123rd MPEG Ljubljana, Slovenia, 16 - 20 July 2018, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 MPEG Media Standardisation Roadmap

The following figure depicts MPEG's current thinking on its roadmap for upcoming media standards.

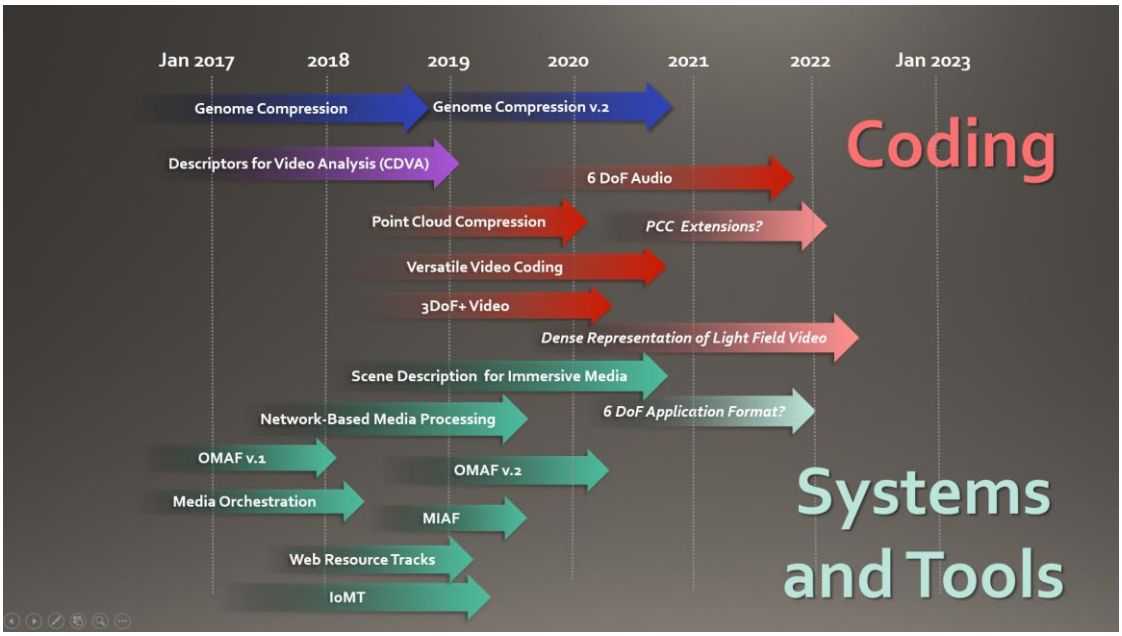

Figure 1 - MPEG Media Standardisation Roadmap.

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.

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.

A description of the current MPEG projects is following in the next section.

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.

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, light field 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. 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

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

Dense Representation of Light Field Video
To reach a high degree of realism in Virtual Reality (VR), high resolution colour images of the scene need to be captured and processed adequately, which involves image fusion and panoramic stitching. 360-degree VR, for instance, provides an immersive feeling by positioning the user in the centre of one or two (for stereoscopy) spherical/cylindrical panoramic texture(s) obtained from multi-camera stitching, out of which a particular viewport is selected from the user's head direction.

Next-generation Cinematic VR should provide an even better, authentic virtual viewing experience, fully indistinguishable from what the user would experience in the real world. Similar to the Matrix bullet effect, it should support free navigation to any position in the scene with correct motion parallax and depth cues, as well as correct eye accommodation and focus to the display, eventually reaching glasses-free 3D all-around viewing. Light Fields, describing light information at all positions and from all viewing directions of the scene, provide the necessary means to reach these goals.

Practical systems include multi-camera acquisitions (possibly including depth sensing cameras) providing sparse light fields, as well as dense light field head-mounted devices and displays. Most targeted next-generation VR effects can be handled with light field image-based approaches, without the need to explicitly 3D model the scene, therefore overcoming a cumbersome processing step for highly natural scenes to render.

To reduce transmission bandwidth requirements in light field video coding, depth-based editing and so-called depth image based rendering techniques, generating additional viewpoints from a small number of transmitted camera views, are studied towards optimal trade-offs in end-to-end system performance (acquisition, coding, rendering) and user experience (reduced latency and cyber sickness, depth-based image editing satisfaction, etc.).

Audio Wave Field Coding

MPEG has provided ground breaking technology that facilitates the delivery of audio-visual media to the user. Such media may be digital cinema or TV programs, both of which assume that the user is at a stationary point when viewing and listening to the audio-visual content. In audio, this is the "sweet spot" located directly between the stereo speakers (or the centreline for 5.1 or a greater number loudspeaker layouts).

However, MPEG is moving to support less restrictive viewing conditions, such as flexible "point of view" media in which a viewer can move to various positions in front of a viewing screen and see a realistic presentation that is true to that point of view. In a similar way, the audio signal needs to change with the varying positions that the viewer might take. This is not just "enlarging" the sweet spot, but rather changing what the user hears as his or her position changes.

Wave Field capture and synthesis is a technology that fully supports this use case, in that it can capture and reproduce the exact acoustic pressure waves at every point in a space in front of the visual display. It is a spatial audio capture and rendering technique in which an array or grid of microphones is used to capture the exact acoustic wave pattern as produced by some sound scene. When rendering, an array or grid of loudspeakers is used to reproduce the same acoustic wave pattern in the listener area as would be created from the real sound scene. Contrary to traditional reproduction techniques such as stereo or surround sound, with Wave Field techniques the localization of synthesized virtual sources remains accurate anywhere within the listening area. This technique can be used to capture a real acoustic sound scene, or to record a synthesized acoustic scene via propagation of modelled acoustic waves from a virtual source to a virtual array or grid of microphones.

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 colors, 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 should support random access to subsets of the point cloud.

Scene Description for Immersive Media

The scene graph for immersive media is a data structure that can be commonly used by vector-based graphics editing applications and modern computer games. Scene graphs are used in the industry to arrange the logical and often spatial representation of a graphical scene that may include elements like still pictures, 2D video, point clouds, light fields, audio, and geometric primitives like meshes and textures. MPEG is currently studying what interfaces need to be defined to commercially available scene description formats, and how these can benefit from MPEG-defined media representations, including where there are multiple MPEG media streams, often requiring tight timeline synchronization.

Using a scene graph enables composition and description of a scene that may consist of both natural-world (captured) and synthetic (e.g. CGI) content, with both compressed and uncompressed formats.

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

The Cloud-based media processing work comprises two activities: generic interfaces for cloud-based media-processing and interfaces for network-distributed video coding.

The Network Distributed Media Processing framework will define external interfaces for uploading and streaming media data to the network for media processing, and access the processed media in a scalable way in real-time or on-demand. It will also define internal interfaces necessary to add new media processing functions, and to compose media processing services from a pipeline of media processing functions.

Network-distributed Video Coding provides interfaces and formats for guided video coding in any kind of network This includes cloud environments, edge networks, and even home networks. The specification may also specify video coding processing, such as video transcoding.

The two specifications are both slated for final approval in Summer 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".

Media Format for Content with 6 Degrees of Freedom

In the context of MPEG immersive media services and applications, especially in context of Virtual, Extended and Augmented Reality, MPEG expects media to be rendered in complex scene environments where users can interact with the scene. MPEG envisages defining a specification for this use case, much like OMAF has been defined for 3 DoF media.

This 6DoF Application Format – and there may be more than one – will enable aggregating and packaging media in MPEG containers for storage, download, streaming and broadcast distribution. The scene may be read from a scene graph/scene description file or it may be inferred from the content (e.g. a scene with a single sphere geometry for 360 video). In the context of this 6DoF media format, the 6DoF media client may be given option to choose between a full 6DoF scene rendering, it may opt for rendering all content itself, or it may delegate part of the scene rendering to the network. In the latter case, the network may simplify the full 6DoF scene into a simplified version, “down convert” it to a 3DoF+ or 3DoF scene, or even render it as a 2D video.

Content may be composed of a wide variety of formats and types. These can either be 2D or 3D, natural or synthetic, compressed or uncompressed, provided by the content provider or captured locally (e.g. in the case of AR) or remotely. Key aspects to be defined in the context of this format include:

- the definition of the spatial environment in which the scene can be consumed,
- information on how to position and render media sources in 6DoF,
- the interaction with the scene based on sensor and/or use input,
- and management of the different timelines in such a scene.

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 were 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.

Output Documents
N17739 - MPEG Standardization Roadmap

2 MPEG Non-Media Standardisation Roadmap

MPEG has been the main contributor to the stellar performance of the digital media industries and MPEG has started applying its compression toolkit to other non-media data such as point clouds and DNA reads from high speed sequencing machines and will soon apply it to neural networks. By proposing to create a new ISO Technical Committee on Data Compression Technologies, seeks to extend the successful MPEG model to the "Data industry" at large, including the media industries.

Handling data is important for all industries: in some cases it is the raison d'être, in other cases it is crucial to achieve the goal and in still others it is like the oil that lubricates the gears.

Data appear in many and various forms: in some cases a few sources create huge amounts of continuous data, in other cases many sources create large amounts of data and in still others a very large number of sources create small discontinuous chunks of data.

Common to all is the need to store, process and transmit data. For some industries that engaged early in digitisation, the need was apparent from the very beginning. For other the need is gradually becoming clear now.

In the following is given a concise non-exhaustive analysis of the need of data compression by some of the industries.

Telecommunication
Because of the nature of their business, telecommunication operators (telcos) have been the first to be affected by the need to reduce the size of digital data in order to ne able to provide better existing services and/or attractive new services. Today telcos are eager to make their data available to new sources of data.

Broadcasting
Because of the constraints posed by finite wireless spectrum on their ability to expand their range of services, broadcasters have always welcome more data compression. They have moved from Standard Definition to High Definition then to Ultra HD and beyond ("8k"), but also to Virtual Reality. More engaging user experiences require the ability to transmit or receive ever more data, and ever more types of data.

Public security
MPEG standards are already universally used to capture audio and video information for security purposes. However, technology progress enables the embedding of more capabilities in (audio and video) sensors, e.g. face recognition, people and vehicle counting etc., and the sharing of that information in a network of sensors to drive actuators. New standard data compression technologies are needed to support the evolution of this domain.

Big Data
In terms of data volume, audio and video, e.g. those collected by security devices or vehicles, are probably the largest component of Big Data, as shown by the Cisco study forecasting that by 2021 video on the internet will account for more than 80% of total traffic. Moving such large amounts of information from source to the processing cloud in an economic fashion requires data compression and their processing requires standards that allow the data to be processed independently of the information source.

Artificial intelligence
Neural networks of different types are typically used in artificial intelligence. Complex application domains usually require large neural networks, where large may mean many Gigabytes of data and require massive computational complexity. To practically move intelligence across networks as required by many consumer and professional use scenarios, standard data compression technologies are required. Compression of neural networks is not only a matter of bandwidth and storage memory, but also of power consumption, timeliness and usability of intelligence.

Healthcare
The healthcare world is already using genomics but many areas will greatly benefit by a 100-fold reduction of the size and the time to access the data of interest. In particular, it will accelerate the coming-of-age of personalised medicine. As healthcare is often a public concern, standard data compression technologies are required.

Agriculture and Food
Almost anything related to agriculture and food has a genomic source. The ability to easily process genomic data thanks to compression opens enormous possibilities to have better agriculture and better food. To make sure that compressed data can be exchanged across users, standardised data compression technologies are required.

Automotive
Vehicles are becoming more and more intelligent devices that drive and control their movement by communicating with other vehicles and fixed entities, sensing the environment, and storing the data for future use (e.g., for assessing responsibilities in a car crash). Data compression technologies are required and, when security is involved, the technologies must be standard.

Industry 4.0
The 4th industrial revolution is characterised by "Connection between physical and digital systems, complex analysis through Big Data and real-time adaptation". Collaborative robots and 3D printing, the latter also for consumer applications, are main components of Industry 4.0. Again, data compression technologies are vital to make Industry 4.0 fly. To support multi-stakeholder scenarios, technologies should be standard.

Business documents
Business documents are becoming more diverse and include different types of media. Storage and transmission of business documents are a concern if bulky data are part of them. Standards data compression technologies are the principal way to reduce the size of business documents now and more so in the future.

Geographic information
Personal devices consume more and more geographic information and, to provide more engaging user experiences. the information itself is becoming "richer", typically meaning "heavier". To manage the amount of data there is a need to apply data compression technologies. Global deployment to consumers requires that the technologies be standard.

Blockchains
Blockchains and distributed ledgers enable a host of new applications. Distributed storage of information implies that more information is distributed and stored across the network, hence the need for data compression technologies. These new global distributed scenarios require that the technologies be standard.

Output/Input Documents
N17893 - Initial identification of non-media data compression areas
M42881 - A future for MPEG

3 MPEG-I Audio Architecture

The MPEG-I Audio Architecture uses the MPEG-H 3D Audio LC Profile for carriage of coded audio data and 3D audio metadata. Additional 6DoF metadata will be developed together with a 6DoF Audio Renderer for rendering the audio content in a 6DoF environment. A draft MPEG-I Audio architecture is provided in Figure 2.

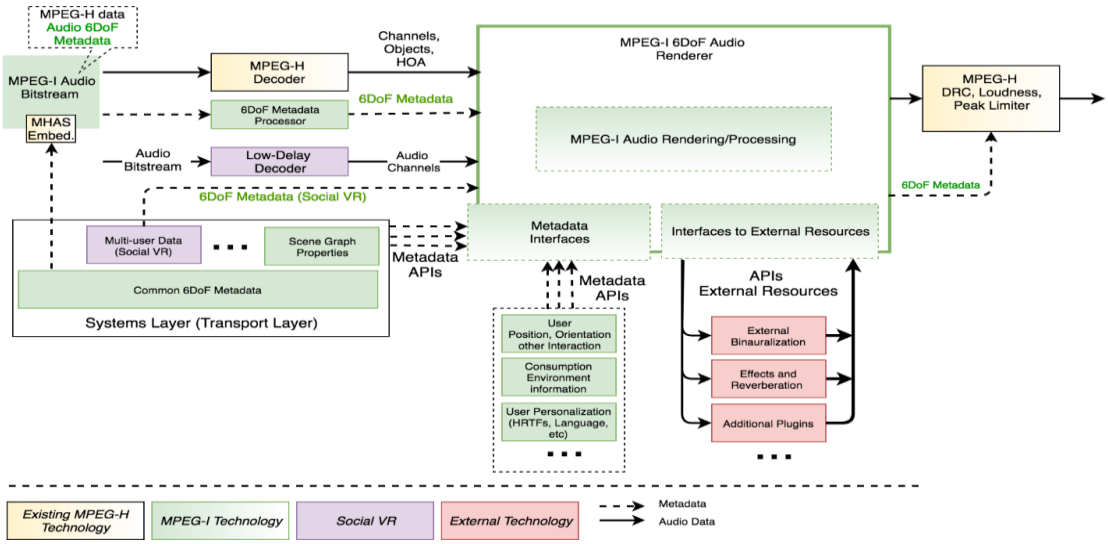

Figure 2 - MPEG-I Audio Reference architecture

Figure 2 provides an overview of the various signal paths for audio data encoded using MPEG-H 3D Audio as well as for a communication path for audio data encoded using a low-delay audio codec (e.g., AAC-ELD). Additionally, various interfaces to the MPEG-I 6DoF Audio Renderer are provided:

- interface for 6DoF metadata delivered in MHAS bitstream
- interface for Scene Graph metadata (if any)
- interface for common 6DoF metadata delivered on system level and shared with other media components (e.g., video and visual)
- interface for User Position, Interactivity, Orientation, Interaction
- interface for Environment information (e.g., room sizes and acoustic properties)
- interface for User Personalization Elements (e.g., HRTFs)
- Interface for external Resources:
    - External Binauralization
    - Effects and Reverberation
    - Additional Plugins

Reusing the metadata delivery concepts from MPEG-H Audio delivery of 6DoF metadata could be done based on the use case either in:

- **The MHAS bitstream**, similar to MPEG-H 3D Audio metadata. This metadata is very important for AR application for example, and depending on the application it might be desirable to have a single bitstream input to the final playback application.
- **The Transport Layer** (e.g., new ISO BMFF boxes) and delivered to the MPEG-I Audio Renderer trough the Metadata interface.
- **The Transport Layer** (e.g., new ISO BMFF boxes) and embedded into the MHAS stream "on the fly" using the MHAS flexibility for new MHAS packets insertion.

Output Documents
N17847 - MPEG-I Audio Architecture
N17848 - Draft MPEG-I Audio Requirements
N17849 - Workplan on MPEG-I Audio

4 Network-based Media Processing

The NBMP framework will define external interfaces between the Media Source/Media Sink and the Media Processing Entity, that will allow users of the framework to access the framework, configure the media processing, upload/stream media data to the network for media processing, and access the processed media and the resulting metadata in a scalable fashion in real-time or in a deferred way.

The media and metadata formats that are used between Media Processing Entities in a media processing pipeline are also within scope. Workflow management that is used to instantiate new media processing entities and to compose media processing services into a pipeline of media processing entities is also in scope.

The following diagram depicts the NBMP architecture that will be used as a reference architecture to scope the NBMP work:

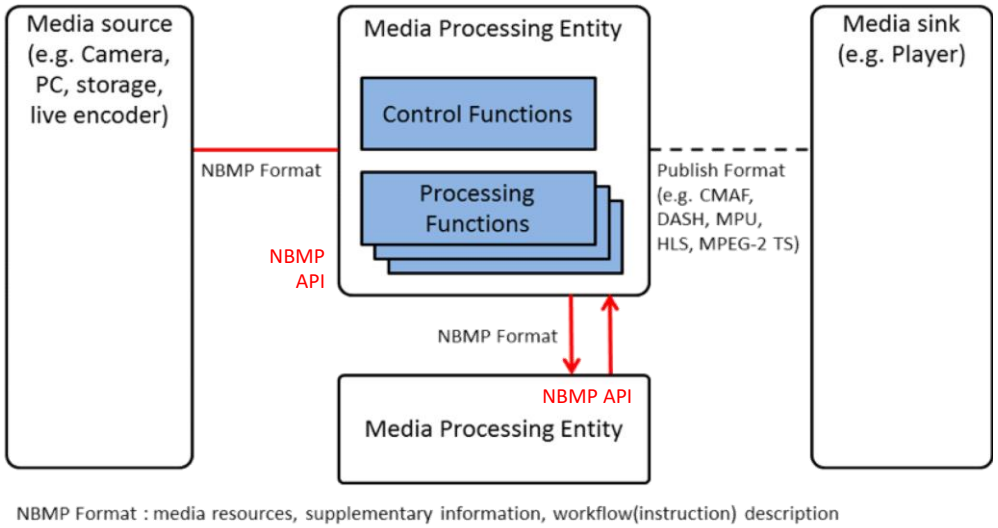

Figure 3 - NBMP Reference Architecture

Network-Based Media Processing (NBMP) is a framework that allows service providers and end users to describe media processing operations that are to be performed by the network-based media processing entities. NBMP describes the composition of network-based media processing services out of a set of network-based media processing functions and makes these network-based media processing services accessible through the NBMP Application Programming Interfaces (APIs). NBMP defines the interfaces that enable the description and execution of media processing functions. NBMP workflow describes the procedure of media processing.

An NBMP media processing entity performs media processing tasks on the input media data and the related metadata. NBMP also provides control functions that are used to compose and configure the media processing functions.

NBMP systems support any type of media content, including the existing MPEG codecs and MPEG formats such as ISO/IEC 13818-1, ISO/IEC 14496-12, ISO/IEC 23008-1 and ISO/IEC 23009-1.

NBMP systems support the delivery over IP-based networks using the different transport protocols. (e.g., TCP, UDP, RTP and HTTP).

NBMP systems support the existing delivery methods such as streaming, file delivery, push-based progressive download, hybrid delivery, multipath and heterogeneous network environments.

NBMP summary of proposals

Proponents	**Summary of proposals**	**Use cases**
m42931	• Focus on the workflow, processing models, task interfaces and APIs • Describes the workflow description format, resource definitions, data plane	360 stitching, 6DoF pre-rendering
m42932	• Focus on the Representation format, scheme, workflow and processing functions • Describes on the Live uplink/downlink data binding and registration API	360 stitching

m42933	• Focus on the Metadata category, new metadata (optical, coordinate alignment, preserved region)	360 stitching
m42936	• Focus on the Conversion instructions, fMP4 metadata, content URL, ingest format, storage access, Live API • Describes the playlist format (e.g. SMIL), profiling of media processing functions	Live streaming, media ingest, temporal content stitching
m43418	• Focus on the video content dependent metadata for 2 type of guided transcoding • Describes the metadata format	Guided transcoding
m43485 & m43486	• Focus on the control method for partial video stitching, QoE-based media processing functions (e.g., upscaling, cropping) • Describes the control method format	stitching, QoE-based upscaling
m43599	• Focus on the abstract model for workflow description based FNL and RVC-CAL • Extension of FNL language deal with dynamical adaption	Generic
m43792	• Focus on the usage of media distribution with loss recovery and SR for upscaling • Describe the list of parameters for NBMP processing	Loss recovery, SR

Output Documents
N17900 - Evaluation Report for CfP on Network-Based Media Processing (NBMP)
N17872 - WD of ISO/IEC 23090-8 Network-based Media Processing
N17874 - Description of Core Experiments on Network-based Media Processing