SMPTE ST 429-5:2017
Revision of SMPTE ST 429-5:2009

SMPTE STANDARD

D-Cinema Packaging — Timed Text Track File

Table of Contents

Approved December 27, 2017

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.

SMPTE ST 429-5 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.

Introduction

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

Many applications, including D-Cinema, rely on XML expressions of timed text material for caption and subtitle essence. An MXF Generic Container mapping for XML text data allows convenient carriage of text resources with related ancillary resources, such as fonts and sub-pictures, in an MXF-oriented media workflow. The mapping optionally allows encryption of these contents for confidentiality.

This work was originally developed to support D-Cinema, but since that time other applications have been developed that have also identified a requirement for carrying XML timed text in MXF. To support these applications, the definition of the timed text GC mapping and the D-Cinema Track File are defined separately so that the GC mapping may be applied to applications other than D-Cinema.

1 Scope

This standard specifies the format of a Generic Container (GC) for XML timed text and Timed Text Track File for the distribution of timed text content using the Material Exchange Format (MXF). The GC and Track File provide for carriage of an XML document and optional supporting resources such as images or fonts. Encryption is optionally available for protecting against unauthorized disclosure of the file contents.

The standard defines data structures for interchange at the signal interfaces of networks or storage media, but does not define internal storage formats for compliant devices.

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, followed by 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 standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

SMPTE ST 377:2004 Television — Material Exchange Format (MXF) — File Format Specification

SMPTE ST 377-1:2012 Television —Material Exchange Format (MXF)

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

SMPTE ST 379-1:2009 —Material Exchange Format (MXF) — MXF Generic Container

SMPTE ST 379-2:2010 —Material Exchange Format (MXF) — MXF Constrained Generic Container

SMPTE ST 410:2008 Television —Material Exchange Format — Generic Stream Partition

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

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

Internet Engineering Task Force (IETF) (1996, November) — [RFC 2045] Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies

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

Internet Engineering Task Force (IETF) (2005, July) — [RFC 4122] A Universally Unique Identifier (UUID) URN Namespace

ISO/IEC —10646:2014 Information Technology — Universal Multiple-Octet Coded Character Set (UCS)

4 Glossary of Acronyms and Terms

Note: See also the list of acronyms and terms in ST 377M:2004 and ST 377-1.

Closed Caption: Timed-text intended for display on a device other than the theater screen.

Composition: A complete artistic or informational motion picture work, such as a feature, a trailer, or an advertisement, etc.

MXF: Material Exchange Format.

Open Caption: Timed-text intended for display on the theater screen, usually written in the same language as the sound essence.

Reel: A unit of essence comprising all or part of a composition.

Resource: An integral unit of data, such as an XML document, a font, or a sub-picture image.

Sub-picture: An ancillary image intended for display over a larger main image.

Subtitle: Timed-text intended for display on the theater screen, usually written in a language other than the language of the sound essence.

Timed-text: Text intended for display over a timeline, in synchronization with image and sound essence.

URI: A Uniform Resource Identifier, defined by [RFC 2396], is a text string which uniquely identifies a resource, such as an XML Schema. URI values often identify objects not accessible via a computer network.

UUID: Universally Unique Identifier, defined by [RFC 4122].

XML: eXtensible Markup Language, an abstract syntax for structured text with metadata.

5 Overview (Informative)

Subtitles, closed-captions or other forms of textual information often accompany sound and picture essence. The display of textual information varies according to purpose, but the essence encoding can be generalized as a timed-text resource (an XML document) that provides content, position and timing information (*i.e.*, essence plus metadata), and optional ancillary resources such as fonts and sub-pictures.

In some cases, the essence is entirely contained in one or more sub-pictures. In these cases, the timed-text resource contains only timing and position information.

Timed-text essence, therefore, consists of one XML timed-text resource plus optional supporting resources. To simplify the delivery of what may potentially be many dozens of files, this specification allows all resources to be wrapped in a single MXF file. The timed-text resource is contained in an MXF Generic Container (GC) in the Body partition, and any ancillary resources are individually contained in their own Generic Stream partitions.

Figure 1 below illustrates a simple Timed Text Track File for cinema containing text-based subtitle essence. The timed-text resource is contained in the Track File along with a font resource used to render the characters when the text is reproduced on the theater screen.

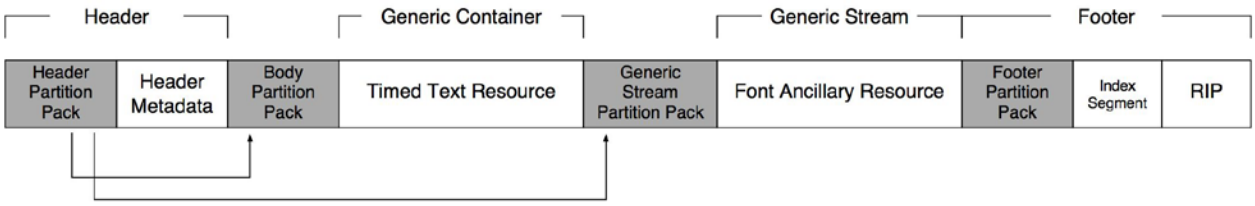

Figure 1 – Example Timed Text Track File Structure using a Font

Figure 2 illustrates a more complex Timed Text Track File containing sub-picture-based subtitles. The timed-text resource is contained in the Track File along with a number of sub-picture resources.

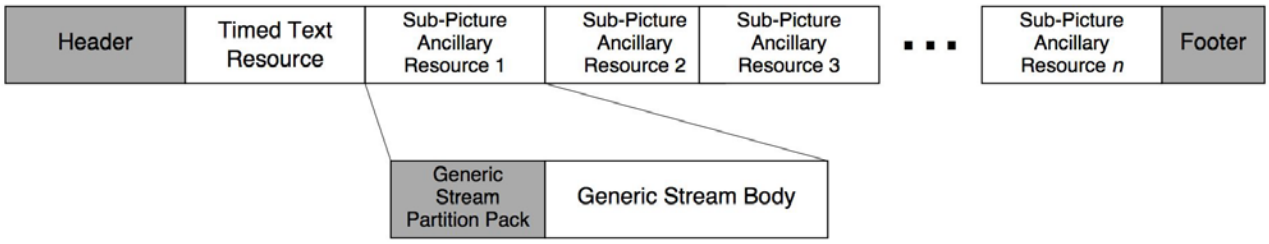

Figure 2 – Example Timed Text Track File Structure using Sub-Pictures

6 Ancillary Resource Wrapping

6.1 Generic Stream Wrapper

Each Ancillary Resource referenced by a Timed Text Generic Container shall be entirely contained within an MXF Generic Stream Partition constructed per ST 410 and located in the same MXF file. Each Generic Stream Partition in the file shall contain exactly one Ancillary Resource. Each Generic Stream Partition shall have a distinct BodySID per ST 410. The Generic Stream Partition shall consist of a Generic Stream Partition Pack

immediately followed by a single KLV packet containing all of the resource data. KLV Fill packets shall not be permitted between the Generic Stream Partition Pack and the resource KLV packet. The actual format of the resource data is beyond the scope of this document. Consult the defining document for the Timed Text Resource for more information. Figure 3 illustrates a Generic Stream Partition containing Ancillary Resource data.

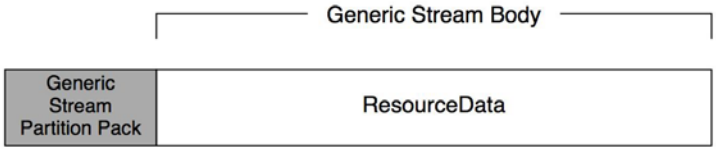

Figure 3 – Ancillary Resource Partition Structure

The Ancillary Resource KLV packet shall be identified by the Default Generic Stream Data Element key (see Generic Stream Data Element coding in ST 410). Data Arrangement bit 1 shall be zero (the KL pair shall not be considered an intrinsic part of the Ancillary Resource data) and bits 3 and 2 shall be one and zero, respectively (the Generic Stream payload is a byte string). Wrapping Signaling bits 1-3 shall be zero (there are no internal access units). Table 1 gives the UL value of the Default Generic Stream Data Element key, set per the above constraints:

Table 1 – Ancillary Resource Key (hexadecimal)

06.0e.2b.34 01.01.01.0c 0d.01.05.09 01.00.00.00

6.2 Generic Stream Repetition

Repetition of the Generic Stream, as defined by ST 410, shall not be used.

6.3 Indexing Generic Stream Data

Generic Stream Data shall not be indexed using MXF Index Tables. Generic Stream Partitions shall be included in the MXF Random Index Pack (RIP). See Section 10 for an informative description of locating an Ancillary Resource using its Resource ID as a lookup value.

7 Mapping XML Timed Text to the MXF Generic Container

7.1 Essence Encoding

The essence container shall contain the Timed Text Resource, an XML document that contains all of the timing and position information for the timed-text instances.

The Timed Text Resource shall be clip wrapped as a single Data Element in a single Data Essence Item of a Generic Container, either ST 379M:2004, ST 379-1 or ST 379-2 as required by the application.

The Timed Text Resource may refer to Ancillary Resources such as fonts and sub-pictures. All Ancillary Resources referenced by the Timed Text Resource shall be contained within the same MXF file in separate Generic Stream Partitions (see Section 6 above). The MXF file shall not contain resources not referenced by the Timed Text Resource.

7.2 Index Table

The index shall comprise a single Index Table Segment pack as defined in ST 377M:2004 or ST 377-1 as required by the application. The Index Table Segment shall contain one entry, pointing to the beginning of the single Data Element in the GC that holds the clip-wrapped Timed Text Resource. Within the segment, the DeltaEntryArray shall be empty, and the value of EditUnitByteCount shall be 0 (zero).

7.3 Timed Text Resource Constraints

While this specification does not define or reference a specific standard for the format of the Timed Text Resource, the following requirements must be met by the resource format for the resource to be used in a Timed Text Generic Container mapping:

1. The resource shall be encoded as an XML document.
2. The resource should be identifiable using an embedded UUID value per RFC 4122.

For example, SMPTE ST 428-7 meets these requirements.

Note: Access to ancillary resources contained within a Timed Text Track File is made via reference to UUID values in the set of Timed Text Resource Subdescriptor items in the file header. In a case where the Timed Text Resource does not directly use UUID values to indicate ancillary resource references, the application must provide an appropriate translation from references to ancillary resources in the Timed Text Resource to the corresponding UUID values in the set of Timed Text Resource Subdescriptor items.

7.4 Header Metadata Construction

The Timed Text Resource shall be described by a top-level File Package per ST 377M:2004 or ST 377-1 as required by the application. The File Package shall contain one Data Essence Track with a single Data Source Clip. A single Material Package shall be present which shall contain one Data Essence Track with a single Data Source Clip referencing the File Package.

If an MXF file constructed per this mapping contains encrypted essence (see Section 8 below), the header shall contain a Cryptographic Framework per ST 429-6.

7.5 Essence Descriptors

The primary File Package in the header metadata shall have a strong reference to a TimedText Descriptor, which shall describe the Timed Text Resource (see Annex A.2).

If the Timed Text Resource references one or more Ancillary Resources, the TimedText Descriptor shall contain the same number of strong references to TimedTextResource Descriptors (one for each Ancillary Resource, see Annex A.3). A TimedTextResource Descriptor contains the resource ID (a UUID) and MIME type (per RFC 2045) of the respective resource, and also the BodySID of the Generic Stream Partition containing the resource data. Figure 4 illustrates the metadata descriptors for a Timed Text Track File containing a Timed Text Resource and two Ancillary Resources (a font and an image).

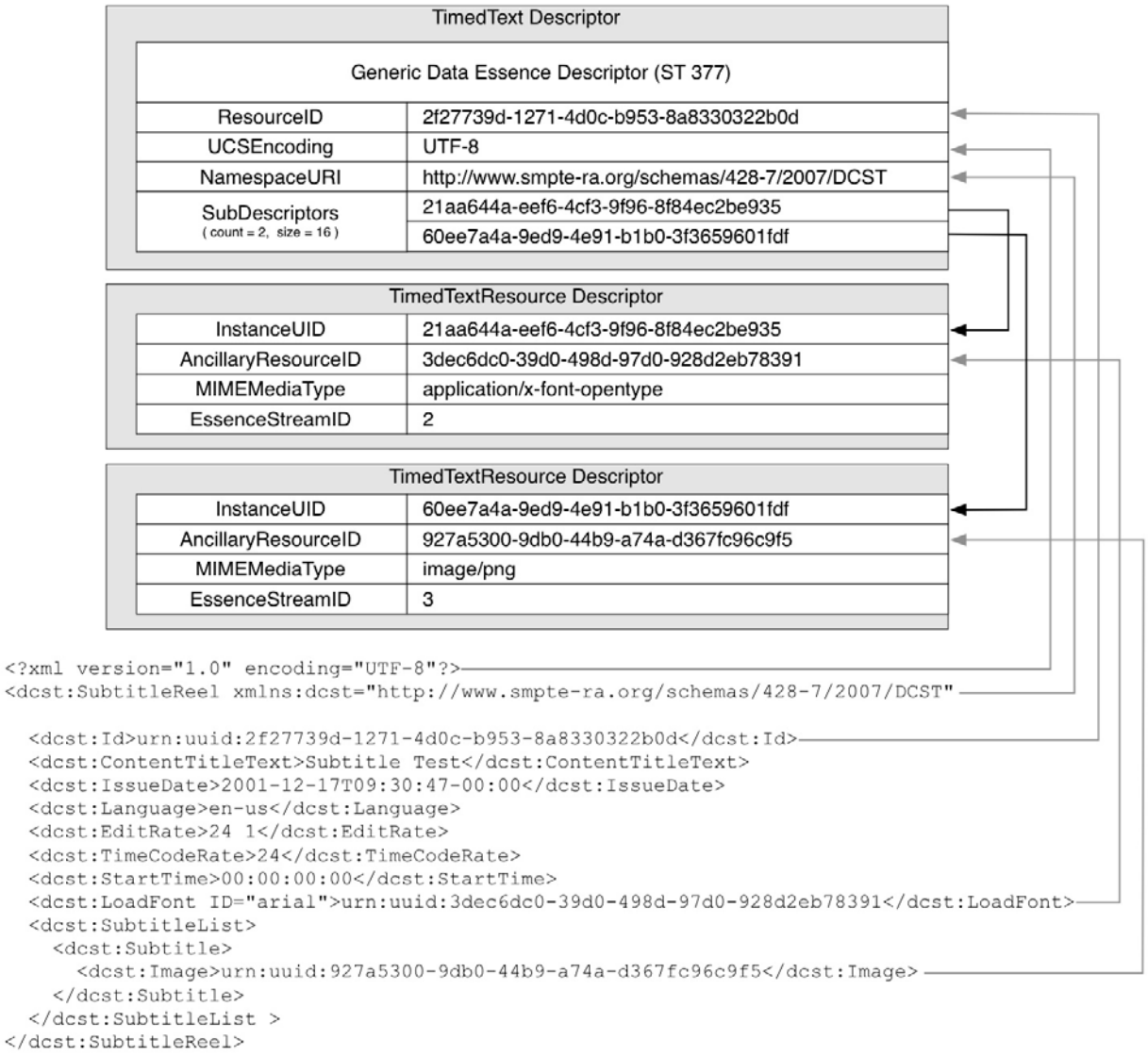

Figure 4 – Essence Descriptor example

8 Essence Encryption

Essence in a Timed Text Generic Container may be encrypted. For this purpose, the Timed Text Resource shall be contained in an Encrypted Triplet per ST 429-6. Ancillary Resources may also be encrypted. Because an Ancillary Resource is a component of the file's essence track, it shall be encrypted using the same Cryptographic Context used to encrypt the Timed Text Resource. Ancillary Resources shall not be encrypted unless the Timed Text Resource is also encrypted.

When encrypting Ancillary Resources, the Generic Stream Data Element KLV packet in the Generic Stream Partition that contains the resource data shall be contained in an Encrypted Triplet per ST 429-6. If the optional MIC value is present in the EKLV packet, the Sequence Number value shall increment with each successive encrypted Ancillary Resource (i.e., the first Ancillary Resource EKLV packet shall have a Sequence Number that is one greater than that of the Timed Text Resource EKLV packet, the second Ancillary Resource EKLV packet shall have a Sequence Number that is two greater than that of the Timed Text Resource EKLV packet, etc.)

Figure 5 is a modified form of the Cryptographic Framework diagram from ST 429-6. This diagram shows the relationship of the Framework instance to the EKLV packets in the Generic Stream Partitions. As specified in ST 429-6, the EKLV packet contains a weak reference to the Cryptographic Framework. Decoders can trace from the Cryptographic Framework to the Generic Stream Partitions by using the set of TimedText Resource subdescriptors in the File Package that references the Framework.

A Timed Text Generic Container that contains Encrypted Triplets shall have a Cryptographic Framework and single Cryptographic Context (i.e., all Encrypted Triplet packets in a Timed Text Generic Container shall be encrypted using the same symmetric key).

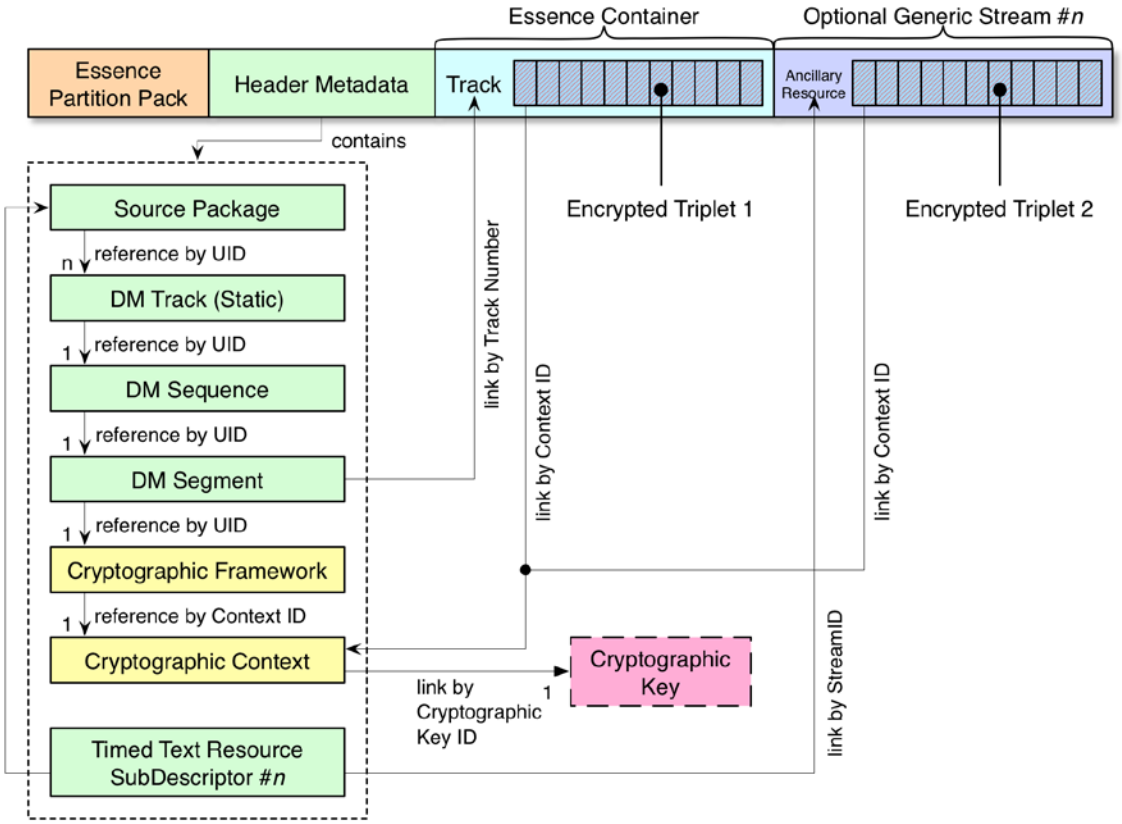

Figure 5 – Cryptographic Framework

9 Synchronization

Synchronization information is contained in the Timed Text Resource. The normative definition of the Timed Text Resource shall specify synchronization with other essence.

10 Random Access to Ancillary Resources (Informative)

During reproduction of the essence encoded in the Timed Text Resource, the decoder will have to retrieve from the MXF file any Ancillary Resources referenced by the Timed Text Resource. This section provides an informative method of efficiently performing this retrieval.

It should be noted that each Generic Stream Partition has a distinct BodySID value. This value is given in the respective TimedTextAncillaryResource sub-descriptor that describes the Ancillary Resource contained in a Generic Stream Partition, and also in the Random Index Pack (RIP). Given this information, and a UUID value identifying an Ancillary Resource, the following algorithm can be used to seek to the location of the Ancillary Resource in the MXF file:

1. Search for the UUID value in the AncillaryResourceID property in the set of TimedTextAncillaryResource sub-descriptors referenced by the TimedTextResource descriptor. If no match is found, the resource does not exist in the file.

2. Using the value of the EssenceStreamID property of the matching sub-descriptor, locate in the Random Index Pack (RIP) the Pair entry having a matching BodySID value.

3. The ByteOffset property in the Pair located in step 2 above gives the location of the Partition Pack at the start of the Generic Stream Partition. The Ancillary Resource data will be contained in the KLV packet immediately following the Partition Pack.

11 D-Cinema Timed Text Track File Structure

11.1 Timed Text Track File Definition

For digital cinema applications the Timed Text Generic Container mapping shall be present in a Track File as defined by ST 429-3, subject to the following exceptions:

11.2 Exceptions to Provisions of ST-429-3

The following exceptions to ST 429-3 shall apply:

a. The Timed Text Track File shall contain exactly one Timed Text Resource in a single Data Track. No other File Packages shall be present.

b. The Timed-Text Resource shall be clip-wrapped, not frame-wrapped.

c. The file shall contain more than three (3) partitions when Ancillary Resources are present.

d. Essence Constraints shall be determined by this document. Note that ST 429-3 references ST 379:2004 and ST 377:2004.

11.3 Generic Container

For digital cinema applications, the Timed Text Generic Container mapping shall use the Generic Container specification referenced by ST 429-3.

11.4 Timed Text Resource

1. The Timed Text resource shall have an embedded UUID value per RFC 4122 which shall be repeated in the ResourceID property of the Timed Text Resource Descriptor.

2. The URI value giving the XML namespace name of the Timed Text resource shall be recorded in the NamespaceURI property of the Timed Text Resource Descriptor.

11.5 Header Metadata Constraints

The MXF Header Metadata of Timed Text Track Files shall conform to the section “Header Metadata Constraints” in ST 429-3.

Timed Text Track Files shall not contain synchronization information other than what is required to create valid MXF header metadata. When present in a Timed Text Track File, MXF header synchronization metadata shall be ignored by the decoder.

11.6 Index Tables

The Index Table Segment required by the Timed Text Generic Container mapping shall appear in the Footer Partition of the Timed Text Track File in accordance with ST 429-3.

Annex A Labels and Descriptor Sets (Normative)

With the exception of InstanceID and GenerationUID, which are already defined in ST 377 (either ST 377:2004 or ST 377-1 as required by the application), all Local Tag values for the descriptors shall be dynamically allocated as defined in 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 ST 377.

A.1 Key UL Values

Table A.1 – Specification of the Timed Text Essence Container Label

Byte No.	Description	Value (hex)	Meaning
1-7	Defined by Generic Container		See ST 379 Element Key
8	Version	0ah	
9-12	Defined by Generic Container		
13	Essence Container Kind	02h	MXF Generic Container
14	Mapping Kind	13h	Timed Text
15	reserved	01h	
16	reserved	01h	

Table A.2 – Key Value for a Timed Text Essence Element

Byte No.	Description	Value Hex	Meaning
1-7	Defined by Generic Container		See ST 379 Container Label
8	Version	01h	
9-12	Defined by Generic Container		
13	Item Type Identifier	17h	Timed Text Item
14	Essence Element Count	01h	Count of XML Resource Elements in the file (always 1)
15	Essence Element Type	0Bh	Clip Wrapped Element
16	Essence Element Number	01h	

A.2 TimedText Descriptor Set

Table A.3 – TimedTextDescriptor

Item Name	Type	Len	UL Designator	Req?	Meaning	Default
TimedText Descriptor	Set UL	16	See table below	Req	Defines the TimedText Descriptor Set	
Length	BER Length	var		Req	Set length	
All items from the MXF Generic Data Essence Descriptor						
ResourceID	UUID	16	06.0E.2B.34 01.01.01.0C 01.01.15.12 00.00.00.00	Opt	A UUID value that identifies this Resource.	
UCSEncoding	UTF16 String	var	06.0E.2B.34 01.01.01.0C 04.09.05.00 00.00.00.00	Req	A text string giving the ISO/IEC 10646-1 encoding of the essence data.	“UTF-8”
NamespaceURI	UTF16 String	var	06.0E.2B.34 01.01.01.08 01.02.01.05 01.00.00.00	Req	A URI that uniquely identifies the defining specification of the top-level XML element in the essence data, *e.g.*, namespace name or profile designator.	

Table A.4 – Key for TimedTextDescriptor

Byte No.	Description	Value (hex)	Meaning
1-7	Structural Header Metadata		Defined in ST 377, “Structural Header Metadata Implementation"
8	Version	01h	
9-13	Structural Header Metadata		
14	Set Kind (1)	01h	TimedTextDescriptor
15	Set Kind (2)	64h	
16	Reserved	00h	Reserved

A.3 TimedText Resource SubDescriptor Set

The TimedTextResourceSubDescriptor is a supplementary Essence Descriptor that can be strongly referenced by the TimedText Descriptor. So that the strong reference can be made, the MXF Generic Descriptor (as defined in ST 377:2004 or ST 377-1 as required by the application) has an additional optional property as defined in Table A.5.

The Local Tag value associated with this additional optional property (called “Sub Descriptors”) shall be dynamically allocated (dynamic) as defined in ST 377:2004 or ST 377-1 as required by the application.

Table A.5 – 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 ST 377							
Sub Descriptors	Array of StrongRef (Sub Descriptors)	8+ 16n	dynamic	06.0E.2B.34. 01.01.01.09. 06.01.01.04. 06.10.00.00	Opt	Array of strong references to sub Descriptor sets	

Table A.6 – TimeTextResourceSubDescriptor

Item Name	Type	Len	UL Designator	Req?	Meaning	Default
TimedTextResource SubDescriptor	Set UL	16	See table below	Req	Defines the TimedTextResource SubDescriptor Set	
Length	BER Length	Var		Req	Set length	
Instance UID	UUID	16	06.0E.2B.34. 01.01.01.01. 01.01.15.02. 00.00.00.00	Req	Unique ID of this instance [The ISO/IEC 11578 (Annex A) 16 byte Globally Unique Identifier]	
Generation UID	UUID	16	06.0E.2B.34. 01.01.01.02. 05.20.07.01. 08.00.00.00	Opt	Generation Identifier [**RP210** Specifies the reference to an overall modification]	
AncillaryResourceID	UUID	16	06.0E.2B.34 01.01.01.0C 01.01.15.13 00.00.00.00	Req	A UUID value that identifies this Ancillary Resource (copied from the set of resource ids in the Timed Text Resource)	
MIMEMediaType	UTF16 String	var	06.0E.2B.34 01.01.01.07 04.09.02.01 00.00.00.00	Req	A MIME Type identifier (see Content-Type Header Field in RFC 2045) which identifies the resource data type	
EssenceStreamID	UINT32	4	06.0E.2B.34 01.01.01.04 01.03.04.04 00.00.00.00	Req	The BodySID of the partition that contains the resource data	

Table A.7 – Key for TimedTextResourceSubDescriptor

Byte No.	Description	Value (hex)	Meaning
1-7	Structural Header Metadata		Defined in ST 377, “Structural Header Metadata Implementation"
8	Version	01h	
9-13	Structural Header Metadata		
14	Set Kind (1)	01h	TimedTextResourceSubDescriptor
15	Set Kind (2)	65h	
16	Reserved	00h	Reserved

Annex B Bibliography (Informative)

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

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

[SMPTE RP 210] Metadata Dictionary Registry of Metadata Element Descriptions

[SMPTE RP 224] SMPTE Labels Register

Annex C Summary of Changes from Version 2009 (Informative)

This document has been revised from its 2009 version in order to accommodate the different `IndexTableSegment` partitioning scheme of OP 1a structures used in ST 2067-5 for Interoperable Master Format (IMF) applications. The purpose of this section is to highlight the main differences for the benefit of the reader. Only brief descriptions are provided in each case, and the intent of this section is not to list every single change.

- The value of the `Ancillary Resource Key` was incorrectly published in Table 1 of Section 6. The correct value has been substituted.
- The definition of the Timed Text GC mapping has been separated from the definition of the Timed Text Track File to facilitate application of this standard to non-cinema applications.
- The UL definitions have been updated to include the version byte.
- The GC mapping may be used with ST 377:2004 or ST 377-1
- The GC mapping may be used with ST 379:2004, ST 379-1 or ST 379-2
- No structural changes have been made to the format written for D-Cinema.