ISO/IEC JTC 1/SC 29/WG 03 N1431

**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:** 2025-02-07

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

**No. of pages:** 9 (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 N1431**

**January 2025 – Geneva, CH**

|  |  |
| --- | --- |
| **Title** | **Technology under Consideration on ISO/IEC 23008-12** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | **24731** |

Abstract

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

Table of Contents

[1 Signaling for pre-derived coded image items 3](#_Toc189729837)

[2 On MPEG/JPEG file embedding (MPEG#141, Issue#87) 3](#_Toc189729838)

[2.1 Discussion 3](#_Toc189729839)

[2.2 Initial text proposal 3](#_Toc189729840)

[3 Generic compression of items (MPEG #147, issue #157)) 4](#_Toc189729841)

[3.1 Abstract 4](#_Toc189729842)

[4 Marking image items as belonging to a session (MPEG#148, Issue#166) 4](#_Toc189729843)

[4.1 Abstract 4](#_Toc189729844)

[4.2 Capture session identifier item property 4](#_Toc189729845)

[4.2.1 Definition 4](#_Toc189729846)

[4.2.2 Syntax 5](#_Toc189729847)

[4.2.3 Semantics 5](#_Toc189729848)

[4.3 Alternatives considered 5](#_Toc189729849)

[5 On user description of Images (MPEG#148, Issue#170) 5](#_Toc189729850)

[5.1 Abstract 5](#_Toc189729851)

[5.1.2 Syntax 6](#_Toc189729852)

[5.1.3 Semantics 6](#_Toc189729853)

[6 On GeoKey Item property (MPEG#148, Issue#172) 9](#_Toc189729854)

[6.1 Abstract 9](#_Toc189729855)

[6.x.y GeoSpatial Keys Property 9](#_Toc189729856)

[6.1.1 6.x.y.1 Definition 9](#_Toc189729857)

[6.1.2 6.x.y.2 Syntax 9](#_Toc189729858)

[6.1.3 6.x.y.3 Semantics 9](#_Toc189729859)

[7 Compact VVC decoder configuration (MPEG#149, Issue#178) 9](#_Toc189729860)

[7.1 Abstract 9](#_Toc189729861)

# 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 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.

# Generic compression of items (MPEG #147, issue [#157](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/157)))

## Abstract

During MPEG #147 the compression of EXIF metadata item was proposed by introducing a new item type for compressed EXIF. However, defining a new item type for a different encoding may not always be a right design choice. Alternative options for meeting the requirements also need to be explored. During MPEG #147 different alternate approaches were discussed in the MPEG GitHub for a generic compression of items(under issue [#157](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/157)) .

1. Define a URI that defines the content encoding and use that with a uri item
2. Add a new infe version
3. Extending the mechanism currently under definition for generic compression of uncompressed image items to apply to any uncompressed items (ISO/IEC 23001-17 CDAM2 - [w23515](https://dms.mpeg.expert/doc_end_user/documents/145_OnLine/wg11/MDS23515_WG03_N01148-v3.zip" \t "_blank))

# Marking image items as belonging to a session (MPEG#148, [Issue#166](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/166))

## Abstract

In m69517 the capture session identifier item property was proposed which provides the session id for images captured in the same session or event.

## Capture session identifier item property

### Definition

|  |  |
| --- | --- |
| Box type: | 'casi' |
| Property type: | Descriptive item property |
| Container: | ItemPropertyContainerBox |
| Mandatory (per associated item): | No |
| Quantity (per associated item): | Zero or more |

The capture session identifier item property allows a file writer to associate an image item with a session identifier. All image items (potentially in multiple files) sharing the same session identifier are marked as belonging to the same session or event.

NOTE Example use cases are multiple files captured from different angles of an object, or the same scene or photographic event captured with multiple cameras or settings.

If two or more image items belonging to different files

* share the same capture session identifier
* are associated with CameraExtrinsicMatrixProperty boxes containing the same coordinate system id

they shall share the same global coordinate system rather than having local coordinate systems per file.

The CreationTimeProperty item property should be used to signal the capture time of the image items associated with a specific session identifier.

### Syntax

aligned(8) class CaptureSessionIdentifierProperty  
extends ItemFullProperty('casi', version = 0, flags) {  
 int has\_uuid = flags & 0x1 == 1;  
 if (has\_uuid) {

unsigned int(8) session\_uuid[16];

}  
 else {  
 utf8string session\_uri;  
 }  
}

### Semantics

version shall be equal to 0.

(flags & 0x1) equal to 1 specifies that the property contains a session\_uuid rather than a session\_uri.

session\_uri specifies a Uniform Resource Name (URI) that identifies the global session or event that this image item belongs to. URI's of type " urn:uuid:XYZ" should instead be directly specified as a session\_uuid.

session\_uuid specifies a UUID that identifies the global session or event that this image item belongs to. Specifying a session\_uuid is equivalent to specifying a session\_uri of type " urn:uuid:XYZ".

## Alternatives considered

One alternative to the proposed solution for linking the coordinate spaces between files would be to add a new version to the CameraExtrinsicMatrixProperty that adds a global session UUID. This would work but would not allow someone to solve the other listed use-cases.

During MPEG #148 different alternate approaches were discussed in the MPEG Git for signalling images captured in the same session (under [Issue#166](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/166))

1. using the Album entity group
2. 'tsyn' entity group to indicate synchronous capture of images
3. unique ids for indicating files that are related to each other (m64321, m65210)

# On user description of Images (MPEG#148, [Issue#170](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/170))

## Abstract

In m70076 extension to User description was proposed to indicate additional information related to video tracks and images associated with a user description.

#### User description

#### Definition

The UserDescriptionProperty permits the association of item(s) or entity group(s) with ~~a~~ user-defined information which is related to image content ~~name, description and tags~~; there may be multiple such properties~~, which shall have different language codes~~.

When several instances of UserDescriptionProperty are associated with the same item or entity group, each instance shall have different value of association\_info\_type. ~~they represent alternatives possibly expressed in different languages and a reader should choose the most appropriate. At most one UserDescriptionProperty with the same alt\_lang value should apply to the same item or entity group~~.

### Syntax

aligned(8) class UserDescriptionProperty  
extends ItemFullProperty('udes', version = 0, flags = 0){

unsigned int(8) association\_info\_type;

unsigned int(1) presentation\_flag;

unsigned int(1) lang\_flag;

bit(6) reserved;

unsigned int(8) num\_info\_entry;

for(i=0; i<num\_info\_entry; i++){

if(language\_flag == 1){

utf8string lang;

}

utf8string association\_info\_string;

}  
 ~~utf8string name;  
 utf8string description;  
 utf8string tags;~~  
}

### Semantics

association\_info\_type is an integer value that specifies the information type of the associated string. When equals to 0, association\_info\_string indicates user defined description of the corresponding item(s) or entity group(s); when equals to 1, association\_info\_string indicates text recognition information of the corresponding item(s) or entity group(s); when equals to 2, association\_info\_string indicates object recognition information of the corresponding item(s) or entity group(s); when equals to 3, association\_info\_string indicates scene understanding information of the corresponding item(s) or entity group(s). Other values are reserved.

presentation\_flag equals to 1 specifies that association\_info\_string shall be presented with the corresponding image; equals to 0 specifies that association\_info\_string shall not be presented with the corresponding image.

lang\_flag equals to 1 specifies that lang shall be indicated; equals to 0 specifies that lang shall not be indicated.

num\_info\_entry indicates the number of associated info entries.

lang is a character string containing an RFC 5646 compliant language tag string, such as "en-US", "fr-FR", or "zh-CN“, representing the language of the text contained in name, description and tags. When lang is empty, the language is unknown/undefined.

association\_info\_string is a null-terminated UTF-8 character string containing human readable text of the corresponding association\_info\_type.

~~name is a null-terminated UTF-8 character string containing human readable name for the item or group of entities. If not present (an empty string is supplied) no name is provided.~~

~~description is a null-terminated UTF-8 character string containing human readable description of the item or group of entities. If not present (an empty string is supplied) no description is provided.~~

~~tags is a null-terminated UTF-8 character string containing comma-separated user-defined tags related to the item(s). If not present (an empty string is supplied) no tags is provided.~~

#### User description sample group

##### Definition

The user description ('udes') sample grouping allows associating user-defined information ~~annotations, such as a name, a description or tags~~, with a group of samples or with a region within a sample.

The association of a UserDescriptionSampleGroupEntry with a region within a sample is defined using a SampleToRegionIdMappingEntry of a sample-to-region-id-mapping sample grouping with an annotation\_container\_type equal to 1 and an annotation\_reference\_type equal to 'udes'.

##### Syntax

class UserDescriptionSampleGroupEntry ()  
 extends MetadataSampleGroupEntry ('udes')  
{  
 unsigned int(8) association\_info\_type;

unsigned int(1) presentation\_flag;

unsigned int(1) lang\_flag;

bit(6) reserved;

unsigned int(8) num\_info\_entry;

for(i=0; i<num\_info\_entry; i++){

if(language\_flag == 1){

utf8string lang;

}

utf8string association\_info\_string;

}  
 ~~utf8string name;  
 utf8string description;  
 utf8string tags;~~  
}

##### Semantics

association\_info\_type is an integer value that specifies the information type of the associated string. When equals to 0, association\_info\_string indicates user defined description of the corresponding sample or the region within the sample; when equals to 1, association\_info\_string indicates text recognition information of the corresponding sample or the region within the sample; when equals to 2, association\_info\_string indicates object recognition information of the corresponding sample or the region within the sample; when equals to 3, association\_info\_string indicates scene understanding information of the corresponding sample or the region within the sample. Other values are reserved.

presentation\_flag equals to 1 specifies that association\_info\_string shall be presented with the corresponding sample or the region within the sample; equals to 0 specifies that association\_info\_string shall not be presented with the corresponding sample or the region within the sample.

lang\_flag equals to 1 specifies that lang shall be indicated; equals to 0 specifies that lang shall not be indicated.

num\_info\_entry indicates the number of associated info entries.

lang is a character string containing an RFC 5646 compliant language tag string, such as "en-US", "fr-FR", or "zh-CN“, representing the language of the text contained in association\_info\_string ~~name, description and tags~~. When lang is empty, the language is unknown/undefined.

association\_info\_string is a null-terminated UTF-8 character string containing human readable text of the corresponding association\_info\_type.

~~name is a null-terminated UTF-8 character string containing human readable name for the sample or the region within the sample. If not present (an empty string is supplied) no name is provided.~~

~~description is a null-terminated UTF-8 character string containing human readable description of the sample or the region within the sample. If not present (an empty string is supplied) no description is provided.~~

~~tags is a null-terminated UTF-8 character string containing comma-separated user-defined tags related to the sample or the region within the sample. If not present (an empty string is supplied) no tags is provided.~~

During MPEG #148, the following drawbacks were identified in the MPEG git (under [Issue#170](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/170))

1. The proposed extension seem to over-complicate the simple 'udes' data structure and also seems to move a bit away from the main purpose of 'udes'
2. The proposed changes were not backward compatible
3. Explore other means for object recognition and scene recognition
4. Consider the use of text item and the associated text layout information for presentation\_flag

# On GeoKey Item property (MPEG#148, [Issue#172](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/172))

## Abstract

In m70222 an item property was proposed which associates geospatial coordinate system to overview images and base image in an Image pyramid entity group.

## 6.x.y GeoSpatial Keys Property

### 6.x.y.1 Definition

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

The GeoSpatialKeysProperty descriptive item property provides the geo spatial mapping data for geospatial images. The GeoSpatialKeysProperty shall be associated with ImagePyramidEntityGroup when the overview images and the base image in the entity group are geospatial images.

### 6.x.y.2 Syntax

aligned(8) class GeoSpatialKeysProperty  
extends ItemFullProperty('gske', version = 0, flags) {

unsigned int(8) geo\_key\_data[];  
}

### 6.x.y.3 Semantics

geo\_key\_data[] specifies the geo spatial mapping data of the associated image item structured according to the GeoKeys information data of the GeoTiFF standard in <http://www.opengis.net/doc/IS/GeoTIFF/1.1>

During MPEG #148, the following concerns were made in the MPEG git (under [Issue#172](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/172))

1. The proposed item property may be limited due to dependency on GeoKeys which are based on GeoTiff standard
2. The HEIF format need not define formats for mapping images to geo spatial systems and it may be left to other derived specifications

# Compact VVC decoder configuration (MPEG#149, [Issue#178](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/178))

## Abstract

In m71337 extension to CompactVvcDecoderConfigurationRecord was proposed to support multi-layer images in SlimHEIF.

During MPEG #149, the following (non-comprehensive) concerns were raised in [Issue#178](https://git.mpeg.expert/MPEG/Systems/FileFormat/HEIF/-/issues/178) about supporting multi-layer images in Compact VVC decoder configuration record to be used in the low-overhead HEIF format.

1. Is it necessary to support multilayer images in SlimHEIF?
2. Can we consider multi-layer images as small and simple files where the full MetaBox would result in considerable overhead compared to the image data payload?

Based on the concerns it was proposed to remove the syntax and semantics related to the support of multi-layer images from CompactVvcDecoderConfigurationRecord.

**L.4.4.3.2** **Syntax**

aligned(8) class CompactVvcDecoderConfigurationRecord {   
   unsigned int(2) lengthSizeMinusOne;    unsigned int(1) ptl\_present\_flag;   
   if (ptl\_present\_flag) {   
       VvcPTLRecord(1) native\_ptl;   
   }   
   unsigned int(1) additional\_nal\_unit\_flag;   
   if (additional\_nal\_unit\_flag) {   
      unsigned int(3) num\_aps\_nal\_unit;   
      unsigned int(3) num\_sei\_nal\_unit;   
   }   
 unsigned int(8) sps\_nal\_unit\_length;  
 bit(8\*sps\_nal\_unit\_length) sps\_nal\_unit;  
 unsigned int(8) pps\_nal\_unit\_length;  
 bit(8\*sps\_nal\_unit\_length) pps\_nal\_unit;

if (additional\_nal\_unit\_flag) {  
 for (i=0; i< num\_aps\_nalus; i++) {  
 unsigned int(8) aps\_nal\_unit\_length;  
 bit(8\*aps\_nal\_unit\_length) aps\_nal\_unit;  
 }  
 for (i=0; i< num\_sei\_nalus; i++) {  
 unsigned int(8) sei\_nal\_unit\_length;  
 bit(8\*sei\_nal\_unit\_length) sei\_nal\_unit;  
 }  
 }

trailing\_bits();   
}

**L.4.4.3.3** **Semantics**

additional\_nal\_unit\_flag: equal to 1 indicates the presence of additional NAL units in the decoder configuration record. additional\_nal\_unit\_flag equal to 0 indicates the absence of additional NAL units in the decoder configuration record.

num\_aps\_nal\_unit: indicates the number of APS NAL units included in the configuration record for the referenced CVS.

num\_sei\_nal\_unit: indicates the number of SEI NAL units included in the configuration record for the referenced CVS.

trailing\_bits: padding bits to ensure payloads are 8-bit aligned. Shall all be 0.

sps\_nal\_unit\_length: indicates the length in bytes of the NAL unit. When equal to 0, the SPS NAL unit is not present.

sps\_nal\_unit: contains the SPS NAL unit as specified in ISO/IEC 23090-3.

pps\_nal\_unit\_length: indicates the length in bytes of the NAL unit. When equal to 0, the PPS NAL unit is not present.

pps\_nal\_unit: contains the PPS NAL unit as specified in ISO/IEC 23090-3.

aps\_nal\_unit\_length: indicates the length in bytes of the APS NAL unit.

aps\_nal\_unit: contains the APS NAL unit as specified in ISO/IEC 23090-3.

sei\_nal\_unit\_length: indicates the length in bytes of the SEI NAL unit.

sei\_nal\_unit: contains the SEI NAL unit as specified in ISO/IEC 23090-3.

The semantics of the other parameters are the same as for VvcDecoderConfigurationRecord as defined in ISO/IEC 14496-15.

##### L.4.4.3.4 Equivalence with the VVC decoder configuration

CompactVvcDecoderConfigurationRecord shall be considered equivalent to VvcDecoderConfigurationRecord as defined in ISO/IEC 14496-15 with the following fields:

* if ptl\_present\_flag is present and set to 1:
  + num\_sublayers is set to 1,
  + constant\_frame\_rate is set to 1,
  + if the codec configuration property is associated with the main image item:
    - chroma\_format\_idc is set to the value of the chroma\_subsampling field from the MinimizedImageBox,
  + if the codec configuration property is associated with the alpha auxiliary image item:
    - chroma\_format\_idc is set to 0,
  + if the codec configuration property is associated with the main image item or with the alpha auxiliary image item:
    - bit\_depth\_minus8 is set to 0 if the value of the high\_bit\_depth\_flag field from the MinimizedImageBox is 0, or to the value of the bit\_depth\_minus9 field from the MinimizedImageBox plus 1,
    - max\_picture\_width is set to the value plus 1 of the width\_minus1 field from the MinimizedImageBox,
    - max\_picture\_height is set to the value plus 1 of the height\_minus1 field from the MinimizedImageBox,
  + if the codec configuration property is associated with the gain map image item:
    - bit\_depth\_minus8 is set to 0 if the value of the gainmap\_high\_bit\_depth\_flag field from the MinimizedImageBox is 0, or to the value of the gainmap\_bit\_depth\_minus9 field from the MinimizedImageBox plus 1,
    - max\_picture\_width is set to the value plus 1 of the gainmap\_width\_minus1 field from the MinimizedImageBox,
    - max\_picture\_height is set to the value plus 1 of the gainmap\_height\_minus1 field from the MinimizedImageBox,
  + avg\_frame\_rate is set to 0,
  + num\_of\_arrays is set to the number of entries below:
    - if sps\_nal\_unit\_length is not 0, there is a SPS NAL unit array with:
      * NAL\_unit\_type set to 15 (SPS\_NUT as defined in ISO/IEC 23090-3),
      * num\_nalus set to 1:
        + nal\_unit\_length set to sps\_nal\_unit\_length,
        + nal\_unit set to sps\_nal\_unit.
    - if pps\_nal\_unit\_length is not 0, there is a PPS NAL unit array with:
      * NAL\_unit\_type set to 16 (PPS\_NUT as defined in ISO/IEC 23090-3),
      * num\_nalus set to 1
        + nal\_unit\_length set to pps\_nal\_unit\_length,
        + nal\_unit set to pps\_nal\_unit.
    - if additional\_nal\_unit\_flag is set to 1 and num\_aps\_nal\_unit is not 0, there is a prefix APS NAL unit array with:
      * NAL\_unit\_type set to 17 (PREFIX\_APS\_NUT as defined in ISO/IEC 23008-2),
      * num\_nalus set to num\_aps\_nal\_unit, and for each prefix APS NAL unit:
        + nal\_unit\_length is set to aps\_nal\_unit\_length,
        + nal\_unit is set to aps\_nal\_unit.
    - if additional\_nal\_unit\_flag is set to 1 and num\_sei\_nal\_unit is not 0, there is a prefix SEI NAL unit array with:
      * NAL\_unit\_type set to 23 (PREFIX\_SEI\_NUT as defined in ISO/IEC 23008-2),
      * num\_nalus set to num\_sei\_nal\_unit, and for each prefix SEI NAL unit:
        + nal\_unit\_length is set to sei\_nal\_unit\_length,
        + nal\_unit is set to sei\_nal\_unit.
* if array\_completeness is set equal to 1 in VvcDecoderConfigurationRecord.
* the other parameters are carried over as is, and repeated if needed.