SMPTE STANDARD

D-Cinema Packaging — DCP Operational Constraints

Table of Contents

Approved
March 11, 2009

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.

SMPTE Standard 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 Standard. 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

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

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

[SMPTE 400M-2004] Television — SMPTE Labels Structure

[SMPTE 422M-2006] Material Exchange Format — Mapping JPEG 2000 Codestreams into the MXF Generic Container

[SMPTE 428-1-2006] D-Cinema Distribution Master (DCDM) — Image Characteristics

[SMPTE 428-2-2006] D-Cinema Distribution Master — Audio Characteristics

[SMPTE 428-3-2006] D-Cinema Distribution Master — Audio Channel Mapping and Channel Labeling

[SMPTE 428-7-2007] D-Cinema Distribution Master —Subtitle

[SMPTE 428-10-2008] D-Cinema Distribution Master — Closed Caption and Closed Subtitle

[SMPTE 429-3-2007] D-Cinema Packaging — Sound and Picture Track File

[SMPTE 429-4-2006] D-Cinema Packaging — MXF JPEG 2000 Application

[SMPTE 429-5-2009] D-Cinema Packaging — Timed Text Track File

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

[SMPTE 429-7-2006] D-Cinema Packaging — Composition Playlist

[SMPTE 429-8-2007] D-Cinema Packaging — Packing List

[SMPTE 429-10-2008] D-Cinema Packaging — Stereoscopic Picture Track File

[SMPTE 429-12-2008] D-Cinema Packaging — Caption and Closed Subtitle

[SMPTE 430-2-2006] D-Cinema Operations — Digital Certificate

[SMPTE 2029-2009] Uniform Resource Names for SMPTE Resources

[ISO/IEC 10646:2003] Information Technology — Universal Multiple-Octet Coded Character Set (UCS)

[ISO/IEC 15444-1:2004] Information Technology — JPEG 2000 Image Coding System: Core Coding System

[ISO/IEC 15444-1:2004/Amd 1:2006] Profiles for Digital Cinema Applications

[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

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 377M] and [SMPTE 429-7].

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.

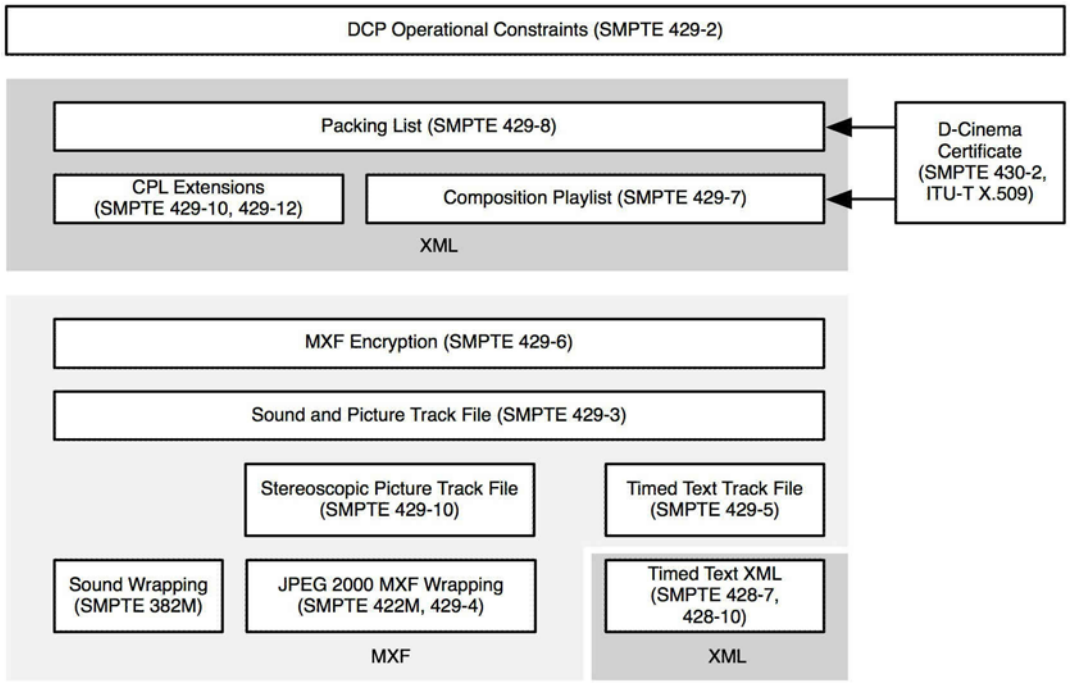

Figure 1 – DCP Family of Specifications

5.1 D-Cinema Package

A D-Cinema Package (DCP) is a set of files consisting of one (1) Packing List [SMPTE 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.

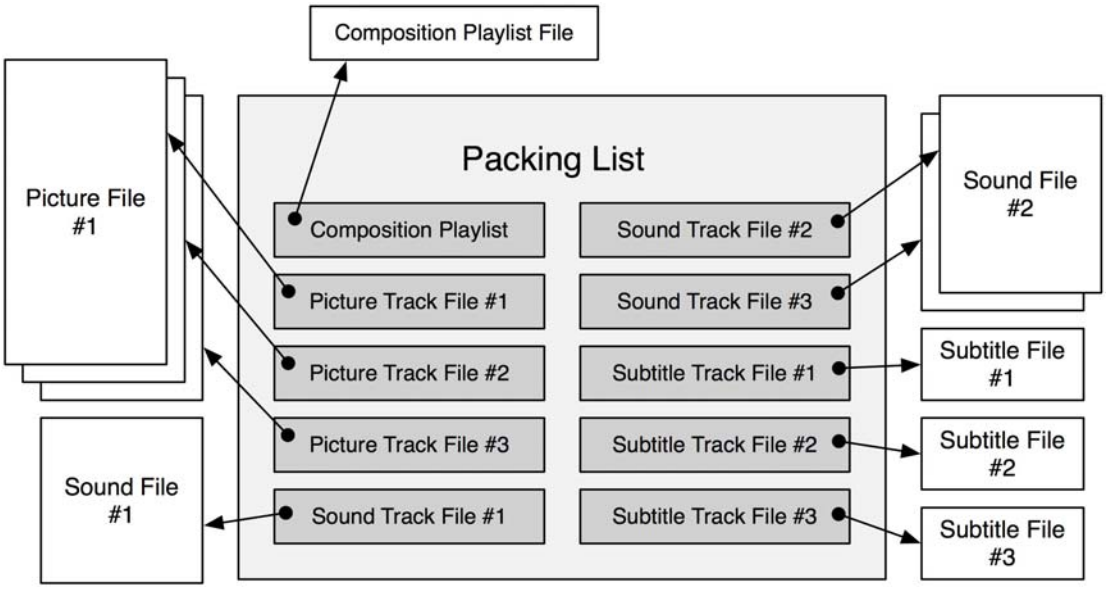

Figure 2 – A D-Cinema Package consists of a Packing List and the files to which it refers

5.2 D-Cinema Composition

A Composition is a set of files consisting of one (1) Composition Playlist document [SMPTE 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.

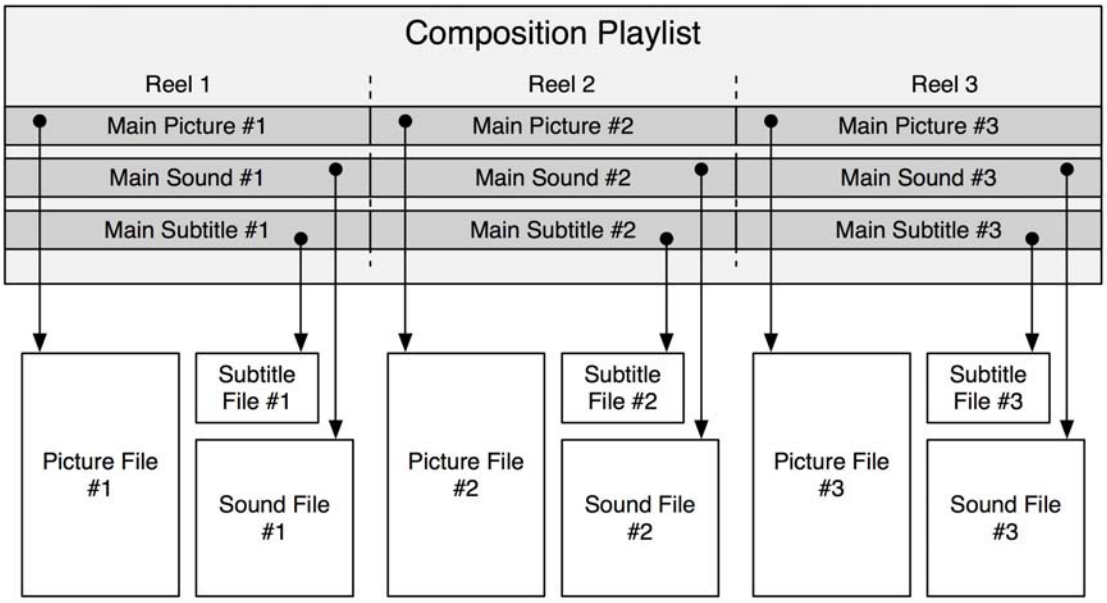

Figure 3 — A Composition consists of a Composition Playlist and the Track Files to which it refers

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' (`0100`b[1]) [RFC 4122].

6.3 XML Character Encoding

XML documents ([SMPTE 428-7], [SMPTE 429-7], [SMPTE 429-8]) in a DCP shall be encoded using the UTF-8 character encoding [ISO/IEC 10646-1].

7 Packing List Constraints

The Packing List document which defines the DCP contents shall be created as specified in [SMPTE 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 429-3] and [SMPTE 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 429-8], digital certificates in the signer's certificate chain shall conform to the provisions of [SMPTE 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.

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

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

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 or 48/1.

8.2 Picture Essence Encoding

Picture essence tracks shall be encoded as specified in [SMPTE 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 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 or 48/1
2K Flat (1.85:1)	1998	1080	24/1 or 48/1
2K Full (1.90:1)	2048	1080	24/1 or 48/1
4K Scope (2.39:1)	4096	1716	24/1
4K Flat (1.85:1)	3996	2160	24/1
4K Full (1.90:1)	4096	2160	24/1

8.3 Sound Essence Encoding

Sound essence tracks shall be encoded as specified in [SMPTE 428-2]. The set of sound channels comprising the Sound essence shall be restricted to the channels and loudspeaker positions defined in [SMPTE 428-3], "Definition of terms".

8.4 Timed Text Essence Encoding

Timed Text essence shall be encoded as XML data as specified in [SMPTE 428-7], and may be constrained per [SMPTE 428-10]. Sub-pictures shall be encoded as Portable Network Graphics (PNG) images as specified in [ISO/IEC 15948:2004].

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.

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 interpreted as sRGB values [IEC 61966-2-1:1999].

8.4.3 Images for On-Screen Timed Text

PNG image resources used per [SMPTE 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. Regardless of the headers present in a PNG image file, the image shall be interpreted as though the sRGB chunk is present, having a `Rendering intent` value of 3 (three*)* (“absolute colorimetric”, see [ISO/IEC 15948:2004]).

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 three (3) 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.

8.5 Sound and Picture Sample Rates

The sample rate of sound essence in a Composition shall be one of the combinations *a*-*d* listed in Table 2.

Table 2 – Sample Rate Constraints

	Sound Sample Rate	Composition Edit Rate	Samples per Edit Unit
a	48 kHz	24/1	2000
b	48 kHz	48/1	1000
c	96 kHz	24/1	4000
d	96 kHz	48/1	2000

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.

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 URN value conforming to one of the following types:

- a Basic UMID [SMPTE 2029]
- an ISAN [RFC 4246]
- a UUID [RFC 4122]

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 shall be equal.

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 429-7] or one (1) `MainStereoscopicPicture` element [SMPTE 429-10]. This element shall refer to a Picture Track File as defined by [SMPTE 429-3]. If the element name is `MainStereoscopicPicture`, the referenced Track File shall also conform to [SMPTE 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 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

Each `Reel` element in a Composition Playlist document may contain one on-screen text track, either `MainSubtitle` as defined by [SMPTE 429-7] or `MainCaption` as defined by [SMPTE 429-12]. When present, the `MainSubtitle` element shall refer to a Timed Text Track file as defined by [SMPTE 429-5], containing an XML resource conforming to [SMPTE 428-7]. When present, the `MainCaption` element shall refer to a Timed Text Track file as defined by [SMPTE 429-5], containing an XML resource conforming to [SMPTE 428-10].

Each `Reel` element in a Composition Playlist document may contain up to six (6) closed text tracks, using any combination of `ClosedSubtitle` and `ClosedCaption` elements as defined by [SMPTE 429-12]. When present, a closed text track element shall refer to a Timed Text Track file as defined by [SMPTE 429-5], containing an XML resource conforming to [SMPTE 428-10]. When more than one closed text track of the same type (`ClosedSubtitle` or `ClosedCaption`) is present, the `Language` attribute shall be used. The `Language` attribute value of each closed text track shall be unique among the set of similarly-typed closed text tracks. The value of the `Language` attribute shall be used to identify material of the same closed 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) closed 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 429-7], digital certificates in the signer's certificate chain shall conform to the provisions of [SMPTE 430-2].

10 Track File Constraints

Essence data shall be contained in MXF files [SMPTE 377M].

10.1 Encryption

When cryptographic protection is required, Track Files shall use KLV encryption per [SMPTE 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 429-3].

10.2.2 Compression

Picture essence shall be compressed using JPEG 2000 [ISO/IEC 15444-1] as constrained by [ISO/IEC 15444-1:2004/Amd 1:2006].

10.2.3 Wrapping

Picture essence shall be frame wrapped according to [SMPTE 422M] and [SMPTE 429-4]. Stereoscopic picture essence shall also conform to [SMPTE 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 429-3].

10.3.2 Wrapping

Sound essence shall be frame wrapped per [SMPTE 382M]. 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 A1) 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 428-7], and may be constrained per [SMPTE 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 429-5].

Annex A (Normative)
Audio Channel Assignment Label

[SMPTE 382M] 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 382M] container. The following tables define how [SMPTE 382M-2007] container channels are mapped to [SMPTE 428-3] labels. These labels, in turn, indicate the intended reproduction location in the theater.

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 of the WaveAudioEssence Descriptor in a Sound Track File, as described in Section 10.3.3 above. 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.

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

It is not required that Sound Track files carry the maximum number of channels defined by the table associated with a ChannelAssignment property. However, if a given container channel is present, it must 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 (HI and VI) are not present.

A.1 Channel Label Sets and ULs

Table A.1 – Channel Configuration 1

Container Channel	SMPTE 428-3 label
1	L
2	R
3	C
4	LFE
5	Ls
6	Rs
7	HI
8	VI-N

Table A.2 – Channel Configuration 2

Container Channel	SMPTE 428-3 label
1	L
2	R
3	C
4	LFE
5	Ls
6	Rs
7	Cs
8	Not Used
9	HI
10	VI-N

Table A.3 – Channel Configuration 3

Container Channel	SMPTE 428-3 label
1	L
2	R
3	C
4	LFE
5	Ls
6	Rs
7	Lc
8	Rc
9	HI
10	VI-N

A.2 Channel Label Set ULs

Table A.4 – Specification of the Channel Configuration Channel Assignment Label

Byte No.	Description	Value (hex)	Meaning
1-7	Registry Designator	See SMPTE 400M	
8	Registry Version Number	0bh	Version of 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 labelling
13	Sound Channel Labeling SMPTE 429-2	03h	Identifies sound channel labels defined in this document (S429-2)
14	Channel Label Sets	01h	Identifies sound Channel Label Sets
15	Channel Configuration	See Table A.5	Identifies sound Channel Configuration
16	Reserved	00h	Reserved

Table A.5 – Values for Table A.4, Byte 15

Channel Configuration	Byte 15 Value
Channel Configuration 1 (Table A.1)	01h
Channel Configuration 2 (Table A.2)	02h
Channel Configuration 3 (Table A.3)	03h

Annex B (Informative)
Additional Frame Rates

Implementers choosing to provide support for additional frame rates, such as those defined in [Proposed SMPTE 428-11, Additional Frame Rates for D-Cinema], are advised to adhere to the provisions of the normative sections of this document as specified, but using instead the alternative frame rate and respective edit rate. For monoscopic images, the edit rate is equal to the frame rate ($r_e = r_f$). For stereoscopic images, the edit rate is equal to half the frame rate ($r_e = r_f / 2$).

Content prepared in this way might not play on all equipment considered to be compliant with this standard. Content owners are advised to consider of the capabilities of intended playout equipment when choosing to use edit rates other than 24/1 and 48/1.

Annex C (Informative)
Bibliography

[SMPTE 336M-2007] Data Encoding Protocol using Key-Length-Value

[SMPTE EG 41-2004] Material Exchange Format (MXF) — Engineering Guideline

[SMPTE RP 224[2]] SMPTE Labels Register

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

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

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

Internet Engineering Task Force (IETF) [RFC 2046] (November 1996) Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types

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