



**ISO/IEC JTC 1/SC 29/WG 11**  
**Coding of moving pictures and audio**  
**Convenorship: UNI (Italy)**

**Document type:** Approved WG 11 document

**Title:** MPEG 127 version of MPEG Standardisation Roadmap

**Status:** Approved

**Date of document:** 2019-07-12

**Source:** WG11

**Expected action:** Publication

**No. of pages:** 18

**Email of convenor:** [leonardo@chiariglione.org](mailto:leonardo@chiariglione.org)

**Committee URL:** <http://mpeg.chiariglione.org>

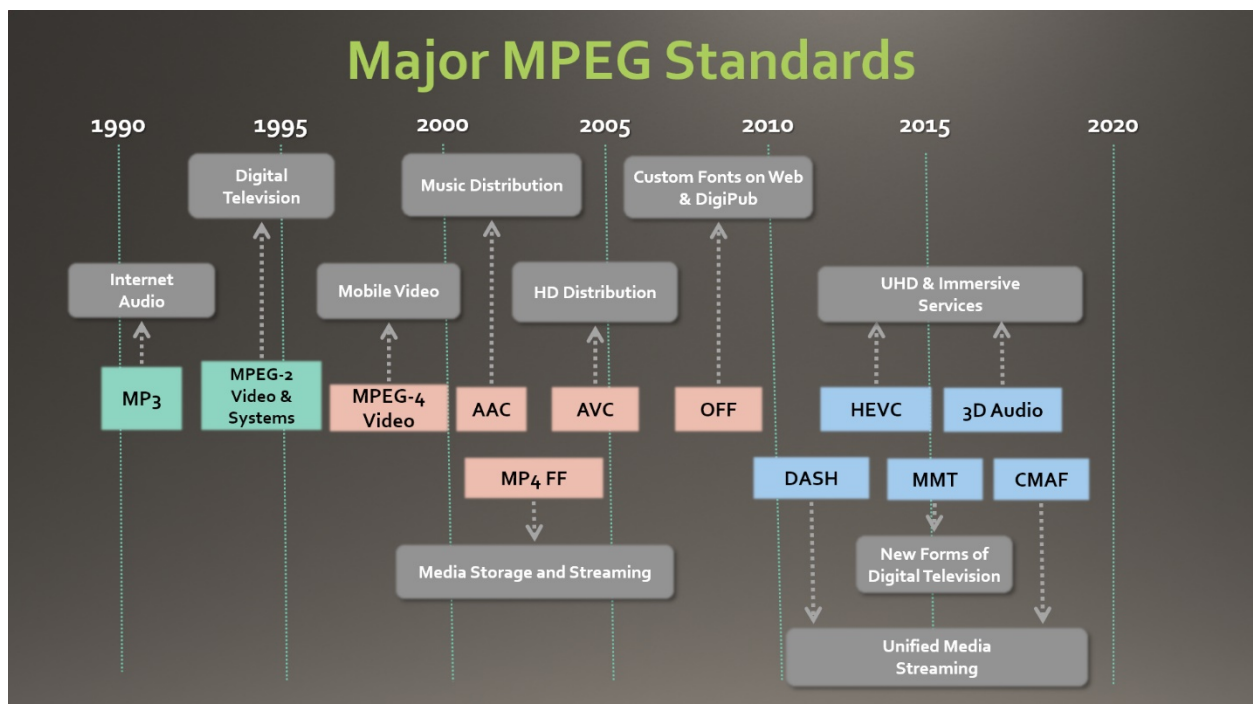
**INTERNATIONAL ORGANISATION FOR STANDARDISATION  
 ORGANISATION INTERNATIONALE DE NORMALISATION  
 ISO/IEC JTC 1/SC 29/WG 11  
 CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 11 N18796**  
 October 2019, Geneva, SE

**Source:** MPEG  
**Title:** MPEG 128 version of MPEG Standardisation Roadmap  
**Public:** Yes

## MPEG Standardisation Roadmap

In this document, MPEG lays out its medium-term Strategic Standardisation Roadmap, aimed at collecting feedback from the broadcasting, content and service provision, media equipment manufacturing and telecommunication industry, and anyone in professional and B2B industries dealing with media.



*Figure 1 - MPEG Standards Enable Industries*

## MPEG Standards Enable Markets to Flourish

MPEG is an ISO/IEC standardisation group that has enabled huge markets to flourish through its standards. MP3 revolutionised the way music is distributed and enjoyed. MPEG-2 enabled the

digital television industry to replace analogue TV and has facilitated the expansion of satellite and Cable TV. MPEG-4 Advanced Video Coding (AVC) and High Efficiency Video Coding (HEVC) have also enabled large scale interactive media distribution. The ISO file format family (“mp4”) has enabled interoperable exchange of media files, while MPEG DASH powers the adaptive and optimised distribution of interactive media. See Figure 1 for an overview of MPEG’s cornerstone standards. (Please refer to Annex A for the acronyms).

MPEG serves not only B2C markets but also B2B content exchange (e.g. between TV studios and surveillance) and consumer-to-consumer communication in all modern smartphones. MPEG standards have enabled and continue to enable the creation and development of markets by providing interoperability while giving technology buyers options to choose from.

In its almost 30 years of existence, MPEG has gathered an increasing number of the world’s best experts in media technologies and accessed the best and most recent R&D results, Through an extremely thorough and competitive process, MPEG has created widely deployed, cross-industry standards that have enabled industries to converge.

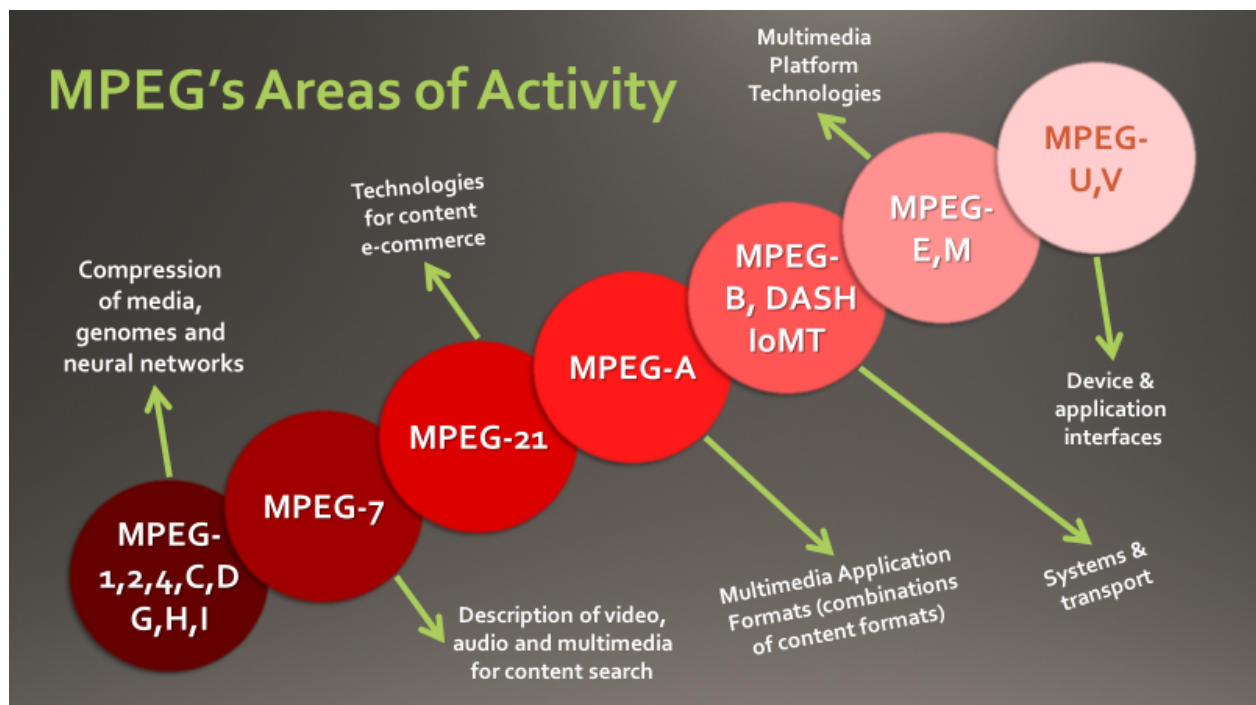


Figure 2 – Areas of MPEG Standardisation

MPEG works by proactively anticipating industry needs for standards. It can do so because it is firmly rooted in industry, with hundreds of industry representatives participating in its standardisation efforts. Once the need for a standard is identified, MPEG defines its requirements by interacting with its constituent industries, related standards committees, and industry fora. MPEG typically issues "Calls for Proposals" to obtain the best technologies which are integrated in its standards through a highly competitive process. This enables MPEG to produce timely standards that not just follow but foment and lead technology developments. MPEG prides itself to adhere to its strict schedules while being able to quickly respond to emerging industry needs, as it did with MPEG DASH and the CMAF distribution format.

Demands from increasingly sophisticated technologies have prompted MPEG to extend the coverage of its standards from audio and video "compression" with support for "transport" (network and storage), to a range of media-related technologies as depicted in Figure 1, above.

MPEG has executed several standardisation projects represented by numbers (e.g. MPEG-4) and letters (e.g. MPEG-H). In some cases, these represent bundles of technologies (e.g. MPEG-2 and MPEG-H) for well-defined services. In general, however, its standards can be independently combined to create specific media experiences.

Figure 3 below shows the delivery timeline of a selection of major MPEG standards organised along 5 categories: audio, video, graphics, media-related and systems.

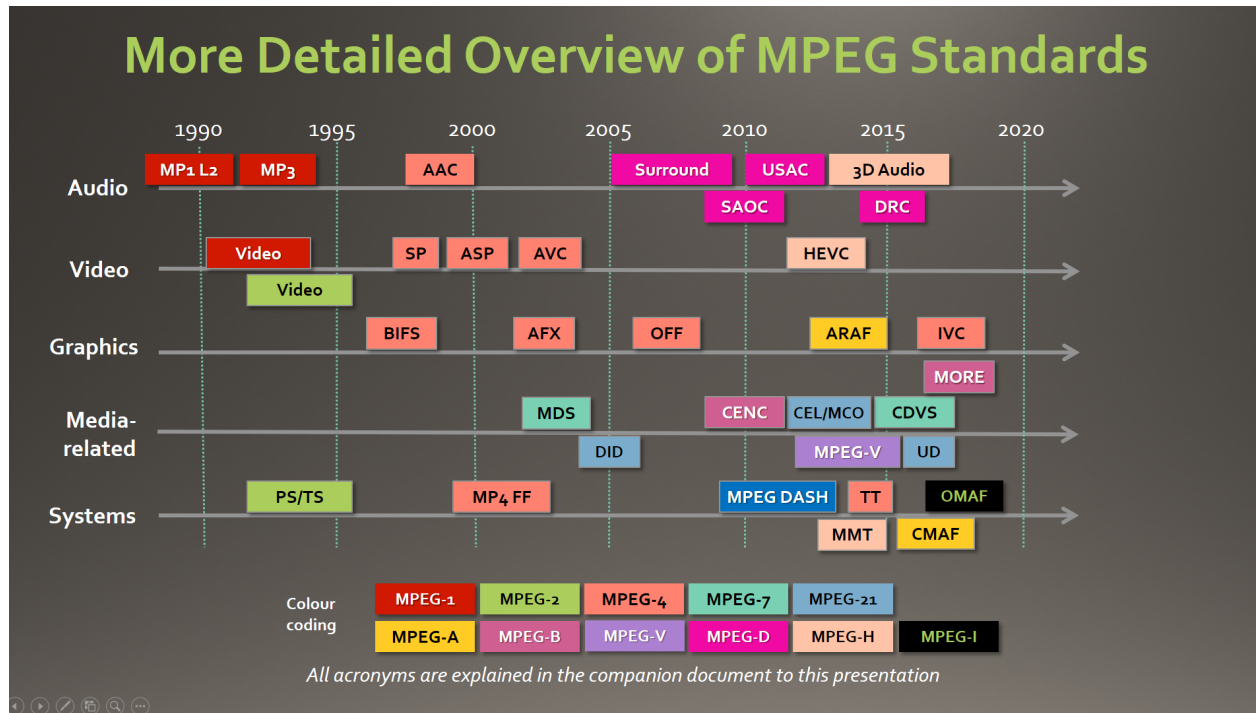


Figure 3 – Overview of MPEG Standards

## MPEG Operates in a Dynamic Environment

The demand for all types of media content continues to grow and media data is expected to dominate communication traffic not just at peak demand times, but at every time of the day in the next few years. IP video will represent 80% of all global traffic in the next three years and three-fourths of all world-wide mobile traffic will be video by the year 2020 [1].

Growing trends in higher space-time resolution, higher dynamic range, wider colour gamut video, and immersive media (VR, AR, 360 panoramic, etc.) require increased bandwidth, low latency and improved services. As media communication industries move towards providing more personalized media experiences, devices must become more personal and must provide more immersive services. Augmented and virtual reality will give us more immersive experiences in film, television, voice and data markets, and these markets are forecast to grow many-fold in the next five years [2].

All types of devices and sensors will be part of the Internet of Things (IoT) and will be able to communicate not just plain data, but also audio-visual information. Of the 20 billion connected 'things' predicted in five years, 65% will be consumer-oriented [3]. This widespread adoption of the IoT will require new machine-to-machine media communication to provide enhanced

capabilities that will augment sectors such as transportation systems (e.g., autonomous vehicles). Cloud computing and Big Data technologies evolve from basic data to rich audiovisual media, and enabling efficient search and discovery with everything connected will be key. This requires high levels of interoperability and efficiencies of communication to fuel market adoption and growth.

## MPEG's Five-Year Roadmap

The following figure depicts MPEG's current thinking on its roadmap for upcoming standards<sup>1</sup>.

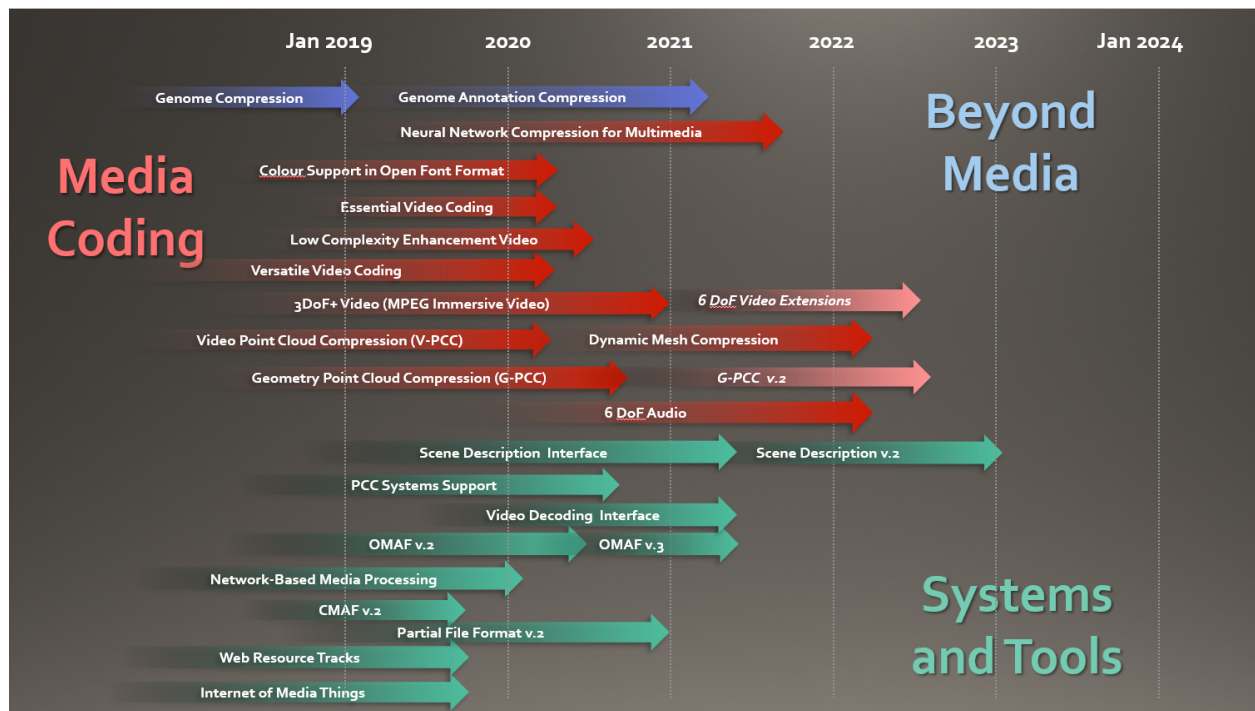


Figure 4 - MPEG Standardisation Roadmap. See Annex B for short project descriptions

MPEG will keep working on interoperable exchange formats for media for a variety of adaptive streaming, broadcast, download, and storage delivery methods, such as the Common Media Application Format (CMAF) that was finalized in 2017.

MPEG is further working on supporting better pixels (more and brighter colours and more contrast, also known as wide colour gamut and high dynamic range) in its existing HEVC standard. MPEG is now researching the next generation codec, suitable for ever higher resolutions, for larger screens, VR applications and for modern distribution models including OTT and for next-generation networks, including 5G.

A major focal point for MPEG is to enable personal, immersive experiences. This includes augmented and virtual reality entertainment with immersive video and audio, as well as immersive media communication in real and virtual environments. With this work, MPEG caters to the needs of social media moving from text, via multimedia, to fully immersive experiences. Industry has told MPEG loud and clear that it urgently needs standards for efficient representation, communication and distribution of immersive content and services.

<sup>1</sup> The elements in italics are under consideration; MPEG has not yet formally committed to develop such standards, and the timing of them is indicative.

MPEG's upcoming standards empower new types of devices, like head-mounted displays and sensors, to be interoperable with services, and affordable to consumers. They will also enrich the *capture* of new types of media, including more immersive media, including 360-degree recording of audio and video producing MPEG-encoded surround sound, object-based audio and video formats. MPEG has researched various forms of immersive TV, e.g. enabling users to freely select their viewpoint in their interaction with media content, and expects to publish a standard. Further, new MPEG standards will allow streaming non-AV environmental dimensions, like GPS coordinates. In doing so, it will become possible to integrate media from many different and heterogeneous sources into a single, coherent media experience.

MPEG doesn't just optimise its coding standards to new delivery networks (e.g., 5G) and environments (e.g., automotive), it also provides the delivery methods attuned to the opportunities and constraints of these environments. It has recently become clear that the interaction between coding and networking requires attention. As both become more intelligent and adaptive to the environment, the need arises for network-aware content representation, or representation-aware networking. Absent such solutions, both the network and the delivery layers will respond dynamically and independently to the same changing conditions, which will result in unpredictable and undesirable behaviour. A recent effort in this field is the MPEG DASH "SAND" specification, which allows for network-aware DASH delivery.

With its work on compact descriptors for search (CDVS) and on video analysis (CDVA), MPEG enables content identification and search in vast media databases ("Big Media"). These standardised descriptors allow simple querying across diverse and heterogeneous databases empowering the automatic understanding of what is actually contained in the media itself.

Last but not least, upcoming MPEG standards will support an Internet of Media Things and Wearables, where MPEG standards facilitate the interoperable and efficient exchange of media data between these "Things". This work will also allow automating the extraction of information from media data that machines and humans can understand, and act upon.

## Share your Thoughts and Requirements

MPEG keeping industry informed of its plans, providing the opportunity to influence the direction of international digital media standardisation. If you represent an industry that relies on standards-based interoperability in audiovisual products, services and applications, MPEG is very interested to hear about your needs and vision, for example by answering the following questions:

- Which needs do you see for media standardisation, between now and 5 years out?
- What MPEG standardisation roadmap would best meet your needs?
- To accommodate your use cases, what should MPEG's priorities be for the delivery of specific standards?
  - For example, do you urgently need something that may enable basic functionality now, or can you wait for a more optimal solution to be released later?

When providing feedback, please note that MPEG is flexible in providing not only the desired technology, but also the level of integration across multiple technologies to meet an industry vertical's specific needs. Note that MPEG plays no role in when and how standards become available on the market in the form of products and services (these are specific industries' and companies' decisions), but that we do have liaisons with relevant trade organisations. MPEG also plays no role in when and how relevant patents are licensed, as ISO rules prevent MPEG from handling licensing matters.

## References

- [1] Cisco Visual Networking Index Report February 2016
- [2] Digi-Capital Augmented/Virtual Reality Report 2015
- [3] Gartner Symposium/ITxpo, Barcelona, Spain November 2015

## Annex A - Acronyms

For more information, please refer to the MPEG website: <http://mpeg.chiariglione.org/>

AAC	Advanced Audio Coding
AFX	Animation Framework eXtension
ARAF	Augmented Reality Application Format
ASP	MPEG-4 Advanced Simple Profile
BIFS	Binary Format for Scenes
CDVA	Compact Description for Video Analysis
CDVS	Compact Description for Visual Search
CEL	Contract Expression Language
CENC	Common Encryption
CMAF	Common Media Application Format
DASH	Dynamic Adaptive Streaming over HTTP
DID	Digital Item Declaration
DRC	Dynamic Range Control (Audio)
FF	File Format
HDR	High Dynamic Range (Video)
HEVC	High Efficiency Video Coding
IoMT	Internet of Media Things
IVC	Internet Video Coding
MCO	Media Contract Ontology
MDF	Multimedia Description Scenes
MORE	Media Orchestration
IoTM	Internet of Media Things
MLAF	Media Linking Application Format
MMT	MPEG Media Transport
MP1 L2	MPEG 1 Layer 2 Audio
MP3	MPEG 1 Layer 3 Audio
OFF	Open Font Format
OMAF	Omnidirectional Media Format
PS/TS	(MPEG-2) Program Stream
SAND	Server And Network-assisted DASH
SAOC	Spatial Audio Object Coding
SP	(MPEG-4 Visual) Simple Profile
TT	Timed Text
TS	(MPEG-2) Transport Stream
UD	User Description
USAC	Unified Speech and Audio Coding



## Annex B – Project Descriptions

### Versatile Video Coding

The expanding use of more information rich digital video in diverse and evolving context and the still limited transmission and storage capabilities demand more powerful compression schemes. Reasons for this include the increases in fidelity, frame rate, and picture resolution for both stationary and mobile devices, the increasing number of users, the increased amount of video used by each user, and the increasing tendency towards individual use of download, upload, streaming, and security-related video services. Future standardization will address existing markets for video coding including terrestrial and satellite broadcasting, cable services, managed IPTV via fixed telecommunication services, Over-the-top services, professional content production and primary distribution, digital cinema and packaged media as well as surveillance, screen content and gaming.

It is anticipated that a new generation of video compression technology will be needed by the beginning of the next decade that has sufficiently higher compression capability than the HEVC standard and can particularly support professional high quality, high resolution and extended dynamic/colour-volume video, as well as user generated content. For some of the markets listed above, the new standard will strive to reduce the bitrate for storage and transport of video by 50%.

Special attention will also be given to support developing markets like augmented and virtual reality, unicast streaming, automotive applications and media centric Internet of Things. For these markets special attention will be given to seamlessly enable the required functionality by close integration with transport and storage to provide efficient personalized interactive services, as well as appropriate projections to enable VR applications. Therefore, though the further improvement of compression performance is expected to play the major role in this development, adaptation capabilities for usage in various network environments, a variety of capturing/content-generation and display devices are considered important as well.

### MPEG-5 Video

MPEG-5 will be a video coding standard to address business and technology needs in use cases that are not well served by existing ISO standards, such as HEVC.

The goal is to develop a new video coding standard that meets a combination of business and technology requirements:

- a Baseline profile containing technologies which are over 20 years old or which are accompanied only by Type 1 IPR declarations
- a Main profile containing a small number of additional tools, each of which is capable of being cleanly switched off or switched over to Baseline tools on an individual basis
- Encouragement of the timely publication of licensing terms\
- Coding efficiency at least as good as that of HEVC\
- Complexity suitable for real time encoding and decoding

### Tools for Virtual Reality

The market is rapidly adopting to Virtual Reality (VR) to provide immersive experiences that go beyond what a UHD TV can offer. To enable a true-to-life VR experience, immersive video is essential. Interactivity between the user and the content, high quality video (HDR, increased spatial resolution), and efficient delivery over existing networks are required.

There is a strong market momentum and interest in VR with a rapid increase in number of companies providing cameras, content and devices for VR. Since there are no existing standards for VR, the lack of interoperability is a significant challenge in the current marketplace. Therefore, MPEG has started standardization of essential VR technologies, to prevent fragmentation in VR marketplace and to enable VR services and applications to thrive in the mass market.

MPEG recently finished its Omnidirectional Media Format, OMAF, in a first version. It covers 3D projection methods and related metadata and signalling for “VR360”. Further work is ongoing on optimized video coding technologies for VR, delivery mechanisms, lightfield video coding and point cloud compression – see below.

### **Omnidirectional Media Format (OMAF)**

After doing a survey in 2016, MPEG decided to have a number of phases in providing support for immersive media. The first of these phases (“phase 1a”) was the OMAF specification, see above. This specification was approved in its final version in October 2017. Also already in 2016, MPEG planned a “Phase 1b” to follow 1a, in the form of a new version of OMAF and anything that would need to be added to other specifications, like the file format and the specification on Metadata for Immersive Media.

MPEG is now working to add a number of things to OMAF v.2. According to the Requirements Document that guides the work, this includes:

- Providing a more natural viewing experience by supporting head motion parallax (also referred to as 3DoF+)
- Providing for interactivity (e.g., switching between different camera positions, or adaptively changing the audio rendering depending on user actions)
- More efficiency, notably in viewport-adaptive streaming
- Temporal navigation (“trick play”) and zooming in an immersive experience
- Providing overlays for things like logos, a sign interpreter, or subtitles/closed captions

### **Immersive Audio with 6 Degrees of Freedom**

MPEG-I Audio seeks to provide an immersive virtual audio and visual experience to users wearing a head-mounted visual display (HMD) with audio headphones, or using loudspeakers. The user’s view and perception of the virtual reality environment (VR) is responsive to user motion (i.e. position and orientation), where the user is free to move around in a certain bounding box. The audio rendering will also be responsive to interaction with selected virtual items. In a similar way, MPEG-I Audio will support augmented reality (AR) in which virtual objects and associated sounds are placed in the user’s view of the physical world. The project will use the existing MPEG-H 3D Audio Low Complexity Profile (ISO/IEC 23008-3) as a compression engine, and the to be standardized technology MPEG-I Immersive Audio (ISO/IEC 23090-4), as an audio rendering engine.

### **Web Resource Tracks**

The Web Tracks specification “ISO/IEC 23001-15: Carriage of Web Resource in ISO/BMFF” defines how to carry web resources in the ISO Base Media File Format (ISO/IEC 14496-12, or “ISO/BMFF”). It also defines associated brands. (“Brands” are used to signal compatibility to the constraints and requirements in specifications based on the ISO/BMFF.) ISO/BMFF specifies

a format for the storage of timed resources such as media streams and resources for which no timed stream structure exists or when the timed stream structure does not need to be exposed. By the addition of web resources to an ISOBMFF file, the audiovisual content can be augmented with interactive features, overlays and other web-based features. The specification will enable playback of the file and the timed web resources in a browser. It also specifies how references from these Web resources to the file that carry them are handled. The specified storage enables the delivery of synchronized media and web resources as supported by the file format: file download, progressive file download, streaming, broadcast, etc.

### Multi-Image Application Format

The MIAF specification is fully compliant to the High Efficiency Image File (HEIF) format, and defines additional constraints to ensure a higher level of interoperability. While the HEIF specification defines the file format and general requirements for the included coding formats, HEIF does not define specific interoperability points by which capturing devices, editing applications, storage systems, cloud and delivery networks, and playback devices and applications can interoperate with each other.

The Multi-Image Application Format (MIAF) defines precise interoperability points for creating, reading, parsing and decoding images with associated metadata, including support for enhanced compression, High Dynamic Range, broader colour palette and object identification/portraiture depth maps for still, burst, sequence and live photos.

MIAF also limits the supported encoding types to a set of specific profiles and levels. It requires specific metadata formats, and defines a set of brands for signalling such constraints.

### Point Cloud Compression

We are witnessing a major paradigm shift in capturing the world. If some time ago 2D pictures and videos were enough and content consumption was restricted to 2D displays, there are now more and more devices for capturing and presenting 3D representations of the world. The easiest way of capturing the 3D information is by associating the depth with each captured pixel. By doing this from various angles, it is possible to reconstruct 3D points. An object or a scene is then a composition of such objects forming what is known as “point clouds”. These structures can have attributes such as colours, material properties and other attributes.

Point clouds typically use thousands up to billions of points to represent scenes that can be realistically reconstructed. MPEG’s Point Cloud Compression standard targets both lossy compression for, e.g., real-time communications, and lossless compression, for GIS, CAD and cultural heritage applications. Point clouds are typically captured using multiple cameras and depth sensors in various setups, but as usual in MPEG, acquisition is outside of the scope of the standard. The standard targets efficient geometry and attribute compression, scalable/progressive coding, as well as coding of sequences of point clouds captured over time. The compressed data format will support random access to subsets of the point cloud.

MPEG is developing two approaches to compress point clouds:

- **Video-based Point Cloud Compression (V-PCC)** leverages “traditional” (existing and future) video codecs and ecosystems (including hardware acceleration, transmission services and infrastructure) while enabling new applications. V-PCC covers lossless and lossy coding of 3D point clouds with associated attributes such as color. The current V-PCC test model encoder implementation shows compression performance of 125:1 while achieving good perceptual quality.

- **Geometry-based Point Cloud Compression (G-PCC)** uses 3D geometry-oriented coding methods, not relying on existing video coding technologies, for the efficient compression of sparse dynamically varying point clouds such as those used in vehicular LiDAR, 3D mapping, cultural heritage, and industrial applications.

The work on V-PCC is scheduled for completion in October 2019, and G-PCC follows one meeting later in April 2020. It is likely that second editions of these standards will follow.

### Scene Description for Immersive Media

MPEG Scene Description will enable the creation of 6DoF experiences based on MPEG media. The Scene Description is usually a graph that describes the 3D space as a scene that includes audio and visual objects in that scene. The 3D objects are defined using MPEG media like still pictures, 2D videos, point clouds, light fields, audio, and geometric primitives like meshes. The scene graph enables the composition of a scene that may consist of both natural-world (captured) and synthetic (e.g. CGI) content, with both compressed and uncompressed formats.

MPEG will develop a solution for scene description based on the widely adopted glTF 2.0 standard that was developed by Khronos. The solution will be specified as a set of extensions to add support for timed, MPEG represented media to glTF. The interfaces for exchange of information extracted from the scene description will also be defined to enable efficient media retrieval.

### Internet of Media Things (IoMT)

The phrase "Internet of Things" (IoT) encompasses a large variety of research, development and market efforts related to the communication between smart objects. The definition may be fuzzy, but the market reality is very clear: the number of devices connected to the Internet will reach 50 billion by 2020. An important factor contributing to the growing adoption of IoT (Internet of Things) and IoE (Internet of Everything) is the emergence of wearable devices, a category with high growth market potential. Wearable devices are commonly understood to be devices that can be worn by, or embedded in, a person, and that have the capability to connect and communicate to the network either directly through embedded wireless connectivity or through another device (primarily a smartphone) using Wi-Fi, Bluetooth, or another technology.

In order to offer interoperability in such a dynamic market, several international consortia have emerged, like the Internet Industrial Consortium (IIC), the Alliance for Internet of Things Innovation (AIOTI), the Internet of Things Architecture (IoTA), the WSO2 reference architecture for the IoT, oneM2M and OIC, to mention but a few. As these consortia focus on specific challenges, following their own specific requirements, MPEG identified the need for ensuring the interoperability among IoT systems, where MPEG focuses on multimedia content processing to enable an "Internet of Media Things".

MPEG's specific aim is to standardize the interaction commands from the user to the "Media Thing", the format of the aggregated and synchronized data sent from the Media Thing to external connected entities, as well as identify a focused list of Media Things that are "Wearables", to be considered for integration in multimedia-centric systems.

### Genomic Information Representation

The development of Next Generation Sequencing (NGS) technologies enable the usage of genomic information as everyday practice in several fields, but the growing volume of data generated becomes a serious obstacle for a wide diffusion. The lack of an appropriate

representation and efficient compression of genomic data is widely recognized as a critical element limiting its application potential. Beside compression which is at the base of any efficient processing of genomic information, there are several other requirements that the current data formats for raw and aligned data do not fulfil. ISO/TC 276 and MPEG have combined their respective expertise and missions and are jointly working to develop a new compression standard capable of providing new effective solutions for genomic information processing applications.

Technologies and use cases under consideration include:

1. The compression of genomic data generated by genomic sequencing machines.
2. The compression of aligned genomic data.
3. The definition of a data format for efficient storage, transport and access that is capable of carrying and integrating in compressed form genomic data and metadata generated during the "genomic information life cycle".

The standard is planned to be finished by early 2019.

### **Network-based Media Processing**

Network-based Media Processing ("NBMP", MPEG-I part 8) addresses advanced media processing use cases such as network stitching for virtual reality (VR) services, super-resolution for enhanced visual quality, transcoding by a mobile edge cloud, or viewport extraction for 360-degree video within the network environment such as 5G.

MPEG selected three technologies as starting points for the (i) workflow, (ii) metadata, and (iii) interfaces for static and dynamically acquired NBMP. The NBMP framework intends to leverage the Cloud/5G infrastructure and its media processing services – and could significantly improve the performance and efficiency of that infrastructure. NBMP allows service providers and end-users to describe media processing operations that are to be performed by the entities in the networks. NBMP will specify the composition of network-based media processing services out of a set of NBMP functions, and makes these NBMP services accessible through Application Programming Interfaces (APIs). NBMP will support the existing delivery methods such as streaming, file delivery, push-based progressive download, hybrid delivery, and multipath delivery within heterogeneous network environments. The specification is slated for final approval in 4Q 2019.

### **Compact Descriptors for Video Analysis (CDVA)**

MPEG has recently completed the work on Compact Descriptors for Visual Search (CDVS), which enables efficient search in large-scale image collections. While this standard is an important step forward, there are still open challenges from the large and quickly growing amount of video, for example, in the media and entertainment industry, in the automotive industry and in surveillance applications. Video is more than a collection of images, and the temporal redundancy of video as well as the spatiotemporal behaviour of objects in the video need to be taken into account.

CDVA aims at developing tools to analyse and manage video content, including search for object instances in video, categorisation of scenes and content grouping, based on compact descriptors for video, which can be efficiently matched and indexed for large-scale video collections. The ongoing work on CVDA targets search and retrieval applications, aiming to find a specific object instance in a very large video database (e.g., a specific building, a product). Applications include for example content management in media production, linking to objects in interactive media services and surveillance.

The data captured in CDVA descriptors will make it possible to include media streams and content in Big Data analyses – MPEG refers to this as "Big Media".

### **Low Complexity Video Codec Enhancements**

Despite surging demand for video, it is often difficult to deliver the high video quality that most end users expect. For instance, prime-time bandwidth availability is often insufficient for delivery of high-quality OTT video. Further, it is often difficult to upgrade video services to higher qualities without either ignoring customers with legacy video devices or creating costly duplicate services for new devices.

The objective of the Low Complexity Video Codec Enhancement activity is to develop a standard data stream that can be added to a hardware codec to provide new features such as enhanced resolution and lower encoding and decoding complexity and that is capable of being implemented using software processing with sustainable power consumption.

The objective of the Low Complexity Video Coding Enhancements activity is to develop a data stream structure defined by two component streams, a base stream decodable by a hardware decoder, and an enhancement stream suitable for software processing implementation with sustainable power consumption. The enhancement stream will provide compression capability extension to existing codecs at low encoding and decoding complexity.

The new data stream must enhance an  $N^{\text{th}}$  generation video codec in such a way that the compression performance is as close as possible to that of an  $(N+1)^{\text{th}}$  generation video codec while at the same time keeping its encoding/decoding complexity at a level similar to that of the  $N^{\text{th}}$  generation codec. It should be possible to implement an enhancement data decoder using software processing on existing devices, allowing timely deployment.

The project is codec-agnostic; the base encoder and base decoder could be AVC, HEVC, or any other codec in the market.

### **MPEG-5 Video**

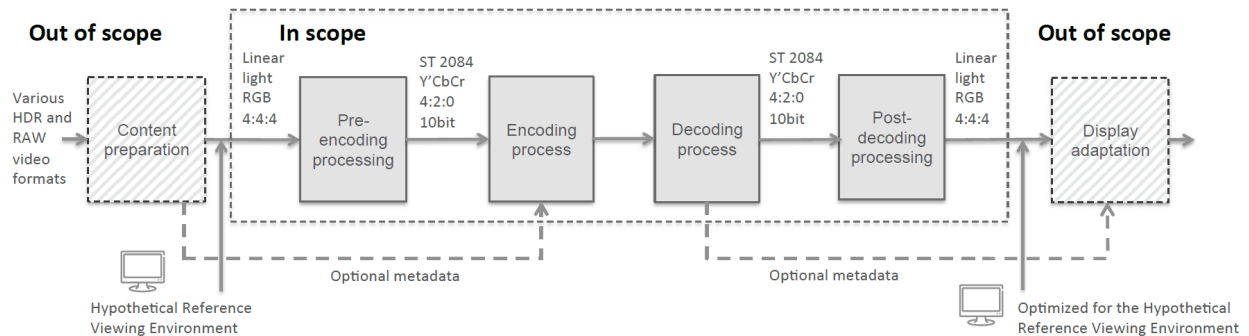
There is a constant demand for more efficient video coding technologies, but coding efficiency is not the only factor which determines the industry choice of video coding technology for products and services. In particular, it has sometimes been difficult to ensure that standardised video coding technologies are attractive options to adopt from a business perspective.

One focus of the MPEG-5 Video Coding standard is on use cases such as offline encoding for streaming VOD and live OTT streaming. The new video coding standard should provide a video compression solution which combines coding efficiency similar to that of HEVC with a level of complexity suitable for real-time encoding and decoding and the timely availability of suitable licensing terms.

## Annex C - Recently Completed Specifications

### HDR Video Coding

The HDR Technical Reports provide guidance on processing of consumer distribution high dynamic range video including conversions steps for going from a linear light RGB representation with BT.2020 colour primaries to a 10-bit, narrow range, ST 2084, 4:2:0, non-constant luminance Y'CbCr representation. The HDR System consists of four major stages; Pre-encoding processes, an Encoding process, a Decoding process, and Post-decoding processes as being shown in the figure below.



These four stages are applied sequentially with the output of one stage being used as input to the next stage according to the above-mentioned order. The primary purpose of the pre-encoding process is to convert the video input from its 4:4:4 RGB linear light, floating-point signal representation to a signal that is suitable for a video encoder. Encoding and decoding is performed in a 4:2:0, 10-bit representation. An encoder is expected to make the best use of the encoding tools available according to a particular specification, profile, and level, given also the characteristics of the content and the limitations of the intended application and implementation. The decoding process on the other hand is fully described in the respective HEVC and AVC decoder standards, where a decoder must fully comply to the intended profile and level to output precisely reconstructed video samples from a given input bitstream according to a deterministic decoding process, nominally over a time window indicated in the bitstream. The steps in the post-decoding process are aligned with what is commonly referred to as the non-constant luminance representation (NCL) in which colour conversion, to R'G'B', is performed prior to applying the transfer function to produce linear RGB.

### Internet Video Coding

From the latest paradigm shifts from analogue to digital, from one standard codec to many competing codecs, from broadcast to internet streaming, there is an increasing demand for a video codec standard that can be made, used, and sold free of charge ("Type-1" in ISO/IEC language). MPEG's Internet Video Coding (IVC) project is one of the first bottom-up approaches to design such a video coding standard in a standards body. MPEG aims to achieve a compression performance similar to the Advanced Video Coding (AVC) High Profile.

To achieve this goal, IVC employs a set of Type-1 tools that are either drawn from existing patents or published papers that are 20 years or older, or contributed by those who express their willingness to grant Type-1 licensing on their patents. Whenever MPEG becomes aware that certain tools are not Type-1, a profile without such tools will be developed.

## Common Media Application Format (CMAF)

The Common Media Application Format (CMAF) sets a clear standard for a format optimized for large scale delivery of a single encrypted, adaptable multimedia presentation to a wide range of devices. The format is compatible with a variety of adaptive streaming, broadcast, download, and storage delivery methods.

The segmented media format, which has been widely adopted for internet content delivery using DASH, Web browsers, commercial services such as Netflix and YouTube, is derived from the ISO Base Media File Format, using MPEG codecs, Common Encryption, etc. The same components have already been widely adopted and specified by many application consortia, but the absence of a common media format, or minor differences in practice, mean that slightly different media files must often be prepared for the same content. The industry greatly benefits from a common format, embodied in an MPEG standard, to improve interoperability and distribution efficiency.

CMAF defines a standard for encoding and decoding of segmented media. While CMAF defines only the media format, CMAF segments can be used in environments that support adaptive bitrate streaming using HTTP(S) and any presentation description, such as the DASH MPD, the Smooth Streaming Manifest, and the HTTP Live Streaming (HLS) Manifest (m3u8). MPEG's CMAF specification is addressing the most common use-cases and defining a few CMAF profiles that would help industry and consortia to reference this specification and avoid fragmentation of media formats.

Some of the major use cases for CMAF include OTT adaptive bitrate streaming, broadcast/multicast streaming, hybrid network streaming of live content, download of streaming files for local playback, and server-side and client-side ad insertion.

## Media Orchestration

The amount of multimedia capture and display devices is still growing fast – every phone or tablet can record and play multimedia content. Applications and services move towards more immersive experiences, and we need tools to be able to manage such devices over multiple, heterogeneous networks, to create a single experience. In other words: we need tools to coordinate media that is recorded by many devices simultaneously and that can be consumed on different devices, simultaneously. An example is a TV and a tablet that show different views of the same event, in sync. We call this process Media Orchestration: orchestrating devices, media streams and resources to create such an experience. Media orchestration:

- Applies to capture as well as consumption;
- Applies to fully offline use cases as well as network-supported use, with dynamic availability of network resources;
- Applies to real-time use as well as media created for later consumption;
- Applies to entertainment, but also communication, infotainment, education and professional services;
- Concerns temporal (synchronization) as well as spatial orchestration;
- Concerns situations with multiple sensors (“Sources”) as well as multiple rendering devices (“Sinks”), including one-to-many and many-to-one scenarios;
- Concerns situations with a single user as well as with multiple (simultaneous) users, and potentially even cases where the “user” is a machine. There is an obvious relation with the notion of “Media Internet of Things” that is also discussed in MPEG.

A first version of the Media Orchestration (MORE) specification was published in 2018.



## **Media Linking Application Format (MLAF)**

The “Media Linking Application Format” standard has been prompted by many examples of existing services where media transmitted for consumption on a primary device give hints to users to consume related media on a secondary or companion device. Interoperability of such services is facilitated if there is a data structure (a “format”) that codifies the relationship between these two media. MLAF defines a standard representation for these relationships and calls them “bridgets”, i.e. links between a source content item and one or more destination content items. This representation is based on the MPEG-21 Digital Item.

Bridgets can be products of an editorial decision, and can be the output of a workflow which involves different roles taking care of finding, organising and finally crafting the data that constitute them. In this workflow actors with different roles define bridgets from different perspectives. Authors of TV programmes will define bridgets following criteria matching the editorial intention of the programme, the main distribution channel or the target audience of the programme. At the same time marketing and commercial operators (e.g., advertisement agents, sales houses) will define bridgets following their own objectives, which may be independent from the authorial perspective. Last, but definitely not least, end users can define their own bridgets through social media interaction. All the above approaches can include not only the generation of the linking information but also of information related to how referenced content have to be presented graphically or should interact with the user.

Therefore, the Media Linking Application Format (MLAF) offers a standard format for representing and exchanging bridget-related information that fosters integration of all those systems playing a role in generating bridget information in the different and heterogeneous aforementioned domains.