**A black and white logo

Description automatically generated** **ISO/IEC JTC 1/SC 29/WG 03 N1711**

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

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

**Document type:** Output Document

**Title: Technologies under consideration on CMAF**

**Status:** Approved

**Date of document:** 2026-01-23

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

**Expected action:** ACT

**Action due date:**

**No. of pages:** 17 (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 N1711**

**Virtual – January 2026**

|  |  |
| --- | --- |
| **Title** | **Technologies under consideration on CMAF** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | **25900** |

# Introduction

This document captures the technologies under consideration for future amendments of the Common Media Application Format (CMAF), ISO/IEC 23000-19 specification.

The current technologies under consideration include:

* On subsample encryption of CMAF tracks.
* Support for AES-256 in encrypted profiles.

# On subsample encryption of CMAF tracks

# 2.1 Abstract

Proposal to explicitly allow partial encryption for non-video tracks

# 2.2 Introduction

Media encryption in ISOBMFF can be applied either as full sample encryption (where an entire media sample is encrypted) or subsample encryption (where only portions of a sample are encrypted). Subsample encryption is particularly useful when certain parts of the media sample need to remain in the clear for processing, such as frame headers, while still protecting the actual media content. This allows for more efficient processing while maintaining content protection.

# 2.3 Motivation

Subsample encryption is useful when configuration and format information is embedded in the stream and needs to be available in the clear even when the content is protected.

In the context of video, this information is often carried in separate NAL units, or other packetized formats (e.g. OBUs with AV1).

# 2.4 Problem Description

The current CMAF specification (ISO/IEC 23000-19:2024) appears to disallow subsample encryption for non-video content through the following requirement:

"Encrypted *non-video* tracks *shall use* the schemes specified in ISO/IEC 23001-7, which define a *full media sample encryption* method for each scheme."

This wording can lead to confusion among implementers and potentially unnecessary restrictions in deployments. It is unclear whether it requires that non-video (e.g. audio) tracks use full sample encryption, or whether it requires that encryption uses one of the schemes specified in ISO/IEC 23001-7 but only those schemes that describe full sample encryption. In either case, implementers might conclude that partial encryption is disallowed for audio. It seems unlikely that this was the original intention.

# 2.5 Proposal

*In clause 8.2.1, replace the following sentence:*

Encrypted non-video tracks shall use the schemes specified in ISO/IEC 23001-7, which define a full media sample encryption method for each scheme.

*with:*

Encrypted non-NAL structured tracks shall use the schemes specified in ISO/IEC 23001-7. These tracks may use either full sample encryption or subsample encryption as defined in ISO/IEC 23001-7. Subsample encryption shall only be used with tracks those CMAF media profiles include a definition of subsamples.

NOTE Examples of formats that may benefit from subsample encryption include those with frame headers that need to remain in the clear for processing, similar to how NAL headers remain in the clear for NAL structured video content.

This change would clarify that subsample encryption is permitted for audio content while maintaining alignment with CENC's security model.

The yellow-highlighted change would clarify that subsample encryption is permitted for non-NAL structured tracks including audio content while maintaining alignment with CENC's security model.

With the newly proposed text (as highlighted in yellow), it is guaranteed that existing CMAF media profiles without a proper definition of subsamples (example: ‘caac’) won’t utilize subsample encryption and therefore any potential issues with deployed players are avoided. On the other hand, already existing CMAF media profiles including non-NAL structured video profiles (example: ‘av01’) can still use subsample encryption if such a definition was previously in place. This holds true for both MPEG-defined as well as for externally defined CMAF media profiles.

# 2.6 Alternatives

If there are concerns with a retroactive bugfix, the fix could be conditional for the new “cmfm” profile. Alternatively, the fix could be introduced only into some, not all, of the existing profiles.

# 2.7 Other Specifications

Even for video, the current CENC specification (ISO/IEC 23001-7:2023) raises questions about use of subsample encryption for video that is not based on NAL structures as defined in ISO/IEC 14496-15. The CENC spec puts enough emphasis on describing NAL structured video encryption that it is difficult to determine if that means subsample encryption may *only* be used with NAL structured video, or if NAL structured video is just used as an example of the kind of bitstreams that subsample encryption can be applied to.

* For 'cenc' scheme: It requires full sample encryption for non-NAL structured content. However, it allows derived specifications to relax this constraint.
* For 'cbcs' scheme: It allows pattern encryption for non-NAL content, which implicitly requires subsample encryption, yet also contains language about "relaxing constraints" for subsample encryption.
* Only the AV1 ISOBMFF specification [1] has extended subsample encryption support to OBU-based structures, setting a precedent for extending beyond NAL-only structures.

It is advised to review this contribution in conjunction with contribution [m72214](https://dms.mpeg.expert/doc_end_user/current_document.php?id=98542&id_meeting=202) (“On subsample encryption”) addressing this issue in CENC.

# 2.8 References

[1] AV1 Codec ISO Media File Format Binding,  
<https://aomediacodec.github.io/av1-isobmff/>

# 2.9 Changed proposed text after MPEG#152

Based on Contribution [m74450](https://dms.mpeg.expert/doc_end_user/current_document.php?id=101329&id_meeting=204), the following changes to the CMAF specification are considered:

|  |  |  |
| --- | --- | --- |
| **Existing text** | **Proposed text (including minor improvements to text from previous TuC)** | **Justification for change** |
| *Encrypted non video tracks …* | *Encrypted non-NAL structured tracks (including non-NAL structured video tracks and non-video tracks) …* | This change takes the fact into account that not all video codecs are NAL structured. (The preceding paragraph considers NAL structured video only.) |
| … *shall use the schemes specified in ISO/IEC 23001-7 which define a full media sample encryption method for each scheme* | *… may use either full sample encryption or subsample encryption as defined in ISO/IEC 23001-7.* | This change reflects the recent clarification in ISO 23001-7. |
|  | *Subsample encryption shall only be used with tracks for which the underlying CMAF media profile includes a definition of subsamples.*  or  *Subsample encryption shall not be used with tracks for which the corresponding CMAF media profile does not include a definition of subsamples.*  or  *CMAF media profiles may impose restrictions on how subsamples may be constructed from sample data.* | This addition addresses compatibility concerns with deployed (legacy) CMAF clients by making encryption a feature of individual CMAF media profiles. This aligns with the considered additions to CENC (new subclause 9.5.3), where support for subsample encryption requires format specific definitions.  **Note:** The 2nd proposed change was not captured in the TuC document, but originates from MPEG#151 discussions: [GitLab discussion](https://git.mpeg.expert/MPEG/Systems/ApplicationFormat/CMAF/-/issues/87).  **Note2**: The 3rd proposed change was raised verbally during MPEG#152 discussion. |

**Note**: Further discussion is expected to happen in GitLab:

<https://git.mpeg.expert/MPEG/Systems/ApplicationFormat/CMAF/-/issues/93>

# Support for AES-256 in encrypted profiles

# 3.1 Introduction

This contribution asks CMAF experts to consider the support of AES-256 support in CMAF profile. The File Format group is developing an amendment addressing this support in Common Encryption [1] and there is a need from the industry to have this support also available in CMAF. The need comes from evolution of regulation in the US [2], as explained below.

We would like to hear experts’ advices and recommendations on how to support AES-256 in CMAF.

# 3.2 Context and requirements

**Regulation**

The NSA is defining security requirements for future “quantum-resistant” products in the Commercial National Security Algorithm Suite (CNSA) 2.0 [2]. In particular, **for** **symmetric-key algorithms, using 256-bit keys will become mandatory**. The NSA set a milestone in 2030 from which IP products not compliant to CNSA 2.0 could not be sold anymore in the US.

**Impact on the market**

The market of video surveillance will be impacted by this new regulation. Standardisation organizations or industrial consortia have concerns about this. For example, ONVIF (Open Network Video Interface Forum) defines specifications that makes mandatory the support of **MPEG Common Encryption** and **CMAF** for devices supporting the recording of encrypted content [3]. To keep on using MPEG standards, to guarantee interoperability, there is a need for CMAF to consider AES-256.

**On-going work in Common Encryption**

In July 2024, the MPEG File Format group started an amendment to Common Encryption considering the support of AES-256 [1].

The support for 256-bit keys has been added in the specification text as well as means to indicate which length is actually in use in the encrypted file but also in MIME types [4]. It is proposed to extend the TrackEncryptionBox and use its flags to indicate whether some Common Encryption features are in use or not; for example: 128 or 256 key-length, single of multiple keys, full or subsample encryption, pattern mode or not…. The only new feature compared to previous edition of Common Encryption is the variable length for encryption keys.

# 3.3 Possible support in CMAF

A quick look at CMAF specification lets us think that while nothing is said about the length of the encryption keys, considering different key lengths would not break the CMAF encrypted tracks. (it was the same situation in MPEG Common Encryption). CMAF section 8 defines “encryption constraints” but a priori none prevents using longer encryption keys than 128-bits.

The CMAF presentation profiles (CMFHDc and CMFHDs) in Annex A.1 are respectively referencing the ‘cenc’ and ‘cbcs’ protection schemes. No mention about the length of encryption keys is given. We may assume that it can be any length, or we may interpret that it is 128-bit only as defined in Common Encryption 4th edition.

The question we ask to CMAF experts is then:

* Should we have an **explicit CMAF presentation profile for encrypted tracks using AES-256**? Or
* Should we keep on using existing ones and consider the signaling introduced by the AMD to Common Encryption to determine the key length in use?
  + In this case, we may want to update the section 8 to indicate the 128 or 256-bit keys can be used.

At MPEG#151, it has been suggested to not use presentation profiles since that is not the right mechanism. However, the Section 8.2.2 might need an update based on the agreement in Common Encryption, possibly with a note (language) referring to that. For example (from input m73487):

**Update** **section 8.2.2.2 Track Encryption Box ('tenc')**

A TrackEncryptionBox specified in subclause 7.4.1 shall be present in a CMAF header if any media samples in the track are encrypted.

NOTE: When AES-256 is used, this can be indicated in the TrackEncryptionBox (or any other mechanism defined in the Common Encryption which might be defined at future meeting), as defined in ISO/IEC 23001-7).

# 3.4 Conclusion

We explained the need for AES-256 support in CMAF, after introducing its support in Common Encryption.

We would like CMAF experts to consider how AES-256 can be supported in CMAF.

# 3.5 References

1. MDS24728\_WG03\_N01428, “*WD of ISO/IEC 23001-7:2023 AMD 1 AES-256 Support”*, MPEG#149, January 2025

1. [CNSA Suite 2.0 and Quantum Computing FAQ](https://media.defense.gov/2022/Sep/07/2003071836/-1/-1/0/CSI_CNSA_2.0_FAQ_.PDF)  (.pdf)

1. [ONVIF Recording Control Service Specification](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fwww.onvif.org%2fspecs%2fsrv%2frec%2fONVIF-RecordingControl-Service-Spec.pdf&c=E,1,f0jEn7_W1yp4cllXC4h6H1XnZ32AximpEYVxvym5XU4K3jUlvLx2Jh6gQBAFA6B5EZ0SUULp60VpFXt-BjWbXORQaGuy72hxHUVaZw82&typo=1), v24.12 Dec. 2024 (.pdf)
2. m72413, “*On MIME types for encrypted content*”, MPEG#150, April 2025

# On Entity Group usage in CMAF files

# 4.1 Introduction

The CMAF specification currently contains restrictions on file-level metadata that conflict with legitimate use cases requiring Entity Groups, which are used to not only signal structural properties across multiple tracks but also provide essential insights even within individual tracks.

With the introduction of the 'cmfm' brand allowing multiple tracks in CMAF files, this restriction becomes increasingly problematic and needs clarification.

# 4.2 Problem Description

To enable rich and differentiated user experiences in CMAF-based streaming, signaling meaningful combinations of tracks—whether single or dependent—using standardized metadata structures becomes increasingly important.

One example of such metadata is Preselection signaling, which was introduced for DASH streaming, but is also supported by ISOBMFF through Entity Groups and allows content providers to define selectable experiences including accessibility features across multiple tracks or within a single track.

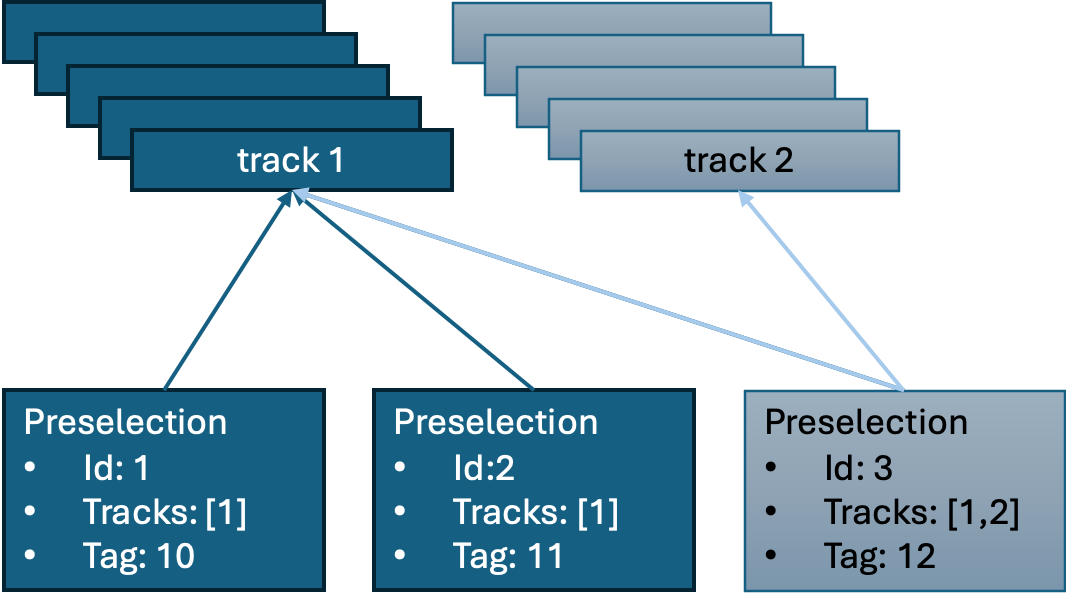


Figure 1: 2 tracks being referenced by 3 preselections; where the first 2 preselections reference the same single track, differentiated by their tag only. This contribution is about enabling these 2 preselections; the 3rd preselection referencing 2 tracks is not in scope since CMAF does not allow for multiple tracks (not considering cases under the ‘cmfm’ brand).

Recent CMAF brand extensions such as 'cmf1'and 'cmfm' introduce capabilities for supplementary metadata and dependent subtracks, further reinforcing the need for robust file-level metadata signaling.

However, a critical conflict exists between the CMAF and ISOBMFF specifications (with added highlighting in **bold** to emphasize important aspects):

* **CMAF** (clause 7.5.2) prohibits file-level MetaBoxes:

*"Metadata, carried in either UserDataBox or MetaBoxes, may be present. When present, they* ***shall not occur at file level****, i.e., they can only be contained in another box, as permitted by ISO/IEC 14496-12."*

* **ISOBMFF** (clause 8.15.1, 8th edition FDIS text) requires file-level MetaBoxes for Entity Groups referencing tracks:

*"Entity groups are indicated in GroupsListBox. Entity groups specified in GroupsListBox of* ***a file level MetaBox refer to tracks*** *or file-level items. Entity groups specified in GroupsListBox of a movie level MetaBox refer to movie level items. Entity groups specified in GroupsListBox of a track level MetaBox refer to track level items of that track."*

In simpler terms: ISOBMFF mandates file-level MetaBoxes for track-referencing Entity Groups, while CMAF explicitly forbids them—creating a specification-level deadlock.

This incompatibility impacts key scenarios:

* **Single-track CMAF files**: Entity Groups cannot reference the single track, making e.g. preselection signaling impossible.
* **Multi-track 'cmfm' files**: Though multiple tracks are permitted, Entity Groups describing their combinations are blocked by the file-level MetaBox restriction.

As a result, implementers are unable to comply with CMAF when legitimate use of Entity Groups is required—limiting user experience signaling, accessibility support, and adaptive streaming logic.

# 4.3 Supposed original intent of prohibiting file-level metadata

The prohibition of file-level metadata in CMAF was likely intended to align with the format's core principle of single-track delivery and streaming optimization. CMAF's original design focused on individual track segments rather than complete media presentations. File-level descriptive metadata (such as title, author, copyright, or overall content descriptions) would be inappropriate as it typically describes complete presentations encompassing all tracks. The restriction was presumably designed to prevent the inclusion of presentation-level metadata that cannot meaningfully apply to individual track files. This encouraged implementers to place such descriptive information at the manifest or presentation level where it belongs in the streaming architecture.

However, this well-intentioned restriction fails to distinguish between descriptive metadata (which may indeed be inappropriate for single-track files) and technical composition metadata like Entity Groups that provide essential structural information not only about track relationships, but may include processing requirements, which remain valid and necessary even in single-track delivery scenarios.

# 4.4 Proposal

It is proposed to modify the CMAF specification to explicitly allow Entity Groups to be contained at file level Meta boxes, with the restriction that they shall only reference tracks that are contained within the same CMAF file.

Current text (for reference):

*"Metadata, carried in either UserDataBox or MetaBoxes, may be present. When present, they shall not occur at file level, i.e., they can only be contained in another box, as permitted by ISO/IEC 14496-12."*

Proposed revised text:

*"Metadata, carried in either UserDataBox or MetaBoxes, may be present. When present, they shall not occur at file level, except for MetaBoxes containing EntityGroups that reference only tracks contained within the same CMAF file, as required by ISO/IEC 14496-12.*

*EXAMPLE: Preselection Entity Groups that reference only one track"*

This change is narrowly scoped to allow only Entity Groups that reference tracks within the same file, ensuring that broader presentation-level metadata remains excluded.

# 4.5 Consideration about compatibility with existing CMAF players

The proposed change to allow file-level metadata for Entity Groups raises important considerations for existing CMAF player implementations that were designed based on the current prohibition.

However, the impact on player compatibility is expected to be minimal for the following reasons:

* graceful degradation:
  + Players that do not recognize or process Entity Groups will typically ignore the file-level MetaBox containing them, as per standard ISOBMFF parsing behavior
  + The core media content (the track containing e.g. video or audio) remains unaffected and playable regardless of Entity Group presence
  + Existing players can continue to function with CMAF files containing Entity Groups, though they may not utilize the enhancement from the grouping information
* optional enhancement rather than breaking change:
  + Entity Groups provide additional structural information but are not required for basic media playback
  + Players that cannot process Entity Groups will fall back to track-by-track processing, which remains functionally equivalent to current behavior
  + The change enables enhanced features for capable players without breaking existing implementations
* implementation path forward:
  + Legacy players continue to work with both old and new CMAF files
  + Updated players can take advantage of Entity Group information for improved user experience
  + The specification change is additive rather than disruptive to the existing ecosystem

# 4.6 Recommendation

It is recommended to update the CMAF specification to resolve the described inconsistency by:

* Explicitly allowing Entity Groups at file level when they reference track(s) within the same CMAF file
* Ensuring alignment with ISO/IEC 14496-12 requirements
* Supporting both traditional single-track CMAF files and the new multi-track 'cmfm' brand files
* This change will maintain the (supposed) intent of the original restriction (preventing unnecessary file-level metadata) while enabling legitimate and standards-compliant use of Entity Groups in CMAF implementations.

This clarification will ensure CMAF remains adaptable to evolving media delivery needs while preserving its core principles.

# Support for Edit Lists

# 5.1 Introduction

Based on industry feedback (e.g. from CTA WAVE) and other contributions, an update to the CMAF 3rd Edition and Amd.1 with respect to Edit Lists is proposed.

# 5.2 Problem Description

Specifically an LS from CTA WAVE was received in [m74856](https://dms.mpeg.expert/doc_end_user/current_document.php?id=101735&id_meeting=204). It suggests as follows:

* + **Based on the potential inconsistent behaviour with edit lists, we suggest creating a profile in MPEG that does not permit edit lists for either, audio and video.**

# 5.3 Proposed Solution (New text is highlighted in yellow)

**5.3.1 In 3rd Edition**

**CMAF switching sets**

**General constraints on a CMAF switching set**

These general CMAF switching set constraints, plus specific constraints defined by a CMAF media profile that the CMAF switching set conforms to, are intended to enable seamless adaptive switching by typical players.

Each CMAF switching set as defined in [3.2.2](#Section_sec_3.2.2) conforms to the following constraints.

1. A CMAF switching set shall contain one or more CMAF tracks, each of which is a synchronized encoding of the same source content.
2. A CMAF switching set shall contain CMAF tracks of only one media type, i.e., audio or video or subtitles.
3. All CMAF tracks in a CMAF switching set shall have the same duration (computed from the media sample durations, optionally stored in a MovieExtendsHeaderBox).
4. All CMAF tracks in a CMAF switching set shall contain the same number of CMAF fragments.
5. If an edit list is used, all CMAF tracks in a CMAF Switching Set shall include an identical edit list.
6. For any CMAF fragment in one CMAF track in a CMAF switching set, there shall be a CMAF fragment with the same decode time in all other CMAF tracks.
7. All CMAF tracks in a CMAF switching set shall have the same value of baseMediaDecodeTime in the first CMAF fragment’s TrackFragmentBaseMediaDecodeTimeBox, measured from the same timeline origin.
8. The presentation time of the earliest media sample of the earliest CMAF fragment in each CMAF track shall be equal.
9. Alternative CMAF fragments in a CMAF switching set with the same baseMediaDecodeTime shall contain alternative encodings of the same source media samples.

NOTE 1 These constraints do not prohibit different frame rates in different CMAF tracks in a CMAF switching set. However, these constraints restrict the combinable frame rates to those that result in identical CMAF fragment durations for alternative CMAF fragments.

1. All CMAF tracks in a CMAF switching set shall conform to one CMAF media profile. However, CMAF tracks included in a CMAF switching set may contain different CMAF media profile brands that all conform to one CMAF media profile. See [subclause 12.1](#Section_sec_12.1) for additional CMAF switching set constraints specified by CMAF media profiles.

NOTE 2 CMAF media profiles specify the maximum limits for encoding parameters, so subset or lower media profiles that can be decoded by a higher media profile decoder conform to the higher media profile.

1. CMAF header parameters shall not differ between CMAF tracks, except as allowed in [Table 11](#Table_tab_11). CMAF media profiles may specify additional constraints, as noted in the table.
2. Additional CMAF header boxes specified in [Table 3](#Table_tab_3) and [Table 4](#Table_tab_4) that are not listed in [Table 11](#Table_tab_11) may contain different boxes and information.

**Table 11 — General constraints on CMAF header boxes in CMAF switching sets**

| **Box** | **General CMAF header constraints in a CMAF switching set** |
| --- | --- |
| ftyp | *Shall be identical except for media profile brands* [see i) in [subclause 7.3.4.1](#Section_sec_7.3.4.1) |
| mvhd | *Shall be identical except for* creation\_time and modification\_time |
| tkhd | *Shall be identical except for* width, height, creation\_time, and modification\_time. See NOTE 1. |
| trex | *Identical* |
| elst | *Shall be identical except for video CMAF track files with a different composition offset* |
| mdhd | *Shall be identical except for* creation\_time and modification\_time |
| mehd | *Identical* |
| meta | *May contain different boxes and data* |
| udta | *May contain different boxes and data* |
| cprt | *Identical* |
| kind | *Identical* |
| elng | *Identical* |
| hdlr | *Identical* |
| vmhd | *Identical* |
| smhd | *Identical* |
| sthd | *Identical* |
| dref | *Identical* |
| stsd | *Sample entries shall have the same* codingname *(four-character code) and conform to other CMAF track format and media profile specified constraints.* See NOTE 2. |
| sgpd | *May differ* |
| pssh | *Identical* |
| sinf | *Identical* |
| schi | *Identical* |
| schm | *Identical* |
| frma | *Identical* |
| tenc | *Shall be identical except for initialization vector values* |
| NOTE 1 Track width and height can differ, but picture aspect ratio is the same for all CMAF tracks.  NOTE 2 Sample entry constraints for CMAF switching sets are defined by each CMAF media profile, its specified or referenced ISO BMFF track format, and its CMAF switching set constraints. A sample description can contain multiple sample entries. | |

There are two functional types of CMAF switching set initialization constraints:

1. Multiple initialization — CMAF switching sets conforming to the general constraints above may rely on decoding parameters stored in each CMAF track’s header to decode a CMAF fragment from that CMAF track. This implies that a player needs to access and process the CMAF header for each CMAF track it switches to.
2. Single initialization — If CMAF headers are constrained to be equivalent within a CMAF switching set, or CMAF tracks only change parameters that are stored in each CMAF fragment, then a player only needs to access one CMAF header and each CMAF fragment to find all the necessary decoding and rendering parameters. A player only needs to initialize the CMAF switching set once, then it can continue decoding CMAF fragments from any CMAF track in the CMAF switching set.

The two types of CMAF switching set constraints and processing are illustrated in [Figure 10](#Figure_fig_10).

**5.3.2 In Amd.1 Update New Clause 7.8**

*Add a new Clause 7.8*

**7.8 The Structural CMAF Brand 'cmf1'**

**7.8.1 General**

A CMAF track conforming to the CMAF structural brand 'cmf1' shall conform to constraints defined in subclauses 7.1, 7.2, 7.3, 7.4, and 7.5 and all remaining constraints and exceptions in this clause 7.8. These constraints are introduced to ensure that the CMAF tracks and CMAF switching set track headers are conforming, including the case where all CMAF Tracks of the presentation conforming to this brand would be included in a single ISO BMFF file, as well as use cases where supplementary metadata is carried alongside primary media data within the same CMAF Track.

**7.8.2 Movie Box (**'moov'**)**

The MovieBox shall contain exactly one TrackBox of the Primary Media Track and may contain one or more additional TrackBoxes of Supplementary Metadata Track(s) with closely associated metadata.

**7.8.3 Track Header Box (**'tkhd'**)**

CMAF TrackHeaderBoxes shall conform to subclause 7.5.4 with the following additional constraints.

— The MovieFragmentBox shall contain exactly one TrackFragmentBox of the Primary Media Track and may contain one or more additional TrackFragmentBox of Supplementary Metadata Track(s) with closely associated metadata.

— If all CMAF tracks of the CMAF presentation are combined into a single ISO BMFF container, then the track header shall be valid. This for example included that:

— The track\_ID is set to a unique identifier over within this CMAF presentation.

— All tracks within one CMAF switching set have alternate\_group shall be set to the same value. Each CMAF switching set in the CMAF presentation shall have a unique value for the alternate\_group.

— For a video track, every decoder output signal shall have its decoded and cropped image size in video spatial samples measured on a uniformly sampled square grid identical to the value of width and height defined in the Track Header.

**7.8.4 Track Reference Box (**'tref'**)**

The TrackReferenceBox shall only be present in Supplementary Metadata Tracks and shall contain a single TrackReferenceTypeBox specifying its relation to the corresponding Primary Media Track.

**7.8.5 Media Header Box (**'mdhd'**)**

CMAF MediaHeaderBoxes shall conform to subclause 7.5.5 with the following additional constraints.

— If one or more Supplementary Metadata Track(s) are present, the value of the timescale should be the same as set in the Primary Media Track. Differences in timescale values may require additional processing to ensure proper synchronization of metadata with media samples.

**7.8.6 Track fragment decode time box (**'tfdt'**)**

The value of the baseMediaDecodeTime in the TrackFragmentBaseMediaDecodeTimeBox of a Supplementary Metadata Track should be set to the same value as the baseMediaDecodeTime in the TrackFragmentBaseMediaDecodeTimeBox of the corresponding Primary Media Track.

**7.8.7 Duration Consistency**

The duration of a Supplementary Metadata Track should match the duration of the corresponding Primary Media Track within a CMAF Track. The sum of sample durations in the TrackRunBox elements of a Supplementary Metadata Track should be equal to the sum of sample durations in the TrackRunBox elements of the corresponding Primary Media Track.

**7.8.8 Metadata Track Switching and Presence in CMAF Fragments**

Supplementary Metadata Tracks, when present, shall also be present in each CMAF Fragment of a CMAF Track. The boundaries of Supplementary Metadata Tracks shall align with the boundaries of the Primary Media Track within a CMAF Track to allow seamless switching at CMAF Fragment boundaries. This ensures consistent availability of metadata across all fragments and enables adaptive switching without playback interruptions.

**7.8.9 Edit List Box (**'elst'**)**

For any CMAF Tracks, the EditBox and in particular the EditListBox shall not be present.

# 5.4 Proposal

It is proposed to consider the updates in the next update of ISO/IEC 23000-19.

# Additional issues not addressed by Amd. 2 of Edition 3

# 6.1 Abstract

Several issues are identified in the Amd-2 of the 3rd edition of the CMAF specification relating to Annex P, “CMAF track and media profiles for MV-HEVC”. Additional issues may also be present beyond those identified, and it is suggested that the published text is further studied and that any issues identified to be carefully addressed.

In particular, the issues identified can be placed in the following categories:

1. Profile indication issues
2. Syntax related issues
3. Layer association issues
4. Conformance issues
5. Editorial issues

They are captured below, outlining the issues and their status.

# 6.2 Detailed description

### 6.2.1 Profile indication issues

#### **6.2.1.1 Profiles for alpha capable profiles:**

##### 6.2.1.1.1 Overview

The CMAF specification only identifies the Multiview Main 10, Multiview Extended, Multiview Monochrome, and Multiview Monochrome 10 profiles for the alpha enhancement layer. There is no mention of the Multiview Extended 10 profile or if, for the stereo + alpha profiles, the profiles for the same type of layers should be identical (which is alluded in Table P.1). These aspects should be addressed.

##### 6.2.1.1.2 Recommendation

1. Update Table P.1, Clause P.3.5, and Clause P.4.5 to explicitly include "Multiview Extended 10" as an allowed profile for enhancement layers when the base layer is HEVC Main 10.
2. Update Clause P.4.5
   1. Explicitly state that both auxiliary alpha layers shall conform to the same profile, aligning the normative text with the "(alpha layers shall match)" requirement in Table P.1.
   2. Add the following bullet point: "If the enhancement layers carrying auxiliary pictures conform to Multiview Monochrome, Multiview Monochrome 10, the enhancement layer with nuh\_layer\_id > 0 carrying the texture picture shall conform to Multiview Extended or Multiview Extended 10."
   3. Add constraint: “All alpha layer NAL units shall have nuh\_layer\_id equal to the alpha nuh\_layer\_id specified in the vps\_extension.”

### 6.2.2 Syntax related issues

#### ****6.2.2.1 SPS video parameter signaling options****

##### 6.2.2.1.1 Overview

Annex P of the 3rd CMAF edition [2] provides various options for signaling video representations using the MV-HEVC extension of the HEVC [1] video example. In particular, it defines profiles for stereoscopic video with and without alpha and for monoscopic video with alpha.

Since the base and the enhancement layers are required to use different profiles (e.g., Main 10 for the base layer and Multiview Main 10 for the enhancement layers), multiple Sequence Parameter Sets (SPS) need to be present in the bitstream, one for each layer. In particular, for the stereoscopic video case, each layer will correspond to a different view (left vs right view), where both views are required to use the same representation format. The signaling of this requirement can be achieved in various possible ways. Different specifications and implementations seem to be using different methods for such indication.

In particular, the CMAF specification says that the syntax element update\_rep\_format\_flag, if present, shall be set to 0 for the MV-HEVC stereo profile, but allows both 0 and 1 values for the other two profiles. On the other hand, TS 26.265 [4] from 3GPP SA4, keeps silent on this aspect, while Apple’s stereo profile implementation suggests that both 0 and 1 values are permitted. It is unclear why the restriction exists for the CMAF specification for only this profile, while this is already in conflict with other specifications and implementations.

The fact is that the same video representation format can be signaled by any of the following options for the enhancement layer:

* **Option A:** Signal identical representation formats in both SPSs. Redundant or duplicated signaling, update\_rep\_format\_flag is not used.
* **Option B:** Use the syntax element update\_rep\_format\_flag and refer to the VPS video representation parameters, with:
  + **Option B.i:** Set the flag value equal to 0 and refer to the default VPS video representation format
  + **Option B.ii:** Set the flag value equal to 1 and explicitly signal the index of the VPS video representation format

For the base layer, full SPS signaling is to be used. It is clear that all of the three options above (A, B.i, B.ii) can be used to achieve the same result for the enhancement layer, either with an absent update\_rep\_format\_flag, or by setting update\_rep\_format\_flag equal to 0 or 1. Still, option B.i makes more sense in terms of least redundant signaling, hence should be promoted to be used consistently.

##### 6.2.2.1.2 Recommendation

1. Add to clause P.2.5 and P.4.5 that both texture layers shall use the same video representation format.
2. Remove the constraints on update\_rep\_format\_flag in clause P.2.5, P.3.5, and P.4.5.

#### ****6.2.2.2**** Restriction on Alpha Channel Inversion

##### 6.2.2.2.1 Overview

Clause P.3.4 of CMAF requires the alpha\_transparent\_value to be set to 0. This constraint prevents "alpha inversion" scenarios (where 0 represents opaque and the maximum value represents transparent), which restricts the flexibility of alpha channel representation allowed by the underlying coding format.

##### 6.2.2.2.2 Recommendation

Remove alpha\_transparent\_value from the list of fields constrained to 0 in Clause P.3.4 (and Clause P.4.4 for consistency).

#### ****6.2.2.3**** 'colr' information in VUI

##### 6.2.2.3.1 Overview

Clause P.2.2 of CMAF states that sample entry boxes such as 'colr' apply to all texture layers and take precedence when "equivalent metadata is carried in Supplemental Enhancement Information (SEI) messages." This is technically imprecise because in HEVC [1], equivalent colour information is carried in the **Video Usability Information (VUI)** parameters within the SPS and is not carried in an SEI message. The text in [2] should reference VUI to accurately reflect where this metadata resides in the bitstream and clarify how this should be utilized.

##### 6.2.2.3.2 Recommendation

Update Clause P.2.2 to read: "When equivalent metadata is carried in VUI parameters or SEI messages..." to accurately cover both the VUI-based colour information (colr) and the SEI-based metadata (clli, mdcv) included in the list.

#### ****6.2.2.4**** 3D Reference Displays Information SEI

##### 6.2.2.4.1 Overview

Clause P.2.4.1 of CMAF mandates the inclusion of the three\_dimensional\_reference\_displays\_info SEI message for the MV-HEVC stereo profile. Usage of more parameters within this SEI message should be clarified.

##### 6.2.2.4.2 Recommendation

Recommend the following existing recommendations in 3GPP TS 26.265 [4], i.e.:

- The value of num\_ref\_displays\_minus1 shall be set to 0.

- The value of the left\_view\_id[ 0 ] shall be set to the corresponding value defined in the view\_id\_val parameter.

- The value of the right\_view\_id[ 0 ] shall be set to the corresponding value defined in the view\_id\_val parameter and shall be different to the left\_view\_id[ 0 ].

- In absence of reference display parameters:

- The prec\_ref\_display\_width is set to 31.

- The ref\_viewing\_distance\_flag is set to 0.

- The exponent\_ref\_display\_width[0] and mantissa\_ref\_display\_width[0] are both set to 0.

- The additional\_shift\_present\_flag[0] and the three\_dimensional\_reference\_displays\_extension\_flag are both set to 0.

### 6.2.3 Conformance issues

#### ****6.2.3.1**** Layer counts and Output Layer Sets (OLS)

##### 6.2.3.1.1 Overview

Clause P.2.2 of CMAF strictly requires that the track "shall contain exactly 2 layers" (or 4 for stereo+alpha). However, it does not address Output Layer Sets (OLS) defined in HEVC [1]. If a bitstream contains 4 layers but all indicated OLSs in the VPS output only 2 layers (e.g., stereo without alpha), the strict "contain exactly" language in [2] may create conformance conflicts with valid MV-HEVC bitstreams.

##### 6.2.3.1.2 Recommendation

1. Remove the requirement "shall contain exactly [2/4] layers" from Clause P.2.2 and Clause P.4.2 to allow for flexibility in bitstream construction provided valid OLSs are defined.
2. Define OLS Requirements:
   1. Clause P.2.5 (Stereo): Add: "The bitstream shall contain one output layer set containing the layer (nuh\_layer\_id = 0). The bitstream shall contain another output layer set containing the layer (nuh\_layer\_id = 0) and a second layer corresponding to a scalability dimension of type Multiview as output layers. The layers with the corresponding nuh\_layer\_id values indicated in the output layer set shall be present in the bitstream."
   2. Clause P.3.5 (Alpha): Add: "The bitstream shall contain one output layer set containing the layer (nuh\_layer\_id = 0). The bitstream shall contain another output layer set containing the layer (nuh\_layer\_id = 0) and a second layer corresponding to a scalability dimension of type Auxiliary as output layers. The layers with the corresponding nuh\_layer\_id values indicated in the output layer set shall be present in the bitstream."
   3. Clause P.4.5 (Stereo + Alpha): Add: "The bitstream shall contain one output layer set containing the layer (nuh\_layer\_id = 0). The bitstream shall contain another output layer set containing the layer (nuh\_layer\_id = 0), a second layer corresponding to a scalability dimension of type Multiview, and two layers corresponding to a scalability dimension of type Auxiliary as output layers. The layers with the corresponding nuh\_layer\_id values indicated in the output layer set shall be present in the bitstream."

#### ****6.2.3.2**** 3D Reference Displays in Mono/Alpha profiles

##### 6.2.3.2.1 Overview

Clause P.3.4 of CMAF [2] states that the three\_dimensional\_reference\_displays\_info SEI message "should not be present" for the HEVC with Alpha profile. Since this profile is defined as monoscopic (P.3.1), this SEI message is irrelevant. The requirement should likely be a normative "shall not" to prevent ambiguity, rather than a recommendation.

##### 6.2.3.2.2 Recommendation

1. Clause P.3.4 (HEVC with alpha):
   1. Remove any mandate on three\_dimensional\_reference\_displays\_info SEI message from this subclause. e.g. rephrase “The constrains from P.2.4 apply, except for the constraint on presence of three\_dimensional\_reference\_displays\_info SEI message, with the additional constraints defined in this clause.”
   2. Remove the bullet point: "The three\_dimensional\_reference\_displays\_info SEI message ... should not be present."
2. Clause P.4.4 (MV-HEVC stereo with alpha): Add an explicit bullet point: "The three\_dimensional\_reference\_displays\_info SEI message shall be present." (This ensures the requirement is explicitly stated for this profile rather than implicitly inherited from Clause P.2).
3. Clause P.2.4.1 (MV-HEVC stereo): Retain the existing requirement: "...it shall include the three\_dimensional\_reference\_displays\_info SEI message..."

### 6.2.4 Miscellaneous issues

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Clause** | **Issue** | **Discussion** |
|  |  | It would be good to have more information on how encryption is applied on layered content.  E.g. could inter-layer prediction be used to reduce the encryption overhead? |  |
|  | Cl 3.3.2 | It is not clear whether you need to include the metadata track in all tracks of one Switching Set or in all tracks of one Selection Set. There are some conceptual issues associated to this and we need to carefully read the definitions. | This is a content authoring question. For sure there is no need to have Supplementary Metadata Tracks for all tracks of a selection set, since the selection is associated with different content characteristics (language, region or something else). Hence, there is no need to mandate having them for a selection set. The question is: do we need to mandate it for all tracks of a switching set? What about the case then that there is a low bitrate representation that does not need the metadata needed by a higher quality rep? |
|  |  | If the cmf1 brand is to be used for combining CMAF tracks into a single ISOBMFF file we need to clarify what happens with the closely associated metadata tracks (Supplementary metadata tracks) as this metadata will be redundant (one for each resolution/quality of the Primary Media data). |  |
|  | 7.5.21 | What is the case if cdsc is not used? | Need to implement some solution |
|  | Cl 7.8.6., 7.8.7, 7.8.8 | “make it a requirement” i.e. change to a shall |  |

## 6.3 Proposal

It is proposed to consider the documented issues for inclusion in the next amendment (or edition) of CMAF.

## 6.4. References

[1] ISO/IEC 23008-2, Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding

[2] ISO/IEC 23000-19:2024, Information technology – Multimedia application format (MPEG-A) – Part 19: Common media application format (CMAF) for segmented media

[3] Apple HEVC Stereo Video Interoperability Profile, <https://developer.apple.com/av-foundation/HEVC-Stereo-Video-Profile.pdf>

[4] 3GPP TS 26.265, Media Delivery: Video Capabilities and Operation Points.