
**Information technology — Coding of
audio-visual objects —**

**Part 30:
Timed text and other visual overlays
in ISO base media file format**

iTeh STANDARD PREVIEW
*Technologies de l'information — Codage des objets audiovisuels —
Partie 30: Texte temporisé et autres recouvrements visuels dans le
format ISO de base pour les fichiers médias*

ISO/IEC 14496-30:2018

<https://standards.iteh.ai/catalog/standards/sist/5883fa28-0e14-4faa-b87e-96d534c081a2/iso-iec-14496-30-2018>



iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 14496-30:2018

<https://standards.iteh.ai/catalog/standards/sist/5883fa28-0e14-4faa-b87e-96d534c081a2/iso-iec-14496-30-2018>



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2018

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

	Page
Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms, definitions and abbreviated terms	1
3.1 Terms and definitions.....	1
3.2 Abbreviated terms.....	2
4 General definitions	2
4.1 Layout.....	2
4.2 Timing.....	3
4.3 Language.....	3
4.4 Resources shared by multiple samples.....	4
4.5 Associating timed text tracks.....	4
5 Timed Text Markup Language (TTML)	4
5.1 General.....	4
5.2 Layout.....	4
5.3 Timing.....	4
5.4 Track format.....	6
5.5 Sample entry format.....	6
5.6 Sample format.....	6
5.7 Additional considerations.....	8
5.8 Codecs parameter.....	8
5.9 Document temporal boundaries.....	8
6 Web Video Text Tracks (WebVTT)	9
6.1 General.....	9
6.2 Layout.....	9
6.3 Timing.....	9
6.4 Track format.....	9
6.5 Sample entry format.....	9
6.6 Sample format.....	10
6.7 Converting to or from a WebVTT text file.....	11
6.7.1 General.....	11
6.7.2 Importing a WebVTT file into the ISO base media file format.....	11
6.7.3 Exporting a WebVTT file from the ISO base media file format.....	12
6.8 Example.....	13
6.8.1 Source file.....	13
6.8.2 Imported format.....	13
Annex A (informative) Captioning information embedded in a media stream of another type	14
Bibliography	15

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1, Information technology, Subcommittee SC 29, Coding of audio, picture, multimedia and hypermedia information.

This second edition cancels and replaces the first edition (ISO/IEC 14496-30:2014), which has been technically revised. It incorporates ISO/IEC 14496-30:2014/Cor.1:2015. The main changes compared to the previous edition are as follows:

- all Clauses (except Clause 1 and Clause 3) and Annex A have been technically revised.

A list of all parts in the ISO/IEC 14496 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document defines a storage format based on, and compatible with, the ISO base media file format (ISO/IEC 14496-12), which is used by the MP4 file format (ISO/IEC 14496-14) and the Motion JPEG 2000 file format (ISO/IEC 15444-3) among others. This document enables timed text and subtitle streams to

- be used in conjunction with other media streams, such as audio or video;
- be used in an MPEG-4 systems environment, if desired;
- be formatted for delivery by a streaming server, using hint tracks; and
- inherit all the use cases and features of the ISO base media file format on which MP4 and MJ2 are based.

iTeh STANDARD PREVIEW (standards.iteh.ai)

[ISO/IEC 14496-30:2018](https://standards.iteh.ai/catalog/standards/sist/5883fa28-0e14-4faa-b87e-96d534c081a2/iso-iec-14496-30-2018)

<https://standards.iteh.ai/catalog/standards/sist/5883fa28-0e14-4faa-b87e-96d534c081a2/iso-iec-14496-30-2018>

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 14496-30:2018

<https://standards.iteh.ai/catalog/standards/sist/5883fa28-0e14-4faa-b87e-96d534c081a2/iso-iec-14496-30-2018>

Information technology — Coding of audio-visual objects —

Part 30:

Timed text and other visual overlays in ISO base media file format

1 Scope

This document describes the carriage of some forms of timed text and subtitle streams in files based on ISO/IEC 14496-12 (the ISO base media file format). The documentation of these forms does not preclude other definition of carriage of timed text or subtitles; see, for example, 3GPP Timed Text (3GPP TS 26.245), or the carriage of captioning information embedded in a media stream of another type (see [Annex A](#)).

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

W3C Recommendation, Timed Text Markup Language 1.0, Second Edition, <https://www.w3.org/TR/ttml1/>

ISO/IEC 14496-12:2015, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

W3C Community Group Report, *WebVTT*, <http://www.w3.org/2013/07/webvtt.html>

3GPP TS 26.245:2017, *Transparent end-to-end Packet switched Streaming Service (PSS); Timed text format*

IETF RFC 2141, *URN Syntax*

IETF RFC 3986, *Uniform Resource Identifier (URI): Generic Syntax*

IETF RFC 6381, *MIME Codecs and Profiles*

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

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

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at <https://www.iso.org/obp>

— IEC Electropedia: available at <http://www.electropedia.org/>

3.1.1

timed text document

file-based representation of textual content, possibly XML, used to produce timed text streams and possibly representing timed text track samples

**3.1.2
timed text stream**

stream of content, which when decoded results in textual content, possibly containing internal timing values, to be presented at a given presentation time and for a certain duration

**3.1.3
subtitle stream**

timed text stream potentially also presenting images

**3.1.4
internal timing value**

value contained in the payload of a timed text stream sample representing a time

EXAMPLE A start time, an end time, or a duration, corresponding to a timed behaviour of a part or the whole of the sample.

**3.1.5
timed text track**

ISOBMFF representation of a timed text stream

**3.1.6
subtitle track**

ISOBMFF representation of a subtitle stream

3.2 Abbreviated terms

TTML Timed Text Markup Language

WebVTT Web Video Text Tracks

ISOBMFF ISO base media file format <https://standards.iteh.ai/catalog/standards/sist/5883fa28-0e14-4faa-b87e-96d534c081a2/iso-iec-14496-30-2018>

iTeh STANDARD PREVIEW
(standards.iteh.ai)

4 General definitions

4.1 Layout

This subclause defines common layout behaviour for processing of timed text or subtitle samples.

Unless specified by an embedding environment (e.g. an HTML page), the track header box information (i.e. `width`, `height`) shall be used to size the subtitle or timed text track content with respect to the associated track(s) as follows:

- 1) If the flag `track_size_is_aspect_ratio` is not set, and the track `width` and `height` are set to values different from 0, the size of the timed text track shall be the track `width` and `height`.
- 2) If the flag `track_size_is_aspect_ratio` is not set, and the track `width` and `height` are set to 0, the size of the timed text track shall match the reference size.
- 3) If the flag `track_size_is_aspect_ratio` is set, it indicates that the content of the track was authored to an aspect ratio equal to the track header `width/height`. In this case, neither `width` nor `height` shall be 0. The timed text track shall be sized to the maximum size that will fit within the reference size and should equal its `width` or `height`, while preserving the indicated aspect ratio.

If only one track is associated with the timed text track, the reference size is the size of the associated track. If multiple tracks are associated, the reference size is the size of the composition of tracks as described by the matrices in the track headers of the associated tracks.

Upon file creation, the `width` and `height` of the subtitle or timed text track should be set appropriately according to the `width` and `height` of the associated track(s), as declared in their track header. A

typical usage is that the timed text or subtitle track has the same `width` and `height` as an associated visual track, and no translation.

If the track it is supposed to overlay is not stored in an ISOBMFF file or if it is stored as a track in a different ISOBMFF file, the values 0x0 may be used; or the `track_size_is_aspect_ratio` flag may be used and the `width` and `height` set to the desired aspect ratio.

For some timed text documents, the region as defined by the `width`, `height` and `track_size_is_aspect_ratio` corresponds to the visual area filled by the rendering of the timed text documents.

When the track `width` and `height` attributes are set to a value different from 0 and the `track_size_is_aspect_ratio_flag` is not used, additional region positioning using the translation values `tx` and `ty` from the track header matrix, as defined for 3GPP Timed Text tracks, may be used (3GPP TS 26.245:2017, 5.7 defines the text track region using `tx`, `ty`, and the track `width` and `height`).

NOTE 1 The 3GPP region is not the same as a WebVTT region.

Unless specified by an embedding environment (e.g. an HTML page), visually composed tracks including video, subtitle, and timed text shall be stacked or layered using the 'layer' value in the track header box. The layer field provides the same functionality as z-index in TTML.

NOTE 2 Timed text and subtitle tracks are normally stacked in front of the associated visual track(s).

4.2 Timing

This subclause defines common timing behaviour for processing of timed text or subtitle samples.

The general processing of timed text or subtitle tracks is that the text content of the sample is delivered to the decoder at the sample decode time, at the latest. The rendering of the sample happens at the composition time, taking into account edit lists if any, and for the whole sample duration, without timing behaviour. However, timed text or subtitle sample data of specific formats may contain internal timing values. Internal timing values may alter the rendering of the sample during its duration as specified by the timed text or subtitle format.

NOTE If an internal timing value does not fall in the time interval corresponding to the sample composition time and sample composition time plus sample duration, the rendering of the sample can be different from the rendering of the same sample data with a composition time such that the internal timing value lies in the associated composition interval.

The subclauses defining the storage of specific formats in the ISOBMFF specify how internal timing values relate to the track time or to the sample decode or composition time (see subclauses 5.3 and 6.3). For instance, start or end times may be relative to the start of the sample, or the start of the track.

For sections of the track timeline that have no associated subtitles or timed text content, 'empty' samples may be used, as defined for each format, or the duration of the preceding sample extended. Samples with a size of zero are not used.

The `timescale` field in the media header box should be set appropriately to achieve the desired timing accuracy. It is recommended to be set to the value of the `timescale` field in the media header box of (one of) the associated track(s).

4.3 Language

Timed text tracks should be marked with a suitable language in the media header box, indicating the audience for whom the track is appropriate. In the case where it is suitable for a single language, the media header must match that declared language. The value 'mul' may be used for a multi-lingual text.

4.4 Resources shared by multiple samples

Common resources, such as images and fonts that are referred to by URLs, may be stored as items in a MetaBox as defined by ISO/IEC 14496-12. These items may be addressed by using the `item_name` as a relative URL in the timed text sample, as defined by ISO/IEC 14496-12:2015, 8.11.9.

NOTE A derived specification, with its applicable brand, can restrict this use of meta boxes for common items.

Fonts not supplied with the content may be already present on the target system(s), or supplied using any suitable supported mechanism (e.g. font streaming as defined in ISO/IEC 14496-18).

4.5 Associating timed text tracks

Timed text tracks may be explicitly or implicitly associated with other tracks in the file. They are explicitly associated with a track when the timed text track uses a track reference of type 'subt' to that track, as defined in ISO/IEC 14496-12, or to a track in the same alternate group. If no 'subt' track reference is used, the timed text track is said to be implicitly associated to all tracks in the file. In particular, if track groups are not used, the timed text track is associated to all tracks in the file. Association is used to indicate which track(s) a timed text track is intended to overlay and may be used to determine the desired rendered size when that information is not provided in the track header of the timed text track, as defined in subclause 4.1. Timed text and subtitle tracks may be associated with any type of track, including visual tracks (e.g. video tracks, graphics tracks, image tracks) or audio tracks as determined by some external context.

5 Timed Text Markup Language (TTML)

ITeH STANDARD PREVIEW
(standards.iteh.ai)

5.1 General

This subclause describes how documents based on TTML, as defined by the W3C, and derived specifications (for example SMPTE-TT), are carried in files based on the ISO base media file format. A TTML Track is a track carrying TTML documents, which can be documents that correspond to a specification based on TTML.

5.2 Layout

Subclause 4.1 defines the general layout behaviour for timed text and subtitle tracks. In particular, this means for TTML tracks that the track width and height provide the spatial extent of the root container, as defined in the TTML Recommendation. Any 'extent' attribute declared on the 'tt' element in the contained TTML document shall match the track width and height. If the 'extent' attribute is not declared on the 'tt' element in the contained TTML document, the track header width and height may be set to 0 or to any desired size.

NOTE This is used when the document is authored in a resolution-independent manner (e.g. using percentage layout).

Alternatively, when a resolution-independent document has been authored to a specific aspect ratio (whether or not the aspect ratio is explicitly signalled in the document) the `track_size_is_aspect_ratio` flag may be used to signal the authored aspect ratio. In this case, the track header width and height shall be set to values that indicate the authored aspect ratio (e.g. 16 by 9).

5.3 Timing

The top-level internal timing values in the timed text samples based on TTML express times on the track presentation timeline – that is, the track media time as optionally modified by the edit list. For example, the begin and end attributes of the <body> element, if used, are relative to the start of the track, not relative to the start of the sample. This is shown in Figure 1, using W3C TTML syntax.

In [Figure 1](#), the sample composition time of each of the samples are 0, 30 minutes, and 1 hour, which correspond to the time at which the decoder will present the TTML content. The first sample, as per W3C Recommendation, Timed Text Markup Language 1.0, will not display any content in the first minute or after 2 minutes, and again, per TTML, will remain as such until the next sample is processed. The second sample contains a document describing the rendering between composition time 0 and 32 minutes. However, since it is provided to the decoder after 30 minutes and since internal timing values are relative to the start of the track, the TTML decoder will display the text as if the decoder sought to 30 minutes into the document. It will not render anything for the first minute from the beginning of the sample, and then render some text for 1 other minute, and then again no rendering until the next sample is processed. The processing of sample 3 is similar, where the top level internal timings on the div elements are handled as relative to the start of the track.

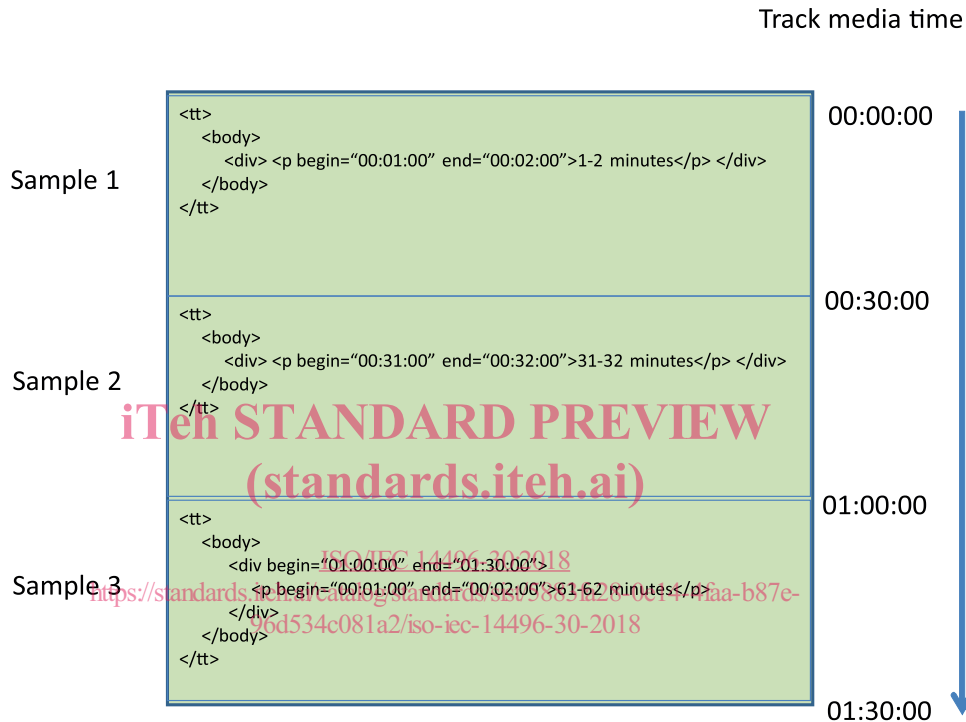


Figure 1 — Example of a TTML track with three samples

No transport layer buffer or timing model is defined to guarantee that subtitle content can be read and processed in time to be synchronously presented with audio and video. It is assumed that users of this track format will define timed text content profiles and hypothetical render models that will constrain content parameters so that compatible decoders may identify and decode those profiles for synchronous presentation.

The following document constraints may need to be specified to define a timed text profile that will guarantee synchronous decoding of conforming content on conforming decoders:

- maximum allowed document size;
- number of document buffers in the hypothetical render model;
- video overlay timing of the hypothetical render model;
- maximum total compressed image size in megabytes per sample;
- maximum total decoded image size in megapixels per sample;
- maximum decoded image dimensions;
- maximum text rendering rate required by a document;