ISO/IEC JTC 1/SC 29/WG 03 N00773

**ISO/IEC JTC 1/SC 29/WG 03  
MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title:** Technology under Consideration on ISO/IEC 23008-12

**Status:** Approved

**Date of document:** 2023-02-01

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

**No. of pages:** 3 (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 N00773**

**February 2023, Virtual**

|  |  |
| --- | --- |
| **Title** | **Technology under Consideration on ISO/IEC 23008-12** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | **22297** |

Abstract

This document collects following candidate technologies for the High Efficiency Image File Format (HEIF) (ISO/IEC 23008-12).

Table of Contents

[Abstract 1](#_Toc126244791)

[1 Region annotations for image sequence or video tracks 2](#_Toc126244792)

[1.1 Region extrapolation (from m60304, MPEG#139, Issue#76) 2](#_Toc126244793)

[1.1.1 Overview 2](#_Toc126244794)

[1.1.2 Text Proposal 2](#_Toc126244795)

[1.2 Region interpolation (from m59508, MPEG#138, Issue#69 comment#60556) 5](#_Toc126244796)

[1.2.1 Text Proposal 5](#_Toc126244797)

[1.2.2 Discussion 6](#_Toc126244798)

[2 Region annotation for image items 6](#_Toc126244799)

[2.1 Combination of regions (from m62028, MPEG#141, Issue#88) 6](#_Toc126244800)

[2.1.1 Summary/Motivation 6](#_Toc126244801)

[2.1.2 Text Proposals 7](#_Toc126244802)

[3 Matrix-based transformation for image items 8](#_Toc126244803)

[4 Signaling for pre-derived coded image items 8](#_Toc126244804)

[5 On HDR signaling alignment with ISO 22028-5 (from m62055, MPEG#141, Issue#86) 9](#_Toc126244805)

[5.1 Motivation 9](#_Toc126244806)

[5.2 Text proposal 9](#_Toc126244807)

[6 On MPEG/JPEG file embedding (MPEG#141, Issue#87) 11](#_Toc126244808)

[6.1 Discussion 11](#_Toc126244809)

[6.2 Initial text proposal 11](#_Toc126244810)

# Region annotations for image sequence or video tracks

## Region extrapolation (from [m60304](https://dms.mpeg.expert/doc_end_user/documents/139_OnLine/wg11/m60304-v1-m60304-Regionextrapolationfortracks.zip), MPEG#139, [Issue#76](https://mpeg.expert/software/MPEG/Systems/FileFormat/HEIF/-/issues/76))

### Overview

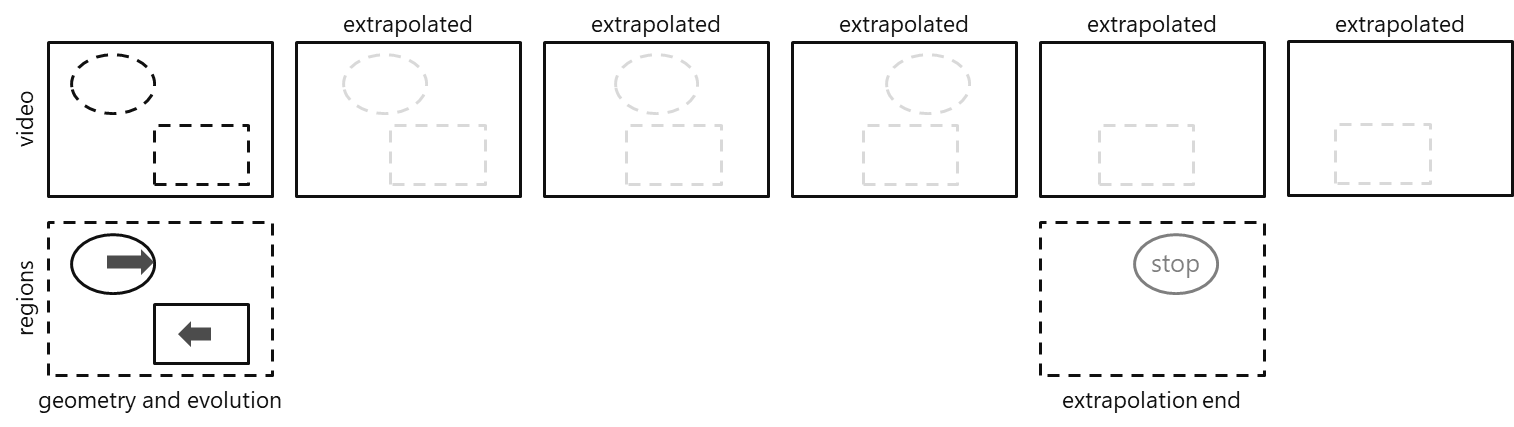


Figure 1: example region description for tracks using extrapolation

Figure 1 shows an example of describing several regions using extrapolation. The video track shown at the top contains two regions, an elliptic one and a rectangular one. The elliptic one is present in the four first samples of the video track and moves to the right of the image. The rectangular one is present in all the samples of the video track and moves to the left of the image.

The region track shown at the bottom describes these two regions. In a first sample, corresponding to the first sample of the video track, these two regions are described with their positions and sizes and the evolution of their respective positions and sizes. There are no region samples corresponding to the three following video samples. The region sample corresponding to the fifth video sample signals that the interpolation of the elliptic region ends.

### Text Proposal

*Update the definition of a region track ( section 7.5.4.1) by adding the following paragraphs:*

The geometry of a region may be defined by specifying the shape, position and size of the region in a sample of the region track. The geometry of a region may also be defined as an initial geometry and its evolution over time by specifying the initial geometry of the region and its evolution in a sample of the region track.

The evolution of a region over time is optional. It can be represented by the evolution speed of some of its parameters inside the reference space. The evolution speed of the parameters is signaled using a scaling factor for increasing its precision. The parameters defining the evolution of a region depend on the geometry of the region as follows:

— When the geometry of a region is represented by a point, the evolution of the region is defined by the evolution of the position of this point.

— When the geometry of a region is represented by a rectangle or an ellipse, the evolution of the region is defined by the evolution of the position and the size of the rectangle or ellipse.

— When the geometry of a region is represented by a polygon or a polyline, the evolution of the region is defined by the evolution of the position of each point of the polygon or polyline. The number of points in the polygon or polyline doesn’t change.

— When the geometry of a region is represented by a mask, the evolution of the region is defined by the evolution of the position of the mask.

The evolution of a region stops when another sample contains a region with the same region identifier. The evolution of a region shall stop for each sync sample of the source track.

*Update the Sample format (section 7.5.4.2.1) with the following paragraph*

When the extrapolate flag is set to 1 for a region inside a sample of a region track, the region is an evolving region defined by an initial geometry and its evolution over time.

The value of each evolving parameter defining the geometry of the region at a given composition time *T* can be computed as follows:

where:

* *param0* is the initial value of the parameter as defined in the initial geometry of the region at time T0.
* *Δparam* is the evolution of the parameter as defined in the evolution of the region.
* *evolution\_scale* is a scaling factor for the evolution values equal to , where is the field\_size and is equal to ((RegionTrackConfigBox.field\_length\_size & 1) + 1) \* 16.
* *T0* is the composition time of the sample defining the evolving region.
* *ΔT* is the duration of the sample defining the evolving region.

*Update the syntax of Sample format (section 7.5.4.2.2) as follows*

aligned (8) class RegionSample {  
 unsigned int field\_size = ((RegionTrackConfigBox.field\_length\_size & 1) + 1) \* 16;  
// this is a temporary, non-parsable variable  
 unsigned int(32) region\_count;  
 for (r=0; r < region\_count; r++) {  
 unsigned int(32) region\_identifier;  
 unsigned int(8) geometry\_type;  
 unsigned int(1) extrapolate;  
 unsigned int(7) reserved;  
 if (geometry\_type == 0) {  
 // point  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 }  
 }  
 else if (geometry\_type == 1) {  
 // rectangle  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 signed int(field\_size) delta\_width;  
 signed int(field\_size) delta\_height;  
 }  
 }  
 else if (geometry\_type == 2) {  
 // ellipse  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) radius\_x;  
 unsigned int(field\_size) radius\_y;  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 signed int(field\_size) delta\_radius\_x;  
 signed int(field\_size) delta\_radius\_y;  
 }  
 }  
 else if (geometry\_type == 3 || geometry\_type == 6) {  
 // polygon or polyline  
 unsigned int(field size) point\_count;  
 for (i=0; i < point\_count; i++) {  
 signed int(field\_size) px;  
 signed int(field\_size) py;  
 }  
 if (extrapolate == 1) {  
 for (i=0; i < point\_count; i++) {  
 signed int(field\_size) delta\_px;  
 signed int(field\_size) delta\_py;  
 }  
 }  
 }  
 else if (geometry\_type == 4) {  
 // referenced mask  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 unsigned int(field\_size) track\_mask\_idx;  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 }  
 }  
 else if (geometry\_type == 5) {  
 // inline mask  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 unsigned int(8) mask\_coding\_method;  
 if (mask\_coding\_method != 0)  
 unsigned int(32) mask\_coding\_parameters;  
 bit(8) data[];  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 }  
 }  
 else if (geometry\_type == 7) {  
 // empty region  
 }  
 }  
}

*Update the semantics of Sample format (section 7.5.4.2.3) with the following text:*

7: the region is an empty region used for signalling the end of the evolution of a previous region with the same region identifier.

Other values are reserved.

extrapolate is a flag indicating whether the geometry changes of the region are specified or not. When equal to 0, it indicates that no geometry changes are specified for the region. When equal to 1, it indicates that both the geometry and the geometry changes are specified for the region.

(…)

evolution\_scale is the scaling factor for the specification of the evolution values, equal to , where is the field\_size and is equal to ((RegionTrackConfigBox.field\_length\_size & 1) + 1) \* 16.

delta\_x, delta\_y specify, in 1/evolution\_scale units of the reference space, the evolution of the x and y fields for the region.

delta\_width, delta\_height specify, in 1/evolution\_scale units of the reference space the evolution of the width and height fields for the region.

delta\_radius\_x, delta\_radius\_y specify, in 1/evolution\_scale units of the reference space the evolution of the radius\_x and radius\_y fields for the region.

delta\_px, delta\_py specify, in 1/evolution\_scale units of the reference space the evolution of the px, py fields for a point of the region.

## Region interpolation (from [m59508](https://dms.mpeg.expert/doc_end_user/documents/138_OnLine/wg11/m59508-v1-m59508-Regionannotationfortracks.zip), MPEG#138, [Issue#69 comment#60556](https://mpeg.expert/software/MPEG/Systems/FileFormat/HEIF/-/issues/69" \l "note_60556))

*[Ed. (FD)This section only contains parts of the contribution requiring further discussion, i.e., the interpolate flag in sample format for region tracks]*

### Text Proposal

**X.X Region track and region annotations for an image sequence or video track**

**X.X.3 Sample format**

**X.X.3.1 Definition**

This subclause defines the sample format for region track. A sample of a region track defines one or more regions.

**X.X.3.2 Syntax**

aligned (8) class RegionSample {  
 unsigned int field\_size = ((RegionTrackConfigBox.field\_length\_size & 1) + 1) \* 16;   
// this is a temporary, non-parsable variable  
 unsigned int(7)reserved;  
 unsigned int(1)interpolate;  
 unsigned int(16) region\_count;  
 for (r=0; r < region\_count; r++) {  
 (…)  
 }  
}

**X.X.3.3 Semantics**

interpolate indicates the continuity in time of the successive samples. When true, the application may linearly interpolate values of the region geometries 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 1 When using interpolation, it is expected that the interpolated samples match the presentation time of the samples in the referenced source track. For instance, for each video sample of a video track, one interpolated region sample is calculated.

(…)

### Discussion

About the interpolate flag: The purpose is to avoid declaring a sample in the region track for each sample of the media track when regions are moving linearly between two positions. Imagine a sample A in the region track with a region at a starting position A and this region is moving linearly to the arrival position B nine samples later. Instead of declaring ten samples in the region track, you can only declare two samples, sample A with a duration corresponding to nine samples in the media track, followed by sample B providing the arrival position B. We should clarify that since the interpolate flag applies to all regions in the sample, the number of regions shall be the same in sample A and B.

# Region annotation for image items

## Combination of regions (from [m62028](https://dms.mpeg.expert/doc_end_user/documents/141_OnLine/wg11/m62028-v1-m62028-Regioncombination.zip), MPEG#141, [Issue#88](https://mpeg.expert/software/MPEG/Systems/FileFormat/HEIF/-/issues/88))

### Summary/Motivation

This contribution proposes two extensions for region annotation feature:

1. The ability to associate annotations to a region defined as the union of several regions

A new type of derived region item (items 100 and 101 in Figure 1) is defined to allow grouping several regions to signal that those regions are part of a group and to be able to annotate this group as a whole (e.g. these persons are part of the same family)

1. The ability to indicate a relation between several regions

A new type of entityToGroup is defined to allow signaling that a region item (person in figure 2 below) represents/covers an area including several other regions (as an aggregation) (left arm, right leg, etc...). In terms of geometry, the inclusion does not need to be total.

|  |  |
| --- | --- |
|  |  |
| Figure 1 - Example of union of regions | Figure 2 – Example of inclusion relationship |

### Text Proposals

#### Proposal 1: Union of regions

*Add the following section in section 6.11 Derived region items*

**6.11.2.2 Union derivation**

An item with an item\_type value of 'cbrg' defines a derived region item that corresponds to the union of all the regions represented by one or more input region items.

The input region items are specified in a SingleItemTypeReferenceBox of type 'drgn' for this derived region item within the ItemReferenceBox. In the SingleItemTypeReferenceBox of type 'drgn', the value of from\_item\_ID identifies the derived region item of type 'cbrg' and the values of to\_item\_ID identify the input region items.

The union derived region item is associated with the image item inside which the regions are defined using an item reference of type 'cdsc' from the union derived region item to the image item.

The region resulting from this derived region item is the union of all the regions of each input region item after being applied to the referenced image item as specified in 6.10.1.1.

#### Proposal 2: relations between region items

*Add the following section in section 6.10 Region items and region annotations*

**6.10.4 Region Entity Group**

**6.10.4.1** 'corg' **Entity Group**

A compound region entity group ('corg') associates one main region item with one or more region items. It indicates an inclusion relationship between a main object covered by regions of a main entity and other objects covered by regions described by one or more other entities, the main object logically including the other objects.

NOTE For example, a compound region entity group can be used to associate a main region corresponding to a car with regions corresponding to wheels to indicate that the car is logically including the wheels.

The entities in a compound region entity group shall be region items. The number of entities in a compound region entity group shall be at least 2. The first entity\_id value shall indicate the main region item. It indicates the region covering the main object that is logically including the objects covered by the regions described by the second and following entity\_ids.

This inclusion relationship does not convey information at the geometry level. A main region signalled as including others regions by a compound entity group may or may not geometrically include the other regions.

*[[ Ed. (FM): MPEG#141: it was questioned:*

* *Can the Figure 1 in the proposal be achieved for example using mask items where all the regions belonging to a group is part of the mask item.*
* *Is there a restriction for any of the proposals that the separate regions must all be derived from a single image item?”]]*

# Matrix-based transformation for image items

*[[ Ed. (FD): MPEG#129: it was questioned:”* Should we also add ‘matrix’ as an image derivation in the HEIF? “. It was warned that “We would need to be clear about the meaning of outputs that don’t have horizontal and vertical sides; if that’s overlaid, the meaning is clear, but what if it’s supposed to be displayed?”*]]*

# Signaling for pre-derived coded image items

*Replace the clause 6.4.7 with the following text:*

**6.4.7** **Pre-derived coded images**

[Ed. (FD): In the following, differences with HEIF 2nd edition (w18310) are highlighted in blue]

If a coded image has been derived from others — for example, a composite HDR image derived from exposure-bracketed individual images, or a panorama derived from a set of images — then it shall be linked to those images by item references of type 'base'. Item references may be from the coded image to all images it derives from, or when unique IDs are used, from the coded image to all entity groups or images it derives from. When unique IDs are used, a to\_item\_ID value in the SingleItemTypeReferenceBox or SingleItemTypeReferenceBoxLarge is resolved to an item identifier whenever the embedding MetaBox contains an item with such identifier, and is resolved to an entity group identifier otherwise.

An image item including a 'base' item reference is referred to as a pre-derived coded image.

NOTE In this version of this document, the exact derivation process used to produce the image is not described.

[[Ed. (FD): At MPEG#129, it was commented that “The slight snag here is defining what it means when the entity group does NOT imply a single output (e.g. a slide show); what does pre-derivation mean? ]]

*Add the following clause as section 6.4.7.1:*

**6.4.7.1 Signaling of the derivation method for pre-derived coded image items**

A pre-derived coded image shall be linked to images it derives from by an item reference of type 'base' to the entity group containing all images the pre-derived coded images derives from. The grouping\_type of the EntityToGroupBox specifies the purpose of grouping and implicitly signals the type of the derivation operation which was applied to generate the pre-derived coded image.

[[Ed. (FM): At MPEG#126, it was commented that “we somehow need to indicate the derivation operation, rather than the nature of the input set”]]

[[Ed. (FD): At MPEG#129, it was commented that “We could allow a pre-derivation of the implied derivation of that entity group.”]]

# On HDR signaling alignment with ISO 22028-5 (from [m62055](https://dms.mpeg.expert/doc_end_user/current_document.php?id=85840), MPEG#141, [Issue#86](https://mpeg.expert/software/MPEG/Systems/FileFormat/HEIF/-/issues/86))

## Motivation

ISO TC42 is developing specification ISO 22028-5 [1], which defines a set of colour image encodings for use in storage, transmission, and display of HDR and WCG digital still images. The purpose of this proposal is to initiate the alignment process with the definitions from the ISO 22028-5 and to provide the signaling for the related metadata at MPEG.

**Reference viewing environment metadata**

To properly interpret the color appearance of images encoded in a color image encoding, ISO 22028-1 specifies the reference image viewing environment that can be used to provide context for interpreting the intended color appearance of the encoded image colorimetry.

For cases where the image is to be viewed in an actual viewing environment significantly different than the specified reference image viewing environment, it might be desirable to use a colour appearance transform to determine corresponding image colorimetry that would produce the intended colour appearance in the actual viewing environment.

ISO 22028-5 [1] specifies parameters to establish a reference viewing environment in which images with display viewing colorimetry are intended to be viewed. These parameters include luminance of surround and periphery as well as color temperature of both.

**Nominal diffuse white luminance**

This data might be used by tone mapping algorithms at display time, to compensate differences between the reference display and the display of the user.

**Additional metadata**

Additional metadata for handling new transfer characteristics will also need to be defined as indicated in [2].

1. ISO TC 42/WG 23, TS 22028-5 "Photography and graphic technology — Extended colour encodings for digital image storage, manipulation and interchange — Part 5: High dynamic range and wide colour gamut image encoding for still images (HDR/WCG)"
2. N. Bonnier, D. Concion, D. Podborski, J. Roland, A. Tourapis: "ISO 22028-5 impact on CICP", [m62260](https://dms.mpeg.expert/doc_end_user/current_document.php?id=86045&id_meeting=193)

## Text proposal

**X..Y Reference viewing environment**

The reference viewing environment applies to display-viewing colorimetry, not to scene-referred colorimetry. It specifies the luminance and chromaticity parameters for the “surround” and “periphery” of the display. The “surround” is the area surrounding a display that can affect the adaptation of the eye, typically the wall or curtain behind the display, while “periphery” is the remaining environment outside of the surround.

**X.Y.1 Syntax**

|  |  |
| --- | --- |
| Box type: | 'reve' |
| Property type: | Descriptive item property |
| Container: | ItemPropertyContainerBox |
| Mandatory (per item): | No |
| Quantity (per item): | At most one |

class ReferenceViewingEnvironmentBox extends ItemFullProperty('reve', 0, 0){  
 unsigned int(32) surround\_luminance;  
 unsigned int(16) surround\_light\_x;  
 unsigned int(16) surround\_light\_y;  
 unsigned int(32) periphery\_luminance;  
 unsigned int(16) periphery\_light\_x;  
 unsigned int(16) periphery\_light\_y;  
}

**X.Y.2 Semantics**

surround\_luminance specifies the luminance of the surround in units of 0.0001 candelas per square metre.

[Ed. note]: disallowing 0 could be considered.

surround\_light\_x and surround\_light\_y specify the normalized x and y chromaticity coordinates, respectively, of the environmental reference surround light in the nominal viewing environment. These parameters are according to the CIE 1931 definition of x and y as specified in ISO 11664-1 (see also ISO 11664-3 and CIE 15) and are in normalized increments of 0.0001. The values of surround\_light\_x and surround\_light\_y shall be in the range of 0 to 10 000, inclusive.

periphery\_luminance specifies the luminance of the periphery in units of 0.0001 candelas per square metre.

[Ed. note]: disallowing 0 could be considered.

periphery\_light\_x and periphery\_light\_y specify the normalized x and y chromaticity coordinates, respectively, of the environmental reference periphery light in the nominal viewing environment. These parameters are according to the CIE 1931 definition of x and y as specified in ISO 11664-1 (see also ISO 11664-3 and CIE 15) and are in normalized increments of 0.0001. The values of periphery\_light\_x and periphery\_light\_y shall be in the range of 0 to 10 000, inclusive.

**X.Z Nominal Diffuse White**

**X.Z.1 Syntax**

|  |  |
| --- | --- |
| Box type: | 'ndwt' |
| Property type: | Descriptive item property |
| Container: | ItemPropertyContainerBox |
| Mandatory (per item): | No |
| Quantity (per item): | At most one |

class NominalDiffuseWhiteBox extends ItemFullProperty('ndwt', 0, 0){  
 unsigned int(32) diffuse\_white\_luminance;  
 unsigned int(16) diffuse\_white\_light\_x;  
 unsigned int(16) diffuse\_white\_light\_y;  
}

**X.Z.2 Semantics**

diffuse\_white\_luminance indicates the default nominal diffuse white luminance in units of 0.0001 candelas per square metre.

[Ed. note]: disallowing 0 could be considered.

diffuse\_white\_light\_x and diffuse\_white\_light\_y specify the normalized x and y chromaticity coordinates, respectively, of the environmental reference periphery light in the nominal viewing environment. These parameters are according to the CIE 1931 definition of x and y as specified in ISO 11664-1 (see also ISO 11664-3 and CIE 15) and are in normalized increments of 0.0001. The values of periphery\_light\_x and periphery\_light\_y shall be in the range of 0 to 10 000, inclusive and should be the same as the chromaticity coordinates of the white point of the content.

[Ed. note]: We could also add a note mentioning that values diffuse\_white\_light\_x = 3 127 and diffuse\_white\_light\_y = 3 290 could be used to signal D65.

# On MPEG/JPEG file embedding (MPEG#141, [Issue#87](https://mpeg.expert/software/MPEG/Systems/FileFormat/HEIF/-/issues/87))

## Discussion

During MPEG 140 (cf. [ISOBMFF/Issue#146](https://mpeg.expert/software/MPEG/Systems/FileFormat/isobmff/-/issues/146)), the potential improvement of ISOBMFF 8th edition was extended with a definition of the UUID (see text in section 6.2 below) to enable embedding an ISO base media file within another file. One of these use-cases would be to embed ISOBMFF in JPEG based on JUMBF ISO/IEC 19566-5, which would also allow HEIF files to be embedded into a JPEG file.

At MPEG#141, it was decided to remove the proposed text from ISOBMFF 8th edition for further study in HEIF. It was pointed out that embedding HEIF into JPEG may lead to sub-optimal encapsulation and compatibility issues. Uses cases were also questioned.

## Initial text proposal

*[Ed.(FM): The text below was initially included into potential improvement of ISOBMFF 8th edition clause 6.8 at MPEG#140 and then removed at MPEG#141 for further study]*

**6.8 UUID value for embedded ISO base media files**

When embedding an ISO base media file into a file compliant to another file format that needs a UUID to identify the format of the embedded file, the UUID to identify the ISO base media file shall be equal to 0x49534F30-0011-0010-8000-00AA00389B71.

NOTE This UUID enables embedding an ISO base media file within a file conforming to the JPEG Universal Metadata Box Format (JUMBF, ISO/IEC 19566-5). The JUMBF Content Type in the JUMBF Description box is set equal to the UUID specified above in this subclause. The JUMBF superbox contains a single content box that contains the ISO base media file.