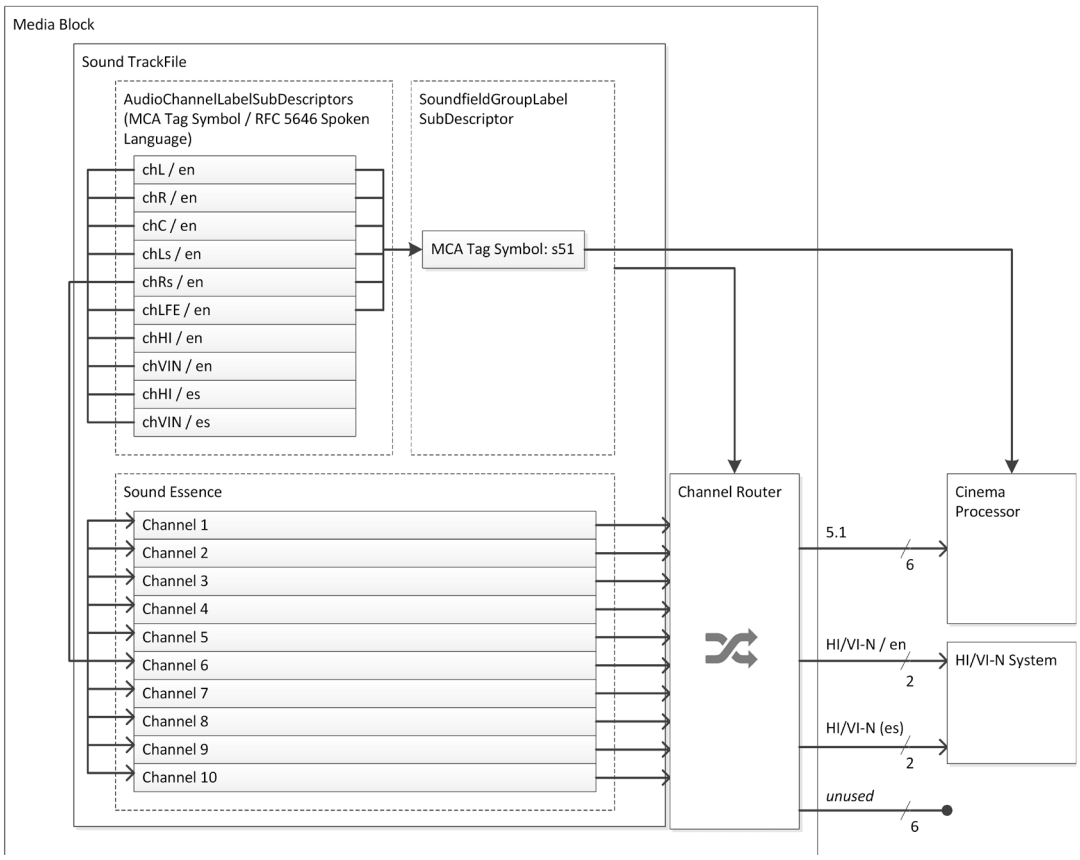

Figure 5 – Illustrative use of AudioChannelLabelSubDescriptor and SoundfieldGroupLabelSubDescriptor for a Sound Track File containing 10 audio channels consisting of a 5.1 soundfield group and associated Hearing Impaired and Visually Impaired-Narrative channels (Informative). The audio channel labeling method defined in this section is not limited to this specific channel count or soundfield configuration.

A.2.1 Configuration Channel Assignment Label

Table A.8 – Specification of the Channel Assignment Label when the MCA Framework is used

Byte No.	Description	Value (hex)	Meaning
1-7	Registry Designator	See SMPTE ST 400	
8	Registry Version Number	0D	Version of SMPTE RP 224 in which this label first appears
9	Parametric	04h	Node used to define parametric data
10	Sound Essence	02h	Identifies sound essence coding
11	Sound Coding Characteristics	02h	Identifies sound coding characteristics
12	Sound Channel Labeling	10h	Identifies sound channel labeling
13	Sound Channel Labeling SMPTE ST 429-2	03h	Identifies sound channel labeling as defined in this document (SMPTE ST 429-2)
14	D-Cinema Application of the MXF Multichannel Audio Framework	02h	Indicates that the D-Cinema Application of the MXF Multichannel Audio Framework is used
15	Reserved	00h	Reserved
16	Reserved	00h	Reserved

A.2.2 AudioChannelLabelSubDescriptor

Each audio channel contained in the Sound Track File shall be associated with zero or one AudioChannelLabelSubDescriptor instance, and each AudioChannelLabelSubDescriptor instance shall be associated with an audio channel.

Implementations shall ignore audio channels not associated with an AudioChannelLabelSubDescriptor instance. These channels should contain silence.

Note: The ChannelCount property of the Wave Audio Essence Descriptor reflects the number of channels in the Sound Track File and not the number of AudioChannelLabelSubDescriptor instances.

In addition to the items required by SMPTE ST 377-4, the following items shall be present in every AudioChannelLabelSubDescriptor instance:

- MCA Channel ID
- MCA Tag Name
- RFC 5646 Spoken Language
- SoundfieldGroupLinkID, if and only if the audio channel referenced by the AudioChannelLabelSubDescriptor instance belongs to a soundfield group associated with a SoundfieldGroupLabelSubDescriptor instance. If present, SoundfieldGroupLinkID shall contain the MCA Link ID value of the associated SoundfieldGroupLabelSubDescriptor instance.

Not all audio channels present in a Sound Track File need to be associated with a soundfield group. For example, Hearing Impaired and Visually Impaired-Narrative channels, if present, do not belong to a soundfield group and, hence, their respective AudioChannelLabelSubDescriptor instances do not reference a SoundfieldGroupLabelSubDescriptor instance.

If an audio channel is associated with a soundfield group, then the value of their respective RFC 5646 Spoken Language items shall be equal.

A.2.2.1 Common D-Cinema Channels

Implementations shall recognize the common D-Cinema audio channels defined in Table 1 of SMPTE ST 428-12.

The presence of such an audio channel shall be indicated by an AudioChannelLabelSubDescriptor instance whose MCA Label Dictionary ID value is equal to a UL value defined by the combination of column 1 of Table 1 and Table 2 of SMPTE ST 428-12.

The MCA Tag Name of such an AudioChannelLabelSubDescriptor instance shall be equal to the Name (as specified in SMPTE ST 428-12) of the audio channel associated with the UL value.

The MCA Tag Symbol item of such an AudioChannelLabelSubDescriptor instance shall be constructed by prepending the string "ch" to the Symbol (as specified in SMPTE ST 428-12) of the audio channel associated with the UL value.

No channel listed in Table 1 of SMPTE ST 428-12 shall appear more than once in a given Sound Track File with the exception of Hearing Impaired and Visually Impaired-Narrative channels. If there are multiple Hearing Impaired or Visually Impaired-Narrative channels in a Sound Track File, they shall be distinguished by the value of their RFC 5646 Spoken Language item.

Furthermore, the RFC 5646 Spoken Language item shall not have the same value in two or more audio channels labeled Hearing Impaired, and the RFC 5646 Spoken Language item shall not have the same value in two or more audio channels labeled Visually Impaired-Narrative.

A.2.2.2 Extension Channels

For extensibility, channels not defined in Table 1 of SMPTE ST 428-12 may be present.

Implementations shall not automatically pre-assign an audio channel with an AudioChannelLabelSubDescriptor instance having a MCA Label Dictionary ID that the implementation does not recognize and, for the purpose of setting appropriate transport flags, should not assume that such an audio channel contains linear PCM audio samples suitable for direct conversion to an analog audio signal.

Implementations may display to the user channels associated with an MCA Label Dictionary ID they do not recognize and offer the user the option to take action on such a channel based on the MCA Tag Name, MCA Tag Symbol and RFC 5646 Spoken Language of the AudioChannelLabelSubDescriptor instance that references it.

A.2.3 SoundfieldGroupLabelSubDescriptor

There shall be one and only one SoundfieldGroupLabelSubDescriptor instance in the Sound Track file.

In addition to the items required by SMPTE ST 377-4, the following items shall be present in the SoundfieldGroupLabelSubDescriptor instance:

- MCA Tag Name
- RFC 5646 Spoken Language

A.2.3.1 Common D-Cinema Soundfield Groups

Implementations shall recognize the common D-Cinema soundfield groups listed in Table 3 of SMPTE ST 428-12.

The presence of such a soundfield group shall be indicated by SoundfieldGroupLabelSubDescriptor instance whose MCA Label Dictionary ID value is equal to one of the UL values defined by the combination of column 1 of Table 3 and Table 4 of SMPTE ST 428-12.

The MCA Tag Name of such a SoundfieldGroupLabelSubDescriptor instance shall match the value of the Name of the soundfield group (as specified in SMPTE ST 428-12) associated with the UL value.

The MCA Tag Symbol item of such an SoundfieldGroupLabelSubDescriptor instance shall be constructed by prepending the string "sg" to the Symbol of the soundfield group (as specified in SMPTE ST 428-12) associated with the UL value.

Not all channels listed in the "Audio Channels" column of a given soundfield group in Table 3 of SMPTE ST 428-12 need to be present in the sound track file, but only those channels listed in the "Audio Channels" column for a given soundfield group may reference that SoundfieldGroupLabelSubDescriptor instance. Furthermore, if a channel is listed in the "Audio Channels" column of a given soundfield group but absent in the Sound Track File, then implementations shall assume the channel was not intended for reproduction by the content provider.

Note: Implementations may indicate to the user if a channel listed in the “Audio Channels” column for a given soundfield group is not present.

A.2.3.2 Extension Soundfield Groups

For extensibility, soundfield groups not defined in Table 3 of SMPTE ST 428-12 may be present. However, implementations shall take no action with a SoundfieldGroupLabelSubDescriptor instance having a MCA Label Dictionary ID that the implementation does not recognize or if a channel that is not listed in the "Audio Channels" column for a given soundfield group references that SoundfieldGroupLabelSubDescriptor instance.

Note: Implementations can use the SoundfieldGroupLabelSubDescriptor instance for display to the user and to appropriately configure the B-Chain for the intended soundfield reproduction.

Annex B SMPTE ST 377-4 Provisions (Normative)

Within the scope of this document, the provisions of SMPTE ST 377-4 shall apply with the following exceptions:

- SMPTE ST 377-4, Introduction Section:
  - In the third paragraph, the following sentence:

    The MCALabelSubDescriptor and its three subclasses are the core of the Multichannel Audio Labeling Framework and are all derived from the MXF SubDescriptor as defined in SMPTE ST 377-1, Annex B.3.

    shall be replaced with:

    The MCALabelSubDescriptor and its three subclasses are the core of the Multichannel Audio Labeling Framework.
- SMPTE ST 377-4, Section 3:
  - The reference to SMPTE ST 377-1:2011 shall be replaced with the reference to SMPTE ST 377:2004
  - The reference to Amendment 1:2012 to SMPTE ST 377-1:2011 shall be disregarded
  - The reference to Amendment 1:2012 to SMPTE ST 382-2007 shall be disregarded
- SMPTE ST 377-4, Section 4:
  - The reference to SMPTE ST 377-1:2011 shall be replaced with reference to SMPTE ST 377:2004
- SMPTE ST 377-4, Section 5.3:
  - The entire paragraph shall be replaced with the following one:

    The specification defines only a minimal set of items for each MCALabelSubDescriptor subclass. Applications using the framework may therefore define additional items, and MCALabelSubDescriptors may therefore contain items that are particular to only that application. MCALabelSubDescriptor extension shall be accomplished in one of two ways: By registering additional metadata items and associated UL's for the MCALabelSubDescriptors defined herein, extending the subclasses, or by creating new MCALabelSubDescriptor subclasses by registering new KLV keys as specified in Table 1 and Table 2. These additional items shall be ignored if they are not recognized by a different implementation.
- SMPTE ST 377-4, Section 6:
  - The following shall be added before the start of Section 6.1:

    In order to allow Essence Descriptors to reference MCALabelSubDescriptors, the MXF Generic Descriptor (as defined in SMPTE ST 377:2004) has an additional optional property as defined in the table below. The Local Tag value associated with this additional optional property (called "Sub Descriptors") shall be dynamically allocated (dynamic) as defined in SMPTE ST 377:2004. 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 377:2004. The full 16-byte UL value is defined in SMPTE RP 210.

Additional Optional Property for the MXF Generic Descriptor

Item Name	Type	Len	Local Tag	Item Designator	Req ?	Meaning	Default
All elements from the Generic Descriptor defined in SMPTE ST 377:2004, Table 17							
Sub Descriptors	Array of StrongRef (Sub Descriptors)	8+ 16n	dynamic	06.01.01.04. 06.10.00.00	Opt	Array of strong references to Sub Descriptor Sets	

Note: The method of adding a Sub-Descriptor was first used in SMPTE ST 422:2006.

- SMPTE ST 377-4, Section 6.2:
  - The reference to SMPTE ST 377-1:2011 shall be replaced with reference to SMPTE ST 377:2004
- SMPTE ST 377-4, Section 6.3, Table 3:
  - "All elements from the SubDescriptor set defined in SMPTE ST 377-1, Annex B.3" shall be replaced with the two following items:

Instance UID	UUID	16	3C.0A	01.01.15.02. 00.00.00.00	Req	Unique ID of this instance SMPTE RP 210 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 SMPTE RP 210 Specifies the reference to an overall modification

- SMPTE ST 377-4, Section 6.3.5:
  - The reference to "Amendment 1:2012 to SMPTE ST 377-1:2011" shall be replaced with the reference to "SMPTE ST 382:2007, Material Exchange Format - Mapping AES3 and Broadcast Wave Audio into the MXF Generic Container"
- SMPTE ST 377-4, Annex B:
  - The entire content of Annex B shall be replaced with the following paragraphs:

    The Multichannel Audio Framework ("MCA Framework") is a hierarchical class system. It is based on the concept of Descriptors and Sub Descriptors as defined in SMPTE ST 377:2004, Section 9.5, and in this document, Section 6. The hierarchical classes and the reference concepts laid forth in these documents are the basis for the MCA Framework object model. An informative summary is presented here as a means to introduce the specifics of the MCA Framework object model and tie it into existing MXF structure.

    In the MXF structure, Descriptors are derived from the Generic Descriptor, which is defined in SMPTE ST 377:2004, Table 17. The Generic Descriptor is never used directly (*i.e.*, it is an abstract superclass). It is required for each individual Descriptor that is derived directly or indirectly from the Generic Descriptor that the Item Designator, Set Keys and Properties be defined. Section 6 of this document adds an array of strong references to Sub Descriptors to the Generic Descriptor.

    It should be noted that Generic Descriptor::Sub Descriptors and Multiple Descriptor::Sub Descriptor UIDs are distinct. The Multiple Descriptor::Sub Descriptor UIDs describes multiple interleaved tracks, whereas Generic Descriptor::Sub Descriptors is used to collect subdescriptors for a single Track.

    The MCA Framework object model defines a parent abstract superclass named "MCALabelSubDescriptor". The concrete subclasses are the AudioChannelLabelSubDescriptor, the SoundfieldGroupSubDescriptor, and the GroupOfSoundfieldGroupsSubDescriptor. The MCALabelSubDescriptor and its three subclasses are the core of the Multichannel Audio Labeling

structure. The metadata elements associated with each are considered the “set” of metadata elements for that subclass and are referred to as such in this document.

Thusly each component of a 5.1 program, whether compressed or uncompressed, is associated with an AudioChannelLabelSubDescriptor and the program as a whole with a SoundfieldGroupLabelSubDescriptor, which is then handled as an entity. The GroupOfSoundfieldGroupsLabelSubDescriptor can be used, for instance, to label a transmission of 5.1, Lt-Rt, mono VI and mono SAP audio essence carried as independent soundfield groups in a single transport file or in individual packages. The GSG labels could also be used if the audio essence is carried in a single encoded bitstream, such as Dolby E, multiple audio channels within a single file, as with D-Cinema Track Files, or in individual packages. The soundfield groups might be played out separately or simultaneously depending on the specific implementation.

The Group of Soundfield Groups could also be used for applications in simultaneous ingest, mix-down or playout, among others. One such use might be to associate multiple soundfield groups to form an “English” version or a “French” version or a “Profanity-free” version.

Annex C Additional Timed Text Essence Descriptor Items

The items listed below are additional properties that may be use in the Timed Text Essence Descriptor when creating MXF Timed Text Track Files as defined by SMPTE ST 429-5. The usage of these items is detailed in the main body of this document.

Table D.1 – Timed Text Essence Descriptor - Additional Properties

Item Name	Type	Len	UL Designator	Req ?	Meaning	Default
Display Type	UTF16 String	var	06.0E.2B.34 01.01.01.0E 06.01.01.02 04.00.00.00	Opt	A text string giving an application specific means to indicate the intended use of the content of the XML document	none
Intrinsic Picture Resolution	UTF16 String	var	06.0E.2B.34 01.01.01.0E 06.01.01.02 05.00.00.00	Opt	Indicates the resolution of the primary picture on which Sub-Picture Ancillary Resources are to be rendered	none
RFC 5646 Language Tag List	UTF16 String	var	06.0E.2B.34 01.01.01.0E 03.01.01.02 02.16.00.00	Opt	An RFC 5646 Language Tag List that identifies one or more languages as a comma-separated list of RFC 5646 language tags.	empty
Z-Position In Use	UInt8	1	06.0E.2B.34 01.01.01.0E 06.01.01.02 06.00.00.00	Opt	When non-zero, indicates that one or more subtitle instances in the enclosed XML resource make use of stereoscopic positioning features.	00h

Bibliography (Informative)

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 336:2007) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 336M-2007). Documents with the same root number (e.g. 336) and publication year (e.g. 2007) are functionally identical.

SMPTE RP 224[6] SMPTE Labels Register

[6] The labels registry is a dynamic document and the version number and date are those at time of access.