**** **ISO/IEC JTC 1/SC 29/WG 03 N1286**

**ISO/IEC JTC 1/SC 29/WG 03**

**MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title: Working Draft of ISO/IEC 23000-24 Messaging media application format**

**Status:** Approved

**Date of document:** 2024-07-20

**Source:** ISO/IEC JTC 1/SC 29/WG 03

**No. of pages:** 11 (with cover page)

**Email of Convenor:** young.L@samsung.com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 03 MPEG SYSTEMS**

**ISO/IEC JTC 1/SC 29/WG 03 N1286**

**July 2024, Sapporo, JP**

|  |  |
| --- | --- |
| **Title** | **Working Draft of ISO/IEC 23000-24 Messaging media application format** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | 24132 |

**ISO 23000-XX:2023**

ISO/IEC JTC1 /SC 29 /WG 03 /N1082

Secretariat: XXXX

Information technology — Multimedia application format (MPEG-A) — Part XX: Messaging Media Application Format (MeMAF)

WD stage

**Warning for WDs and CDs**

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

© ISO 2020

All rights reserved. Unless otherwise specified, 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

Case postale 56 • CH-1211 Geneva 20

Tel.  + 41 22 749 01 11

Fax  + 41 22 749 09 47

E-mail  copyright@iso.org

Web  www.iso.org

Published in Switzerland.

# Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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 ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2. [www.iso.org/directives](http://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 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. [www.iso.org/patents](http://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 on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: [Foreword - Supplementary information](http://www.iso.org/iso/home/standards_development/resources-for-technical-work/foreword.htm)

The committee responsible for this document is ISO/IEC JTC1 SC29.

Information technology — Multimedia application format (MPEG-A) — Part XX: Messaging Media Application Format (MeMAF)

# Scope

The scope of this part of the standard is the definition of a Media Application Format that can be used for Messaging Services and other download/upload and media storage and exchange applications. Hence, it is referred to as Messaging Media Application Format (MeMAF). The main target of the messaging application format are:

* The packaging of media assets into a single container file for transactional exchanges (upload, download, storage, etc.)
* The instructions on how the media assets are to be processed for a desired experience, providing an entry point as well as a processing of the media assets
* Permit playback on a wide variety of end devices, preferably on existing ones.
* Allow generation and recording of the messages on constrained devices (e.g., mobile devices, etc.), but also allow generation of professionally produced messages.

To support basic interoperability, a baseline profile is defined which minimizes options. Beyond the baseline profile, additional profiles are defined that enable richer experiences.

MeMAF is a profile of ISO BMFF and does not define any new functions or boxes. MeMAF defines specific profiles based on the ISO Base Media File Format (ISOBMFF), each defining a conformance point that provides interoperability between MeMAF conformant devices (MeMAF players) and MeMAF presentations/messages that support that profile.

MeMAF supports different entry points that enable composition and rendering of different messaging experiences. Among others, the following experiences are defined:

* An A/V multimedia message that includes video, audio and possibly text to be presented jointly in a synchronized manner.
* A message for which an image is the primary entry point.
* A message that enables 3D and immersive experiences.

# 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.

ISO/IEC 14496‑1, Information technology — Coding of audio-visual objects — Part 1: Systems

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

ISO/IEC 14496‑14, Information technology — Coding of audio-visual objects — Part 14: MP4 file format

ISO/IEC 14496‑15, Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format

ISO/IEC 14496‑30, Information technology — Coding of audio-visual objects — Part 30: Timed text and other visual overlays in ISO base media file format

ISO/IEC 23001‑7, Information technology — MPEG systems technologies — Part 7: Common encryption in ISO base media file format files

ISO/IEC 23000‑19, Information technology — Multimedia application format (MPEG-A) — Part 19: Common Media Media Application Format (CMAF) for segmented media

IETF RFC 5234, Augmented BNF for Syntax Specifications: ABNF, https://tools.ietf.org/html/rfc5234

IETF RFC 6381, The ‘Codecs’ and ‘Profiles’ Parameters for “Bucket” Media Types, <https://tools.ietf.org/html/rfc6381>

# Terms, definitions, symbols, abbreviated terms and conventions

## 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:

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

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

To be done

## Abbreviated terms

|  |  |
| --- | --- |
| ABNF | augmented backus-naur form |
| URI | uniform resource identifier |
| URL | uniform resource locator |
| URN | uniform resource name |
| USAC | unified speech and audio coding |
| UTC | coordinated universal time |
| UUID | universally unique identifier |
| VOD | video-on-demand |
| VCL | video coding layer |
| VPS | video parameter set |
| VUI | video usability information |
| VVC | versatile video coding | |
| XML | eXtensible Mark-up Language |

## Conventions

# Overview of Messaging Media Application Format

## Background and Assumptions

*Editor’s Note: this text is specific for Messaging Applications and may be used to an Annex to show how the formats can be used in messaging services.*

Messaging services typically define a message container. Such a container typically carries one or more body parts with the actual message content, for example, an emoji used in a reaction, a plain text or rich text message of the reply, a link, or richer media types. An important feature of messages are body parts that include media content, such as still images, graphics, voice messages, audio, video, and other media types. These types of media play a crucial role in enhancing the multimedia aspect of the messages.

The focus of this part of the standard is the definition of the *message body* part that includes the multimedia content, referred to as *MPEG multimedia messaging body part* (3MBP) and the definition of a 3MBP is based on the ISO Base Media File format [13]. The 3MBP is either a single ISO BMFF file with associated metadata for identification of the file, or a collection of files describing the relationship of the files for combined playback experience.

This specification does not prescribe a specific encoding method for the 3MBP; it allows for the utilization of existing functionalities and standards, for example the ones defined in OMA MMS PDUs [7][8][9][15] or MIMI message containers [6]. This standard may be used with a fully specified Messaging Service, but it may as well be used as part of a proprietary messaging services as message body, or more specifically as 3MBP. It may also serve to support interoperability across different messaging services. It is assumed that the identification of the file in the system is primarily addressed by a *media type* that is compatible to the IANA media type, subtype, and parameters as defined in IETF RFC 2046 [12]. For example, it may tell if the content is video or audio and it may provide parameters such as the codec in use. This specification defines, or at least assigns to each defined 3MBP a media type.

In order to use 3MBPs as defined in this specification as part of a message container format, it is expected that the message container format supports the following functionalities:

1. It can carry an octet string representing the *content* of the 3MBP
2. It can signal the *media type* of the content.
3. The content and media type of the content is not restricted but allows to include formats that are not defined in the core container format.

In addition, a message container format may support one or more of the following functionalities in alignment of definitions in IETF MIMI [6] and IETF RFC 2046 [12]:

* the body can have multiple, possibly nested parts, referred to as sub-parts, that share a relation, for example, alternatives, related, or parallel.
* it may have body parts that reference external content via a URI that will be processed automatically. It includes a media type and may optionally include the size of the data, an expiration timestamp or other parameters. The content may be rendered with the other parts of the message, or a be downloaded or rendered separately.
* it may have body parts for which the content is encrypted.

## System Description (Informative)

*Editor’s Note: this text is specific for Messaging Applications and may be used to an Annex to show how the formats can be used in messaging services.*

Based on the background and assumptions in clause 4.1, Figure 4.2-1 provides an example system for a messaging services and highlights scope of this part of the standard, namely the definition of a *MPEG* *multimedia messaging body part* (3MBP), i.e. the ISO BMFF files as well as the associated metadata including the media type.



Figure 4.2-1 Example system for MPEG Multimedia Message exchange

*Editor’s Note: More details on the exact interoperability of the standard needs to be added.*

A Messaging Service Sender instructs a 3MBP generator to generate an 3MBP, for example using an API. This for example allows to define configurations on codecs, size, experiences or other attributes of the 3MBP. The details of such an API are outside the scope of this standard. The sender adds the 3MBP to a Container Message, including the 3MBP metadata parameters that provide information about the 3MBP. Metadata includes, but is not limited to:

* The media type of the 3MBP, including subtypes and parameters for codecs, etc.
* The size of the 3MBP
* Accessibility or language information about the 3MBP
* processing requirements of recommendations of the files included in the 3MBP

The client of the messaging service receives the container message that includes the above information. The client communicates with a 3MBP player its capabilities whether the 3MBP can be played back, and if multiple alternatives are present, which of those are to be selected. Then the messaging service client instructs the 3MBP player to playback the 3MBP as part of the messaging service, based on the processing requirements and instructions. Playback may be combined with additional instructions for a player, including play, pause, seek, etc.

## Client Model (Informative)

*Editor’s Note: this text is specific for Messaging Applications and may be used to an Annex to show how the formats can be used in messaging services.*

The Figure 4.3-1 provides an example of the MeMAF player model in the the Messaging Service Client for the formats defined in this specification. In this figure, the the 3MBP parser receives the 3MBP, and playback instructions. The Messaging Service Client may use metadata provided in a container message for playback selection. Such metadata may for example include codec capability information, language codes, accessibility information and other information for the selection of alternative parts in the 3MBP.

The Messaging Service Client then provides the sub-parts for processing and decoding to the related sub-part processors, and controls those for playback. The rendered message output may be handed back to the Messaging Service client for inband rendering or may be rendered directly.

In the context of this document, the sub-parts and file parser conform to formats defined in the ISO Base Media File format [13].



Figure 4.3-1 — MeMAF Player model

## Hypothetical Application Data Model

### Introduction

The application format defines a hypothetical application data model so that encoding to that model results in consistent MeMAF track encoding, representation in container messages, track selection, late binding, synchronization, decoding, and rendering of MeMAF presentations.

Decoding requirements can be inferred from encoding constraints and the hypothetical application data model but are not directly specified by this standard. However, Annex A provides a hypothetical playback model. The application format does not specify message container formats. However, it MeMAF addressable media objects, which can be added or referenced as resources in a container.

External specifications can define how the container describes a MeMAF presentation, including identifying MeMAF addressable media objects as resources and representing their logical relationships determined by the MeMAF structures.

Logical Structures:

* Single file with one or multiple tracks or items
* A collection of multiple tracks and items
  + Parallel – the tracks provide a common Presentation
    - Alternative The tracks are alternatives
  + Related
    - A manifest or presentation document provides the relation of the tracks and items

[Background: will go to Annex An 3MBP itself is identified by a media type.

The 3MBP may be a single content with a media type.

The 3MBP may include multiple additional 3MBPs. The following multi-part 3MBPs are defined:

* *~~mixed 3MBP~~*~~: multiple 3MBPs are associated with the mixed 3MBP that shall be rendered together. Each 3MBP is identified by a media type. The receiver is expected to process as many as possible of the included 3MBPs based on its capabilities. Real-time 3MBPs included in a mixed 3MBP share the same 3MBP presentation timeline, which has a value of zero at the earliest media sample intended for presentation. If presented jointly, they shall be presented using this common 3MBP presentation timeline.~~
* *parallel 3MBP*: multiple 3MBPs are associated with the parallel 3MBP that shall be rendered together. Each 3MBP is identified by a media type. Real-time 3MBPs included in a parallel 3MBP share the same 3MBP presentation timeline, which has a value of zero at the earliest media sample intended for presentation. If presented jointly, they shall be presented using this common 3MBP presentation timeline.
* *alternative 3MBP*: multiple 3MBPs are associated with the alternate 3MBP. Each 3MBP is identified by a media type. The receiver is expected to process exactly one based on its capabilities.
* *related 3MBP*: multiple objects are associated with the process 3MBP. One object is identified as a root 3MBP. The root 3MBP is identified by a media type. The root 3MBP is processed and identifies if any, several or all of the remaining objects are used as well. Hence, all other objects are typically also identified by media types, and a URL that links the objects being part of the related 3MBP. The processor of the root 3MBP also controls the selection, presentation and timing of the other objects.

3MBPs are a recursive structure. Hence, a receiver shall expect that multi-part 3MBPs contain other multi-part 3MBPs.]

### Single File

### 

## Defined Brands and Profiles

# Common Track Constraints

## General Track Constraints

### Overview

### Boxes

MeMAF tracks shall include the following ISO BMFF boxes with nesting, optionality, and cardinality specified in Table X, with box definitions incorporated by reference to the clauses listed in the “Specification” column that reference ISO/IEC 14496-12, ISO/IEC 23001-7, or ISO/IEC 23009-1.

Some boxes are additionally constrained by MeMAF, as specified by clauses in this document referenced in the “Constraints” column of Table X. The versions and flags of boxes used in a CMAF track shall conform to the versions and flags of the same boxes as defined by the 'isob' brand unless additionally constrained by this document.

MeMAF addressable media objects are derived from the MeMAF track.

In Table X the “Format Req.” column indicates the number of boxes that are required to be present in a MeMAF track, where:

“\*” means “zero or more” may be present;

“+” means “one or more” shall be present;

“0/1” indicates only one box may be present, and it is optional;

“1” indicates one box shall be present;

“CR” indicates that a box is conditionally required as specified in the CMAF track format or a specific media profile and shall be present under those conditions (see “Constraints” column).

Editor’s Note:

1. We will keep only fragmented files in order to keep it simple
2. We will use the requirements to develop the subsetting
3. Non-mentioned boxes are permitted, but are expected to be ignored by a compatible MeMAF track player.

Table 3 — MeMAF track header boxes

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NL 1 | NL 2 | NL 3 | NL 4 | NL 5 | NL 6 | Format Req | 14496-12 | Clause | Description |
| trak |  |  |  |  |  | \* | 8.3.1 |  | *container for an individual track or stream* |
|  | ttyp |  |  |  |  |  | 8.3.6 |  | *track type* |
|  | tkhd |  |  |  |  | \* | 8.3.2 | 5.1.3.1 | *track header, overall information about the track* |
|  | ~~tref~~ |  |  |  |  |  | ~~8.3.3~~ |  | *~~track reference container~~* |
|  | ~~trgr~~ |  |  |  |  |  | ~~8.3.4~~ |  | *~~track grouping indication~~* |
|  |  | ~~msrc~~ |  |  |  |  | ~~8.3.4.4.1~~ |  | *~~multi-source presentation track group type box~~* |
|  |  | ~~ster~~ |  |  |  |  | ~~8.3.4.4.2~~ |  | *~~stereoscopic pair track group type box~~* |
|  | edts |  |  |  |  |  | 8.6.5 |  | *edit list container* |
|  |  | elst |  |  |  |  | 8.6.6 |  | *an edit list* |
|  | ~~meta~~ |  |  |  |  |  | ~~8.11.1~~ |  | *~~Metadata~~* |
|  | mdia |  |  |  |  | \* | 8.4 |  | *container for the media information in a track* |
|  |  | mdhd |  |  |  | \* | 8.4.2 |  | *media header, overall information about the media* |
|  |  | hdlr |  |  |  | \* | 8.4.3 |  | *handler, declares the media (handler) type* |
|  |  | elng |  |  |  |  | 8.4.6 |  | *extended language tag* |
|  |  | minf |  |  |  | \* | 8.4.4 |  | *media information container* |
|  |  |  | vmhd |  |  |  | 12.1.2 |  | *video media header, overall information (video track only)* |
|  |  |  | smhd |  |  |  | 12.2.2 |  | *sound media header, overall information (sound track only)* |
|  |  |  | ~~hmhd~~ |  |  |  | ~~12.4.3~~ |  | *~~hint media header, overall information (hint track only)~~* |
|  |  |  | sthd |  |  |  | 12.6.2 |  | *subtitle media header, overall information (subtitle track only)* |
|  |  |  | nmhd |  |  |  | 8.4.5.2 |  | *Null media header, overall information (some tracks only)* |
|  |  |  | dinf |  |  | \* | 8.7.1 |  | *data information box, container* |
|  |  |  |  | dref |  | \* | 8.7.2 |  | *data reference box, declares source(s) of media data in track* |
|  |  |  | stbl |  |  | \* | 8.5.1 |  | *sample table box, container for the time/space map* |
|  |  |  |  | stsd |  | \* | 8.5.2 |  | *sample description box (codec types, initialization etc.)* |
|  |  |  |  | stts |  | \* | 8.6.1.2 |  | *(decoding) time-to-sample* |
|  |  |  |  | ~~ctts~~ |  |  | ~~8.6.1.3~~ |  | *~~(composition) time to sample~~* |
|  |  |  |  | ~~cslg~~ |  |  | ~~8.6.1.4~~ |  | *~~composition to decode timeline mapping~~* |
|  |  |  |  | stsc |  | \* | 8.7.4 |  | *sample-to-chunk, partial data-offset information* |
|  |  |  |  | stsz |  |  | 8.7.3.2 |  | *sample sizes (framing)* |
|  |  |  |  | stz2 |  |  | 8.7.3.3 |  | *compact sample sizes (framing)* |
|  |  |  |  | stco |  | \* | 8.7.5 |  | *chunk offset, partial data-offset information* |
|  |  |  |  | co64 |  |  | 8.7.5 |  | *64-bit chunk offset* |
|  |  |  |  | stss |  |  | 8.6.2 |  | *sync sample table* |
|  |  |  |  | ~~stsh~~ |  |  | ~~8.6.3~~ |  | *~~shadow sync sample table~~* |
|  |  |  |  | ~~padb~~ |  |  | ~~8.7.6~~ |  | *~~sample padding bits~~* |
|  |  |  |  | ~~stdp~~ |  |  | ~~8.7.6~~ |  | *~~sample degradation priority~~* |
|  |  |  |  | sdtp |  |  | 8.6.4 |  | *independent and disposable samples* |
|  |  |  |  | sbgp |  |  | 8.9.2 |  | *sample-to-group* |
|  |  |  |  | sgpd |  |  | 8.9.3 |  | *sample group description* |
|  |  |  |  | subs |  |  | 8.7.7 |  | *sub-sample information* |
|  |  |  |  | saiz |  |  | 8.7.8 |  | *sample auxiliary information sizes* |
|  |  |  |  | saio |  |  | 8.7.9 |  | *sample auxiliary information offsets* |
|  | udta |  |  |  |  |  | 8.10.1 |  | *user-data* |
|  |  | cprt |  |  |  |  | 8.10.2 |  | *copyright etc.* |
|  |  | ~~tsel~~ |  |  |  |  | ~~8.10.3~~ |  | *~~track selection box~~* |
|  |  | kind |  |  |  |  | 8.10.4 |  | *track kind box* |
|  |  | ~~strk~~ |  |  |  |  | ~~8.14.3~~ |  | *~~sub track box~~* |
|  |  |  | ~~stri~~ |  |  |  | ~~8.14.4~~ |  | *~~sub track information box~~* |
|  |  |  | ~~strd~~ |  |  |  | ~~8.14.5~~ |  | *~~sub track definition box~~* |
|  |  | ~~ludt~~ |  |  |  |  | ~~12.2.7~~ |  | *~~audio stream loudness~~* |

Table 4 — MeMAF track media boxes

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NL 0 | NL 1 | NL 2 | NL 3 | NL 4 | NL 5 | NL 6 | Format Req | 14496-12 | Clause | Description |
| moof |  |  |  |  |  |  | + | 8.8.4 |  | *movie fragment* |
|  | mfhd |  |  |  |  |  | 1 | 8.8.5 |  | *movie fragment header* |
|  | meta |  |  |  |  |  | \* | 8.11.1 |  | *metadata* |
|  | traf |  |  |  |  |  | 1 | 8.8.6 |  | *track fragment* |
|  |  | tfhd |  |  |  |  | 1 | 8.8.7 |  | *track fragment header* |
|  |  | trun |  |  |  |  | 1 | 8.8.8 |  | *track fragment run* |
|  |  | sbgp |  |  |  |  | \* | 8.9.2 |  | *sample-to-group* |
|  |  | sgpd |  |  |  |  | \* | 8.9.3 |  | *sample group description* |
|  |  | subs |  |  |  |  | CR | 8.7.7 |  | *sub-sample information* |
|  |  | saiz |  |  |  |  | CR | 8.7.8 |  | *sample auxiliary information sizes* |
|  |  | saio |  |  |  |  | CR | 8.7.9 |  | *sample auxiliary information offsets* |
|  |  | tfdt |  |  |  |  | 0/1 | 8.8.12 |  | *track fragment decode time* |
|  |  | ~~meta~~ |  |  |  |  |  | ~~8.11.1~~ |  | *~~metadata~~* |
|  |  | ~~udta~~ |  |  |  |  |  | ~~8.10.1~~ |  | *~~user-data~~* |
|  | udta |  |  |  |  |  |  | 8.10.1 |  | *user-data* |
| mfra |  |  |  |  |  |  | 0/1 | 8.8.9 |  | *movie fragment random access* |
|  | tfra |  |  |  |  |  |  | 8.8.10 |  | *track fragment random access* |
|  | mfro |  |  |  |  |  | \* | 8.8.11 |  | *movie fragment random access offset* |
| mdat |  |  |  |  |  |  | + | 8.2.2 |  | *media data container* |

### Constraints on ISO Base Media File Format boxes

#### Track Header Box ('tkhd')

MeMAF TrackHeaderBoxes shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The field duration shall be set to a value of zero ('0'), indicating no media samples are referenced from the TrackBox ('trak').

- The field matrix shall be set to their default values as defined in ISO/IEC 14496-12, except to indicate video orientation (i.e., portrait or landscape orientation relative to the captured scene).

- The following fields shall be set to default values as defined in ISO/IEC 14496-12, unless specified otherwise in this document.

- The layer field should equal 0 or greater for normally presented video tracks.

- The layer field should equal −1 for subtitle tracks so they are normally presented over the video.

- The width and height fields for a non-visual track (e.g. audio) shall be 0.

- As defined in ISO/IEC 14496-12, the width and height fields for a CMAF video track shall specify the track’s normalized presentation size as fixed-point 16.16 values expressed in square pixels after decoder cropping, and in the case of video encoded with a non-square video spatial sample shape, after horizontal scaling has been applied. See [subclause 9.2.3](#Section_sec_9.2.3) for normalized width and height calculation.

- Subtitle tracks may set width and height to an intended layout size, in which case the text layout engine or graphics engine can scale the width and height to match the video display aperture (player implementation dependent).

- As defined in ISO/IEC 14496-30, subtitle tracks encoded as text may use relative position coordinates and font sizes so that the text layout engine can adjust glyph and layout size to match the final video display aperture without relying on image scaling. For such tracks, the value of zero width and height should be used to indicate that the data can be rendered at any size, and the layout size may be determined by matching the size of the video display aperture.

- For scalable text and subtitle tracks, the flag track\_size\_is\_aspect\_ratio may also be used.

#### Media Header Box ('mdhd')

MeMAF MediaHeaderBoxes shall conform to ISO/IEC 14496-12 with the following additional constraints:

- The value of the duration field should be set to a value of zero ('0') (see [subclause 7.5.1](#Section_sec_7.5.1)).

- The value of the timescale field should be chosen such that when the sampling rate is constant, the value of the media sample duration is also constant.

- All tracks that are language-specific should identify the language as precisely as possible (e.g. a text track whose language can be written in different scripts should identify which script is used). When the language is not relevant or not known, the 'und' (undetermined) language tag should be used.

#### Video Media Header Box ('vmhd')

The VideoMediaHeaderBox shall conform to ISO/IEC 14496-12 with the constraints specified in [subclause 9.2.2](#Section_sec_9.2.2).

#### Sound Media Header Box ('smhd')

The SoundMediaHeaderBox shall conform to ISO/IEC 14496-12 with the constraints specified in subclause XXX.

The field balance shall equal 0 (centre).

#### Subtitle Media Header Box ('sthd')

The SubtitleMediaHeaderBox shall conform to ISO/IEC 14496-12 with constraints specified in ISO/IEC 14496-30.

NOTE Subtitle media uses the 'subt' handler\_type in the HandlerBox in the MediaHeaderBox.

#### Data Reference Box ('dref')

DataReferenceBoxes shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The DataReferenceBox shall contain a single entry with the entry\_flags field set to 0x000001 (which means that the media data is in the same file as the MovieBox containing this data reference).

#### Sample Description Box ('stsd')

The SampleDescriptionBox in shall conform to version 0 as defined in ISO/IEC 14496-12 with the following additional constraints:

- Sample entries for encrypted tracks (those containing any encrypted media sample data) shall encapsulate the existing sample entry with the appropriate four-character-code listed in ISO/IEC 14496-12 and include a ProtectionSchemeInfoBox ('sinf') that conforms to ISO/IEC 14496-12 and [subclause 7.5.11](#Section_sec_7.5.11).

Constraints on visual sample entries are specified in [subclause 9.2.4](#Section_sec_9.2.4).

Constraints on audio sample entries are specified in [subclause 10.2.5](#Section_sec_10.2.5).

Constraints on subtitle sample entries are specified in ISO/IEC 14496-30.

NOTE A CMAF SampleDescriptionBox can contain multiple sample entries.

#### Protection Scheme Information Box ('sinf')

CMAF shall use common encryption (ISO/IEC 23001-7) for CMAF tracks containing one or more encrypted CMAF fragments and use Scheme Signalling as defined in ISO/IEC 23001-7. An encrypted CMAF track shall include at least one ProtectionSchemeInfoBox ('sinf') containing a TrackEncryptionBox ('tenc') identifying a scheme specified in ISO/IEC 23001-7.

#### Track contained media sample information boxes

All boxes in the SampleTableBox shall have a sample count of 0 as MeMAF does not reference media samples from the TrackBox. The mandatory boxes of ISO/IEC 14496-12 are mandatory, even if they document no samples.

The following boxes therefore shall have an entry\_count of zero:

TimeToSampleBox ('stts');

SampleToChunkBox ('stsc');

ChunkOffsetBox ('stco');

SampleSizeBox or CompactSampleSizeBox ('stsz' or 'stz2');

SyncSampleBox ('stss'), if present.

NOTE 1 The presence of an empty SyncSampleBox in a CMAF header indicates that not all media samples in the CMAF track are sync samples.

NOTE 2 Media sample size, duration, and dependency information can be found in the TrackRunBox(es) in each CMAF fragment or CMAF chunk.

#### Edit List Box ('elst')

If the Edit List Box ('elst') is present, the following conditions apply:

- The EditBox shall contain a single EditListBox.

- The value of entry\_count field in the EditListBox shall be set to 1.

- The value of the media\_rate\_integer field shall be set to 1 and the value of the media\_rate\_fraction field shall be set to 0.

- The value of the segment\_duration field shall be set to 0.

Such conditions define an offset edit or offset edit list.

NOTE Since there is no media in the MovieBox, the duration in the edit is zero. However, as noted in ISO/IEC 14496-12, movie fragments implicitly extend any edit represented by an edit list in the MovieBox. This means that any offset edit list in the CMAF header is applied to the composition times of all samples to determine each sample’s presentation time.

According to ISO BMFF, samples with a sum of movie presentation time and duration less than zero are not meant to be presented.

Using offset edits to set the presentation time of the earliest presented sample of each CMAF track to an equivalent CMAF track presentation time enables synchronization of all CMAF tracks to a common CMAF presentation timeline. Additional constraints specific to certain media types are in the following subclauses:

audio edit lists in [subclause 10.2.6](#Section_sec_10.2.6);

video edit lists in [subclause 9.2.7](#Section_sec_9.2.7).

#### Track Extends Box ('trex')

A TrackExtendsBox shall be present as defined in ISO/IEC 14496-12.

#### Movie Fragment Header Box ('mfhd')

A MovieFragmentHeaderBox shall conform to ISO/IEC 14496-12.

#### Track Fragment Header Box ('tfhd')

A TrackFragmentHeaderBox shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The base-data-offset-present flag (in the tf\_flags field) shall be set to zero.

- The default-base-is-moof flag (in the tf\_flags field) shall be set to one.

#### Track Run Box ('trun')

A TrackRunBox shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The version field shall be set to either '0' or '1',

- The data-offset-present flag (in the tf\_flags field) shall be set to true in order to indicate that the data\_offset field is present and contains the byte offset from the start of this fragment’s MovieFragmentBox to the start of the first media sample in the following MediaDataBox.

NOTE This is called movie fragment relative addressing in ISO/IEC 14496-12.

#### Sample Group Description Box ('sgpd')

As required by ISO/IEC 14496-12, when sample group descriptions can change within a CMAF track, a SampleGroupDescriptionBox shall be stored in each CMAF fragment that references that sample group description. If sample group information is the same for all CMAF fragments in a CMAF switching set, it may be stored in a SampleGroupDescriptionBox in the CMAF header SampleTableBox ('stbl'). Either version 1 or 2 may be used (version 0 is deprecated by ISO BMFF).

EXAMPLE When common encryption is used and KID values can change per CMAF fragment, a SampleToGroupBox stored in each TrackFragmentBox will reference a SampleGroupDescriptionBox containing the KID for that sample group, and both are stored in the TrackFragmentBox in order to support random access.

Pre-roll sample groups can be used for some audio as described in [subclause 10.3.3.1](#Section_sec_10.3.3.1).

#### Media Data Box ('mdat')

Each fragment shall contain one or more MediaDataBox(es) ('mdat') containing media samples. The MediaDataBox conforms to the definition in ISO/IEC 14496-12.

Each MediaDataBox in a chunk shall be immediately preceded by the MovieFragmentBox that references the media samples it contains.

#### Sub-sample Information Box ('subs')

Each CMAF fragment in a TTML image subtitle track of CMAF media profile 'im1i', as specified in [subclause 11.3.4](#Section_sec_11.3.4), shall contain a SubSampleInformationBox in the TrackFragmentBox that indexes any images contained in the subtitle media sample. The field entry\_count shall equal 1.

When one or more images are present in the subtitle media sample, the value of subsample\_count shall equal 1 for the first image sub-sample, and the subsample\_count of the TTML document shall equal 0.

When no images are present in the subtitle media sample, the value of subsample\_count shall equal 0.

Version 0 of this box should be used, unless subsample\_size exceeds a 16-bit byte address size (65 kibibytes), in which case version 1 should be used.

The fields subsample\_priority, discardable, and codec\_specific\_parameters are undefined.

## Video Track Constraints

## Audio Track Constraints

## Text/Subtitles Track Constraints

# Common Item Constraints

# Single File Constraints

*Editor’s Note: this defines the case, were multiple or all tracks are included in a single file*

# File Constraints in Multi-Part 3MBPs

*Editor’s Note: this defines the case, for which a track is added to an existing file.*

# Encryption and Security

# Profiles

## Baseline Profile

## Image Profile

## Immersive Profile

# Media Profiles

1. (informative)  
     
   Hypothetical Playback Model for a MeMAF Player