INTERNATIONAL STANDARD

ISO/IEC 23001-10

Second edition
2020-04

AMENDMENT 1
2021-09

Information technology — MPEG systems technologies —

Part 10: Carriage of timed metadata metrics of media in ISO base media file format

AMENDMENT 1: Support for content-guided transcoding and spatial relationship of immersive media

Technologies de l'information — Technologies des systèmes MPEG —

Partie 10: Transport de métriques de métadonnées de temporisation de supports au format de fichier de support en base ISO

AMENDEMENT 1: Support pour transcodage guidé par le contenu et relation spatiale des média immersifs

Reference number
ISO/IEC 23001-10:2020/Amd.1:2021(E)

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

Published in Switzerland

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 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 or 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) or the IEC list of patent declarations received (see 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 www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

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

A list of all parts in the ISO/IEC 23001 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 and www.iec.ch/national-committees.

Information technology — MPEG systems technologies —

Part 10:
Carriage of timed metadata metrics of media in ISO base media file format

AMENDMENT 1: Support for content-guided transcoding and spatial relationship of immersive media

5.2.2

Replace:

“The decoder power indication metadata sample entry shall be as follows.“

with:

“The decoder-power indication metadata uses the following sample entry:“

Replace:

“The decoder-power indication sample shall conform to the following syntax:“

with:

“The decoder-power indication samples use the following syntax:“

5.3.2.2

Replace:

“Display power indication metadata shall use the following sample entry:

with:

“The display-power indication metadata uses the following sample entry:

Replace:

“The display power indication sample shall use the following syntax“

with:

“The decoder-power indication samples use the following syntax:“

5.3.3.2

Replace:

“The display fine control metadata sample entry shall store static metadata as follows.“

with:

“The display fine control metadata uses the following sample entry:“

Replace:

“The display fine control metadata sample shall use the following syntax:“

with:

“The display fine control metadata samples use the following syntax:“

Clause 6

Add new Clauses 7 and 8 at the end of Clause 6 as follows:

7 Carriage of timed metadata for content-guided transcoding

7.1 General

The metadata provide metrics on the streams that may be produced by a transcoding task. The metadata are encapsulated into timed metadata tracks of ISOBMFF file. Two types of sample entry are proposed to match two use-cases of content-guided transcoding. For an optimal usage of the metadata, the metadata producer needs to be informed of the transcoding function that is used in the network. The detailed description of the transcoding function is out of scope of this specification.

The timed metadata track is linked to the track it describes by means of a `'cdsc'` (content describes) track reference. The synchronisation of metadata when using CMAF standard is based on the same timed metadata tracks.

7.2 Recommended video format metadata

7.2.1 Definition

Sample Entry Type: `'rvrf'`

Container: Sample Description Box (`'stsd'`)

Mandatory: No

Quantity: 0 or 1

This metadata indicates the recommended video resolution and frame rate at different bitrates over the sample duration.

7.2.2 Syntax

Recommended Video Format Metadata use the following sample entry:

```
class RecVideoFormatMetaDataSampleEntry()
     extends MetaDataSampleEntry ('rvrf') {
   RecVideoFormatConfigurationBox();
}
class Device () {
   unsigned int(5) device_class;
   unsigned int(3) reserved = 0;
   unsigned int(32) codec_type;
   unsigned int(8) codec_profile;
}
aligned(8) class RecVideoFormatConfigurationBox
     extends FullBox('rvfc', version = 0, flags = 0) {
   unsigned int(5) num_devices;
```



```
   unsigned int(3) reserved = 0;
   unsigned int(4) num_bitrates;
   unsigned int(3) reserved = 0;
   unsigned int(1) rate_ctl_type;
   unsigned int(8) max_peak_bitrate_var_from_avg;
   for (i = 1; i <= num_devices; i++) {
      Device dev[i];
   }
   for (i = 1; i <= num_bitrates; i++)
      unsigned int(16) bitrate[i];
}
```

The Recommended Video Format samples use the following syntax:

```
class VideoFormat () {
   unsigned int(16) picture_width;
   unsigned int(16) picture_height;
   unsigned int(16) frame_rate;
}
class RecVideoFormatMetaDataSample (int num_devices, int num_bitrates) {
   for (i = 1; i <= num_devices; i++)
      for (j = 1; j <= num_bitrates; j++)
         VideoFormat rec_video_fmt[i][j];
}
}
```

7.2.3 Semantics

`device_class` specifies a service-related identifier for a class of devices. The value 0 is reserved and should be used to signal an unknown device class. It should be used when a service provider does not define and use any particular device classes.

The code points of the device class are not defined in this document, except for the value 0. A service provider is free to define and use a number of arbitrarily defined code points as long as they are correctly handled within its service. The registration of code points commercially relevant across different services is outside of the scope of this version of the document.

`codec_type` specifies the coding standard (using the four character code, defined in ISO/IEC 14496-15) for which metadata apply.

`codec_profile` specifies the profile of the coded representations for which metadata apply.

`num_devices` specifies the number of devices for which metadata apply. `num_bitrates` specifies the number of bitrates for which metadata apply.

`rate_ctl_type` specifies the type of rate control in the transcoding operation: 0 for Constant Bitrate, 1 for Variable Capped bitrate.

`max_peak_bitrate_var_from_avg` specifies the maximum variation in percentage of peak bitrate measured on a segment duration from the average bitrate measured on segment durations over a long period of time (for VOD, the period of time is the full duration; for live/linear, it is about one hour). When `rate_ctl_type` is 0, `max_peak_bitrate_var_from_avg` shall be set to 0.

`dev` specifies the device for which metadata apply and is characterized by its `device_class` and supported `codec_type` and `codec_profile`.

`bitrate` specifies the constant or capped bitrate of the video representation in kbits/s, measured on a segment duration.

`picture_width` specifies the width of the picture in units of luma samples.

`picture_height` specifies the height of the picture in units of luma samples.

`frame_rate` specifies the number of frame pictures over 256 seconds.

`rec_video_fmt` specifies the recommended video format, through `picture_width`, `picture_height` and `frame_rate` indications, for each device and bitrate .

7.3 Bitrate shaping metadata

7.3.1 General

Sample Entry Type: `'vbrs'`

Container: Sample Description Box (`'stsd'`)

Mandatory: No

Quantity: 0 or 1

This metadata indicates the minimum bitrate for a given video format to offer the optimal quality over the sample duration.

7.3.2 Syntax

Bitrate Shaping Metadata use the following sample entry:

```
class BitrateShapingMetaDataSampleEntry()
     extends MetaDataSampleEntry ('brsh') {
   BitRateShapingConfigurationBox();
}
aligned(8) class BitRateShapingConfigurationBox
   extends FullBox('brsC', version = 0, flags = 0) {
   unsigned int(5) num_devices;
   unsigned int(3) reserved = 0;
   unsigned int(3) num_video_formats;
   unsigned int(4) reserved = 0;
   unsigned int(1) rate_ctl_type;
   unsigned int(8) max_peak_bitrate_var_from_avg;
   for (i = 1; i <= num_devices; i++) {
      Device dev[i];
   }
   for (i = 1; i <= num_video_formats; i++) {
      VideoFormat video_fmt[i];
   }
}
```


The Bitrate Shaping samples use the following syntax:

```
class BitrateShapingMetaDataSample
      (int num_devices, int num_video_formats) {
   for (i = 1; i <= num_devices; i++)
      for (j = 1; j <= num_video_formats; j++)
         unsigned int(16) min_bitrate[i][j];
}
```

7.3.3 Semantics

`num_devices` specifies the number of devices for which metadata apply.

`num_video_formats` specifies the number of video formats for which metadata apply. The video formats are described in a decreasing order of luma samples per second.

`rate_ctl_type` specifies the type of rate control in the transcoding operation: 0 for Constant Bitrate, 1 for Variable Capped bitrate.

`max_peak_bitrate_var_from_avg` specifies the maximum variation in percentage of peak bitrate measured on a segment duration from the average bitrate measured on segment durations over a long period of time (for VOD, the period of time is the full duration; for live/linear, it is about one hour). When `rate_ctl_type` is 0, `max_peak_bitrate_var_from_avg` shall be set to 0.

dev specifies the device for which metadata apply and is characterized by its `device_class` and supported `codec_type` and `codec_profile`.

`video_fmt` specifies the video format through `picture_width`, `picture_height` and `frame_rate` indications.

`min_bitrate` specifies, for each device and video format, the minimum bitrate of the video representation in kbits/s, measured on a segment duration, to offer the optimal quality over this duration. This minimum bitrate is set in function of the rate control which applies in the transcoding operation as defined by `rate_ctl_type` and `max_peak_bitrate_var_from_avg`.

8 Timed metadata for spatial relationships of immersive media

8.1 General

This clause defines metadata data structures for spatial sources (e.g., 3D bounding boxes) and regions within the sources of immersive media, especially visual volumetric video-based coding data (V3C), and their carriage in timed metadata tracks for signaling spatial relationship of the sources and regions.

8.2 Signalling of spatial relationship of spatial regions using 3d cartesian coordinates in timed metadata tracks

8.2.1 General

Sample Entry Type: `'3dcc'`

Container: Sample Description Box ('stsd')

Mandatory: No

Quantity: 0 or 1

The 3D Cartesian coordinates sample entry provides spatial information related to the referenced track expressed in a three-dimension Cartesian coordinate system.

Sync samples for ROI metadata tracks are samples for which the `interpolate` value is 0.

8.2.2 Syntax


```
aligned(8) class 3DCartesianCoordinatesSampleEntry
   extends MetadataSampleEntry ('3dcc') {
   unsigned int(16) reference_width;
   unsigned int(16) reference_height;
   unsigned int(16) reference_depth;
}
aligned(8) class 3DCartesianCoordinatesSample(){
   unsigned int(16) top_left_near_x;
   unsigned int(16) top_left_near_y;
   unsigned int(16) top_left_near_z;
   unsigned int(16) width;
   unsigned int(16) height;
   unsigned int(16) depth;
   unsigned int(1) interpolate;
   unsigned int(7) reserved;
}
```

8.2.3 Semantics

`reference_width`, `reference_height` and `reference_depth` give respectively the width, height and depth of the reference cuboid space in which all ROI coordinates (`top_left_near_x`, `top_left_near_y`, `top_left_near_z`, `width`, `height` and `depth`) are computed. These fields allow associating a ROI metadata track with immersive media tracks of different resolutions but representing the same visual source.

`top_left_near_x`, `top_left_near_y`, `top_left_near_z` give respectively the horizontal, vertical and stacked coordinate of the top-left-near corner of the cuboid region associated with the 3D media sample of the referenced track.

`width`, `height` and `depth` give respectively the width, height and depth of the cuboid region associated with the media sample of the referenced track.

`interpolate` indicates the continuity in time of the successive samples. When true, the application may linearly interpolate values of the ROI coordinates between the previous sample and the current sample. When false, there shall not be any interpolation of values between the previous and the current samples.

NOTE When using interpolation, it is expected that the interpolated samples match the presentation time of the samples in the referenced track. For instance, for each visual sample of an immersive media track, one interpolated 3D Cartesian coordinate sample is calculated.