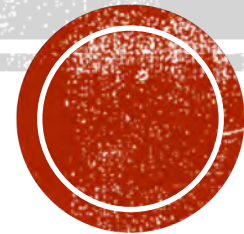




AN OVERVIEW OF THE ISO BASE MEDIA FILE FORMAT

... more than just a collection of Boxes



Mile High Video Workshop 2018

SET THE EXPECTATION



~~From: Thomas Stockhammer~~

Sent: Montag, 30. Juli 2018 08:42

To: David Singer <singer@apple.com>; Cyril Concolato <cconcolato@netflix.com>; Waqar Zia <zia@nomor.de>; Jean Le Feuvre <jean.lefeuvre@telecom-paristech.fr>; 'Kevin Murray (kevimurr)' <kevimurr@cisco.com>; 'Kilroy Hughes' <kilroy_hughes@hotmail.com>; 'Thomas, E.D.R. (Emmanuel)' <emmanuel.thomas@tno.nl>; Miska Hannuksela <miska.hannuksela@nokia.com>; Per Fröjdh <per.frojdh@ericsson.com>; 'Vishy Swaminathan' <vishy@adobe.com>; Imed Bouazizi <ibouazizi@gmail.com>; Michael Dolan <mike@dolan.tv>

Cc: 'Ye-Kui Wang' <yekui.wang@huawei.com>; Begen, Ali (Contractor) <Ali_Begen@comcast.com>; 'Giladi, Alex' <Alex_Giladi@comcast.com>

Subject: File Format at Mile High Workshop **URGENT HELP NEEDED**

Importance: High

All,

at the mile high video workshop, Yekui in CC was supposed to give a presentation on the File Format. Unfortunately, he is sick (fever, cold, too much party in Ljubljana, did China win the World Cup?, anyways).

Ali asked me if I can step in and do a presentation on the file format. However, it is impossible for me to get slides done in such a short amount of time. I have some slides on segment formats, DASH integration, CMAF.

So I am reaching out:

- Do you have tutorial material on the file format that can be used in a public conference? If so, can you send it to me and share it.

Answers:

- A few slides (pages 12-17) on the file format in my presentation three years ago
- Damn - when is the workshop?
- Sorry – most of my stuff is hideously out of date as to be useless for what you want (trust me!)
- An answer from Kilroy that would fill three presentations for itself

I am not aware of a recent presentation on the file format



OVERVIEW

- Basics and History
- Structures and Principles
- More than just a paper spec – Tools and Deployments
- ISO BMFF and streaming
- Other recent application formats
- Crystal Ball – What's next?
- Summary



BASICS

- The ISO Base Media File Format contains structural and media data information principally for timed presentations of media data such as audio, video, etc.
- There is also support for un-timed data, such as meta-data.
- By structuring files in different ways the same base specification can be used for files for
 - capture;
 - exchange and download, including incremental download and play;
 - local playback;
 - editing, composition, and lay-up;
 - streaming from streaming servers, and capturing streams to files.

ISO base media file format (MPEG-4 Part 12) also known as ISO BMFF	
Developed by	ISO
Type of format	Media container
Container for	Audio, video, text, data
Extended from	QuickTime .mov
Extended to	MP4 , 3GP , 3G2, .mj2 , .dvh, .dcf, .m21, .cmf
Standard	ISO/IEC 14496-12, ISO/IEC 15444-12
Website	https://www.iso.org/standard/68960.html



HISTORY

- ISO BMFF is directly based on Apple's QuickTime container format.
- It was developed by **MPEG** (ISO/IEC JTC1/SC29/WG11).
- first MP4 file format specification was created on the basis of the QuickTime format specification published in **2001**.
- The MP4 file format known as "version 1" was published in 2001 as ISO/IEC **14496-1:2001**, as revision of the MPEG-4 Part 1: Systems.
- In 2003, the first version of MP4 file format was revised and replaced by MPEG-4 Part 14: MP4 file format (ISO/IEC **14496-14:2003**), commonly known as MPEG-4 file format "version 2".[13]
- The MP4 file format was generalized into the ISO Base Media File format (ISO/IEC **14496-12:2004** or ISO/IEC 15444-12:2004), which defines a general structure for time-based media files.



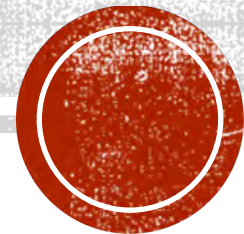
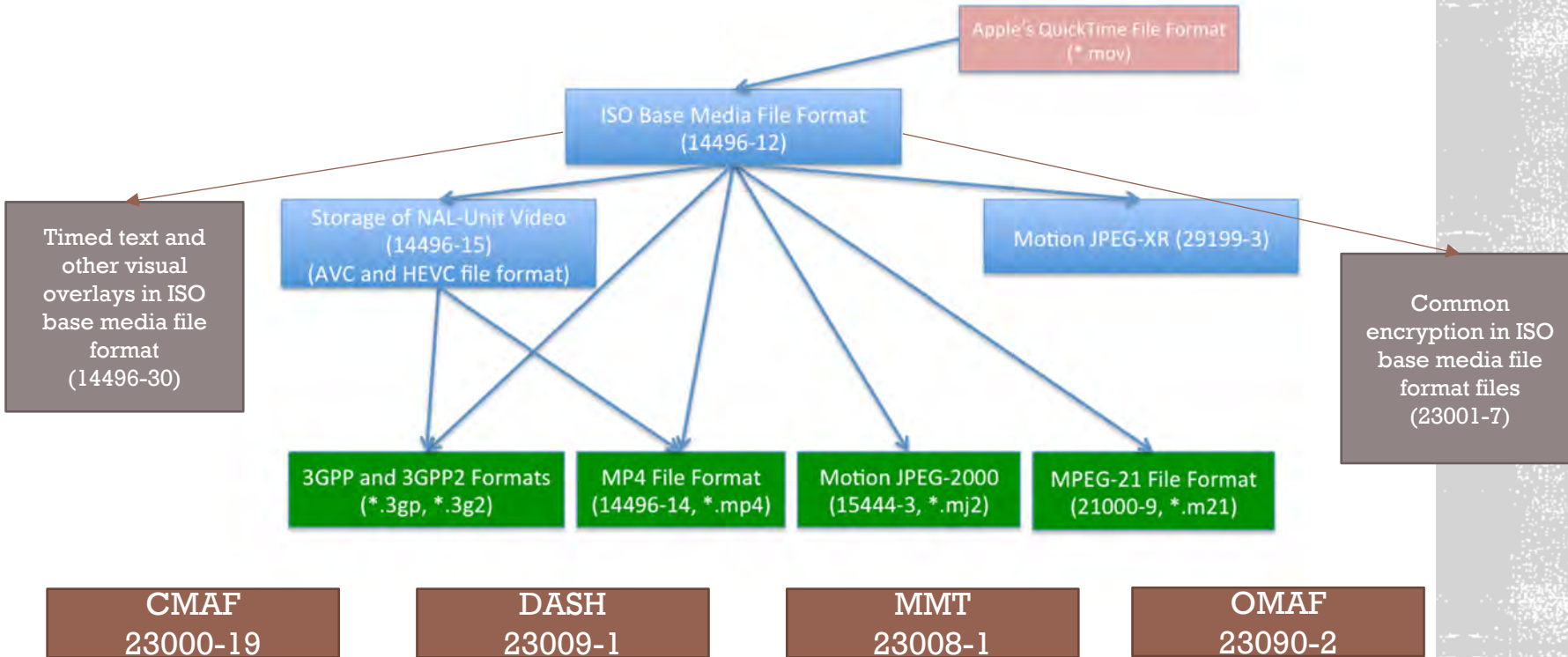
SPEC RELEASES 14496-12

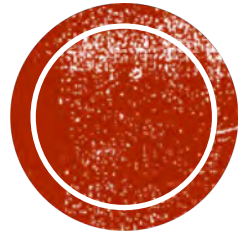
MPEG-4 Part 12 / JPEG 2000 Part 12 editions			
Edition	Release date	Standard	Main Features
First edition	2004	ISO/IEC 14496-12:2004, ISO/IEC 15444-12:2004	Initial base spec
Second edition	2005	ISO/IEC 14496-12:2005, ISO/IEC 15444-12:2005	???
Third edition	2008	ISO/IEC 14496-12:2008, ISO/IEC 15444-12:2008	???
Fourth edition	2012	ISO/IEC 14496-12:2012, ISO/IEC 15444-12:2012	Font streams, subtracks and colors, DASH, reception hint tracks
Fifth edition	2015	ISO/IEC 14496-12:2015, ISO/IEC 15444-12:2015	Timed text and better audio
Sixth edition	2018 (expected)		DRC and HEIF

Supported by Amendments and Corrigendas



THE WHOLE SUITE





STRUCTURES AND PRINICIPLES

Logical, Timing and Physical Structures

BASIC STRUCTURES

- The files have
 - a *logical* structure: a *movie* that in turn contains a set of time-parallel *tracks*.
 - a *time* structure: the *tracks* contain sequences of *samples* in time, and those sequences are mapped into the timeline of the overall movie by optional *edit lists*.
 - a *physical* structure; a series of boxes (sometimes called atoms), which have a size and a type.
- These structures are not required to be coupled.



LOGICAL STRUCTURES

- Each media stream is contained in a *track* specialized for that media type (audio, video etc.), and is further parameterized by a *sample entry*.
- The sample entry
 - contains the '*name*' of the exact media type (i.e., the type of the decoder needed to decode the stream) and any parameterization of that decoder needed.
 - The name also takes the form of a four-character code.
 - There are defined sample entry formats not only for MPEG-4 media, but also for the media types used by other organizations using this file format family.
 - They are registered at the MP4 registration authority.
- Tracks (or sub tracks) may be identified as alternatives to each other, and there is support for declarations to identify what aspect of the track can be used to determine which alternative to present, in the form of *track selection* data.



File

- Contains
 - timed data in tracks of a movie
 - Other data (untimed) in items
 - Or a combination of both
- Defines a common timeline for all tracks to be synchronized

Track

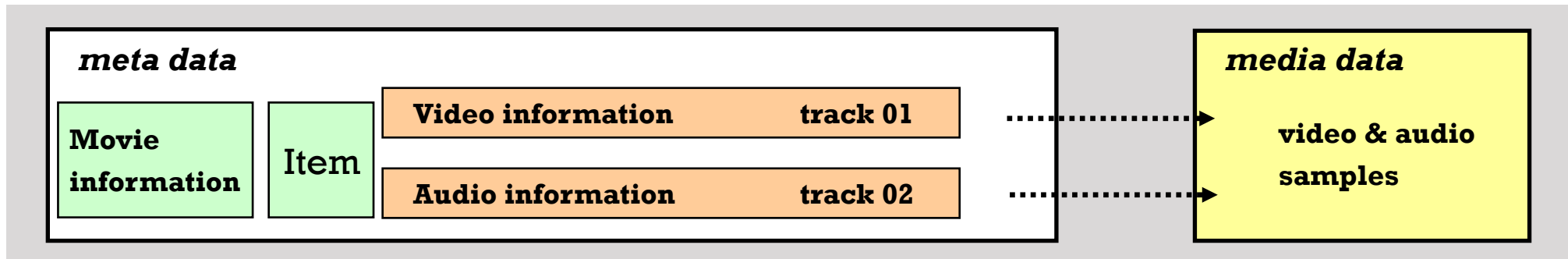
- Corresponds to a specific media type (codec),
- Is associated to a single decoder (except for scalable codecs),
- May be linked, grouped or alternative to other tracks
- May have associated untimed data in items
- May be encrypted
- Is decomposed into samples

Sample

- Represents contiguous data used by a decoder at given times (DTS, CTS)
- Has properties (size, position, random access, decoder configuration...)
- May be described in terms of sub-samples
- May be associated to other similar samples in sample groups
- May have sample-specific auxiliary information

Item

- Represents data consumed as a whole and valid for the entire duration of the movie,
- Has properties (type, position, size ...)
- May be encrypted, compressed, ...

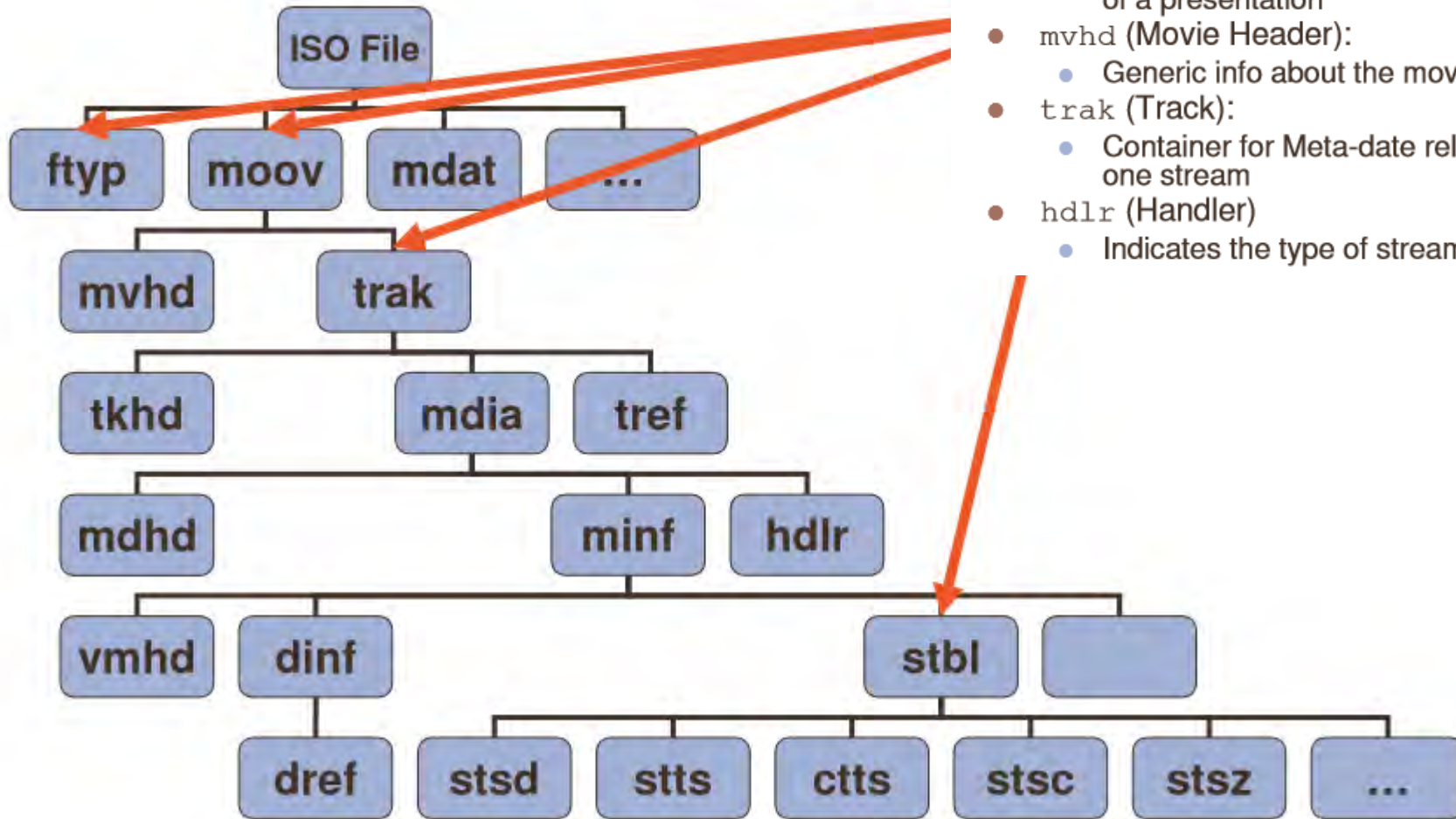


PHYSICAL ORGANIZATION

- Data is stored in a basic structure called box
 - No data outside of a box
- Each box has length, type (4 printable chars), possibly version and flags, and data
- Extensible format:
 - Unknown boxes can be skipped (syntactically)
 - Header information is a hierarchical set of boxes (typically 'moov' or 'meta')
 - Media data is stored unstructured, in boxes (mainly 'mdat', or 'idat') in the same file as the header or may be stored in a separate file



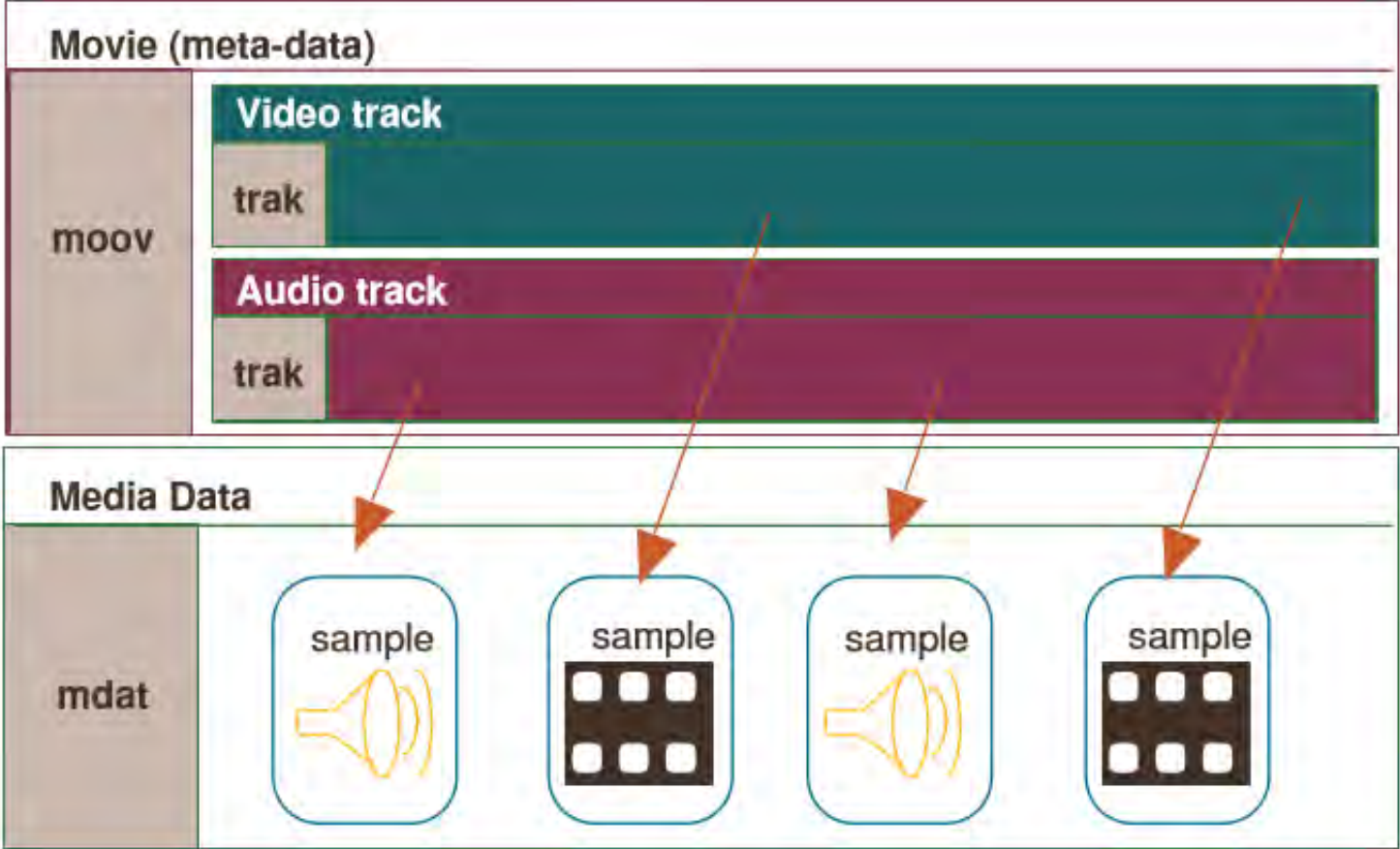
TYPICAL STRUCTURE



- **ftyp (File Type):** 1 per file
 - File type
 - File version
 - Compatibility with other ISO files
- **mdat (Media Data):**
 - Contains the media data
 - A file may have several, non contiguous
- **moov (Movie):**
 - Unique container for the metadata of a presentation
- **mvhd (Movie Header):**
 - Generic info about the movie
- **trak (Track):**
 - Container for Meta-data related to one stream
- **hdlr (Handler)**
 - Indicates the type of stream
- **dinf/dref (Data Information/Data Reference)**
 - Indicates the location of the data (current file or remote file)
- **stbl (Sample Table)**
 - Contains the meta data related to samples, sample per sample
- **stsd (Sample Description)**
 - Contains the decoder configuration for the elementary stream
- **stts (Sample To Time)**
 - DTS for each sample
 - Use a predictive coding scheme
- **stsz (Sample To Size)**
 - Size of each sample, run-length coded



EXAMPLE ORGANIZATION



TIMING ORGANIZATION

- Each track is a sequence of timed samples;
 - each sample has a decoding time, and may also have a composition (display) time offset. Edit lists may be used to over-ride the implicit direct mapping of the media timeline, into the timeline of the overall movie.
- Sometimes the samples within a track have different characteristics or need to be specially identified.
 - One of the most common and important characteristic is the synchronization point (often a video I-frame).
- These points are identified by a special table in each track.
 - More generally, the nature of dependencies between track samples can also be documented.
 - Finally, there is a concept of named, parameterized *sample groups*.
- Each sample in a track may be associated with a single group *description* of a given group type, and there may be many group types.



DECODE, COMPOSITION AND MOVIE TIMES

- ISO BMFF has three timelines
 - Decode times
 - Composition times
 - Movie/Presentation time
- ISO BMFF provides
 - Decode deltas/times
 - Composition offsets (may be negative)
 - Edit Lists signaled in movie header
- The presentation time for synchronized presentation is obtained as
 - $DT + CO + EL$

Segment	/--	--	--	--	--	--\	/-	--	--	--	---	--\			
	I3	P1	P2	P6	B4	B5	I9	P7	P8	P12	B10	B11			
Presentation Order	==	P1	P2	I3	B4	B5	P6	==	P7	P8	I9	B10	B11	P12	==
Base media decode time	0							60							
Decode Delta	10	10	10	10	10	10		10	10	10	10	10	10		
DT	0	10	20	30	40	50		60	70	80	90	100	110		
EPT	10							70							
Composition time offset	30	0	0	30	0	0		30	0	0	30	0	0		
CT	30	10	20	60	40	50		90	70	80	120	100	110		

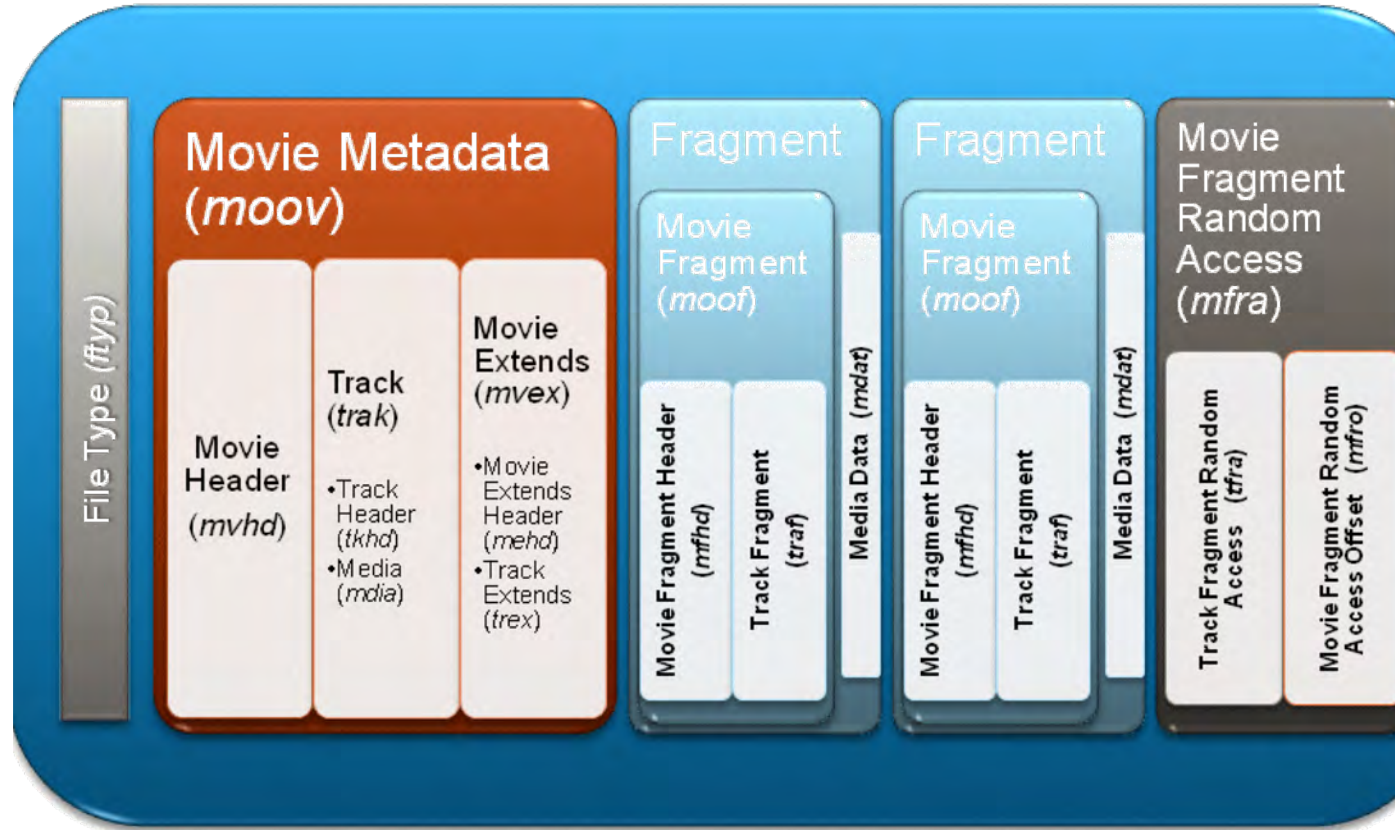
Segment	/--	--	--	--	--	--\	/-	--	--	--	---	--\			
	I3	P1	P2	P6	B4	B5	I9	P7	P8	P12	B10	B11			
Presentation Order	==	P1	P2	I3	B4	B5	P6	==	P7	P8	I9	B10	B11	P12	==
Base media decode time	0							60							
Decode Delta	10	10	10	10	10	10		10	10	10	10	10	10		
DT	0	10	20	30	40	50		60	70	80	90	100	110		
EPT	0							60							
Composition offset	20	-10	-10	20	-10	-10		20	-10	-10	20	-10	-10		
CT	20	0	10	50	30	40		80	60	70	110	90	100		



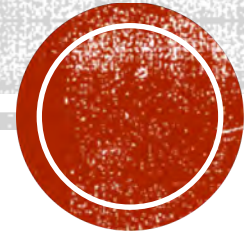
METADATA — TWO FORMS

- First, timed meta-data may be stored in an appropriate track, synchronized as desired with the media data it is describing.
 - See for example for 23001-10 for timed metadata, e.g. Region of interest, location, etc.
- support for non-timed collections of metadata *items* attached to the movie or to an individual track.
 - The actual data of these items may be in the metadata box, elsewhere in the same file, in another file, or constructed from other items.
 - these resources may be named, stored in *extents*, and may be protected.
 - These metadata containers are used in the support for file-delivery streaming, to store both the ‘files’ that are to be streamed, and also support information such as reservoirs of pre-calculated forward error-correcting (FEC) codes (e.g. hint tracks)
- The generalized meta-data structures may also be used at the file level,
 - above or parallel with or in the absence of the movie box.
 - In this case, the meta-data box is the primary entry into the presentation.





FRAGMENTED MOVIES



EXTENSIBILITY

- **Simple extensions:**
 - New codec for temporal data for which you own the sample format (e.g. AV1 in MP4)
 - New sample groups for (codec-specific) annotation of samples (e.g. HEVC CRA/BLA)
 - New sample auxiliary data , for (codec-specific) per-sample data (e.g. init vector, ...)
 - New untimed data format (e.g. EXIF, XMPP ...)
 - New user-, vendor-specific data (use 'meta', 'udta', 'free', 'skip', or 'uuid' boxes)
- **Harder extensions**
 - Beware of backwards compatibility !
 - Only if all other options have been exhausted
 - Extending existing boxes: Use versioning and/or flags
- **New boxes (almost always the wrong option!)**
 - Check for name clashes (www.mp4ra.org)
 - Define box syntax and semantics
 - Choose box location and cardinality
 - Timed/Untimed information
 - File level, segment level, movie level, track level, sample level, ...
 - Define new brand if it implies behavior changes/incompatibilities



MPEG VIDEO IN ISOBMFF (14496-15)

- Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format
- Defines not only what a sample is, but also has various options
 - Parameter sets in the sample entry (initialization), or in-stream
 - Out-of-band mechanism: identified by the use of 'avc1' or 'hvc1'
 - Inband parameter sets: identified by 'avc3' or 'hev1'
 - Sample groups to describe samples (random access etc.)
- Defines carriage of both scalable and multi-view extensions to AVC & HEVC
 - Single-track or multi-track
 - Sample groups etc. to help choose which track(s) to consume

```
aligned(8) class AVCDecoderConfigurationRecord {
    unsigned int(8) configurationVersion = 1;
    unsigned int(8) AVCProfileIndication;
    unsigned int(8) profile_compatibility;
    unsigned int(8) AVCLevelIndication;
    bit(6) reserved = '111111'b;
    unsigned int(2) lengthSizeMinusOne;
    bit(3) reserved = '111'b;
    unsigned int(5) numOfSequenceParameterSets;
    for (i=0; i< numOfSequenceParameterSets; i++) {
        unsigned int(16) sequenceParameterSetLength;
        bit(8*sequenceParameterSetLength) sequenceParameterSetNALUnit;
    }
    unsigned int(8) numOfPictureParameterSets;
    for (i=0; i< numOfPictureParameterSets; i++) {
        unsigned int(16) pictureParameterSetLength;
        bit(8*pictureParameterSetLength) pictureParameterSetNALUnit;
    }
}
```



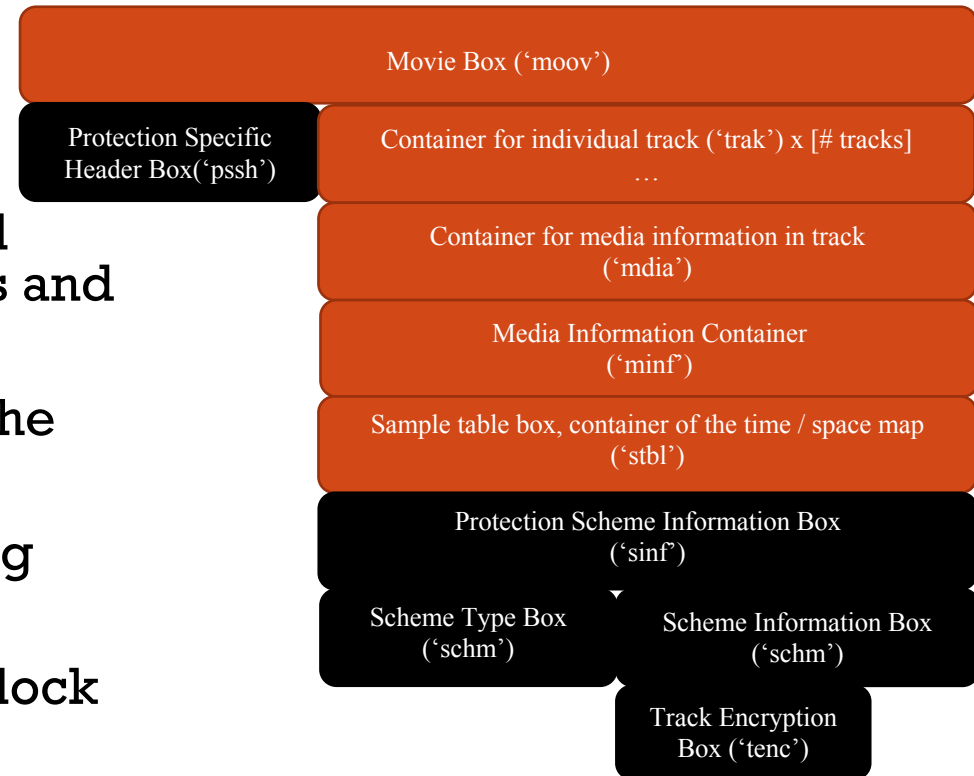
OTHER MEDIA

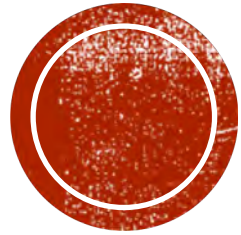
- Audio:
 - 'mp4a' defines the set of MPEG-4 audio in the MP4 spec 14496-14
 - Other audio technologies define the sample entry and track mapping in their media specs
- Subtitles
 - IMSC1 and WebVTT see 14496-30
- External media can be added to the ISO BMFF as well
- The codecs parameter is defined in RFC6381
 - The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types
 - Permits signaling sample entries plus additional information
 - Currently under discussion – how much needs to be there for capability



COMMON ENCRYPTION 23001-7

- specifies elementary stream encryption and encryption parameter storage to enable a single ISO Media file that support different Digital Rights Management systems (DRM) to manage keys and securely decrypt the media.
- Clear and encrypted byte ranges are identified in the track metadata as “subsamples”
- First edition: ‘cenc’ - single encryption scheme using AES-128 counter mode cipher
- Second edition: ‘cbcl’ using AES-128 with Cipher Block Chaining mode (CBC)
- Third edition: two pattern encryption schemes, identified as ‘cbcs’ and ‘cens’



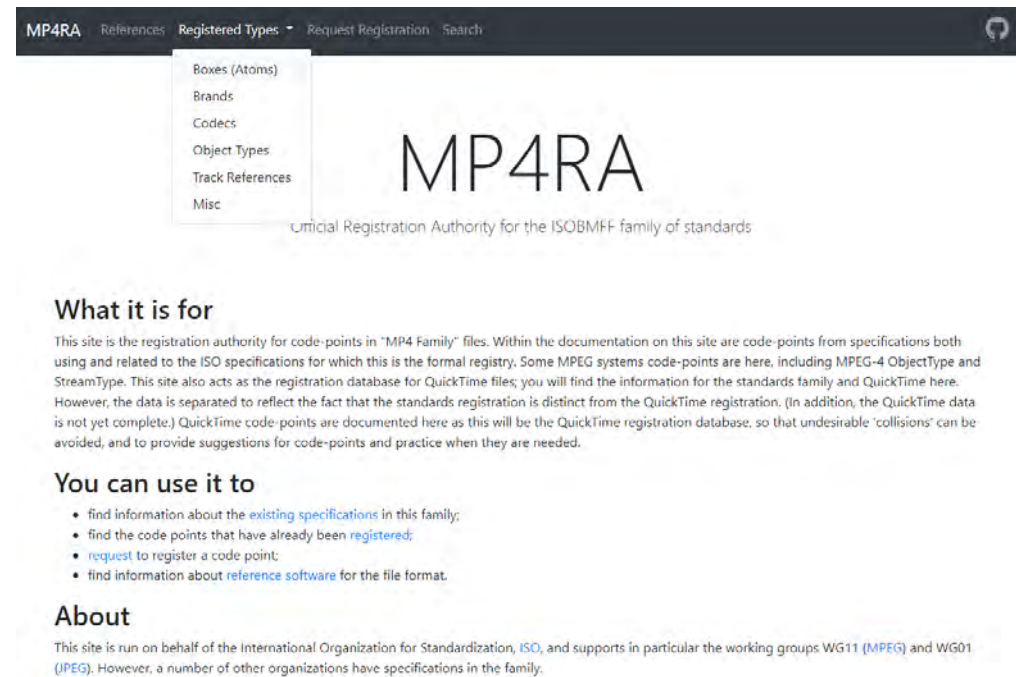


MORE THAN JUST A PAPER SPEC

Tools and Software

MPEG'S SUPPORTING TOOLS

- Conformance bit streams
 - ISO/IEC 14496-4
 - Some streams are freely available
 - <http://standards.iso.org/ittf/PubliclyAvailableStandards/>
 - More are welcome
- Software
 - ISO/IEC 14496-5
 - Reference software, freely available
 - C, ISO Licence
 - Read/Write MP4 files
 - Contributions are welcome
- MP4 Registration Authority <http://www.mp4ra.org>
 - There is a registration authority which registers and documents the four-character-code code-points used in this file-format family, as well as some other code-points related to MPEG-4 systems.
 - The database is publicly viewable and registration is free.



The screenshot shows the MP4RA website interface. At the top, there is a dark navigation bar with the following items: "MP4RA", "References", "Registered Types" (with a dropdown arrow), "Request Registration", and "Search". A dropdown menu is open under "Registered Types", listing: "Boxes (Atoms)", "Brands", "Codecs", "Object Types", "Track References", and "Misc". The main content area features the "MP4RA" logo in a large, light-colored font, with the tagline "Official Registration Authority for the ISO/BMFF family of standards" underneath. Below the logo, there are three sections: "What it is for", "You can use it to", and "About".

What it is for

This site is the registration authority for code-points in "MP4 Family" files. Within the documentation on this site are code-points from specifications both using and related to the ISO specifications for which this is the formal registry. Some MPEG systems code-points are here, including MPEG-4 ObjectType and StreamType. This site also acts as the registration database for QuickTime files; you will find the information for the standards family and QuickTime here. However, the data is separated to reflect the fact that the standards registration is distinct from the QuickTime registration. (In addition, the QuickTime data is not yet complete.) QuickTime code-points are documented here as this will be the QuickTime registration database, so that undesirable 'collisions' can be avoided, and to provide suggestions for code-points and practice when they are needed.

You can use it to

- find information about the [existing specifications](#) in this family;
- find the code points that have already been [registered](#);
- [request](#) to register a code point;
- find information about [reference software](#) for the file format.

About

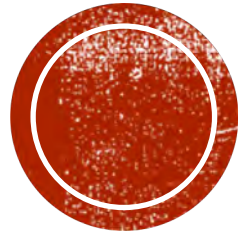
This site is run on behalf of the International Organization for Standardization, [ISO](#), and supports in particular the working groups WG11 ([MPEG](#)) and WG01 ([JPEG](#)). However, a number of other organizations have specifications in the family.



OPEN SOURCE AND COMMERCIAL SERVICES

- Open Source
 - Widely implemented in open source, e.g. FFMpeg, MP4Box
 - Nokia Labs even has a Javascript implementation
- Usage in Commercial Services
 - tbd
- Check here: <http://mp4ra.org/#/brands>



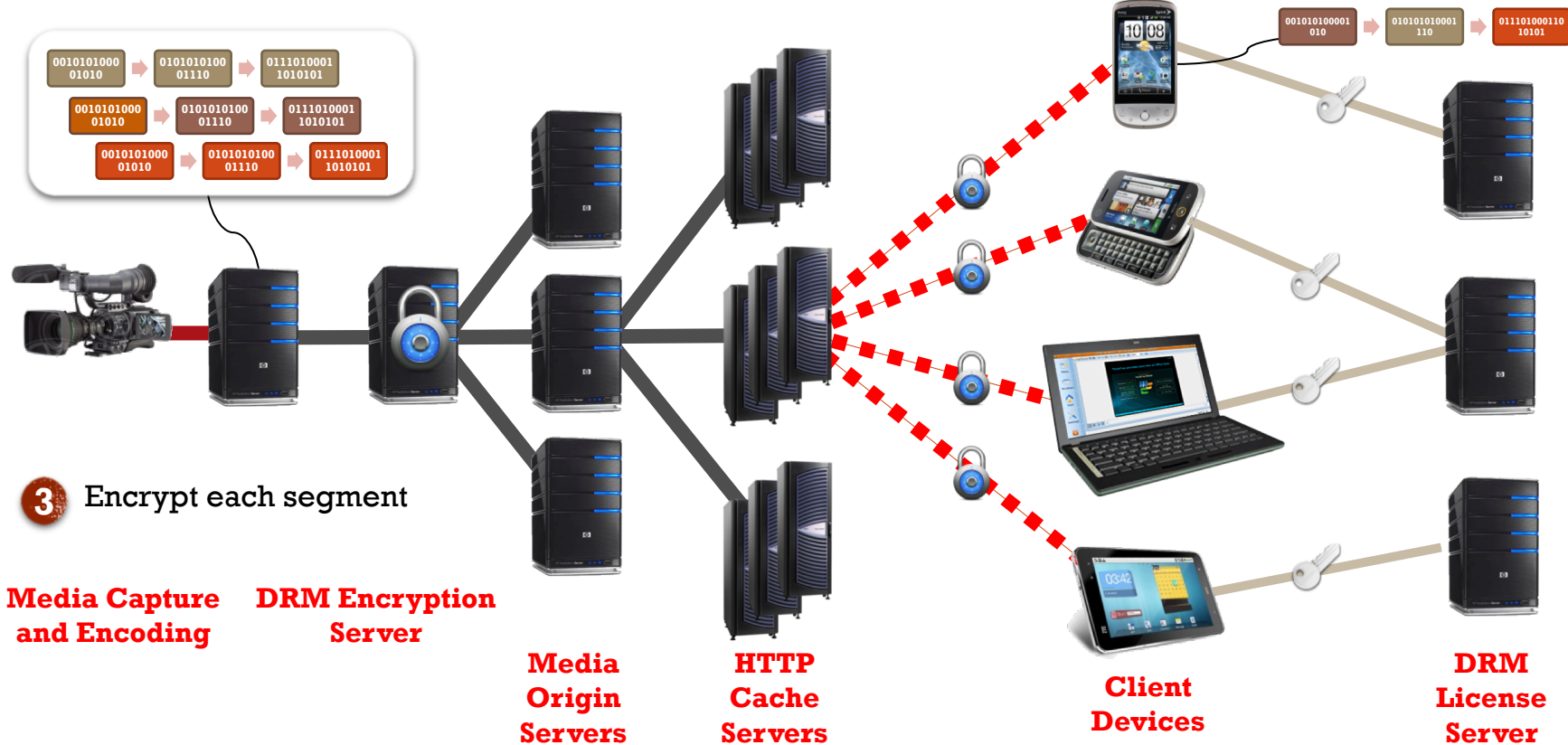


ISOBMPFF AND STREAMING

DASH and CMAF

ADAPTIVE STREAMING

- 1 Encode each video at multiple bitrates
- 2 Split the videos into small segments
- 3 Encrypt each segment
- 4 Make each segment addressable via an HTTP-URL
- 5 Client makes decision on which segment to download
- 6 Client acquires a license for encrypted content
- 7 Client splices together and plays back

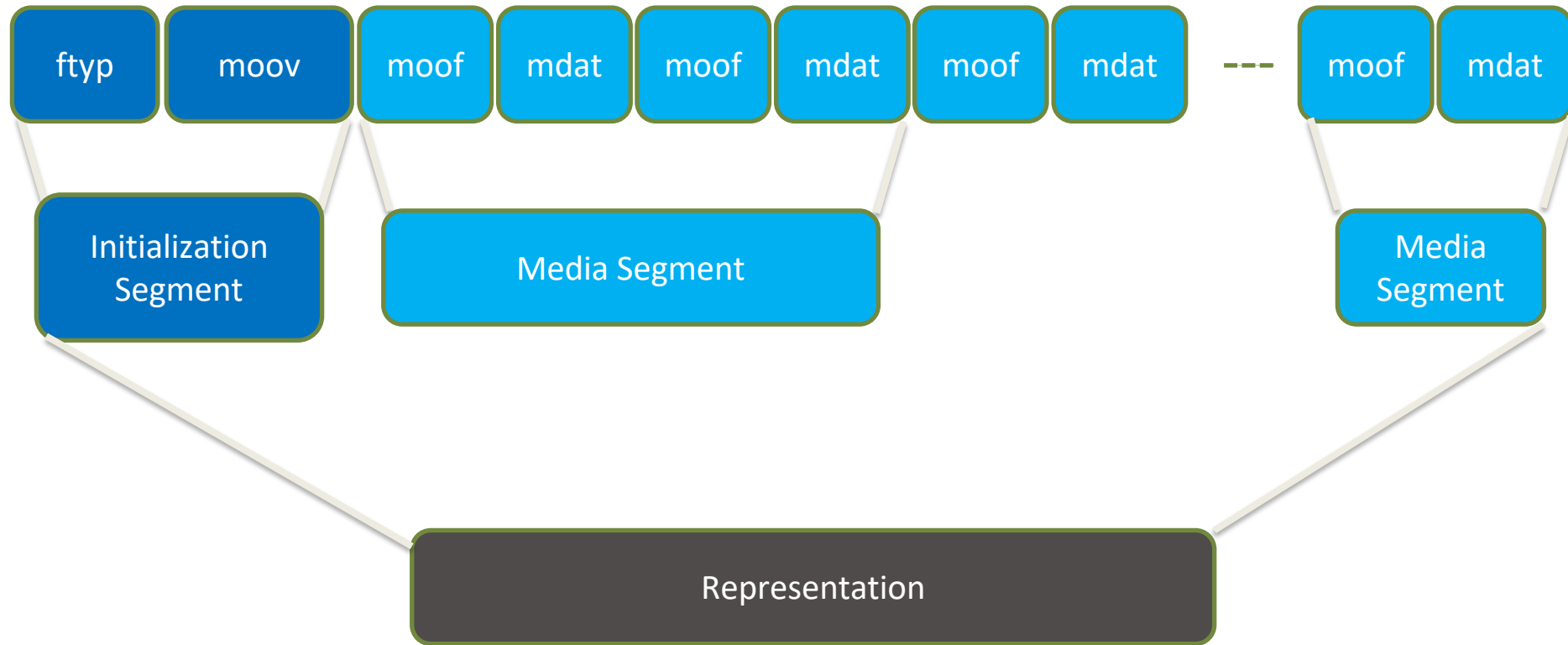


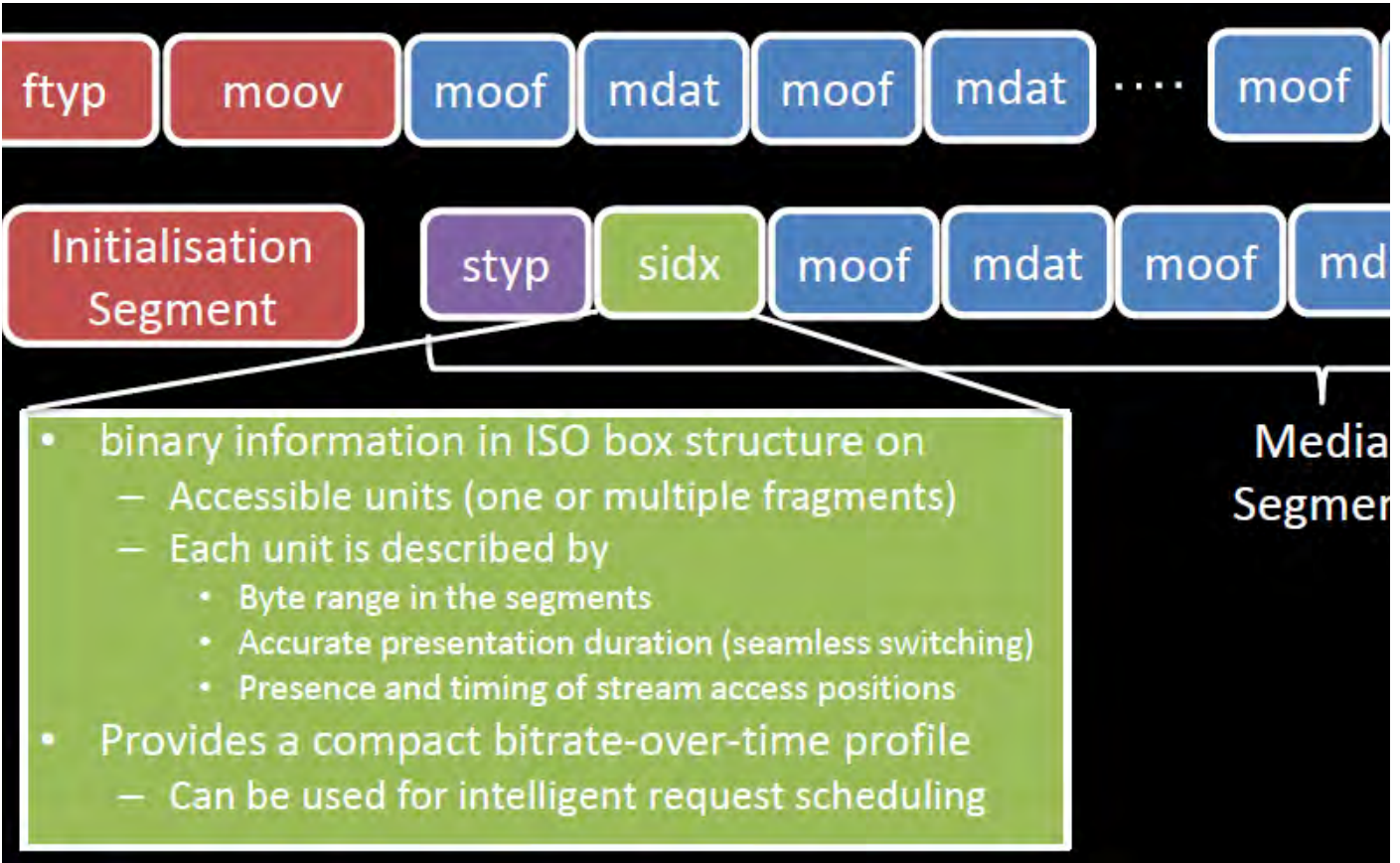
WHY THE FILE FORMAT FOR STREAMING?

- **Object Oriented** – flexible and extensible structures called “boxes” used for sequencing media data along with nested metadata allowed specification of independently decodable “movie fragments” (DASH “Segments”)
- **Extensible metadata model** – that allowed adding information for live streaming, encryption, subtitles, new codecs, etc., separate from media data
- **Extensible timing model** – presentation time is the sum of previous sample durations, allowing time to be calculated on playback ... not a timestamp recorded on each sample
- **Interoperable file “brands”** – identifying sets of new boxes that enable adaptive streaming, Common Encryption, new codecs, live streaming, etc. with well-defined interoperability
- Enabled creation of a Multimedia Presentation Application Model consisting of a **Media Object Model** and **Media Timeline Model** that support **late binding of adaptive multimedia presentations with a single set of media objects** enabling a variety of delivery methods, such as file download, track download, multicast/broadcast, and adaptive streaming



EXAMPLE DASH REPRESENTATION AND SEGMENTS FOR ISOBMFF





SEGMENT INDEX



LATE BINDING

- To avoid combinatorial complexity or useless downloads, tracks are offered individually on cloud
- Client selects relevant tracks and synchronizes playout

Audio Selection Set

English AAC stereo CMAF Switching Set (single Track)

French AAC stereo CMAF Switching Set (single Track)

English multichannel CMAF Switching Set (single Track)

French multichannel CMAF Switching Set (single Track)

Subtitle Selection Set

English WebVTT description CMAF Switching Set (single Track)

English TTML description CMAF Switching Set (single Track)

French WebVTT dub CMAF Switching Set (single Track)

French TTML dub CMAF Switching Set (single Track)

Video Selection Set

SD Media Profile CMAF Switching Set (multiple Tracks)

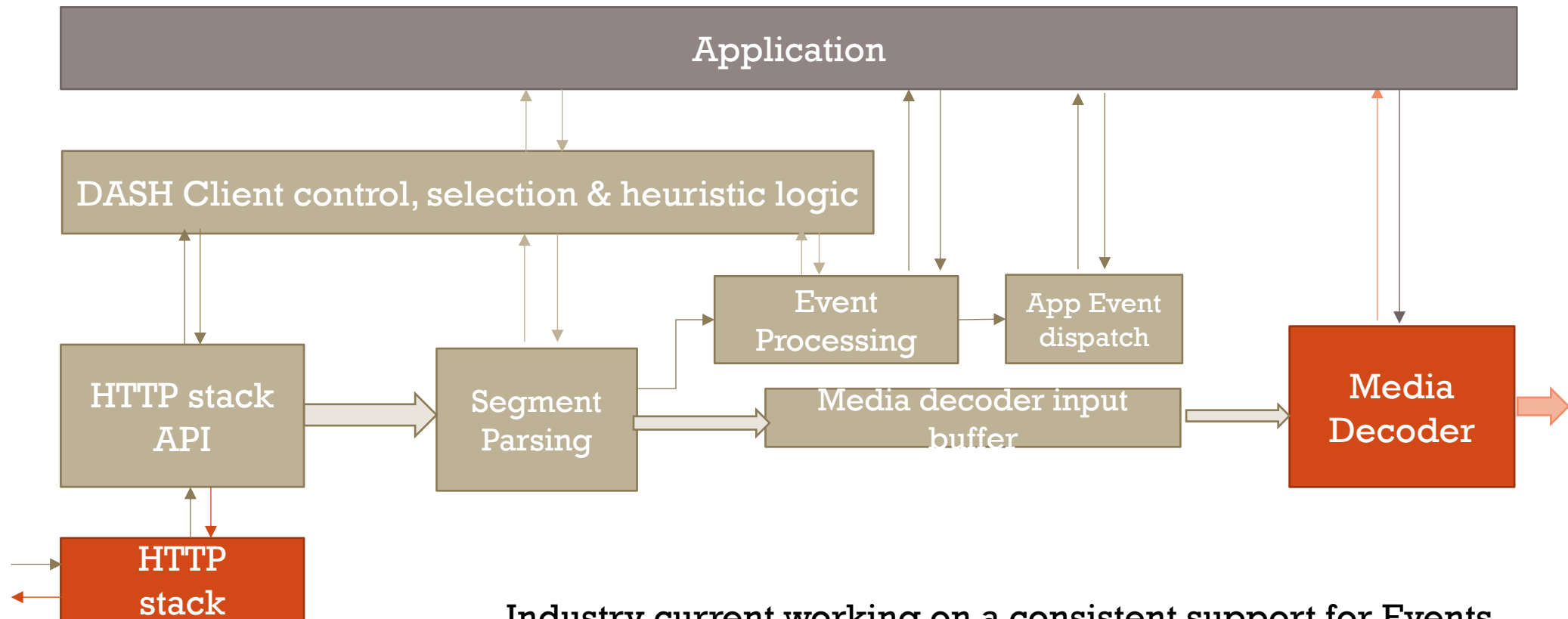
HD Media Profile CMAF Switching Set (multiple Tracks)

UHD10 Media Profile CMAF Switching Set (multiple Tracks)



EVENTS

- Providing the ability that an application can distribute media synchronized events such as SCTE markers, simple overlays, stats, etc.



Industry current working on a consistent support for Events



LOW LATENCY STREAMING

MPD

IS

Encoder

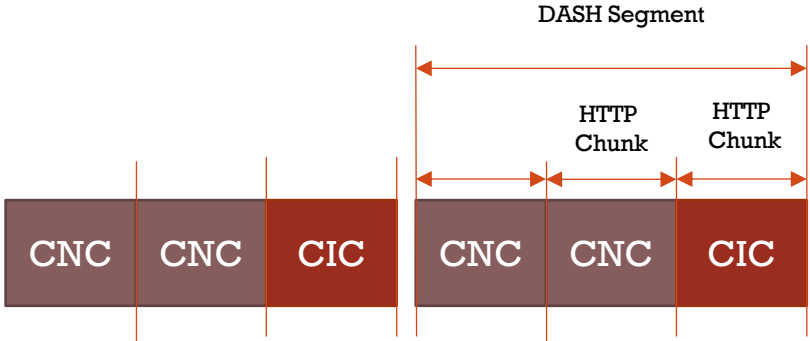


CH = CMAF Header

CNC = CMAF non-initial chunk

CIC = CMAF initial chunk

DASH Packager



Regular DASH Client

10s

Segments

Low-Latency DASH Client

3s

Chunks

CDN stores Segments



MSE AND BYTESTREAM FORMAT

- Media Source Extension
 - This specification extends HTMLMediaElement [HTML51] to allow JavaScript to generate media streams for playback.
 - Allowing JavaScript to generate streams facilitates a variety of use cases like adaptive streaming and time shifting live streams.
- ByteStream Format for ISO BMFF
 - <https://www.w3.org/TR/mse-byte-stream-format-isobmff/>
 - This specification defines a **Media Source Extensions™ [MEDIA-SOURCE]** byte stream format specification based on the ISO Base Media File Format.

```
var ms = new MediaSource();
video.src = window.URL.createObjectURL(ms);
ms.addEventListener('sourceopen', onMediaSourceOpen);

function onMediaSourceOpen() {
  sourceBuffer = ms.addSourceBuffer('video/mp4; codecs="avc1.4d401f"');
  sourceBuffer.addEventListener('updateend', nextSegment);

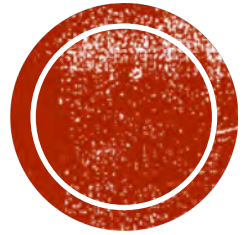
  GET(initUrl, appendToBuffer);

  video.play();
}

function nextSegment() {
  var url = templateUrl.replace('$Number$', index);
  GET(url, appendToBuffer);
  index++;
  if (index > numberOfChunks) {
    sourceBuffer.removeEventListener('updateend', nextSegment);
  }
}

function appendToBuffer(videoChunk) {
  if (videoChunk) {
    sourceBuffer.appendBuffer(new Uint8Array(videoChunk));
  }
}
```



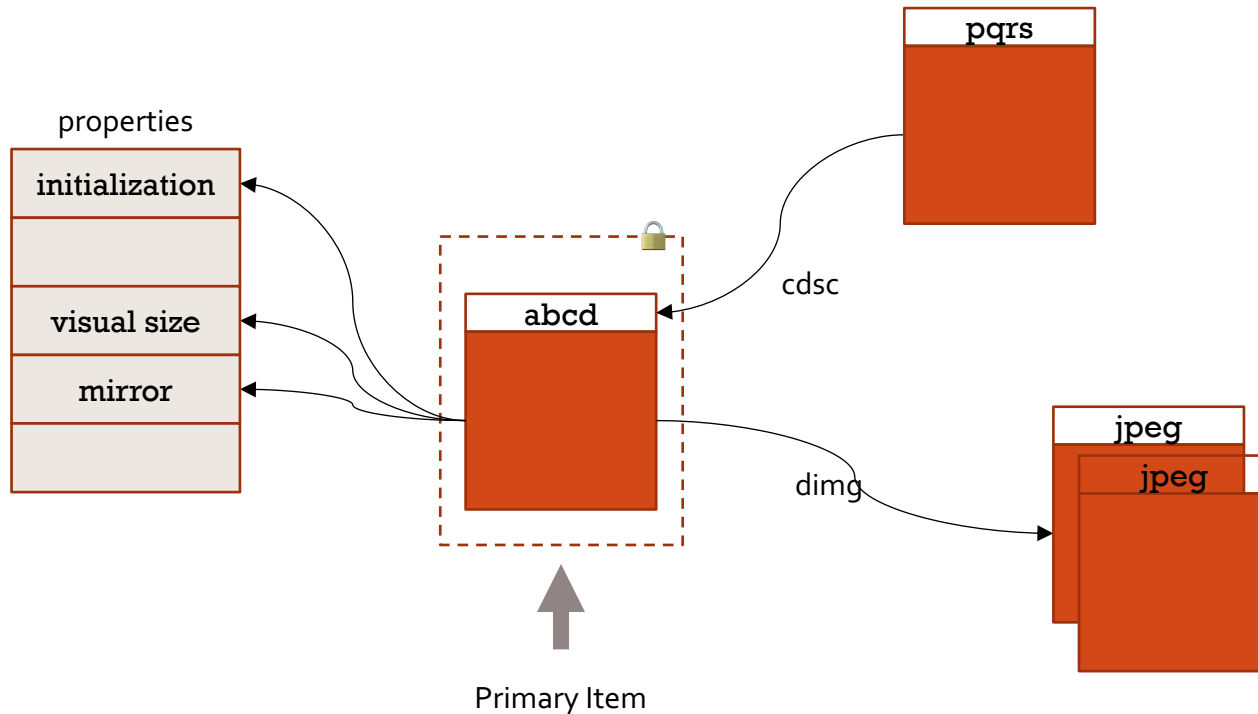


OTHER APPLICATIONS



HIGH EFFICIENCY IMAGE FILE FORMAT

- ISO/IEC 23008-12 permits storage:
 - Sequences (e.g. bursts, brackets): as tracks, MP4-style
 - Images (coded or derived) as Items, MPEG-21-style

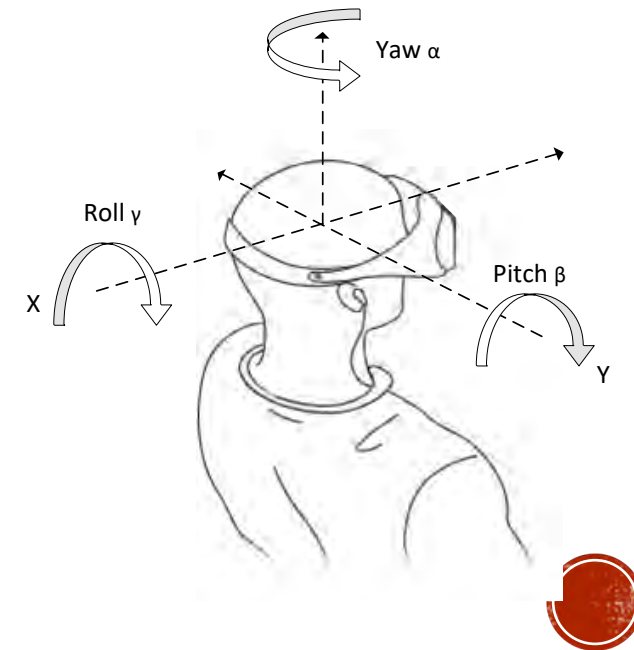
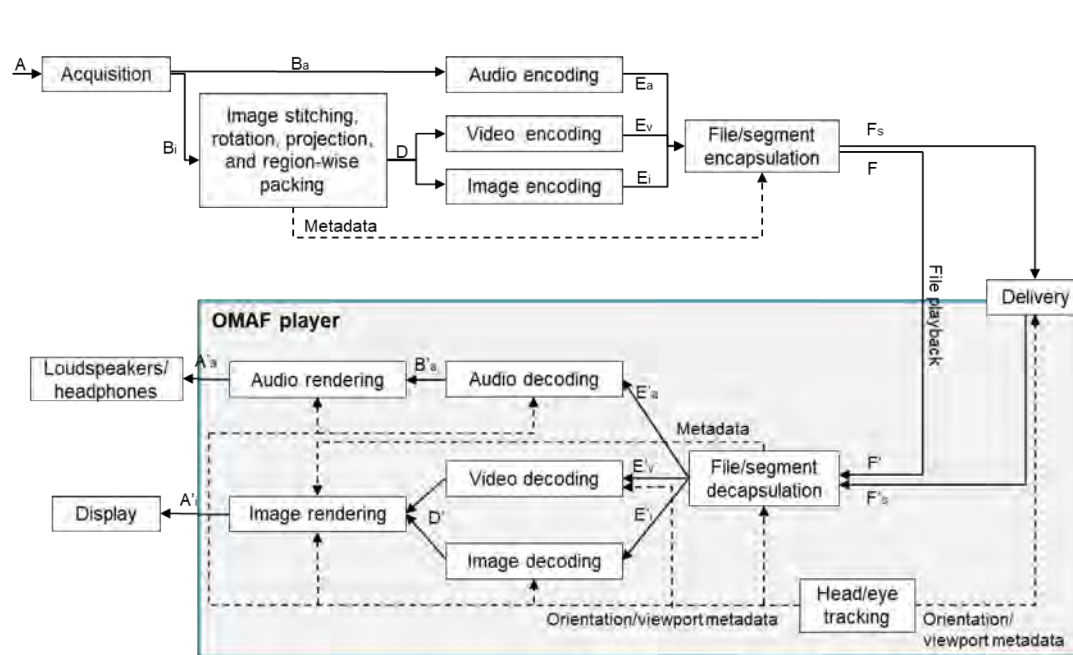


- Coded Items
 - HEVC, AVC, JPEG, (JPEG-XR),...
- Derived items
 - Image overlay (compose)
 - Image Grid
 - ...
- Metadata Items
 - EXiF, XMP, MPEG-7, ...



OMNIDIRECTIONAL MEDIA FORMAT (OMAF)

- 23090-2: Part 2 of MPEG-I Coded Representation of Immersive Media
- It is a **systems standard** developed by **MPEG** that defines a **media format**, enables **omnidirectional media applications**, focusing on **360° video**, images, and audio, as well as associated timed text. ^z



OMAF SIGNALING IN ISO BMFF

- General rules for signalling of important information
- Overall omnidirectional video indication
- Signalling of projection format
- Signalling of region-wise packing and guard bands
- Signalling of rotation
- Signalling of frame packing
- Signalling of content coverage
- Region-wise quality ranking
- Signalling of fisheye video parameters
- Storage and signalling of omnidirectional images
- Storage and signalling of timed text
- OMAF timed metadata



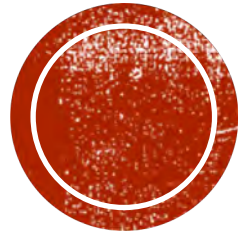
PARTIAL FILE FORMAT 23001-14

The Partial File Format is designed to contain files partially received over a lossy link (with unreceived or corrupted sections), for further processing such as playback or repair.

The file structure is object-oriented; a file can be decomposed into constituent objects very simply, and the structure of the objects inferred directly from their type.

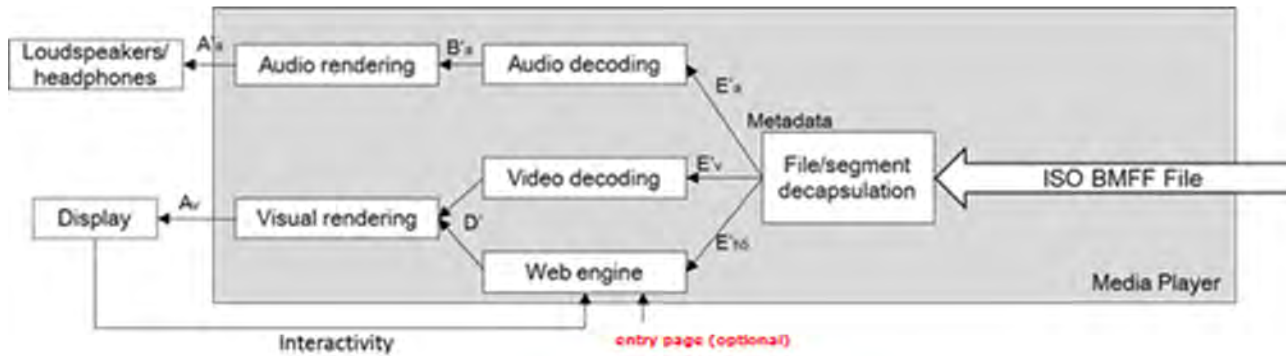
This format contains the correctly received data, missing block identification, and repair information such as location of the file or high-level original indexing information.





CRYSTAL BALL

Some MPEG Activities



- Under development
- specifies how the ISO BMFF format can be used to store web resources (e.g. HTML, JavaScript, CSS, ...)
- specifies hypothetical processing for how these files can be consumed by web browsers, in particular how references from web resources to the file that carry them or to other web resources carried in the same file are handled.
- enables the delivery of synchronized media and web resources as supported by ISO/IEC 14496-12: file download, progressive download, streaming, broadcast, etc.

WEB RESOURCE TRACK 23001-15

Workshop
planned with
3GPP, MPEG, W3C,
ATSC, DVB, CTA
and HbbTV

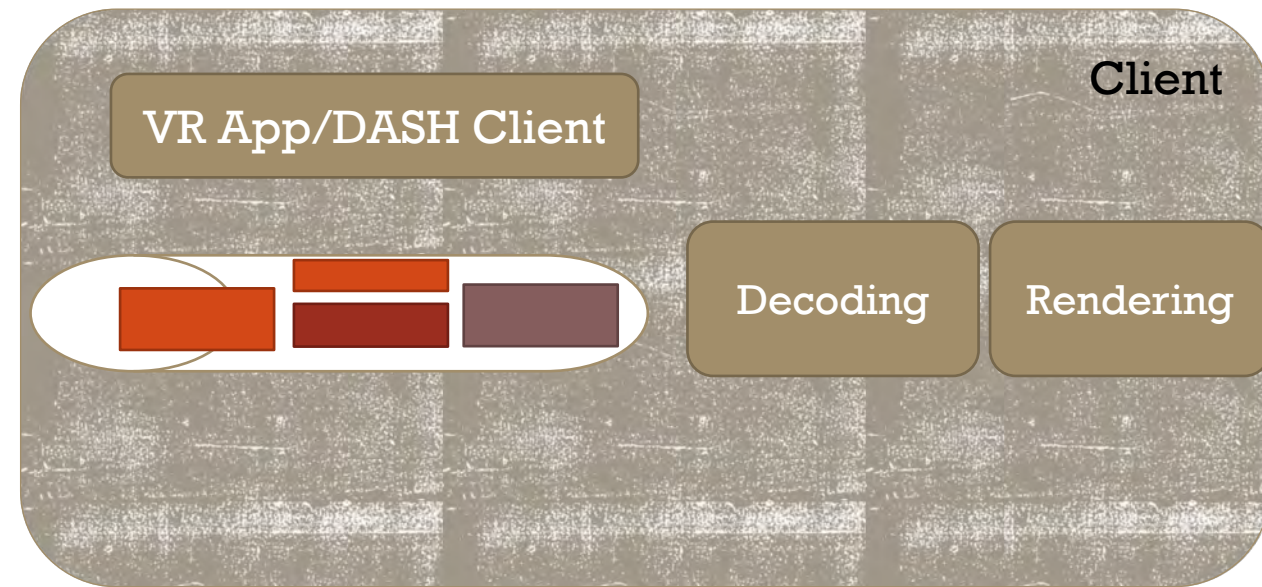
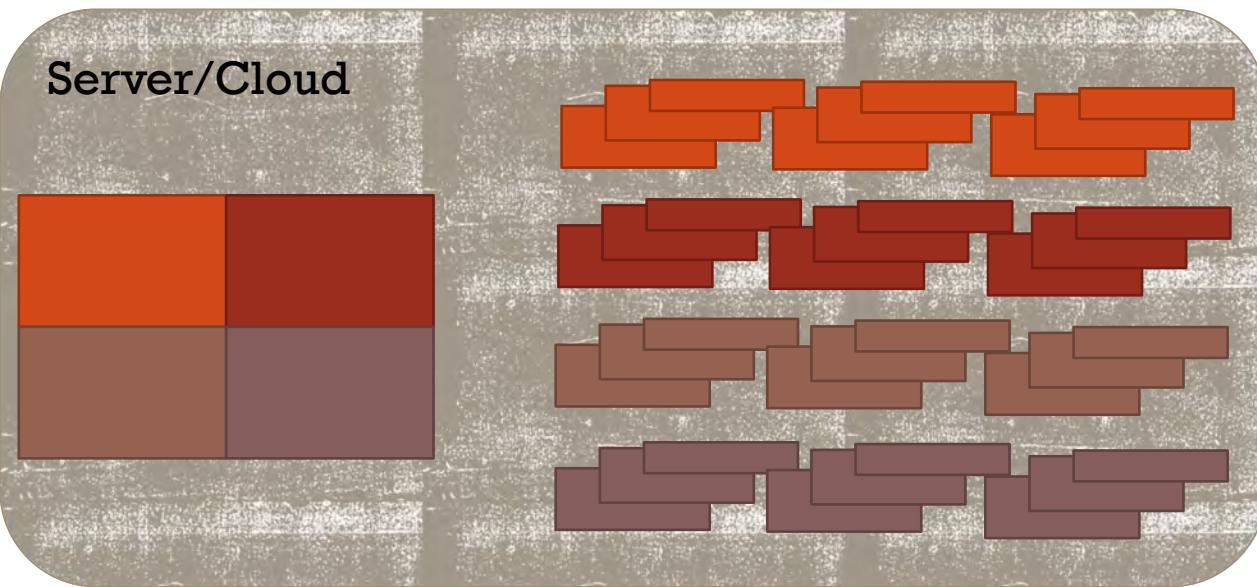
IMMERSIVE MEDIA IN ISO BMFF

▪ Examples

- **Tiled 360 videos** in very high resolution
- Large **Point Clouds** that can be navigated in 6 DoF
- **Lightfields** with lots and lots of small tiles
- A complicated **Scene Graph** with many objects to traverse
- Audio objects can be audible, or beyond the “**audio horizon**” in an immersive experience

▪ Environment

- All likely retrieved from some sort of **cloud infrastructure**
- All of these can be available in multiple quality/bitrate variations
- At the receiver all of those need to **decoded and decrypted** with constrained devices

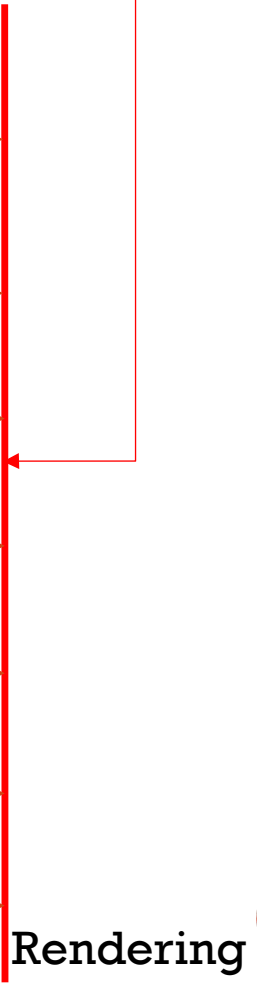
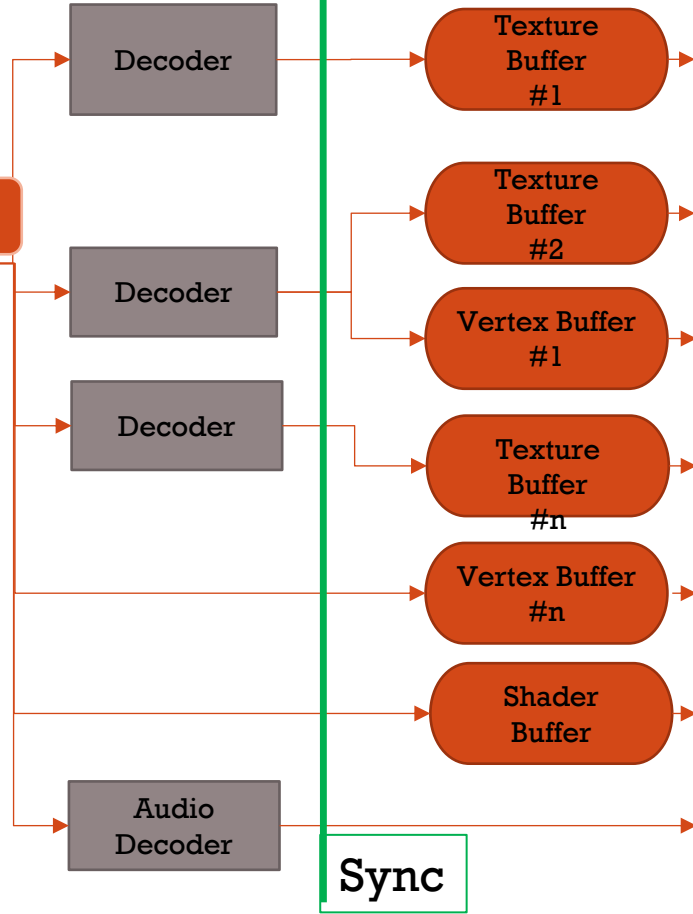
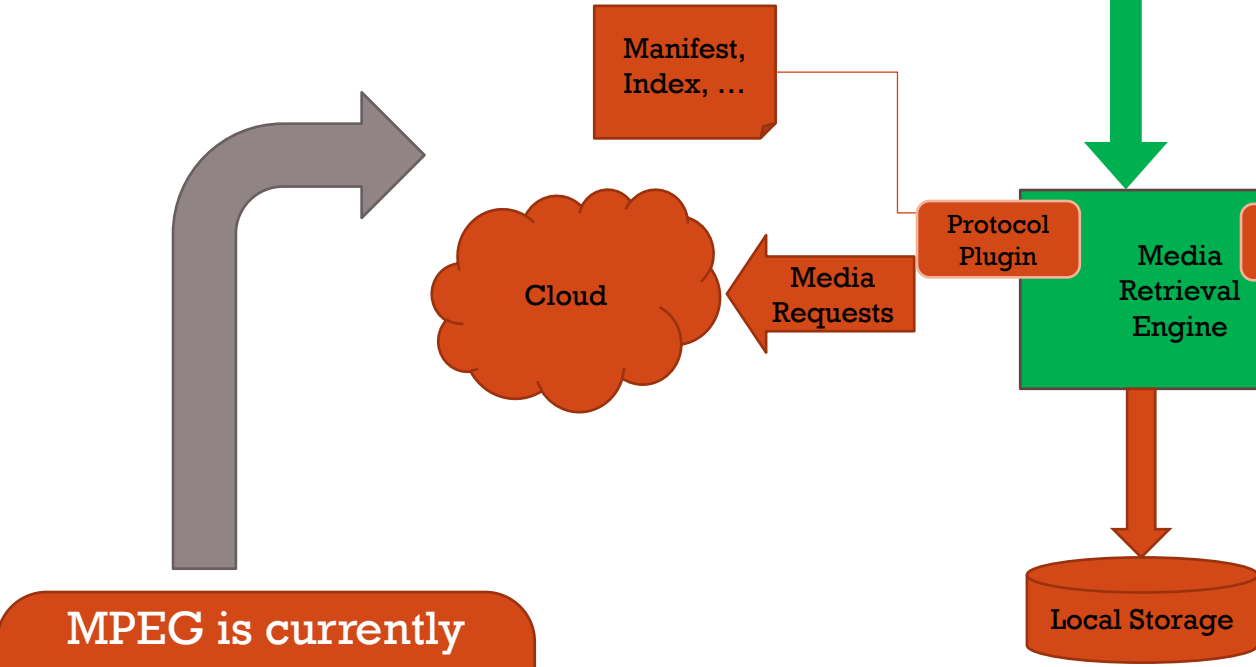


IMMERSIVE CLOUD MEDIA

Media Resource References
Timing Information
Spatial Information
Media consumption information



Sync Information
Shader Information



MPEG is currently investigating storage and streaming formats for immersive media



CHALLENGES

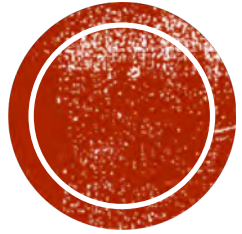
- Flexibly retrieving **parts** of a large body of media data from a **cloud** resource to create a coherent user experience under **constrained resources**
- Where **constraints** exist like bandwidth, access latency, decode resources (and where these can fluctuate dynamically)
- With the client in charge of making **trade-offs** given such constraints
- Where fast **response times** and **efficiency** are crucial for the QoE
- Where inherently, data is accessed and retrieved in multiple **parallel streams**
- Where this data may need to be **protected** and/or encrypted
- Where this data may need to be **cached** close to the user for the best experience
- Where the data is stored in the **cloud** in a distributed manner



ORGANIZATION DIMENSIONS: IMMERSIVE MEDIA

- **Temporal** random access – “as usual”
- **Spatial** random access – retrieving only the relevant parts of the media
 - Depending on user orientation
- Making quality/bitrate trade-offs in switching between **quality** levels
 - Depending on what is visible/audible
 - Depending on retrieval/device and resource constraints, including bandwidth, latency, decoder capability, things like video and audio reproduction capabilities (e.g. screen resolution and color space; speaker config)
- Decoding capabilities, user preferences, etc.
- Addition of static media
- Different timelines
- Scene Descriptions, Nodes, etc.
 - Which objects to retrieve – and which *parts* of objects
- Extend the File Format or do something “NEW”? → ongoing





SUMMARY

SUMMARY

- **Successful file format**
 - Very versatile: from editing to HTTP streaming to broadcasting
 - Very extensible (codecs, usages, applications)
 - Very dynamic (more contributions than ever)
- **Some challenges**
 - Carrying some legacy that is no longer in use
 - Addressing all the use cases while maintain compatibility
 - For certain applications and use cases, the file format principles are suboptimal in terms of overhead or processing efficiency.
- The ISO BMFF is the stable glue between modern media and transport, but will evolve further for new use cases applications.



THANK YOU

A group of men are laughing together outdoors. In the foreground, a man with a shaved head and a blue polo shirt is laughing heartily, his eyes closed and mouth open. He has a black watch on his left wrist. Behind him, another man in a light blue polo shirt is also laughing, with his hands clasped in front of him. To the left, two other men are laughing; one is wearing a grey t-shirt and the other is wearing a white t-shirt. They are all smiling and laughing, creating a joyful atmosphere. The background shows a street scene with buildings and trees.

Thanks to Dave Singer, Kilroy Hughes, Per Fröjdh, Cyril Concolato, Ye-Kui Wang and all other that I stole a picture from