**ISO/IEC 14496-12:2024/AMD 2:202x(E)**

ISO/IEC JTC1/SC 29

Secretariat: JISC

**Information technology — Coding of audio-visual objects — Part 12: ISO base media file format — Amendment 2: Tools for enhanced CMAF and DASH integration**

Preliminary WD stage

**Copyright notice**

This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to 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

Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

Contents

[Foreword iv](#_Toc165489695)

[1 Add restrictions on 'altr' groups 1](#_Toc165489696)

[2 New switching group box 2](#_Toc165489697)

[3 New screen target orientation 5](#_Toc165489698)

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.

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 documents 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](http://www.iso.org/directives) or [www.iec.ch/members\_experts/refdocs](http://www.iec.ch/members_experts/refdocs)).

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](http://www.iso.org/patents)) or the IEC list of patent declarations received (see [patents.iec.ch](https://patents.iec.ch/)).

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 the following URL:  [www.iso.org/iso/foreword.html](http://www.iso.org/iso/foreword.html). In the IEC, see [www.iec.ch/understanding-standards](https://www.iec.ch/understanding-standards)

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

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

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](https://www.iso.org/members.html) and [www.iec.ch/national-committees](http://www.iec.ch/national-committees).

**Information technology — Coding of audio-visual objects — Part 12: ISO base media file format — Amendment 2: Tools for enhanced CMAF and DASH integration**

# Add restrictions on 'altr' groups

In clause 8.18.3.1, change:

'altr': The items and tracks mapped to this grouping are alternatives to each other, and only one of them should be played (when the mapped items and tracks are part of the presentation; e.g. are displayable items or tracks) or processed by other means (when the mapped items or tracks are not part of the presentation; e.g. are metadata). A player should select the first entity from the list of entity\_id values that it can process (e.g. decode and play for mapped items and tracks that are part of the presentation) and that suits the application needs. Any entity\_id value shall be mapped to only one grouping of type 'altr'. An alternate group of entities consists of those items and tracks that are mapped to the same entity group of type 'altr'.

To:

'altr': The items and tracks mapped to this grouping are alternatives to each other, and only one of them should be played (when the mapped items and tracks are part of the presentation; e.g. are displayable items or tracks) or processed by other means (when the mapped items or tracks are not part of the presentation; e.g. are metadata). A player should select the first entity from the list of entity\_id values that it can process (e.g. decode and play for mapped items and tracks that are part of the presentation) and that suits the application needs. Any entity\_id value shall be mapped to only one grouping of type 'altr'. An alternate group of entities consists of those items and tracks that are mapped to the same entity group of type 'altr'. None of the entity\_id values in an 'altr' group shall map to another 'altr' group. An 'altr' group shall not contain both entity groups and items.

In clause 8.18.3.3, change:

entity\_id is resolved to an item, when an item with item\_ID equal to entity\_id is present in the hierarchy level (file, movie or track) that contains the GroupsListBox, or to a track, when a track with track\_ID equal to entity\_id is present and the GroupsListBox is contained in the file level.

To:

entity\_id is resolved to an item, when an item with item\_ID equal to entity\_id is present in the hierarchy level (file, movie or track) that contains the GroupsListBox, or to a track, when a track with track\_ID equal to entity\_id is present and the GroupsListBox is contained in the file level. entity\_id shall not have the same value as group\_id.

# New switching group box

In clause 8.18.3, change:

'prsl': The tracks mapped to this grouping are belonging to a preselection as specified in 8.18.4.1.

NOTE EntityToGroupBox can have grouping\_type specific extensions.

To:

'prsl': The tracks mapped to this grouping belong to a preselection as specified in 8.18.4.1.

'swit': The tracks mapped to this grouping belong to a switching group as specified in 8.18.4.2

NOTE EntityToGroupBox can have grouping\_type specific extensions.

Add the following new subclause after subclause 8.18.4.1:

* + - 1. **Switching group box**
         1. **Definition**

Box Type: 'swit'  
Container: GroupsListBox in a MetaBox on movie level  
Mandatory: No  
Quantity: Zero or more

The SwitchingGroupBox is used to facilitate the generation of adaptive streaming manifests or descriptions such as DASH MPD. Generating DASH or CMAF groupings, such as CMAF Switching Sets or DASH Adaptation Sets, from a set of ISOBMFF tracks may require out-of-band knowledge, e.g. knowing which tracks contain the same source content meant to be used in adaptive streaming switching, or may require deep parsing of the tracks to determine if a decoder can decode all the tracks only with the sample description of a single track. This box allows signaling of generic properties used in adaptive streaming such as switching, time alignment or initialization characteristics, where the precise semantics of the properties are deferred to DASH or CMAF, and identified by identifiers defined in these specifications.

* + - * 1. **Syntax**

aligned(8) class SwitchingGroupBox

extends EntityToGroupBox('swit', version=0, flags)

{

unsigned int(1) switch\_flag;

unsigned int(1) timed\_aligned\_flag;

unsigned int(2) init\_type;

unsigned int(4) reserved;

if (flags & 0x001000) utf8string tag;

if (flags & 0x002000) utf8string int(32) structural\_brand;

if (flags & 0x004000) utf8string int(32) mediaprofile\_brand;

Box boxes[]; // optional other boxes e.g. ExtendedLanguageBox

}

Editor’s note: We should consider using the ‘flags’ of the box instead of defining bit fields in the box payload.

* + - * 1. **Semantics**

switch\_flag equal 1 indicates that the track of this group are alternative encodings of the same source content intended for adaptive streaming switching. The normative requirements applying to tracks belonging to such group are defined by DASH or CMAF and identified by the structural brand and/or media profile brand fields.

time\_aligned\_flag equal 1 indicates the tracks of this group have some timed alignment characteristics. The normative requirements applying to tracks belonging to such group are defined by DASH and CMAF and identified by the structural brand and/or the media profile brand fields.

init\_type with the following values:

* 0: The entity which its entity\_id is first listed in this box can be used to initialize a decoder for decoding any track that directly or indirectly belongs to this group.
* 1: Every track directly or indirectly belonging to this group can be used to initialize a decoder for decoding any track that directly or indirectly belongs to this group.
* 2: reserved
* 3: reserved

tag specifies additional information about the entity group which may be used for selection purposes . Derived specifications define the use of this field. For MPEG-H Audio the value of this field shall contain the whitespace-separated list of mae\_GroupIDs that are contained in the described switching group.

structural\_brand specifies an identifier defined in derived specifications that corresponds to structural constraints of all direct and indirect entities of this group.

mediaprofile\_brand specifies the media profile brand that all direct and indirect entities of this group conform to.

boxes is an array of boxes providing information about the group that can be used to generate DASH or CMAF groupings. Boxes suitable include but are not limited to the following list of boxes defined in this document:

* + ExtendedLanguageBox (subclause 8.4.6)
  + UserDataBox (subclause 8.10.1)
  + KindBox (subclause 8.10.4)
  + LabelBox (subclause 8.10.5)

Editor’s note: Changing the name is under consideration. One candidate:

AdaptiveStreamingEntityGrouping 'aseg'

# New screen target orientation

In clause 12.1.3.2, change:

class VisualSampleEntry(codingname) extends SampleEntry (codingname)  
{  
 unsigned int(16) pre\_defined = 0;  
 const unsigned int(16) reserved = 0;  
 unsigned int(32) pre\_defined[3] = 0;  
 unsigned int(16) width;  
 unsigned int(16) height;  
 template unsigned int(32) horizresolution = 0x00480000; // 72 dpi  
 template unsigned int(32) vertresolution = 0x00480000; // 72 dpi  
 const unsigned int(32) reserved = 0;  
 template unsigned int(16) frame\_count = 1;  
 uint(8) compressorname[32];  
 template unsigned int(16) depth = 0x0018;  
 int(16) pre\_defined = -1;  
 // other boxes from derived specifications  
 CleanApertureBox clap; // optional  
 PixelAspectRatioBox pasp; // optional  
}

To:

class VisualSampleEntry(codingname) extends SampleEntry (codingname)  
{  
 unsigned int(16) pre\_defined = 0;  
 const unsigned int(16) reserved = 0;  
 unsigned int(32) pre\_defined[3] = 0;  
 unsigned int(16) width;  
 unsigned int(16) height;  
 template unsigned int(32) horizresolution = 0x00480000; // 72 dpi  
 template unsigned int(32) vertresolution = 0x00480000; // 72 dpi  
 const unsigned int(32) reserved = 0;  
 template unsigned int(16) frame\_count = 1;  
 uint(8) compressorname[32];  
 template unsigned int(16) depth;  
 int(16) pre\_defined = -1;  
 // other boxes from derived specifications  
 CleanApertureBox clap; // optional  
 PixelAspectRatioBox pasp; // optional  
 // other optional boxes  
 ScreenOrientationBox ornt; // optional

}

Editors’ Note: lisiting the optional boxes is under discussion: whether to list them here, or whether in the definitions of the box, list the boxes in which they can be included. More input is welcomed.

In clause 12.1.3.3, change:

resolution fields give the resolution of the image in pixels-per-inch, as a fixed 16.16 number

frame\_count indicates how many frames of compressed video are stored in each sample. The default is 1, for one frame per sample; it may be more than 1 for multiple frames per sample

compressorname is a name, for informative purposes. It is formatted in a fixed 32-byte field, with the first byte set to the number of bytes to be displayed, followed by that number of bytes of displayable data encoded using UTF-8, and then padding to complete 32 bytes total (including the size byte). The field may be set to 0.

depth takes one of the following values

0x0018 – images are in colour with no alpha  
width and height are the maximum visual width and height of the stream described by this sample entry, in pixels

To:

resolution fields give the resolution of the image in pixels-per-inch, as a fixed 16.16 number

frame\_count indicates how many frames of compressed video are stored in each sample. The default is 1, for one frame per sample; it may be more than 1 for multiple frames per sample

compressorname is a name, for informative purposes. It is formatted in a fixed 32-byte field, with the first byte set to the number of bytes to be displayed, followed by that number of bytes of displayable data encoded using UTF-8, and then padding to complete 32 bytes total (including the size byte). The field may be set to 0.

depth takes one of the following values

0x0018 – the video sequence is in colour with no alpha

0x0028 – the video sequence is in grayscale with no alpha

0x0020 – the video sequence has alpha (gray or colour)

width and height are the maximum visual width and height of the stream described by this sample entry, in pixels

Editors’ Note: Other boxes that are allowed to be inserted in a VisualSampleEntry need a definition table (as we define for all other boxes) which will tell information about the Container, Quantity, Presence (mandatory vs not) etc.

Add the following new subclause after subclause 12.1.9:

12.1.10 Screen Orientation Box

12.1.10.1 Definition

When a video track is intended for a specific screen orientation for consumption, this creator’s intent is signalled using the ScreenOrientationBox. A player is expected to detect the current screen orientation of the device and then select an appropriate track based on this information.

When multiple video tracks are alternative of the same content but for different screen orientations, those tracks may be grouped in the same 'altr' entity group.

A given video track can be suitable for multiple screen orientations.

12.1.10.2 Syntax

[Ed. (MH): Why does 'ornt' extend a Box rather than a FullBox?]

aligned(8) class ScreenOrientationBox extends Box('ornt')   
{  
 bit(1) target\_screen\_orientation\_landscape;  
 bit(1) target\_screen\_orientation\_portrait;  
 bit(1) target\_screen\_orientation\_square;  
 bit(5) reserved;  
}

12.1.10.3 Semantics

target\_screen\_orientation\_landscape equal to 1 indicates that landscape is a suitable screen orientation, target\_screen\_orientation\_portrait equal to 1 indicates that portrait is a suitable screen orientation and target\_screen\_orientation\_square equal to 1 indicates that square is a suitable screen orientation.