116th MPEG Chengdu, China, 17 - 21 October 2016, Meeting Report

Panos Kudumakis and Krishna Chandramouli
qMedia, Queen Mary University of London

Contents

1 MPEG's Five Year Roadmap

MPEG is building its future standardisation roadmap now, and is advertising its plans so that industry can 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 would be 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?

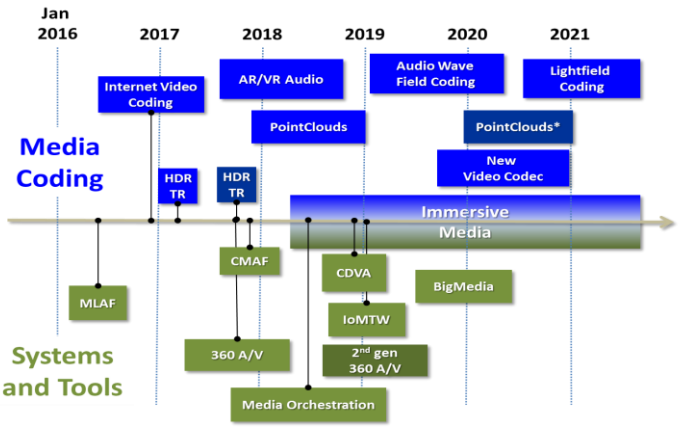

Figure 1 - MPEG's Five Year Standardisation Roadmap.

MPEG is organising a series of short, high level "Industry meets MPEG" workshops around the world, to collect market feedback. The first one of these was held in Chengdu, China on 19 October, 2016. The next one will take place in Geneva on 18 January 2017. Another workshop is planned for the week of February 27, 2017, to take place at the Mobile World Congress in Barcelona, Spain.

The first "Industry meets MPEG" workshop was held on 19 October 2016 in Chengdu, China. The workshop was organized by MPEG delegates, and co-located with the MPEG Meeting in Chengdu in the same week. The program of the workshop was as below. The numbers in the first column refer to the numbers of the presentation on the website. The presentations can be found on the website.

1.	*Opening Address*	Leonardo Chiariglione, Convenor, MPEG
2.	*MP20 Roadmap*	Rob Koenen; José Roberto Alvarez, MPEG
3.	*AVS Standards and Test*	Dr. Wenhui Dong, SARFT
4.	*Current Status and Prospect of K-UHD with New Media Services*	Dr. Chieteuk Ahn, ETRI
	Break	
5.	*8K broadcasting toward Tokyo Olympics*	Mr. Yuji Yamana, NHK
6.	*Future of Video Evolution*	Mr. Toshiyuki Ogura, Sony Corporation
7.	*Future Media Network in the 5G Era*	Prof. Wenjun Zhang, Shanghai Jiao Tong University
	Break	
8.	*Emerging Markets, Specialized Standards*	Mr. Wenfei Jiang, Alibaba Cloud
9.	*Virtual Reality—A new way of entertainment.*	Mr. Wenbo Sun, LetinVR Digital Technology Company
10.	*Virtual Reality: The New Era of the Future World*	Mr. Raymond Pao, HTC
	Panel Discussion	

Output Documents
N16545 - MP20 Roadmap
N16546 - Presentations of the MP20 Workshop on MPEG Roadmap

2 Call for Proposals on Internet of Media-Things and Wearables (IoMTW)

Today, "Internet of Things" (IoT) is an idiom currently encompassing a large variety of research, development and market efforts related to the communication between smart objects. While such an definition is, by its very nature, quite fuzzy, the market reality is very clear: the number of devices connected to the Internet will reach to 50 billion in 2020 and, together with the data they are producing, they will result in the first source of wealth in the world, at least for the IT industry.

An important factor contributing to the growing adoption of IoT (Internet of Things) is the emergence of wearable devices, a category with high growth market potential. Wearable devices, as understood commonly, are devices that can be worn by or embedded in a person and have the capability to connect and communicate to the network either directly through embedded wireless connectivity or through another device (e.g. smartphone) using Wi-Fi, Bluetooth, or another technology.

After a preliminary study, MPEG decided to align the standardization activities related to IoT and wearable devices inside a unique effort, referred to as *Internet of Media-Things and Wearables (IoMT&W)*.

At the 116th meeting, MPEG produced final documents of Use Cases (N16533) and Requirements (N16534) and issued a Call for Proposals on Internet of Media-Things and Wearables (N16535).

Use cases updated, addressing the following areas:
Smart spaces: Monitoring and control with network of audio-video cameras

- Face recognition to evoke sensorial actuations
- Human tracking with multiple network cameras
- Networked digital signs for customized advertisement
- Intelligent firefighting with IP surveillance cameras
- Automatic video clip generation by detecting event information
- Self-adaptive quality of experience for multimedia applications
- Ultra wide viewing video composition

Smart spaces: Navigation

- Blind person assistant system
- Personalized navigation by visual communication
- Personalized tourist navigation with natural language functionalities
- Smart identifier: Face Recognition on Smart Glasses
- Smart advertisement: QR code recognition on smart glasses

Smart environments in smart cities

- Smart factory: Car maintenance assistance A/V system using smart glasses
- Smart museum: Augmented visit museum using smart glasses
- Smart house: Light control and vibrating subtitle
- Smart car: Head-light adjustment and speed monitoring

Smart collaborative health

- Increasing patient autonomy by remote control of left-ventricular assist devices
- Diabetic coma prevention by monitoring networks of in-body / near body sensors
- Enhanced physical activity with smart fabrics networks

Output Documents
N16533 - Use cases for Internet of Media-Things and Wearables
N16534 - Requirements on Internet of Media-Things and Wearables
N16535 - Call for Proposals on Internet of Media-Things and Wearables
N16544 - Liaison on IOMT&W to IEC TC100

3 Common Media Application Format (CMAF)

Common Media Application Format (ISO/IEC 23000-19) is optimized for large scale delivery of a single encrypted, adaptable multimedia presentation to a wide range of devices; compatible with a variety of adaptive streaming, broadcast, download, and storage delivery methods.

Several MPEG technologies have been adopted for much of the video delivered over the Internet and other IP networks (cellular, cable, broadcast, etc.). Various organizations have taken MPEG's core coding, file format and system standards, and combined them into their own specifications for their specific applications. While these specifications share major common parts, their differences result in both unnecessary duplication of engineering effort, and duplication of identical content in slightly different formats that increases storage and delivery costs. The industry would benefit if application consortia could reference a single MPEG specification (a "common media format") that would allow a single media encoding to use across many applications and devices.

CMAF reached DIS stage at the 116th MPEG meeting. This result achieved in a very short time due to high interest by the industry for adopting CMAF albeit a few conflicts arose in the process.

CMAF Hypothetical Application Model and Media Object Model
CMAF defines a Hypothetical Application Model so that encoding to that model results in consistent CMAF Track encoding, representation in manifests, track selection, late binding, synchronization, decoding, and rendering of CMAF Presentations.

Decoding requirements can be inferred from encoding constraints and the Hypothetical Application Model, but are not directly specified by CMAF. CMAF does not specify manifest formats, or associated Resource identification and transport. However, CMAF does specify CMAF Addressable Media Objects derived from encoded CMAF Fragments, which can be referenced as Resources by a Manifest. External specifications can define how a manifest describes a CMAF Presentation, including identifying CMAF Addressable Media Objects as Resources, and representing their logical relationships equivalent to CMAF Tracks, CMAF Switching Sets, CMAF Selection Sets, and CMAF Presentations.

Figure 2 illustrates the media objects that are specified by CMAF, starting with the encoded CMAF Fragments that form CMAF Tracks, then logical CMAF Track Sets determined by CMAF Track encoding constraints, and derived CMAF Addressable Media Objects that can package encoded CMAF Fragments or their Samples for storage and delivery.

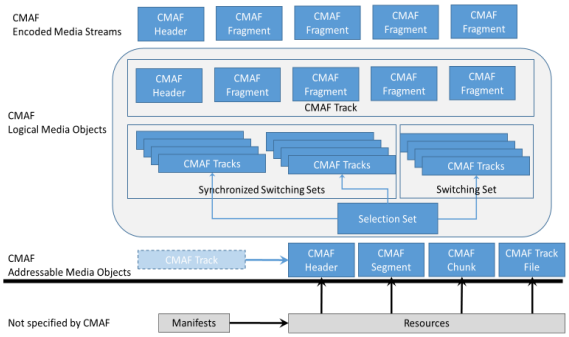

Figure 2 - Media Objects specified in CMAF, and presented by externally specified applications, such as adaptive streaming

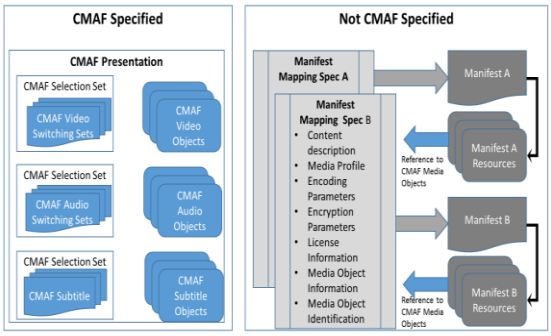

Figure 3 - CMAF Hypothetical Application Model using externally defined Manifests that describe the CMAF Presentation and Media Objects

Figure 3 illustrates the mapping between CMAF specified CMAF Presentations, and externally specified Manifests and Resources. Multiple Manifests may reference the same CMAF Presentation and CMAF Addressable Media Objects.

To accurately represent a CMAF Presentation, a Manifest must describe CMAF Track relationships determined by each Track's source content and CMAF Track encoding constraints. CMAF groups CMAF Tracks based on their encoding constraints in logical Media Objects called Selection Sets and Switching Sets that also determine intended use in late binding, track selection, seamless switching, and synchronization. Additional CMAF Track metadata such as CMAF Media Profile brands, "codecs" parameters, language fields, etc. can be included in Manifests to enable track selection and playback.

Manifests can reference CMAF Addressable Media Objects by Resource Identifiers used by Manifests and servers to select the identified CMAF Addressable Media Objects for delivery and playback. Multiple CMAF Addressable Media Object types are specified for different delivery use cases, such as pre-recorded content that is download or streamed as files, adaptive live and on demand streaming over Internet with efficient CMAF Segment sizes, or low latency live streaming of short duration CMAF Chunks, e.g. over broadcast and managed networks.

Output Documents
N16435 - DoC on ISO/IEC CD 23000-19 Common Media Application Format
N16436 - Text of ISO/IEC DIS 23000-19 Common Media Application Format
N16437 - Technologies under Consideration for ISO/IEC 23000-19 AMD 1
N16438 - Workplan for CMAF conformance

4 Explorations - MPEG vision, objectives, and plan for Big Media

In the current world, a huge amount of data is generated by audiovisual sources or has a multimedia nature. However current audiovisual data are not incorporated in the Big Data standardization paradigm. The objective of the Big Media activity within MPEG is to develop the required standards specific to this kind of data in order to make them exploitable and (re)usable for different application use cases.

Currently, ISO/IEC JTC 1 WG 9 on Big Data is working on the definition of Big Data Reference Architecture (BDRA). Within this framework, it has already been identified a number of existing MPEG standards that can handle different parts/aspects of this Reference Architecture.

As ISO/IEC JTC 1 WG 9 BDRA is still in early development phase (WD 1.0), this analysis is based on the NIST Big Data Reference Architecture (NBDRA) [1]. An important part of this reference architecture is the Big Data Application Provider that executes a set of operations along the data lifecycle to meet the requirements established by the System Orchestration. While a full description can be found in [1], a brief description it follows in relationship with existing MPEG technologies.

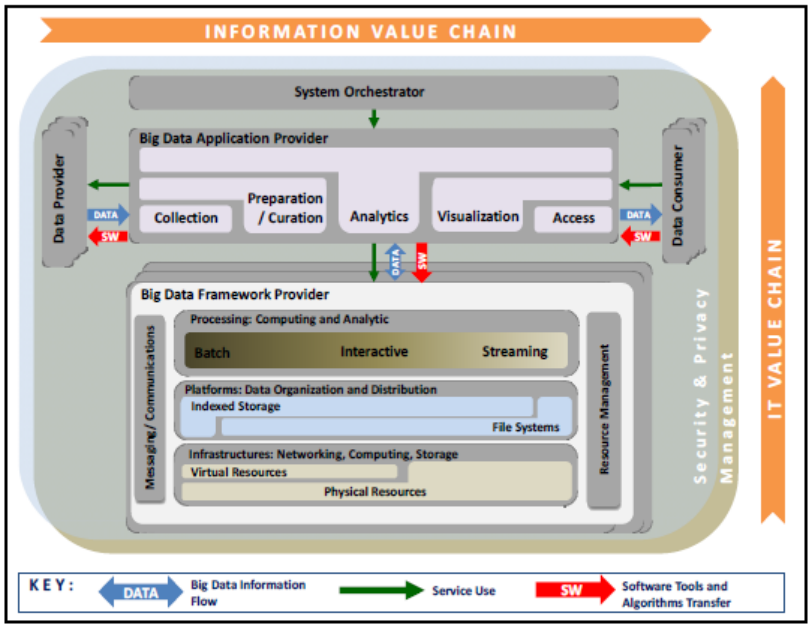

Figure 4 - NIST Big Data Reference Architecture (NBDRA), [1].

Collection activity handles the interfaces with Data Provider. At this level, MPEG has not only technologies to represent audio and video captured by sensors and prepared for transmission but also a large variety of signals captured by many other types of sensor (temperature, pressure, noise …), like in MPEG-V. MPEG offers also technologies for media transport like DASH and MMT.

Curation activity aims mainly to transform data to prepare it for the analytics. This can include cleansing but also other optimization manipulation as preparatory phase for analytics. At this level, MPEG developed a very rich set of descriptors that can be extracted from audio-visual data. MPEG-7 in general, and more recently the works related to Compact Descriptors for Visual Search (CDVS) and Compact Descriptors for Visual Analysis (CDVA), are good examples for that. Analytics activity is about the techniques to extract knowledge from the data based on the requirements of the vertical application. MPEG-7 descriptors can be used to prepare the analysis space.

Visualization activity prepares elements of the processed data and the output of the analytics activity for presentation to the Data Consumer. At this level, MPEG also defined tools to represent rich media scenes, such as BIFS, enabling interactive navigation in 2D/3D complex environments and such tools can be used to visualize the results of Big Data analytics component.

At the System Orchestration level, MPEG has defined the MPEG-M set of specifications with the objective to provide APIs in order to access MPEG components. These APIs can be used in the Big Data context since there is no restriction related to the place where the data is located or how it is transmitted. In general, the APIs ensure access in load/save modes.

Currently MPEG is conducting standardization work in the field of IoT by addressing media related Things, therefore facilitating the integration of such Things in Big Media applications. This new activity is intended to consider collection, processing/analysis, and visualization of media data acquired by, or used to control IoT entities. Therefore this covers key activities within the Big Data Application Provider in the NBDRA (Figure 4). However, the IoMT framework is not covering activities such as advanced analytics.

Current status

MPEG currently has started a work on analyzing the need in terms of Big Media standards. The following non-exhaustive set of use cases has been collected and analyzed with regards to the existing MPEG standards:

- Use case 1: Media recommendation based on n-dimensional usage information
- Use case 2: Media-centric IoTs and Big Media for smart cities
- Use case 3: Big Data for personalized user experience in Augmented Reality
- Use case 4: Big Data platform for 3D reconstruction and diffusion

The full description of these use cases with the corresponding analysis can found in the document [2].

Objectives
Following the analysis conducted and the current vision on the Big Data Reference Architecture (Figure 4), the following objectives are considered by MPEG for future standardization activities:

- MPEG work on the definition of standardized interfaces (APIs) between the main functional architecture components of the Big Data Application Provider in the NIST model. APIs between the Analytics component and the Collection, Curation, Visualization, and Access could be standardized, and also between these components and the System Orchestration.
- Develop functionalities and technologies related to Media for the System Orchestration.

MPEG is also currently developing the IoMT&W (Internet of Media Things and Wearables) conceptual model. The link between IoMT&W and Big Media can be considered at the level of advanced processing of signals obtained from IoMT devices when they are used in large scale systems. By adding such link IoMT systems would be able to expose more relevant semantic information to the external world. An example of potential activity of Big Media will be to define interfaces on how this information is exposed.

While MPEG can provide technology expertise to develop any necessary industry standard, such work can be successful only if it is driven collaboratively with other organizations and groups, such as ISO/IEC JTC 1 WG 9.

References
[1] NIST Special Publication 1500-6, "NIST Big Data Interoperability Framework: Volume 6, Reference Architecture", September 2015.
[2] ISO/IEC JTC 1/SC 29/WG 11 document N16353, "Thoughts and Use cases on Big Media", June 2016, Geneva, Switzerland.

Output Documents
N16540 - MPEG Vision, Objectives, and Plan for Big Media

5 Explorations - Thoughts on Augmented Reality/Virtual Reality and 3D Audio

Some first thoughts of Audio experts captured in a document on what audio technology might be needed to address Augmented Reality/Virtual Reality use cases. Providing Audio standards in support of Augmented Reality and Virtual Reality applications, services and systems will occur in stages:

- A first stage is to standardize technology to support 360-degree movies. This application permits *3 degrees* (x, y, z) of freedom by the user, and the required audio technology has already been standardized in MPEG-H 3D Audio.

- A next stage is to extend MPEG-H 3D Audio to *6 degrees* (incl. pitch, yaw, roll sensors detected orientation) of freedom. Technologies to explore may be point clouds that contain environmental meta-data (e.g. pertaining to acoustic characteristics) and compressed audio signals (e.g. ambient audio scenes or local audio sound sources).

- A more forward-looking stage is to investigate whether capture, compressed representation and reproduction of audio *sound fields* is a means to a more realistic user experience of virtual audio presentations.

Output Documents
N16385 - Thoughts on AR/VR and 3D Audio

6 Explorations - Coded Representation of Immersive Media

The VR Survey resulted in the following understanding of timing of deployment of VR 360 services:

- Commercial Trials: 2016 and 2017, then levelling off
- Initial Commercial Launch: 2017/2018
- Mainstream: 2018 to 2020

The Requirements Group concluded that a first set of specifications is required in time for a market launch of products and services in 2018. This is aligned with the timeframe for OMAF, but it is highly unlikely that MPEG can deliver solutions that are optimised in this time frame. Experiments leading to standardisation of optimised solutions will take longer. Since a large constituency believes that major market launch will be in 2020, MPEG can deliver a next set of specifications – or on update of the first OMAF spec – in 2019. At the same time it is clear that there is a strong need for longer term work, notably in the video area, but possibly also in the Audio space, on 6-degrees-of-freedom content.

Given these results, the Requirements Group currently assumes that standards in support of Immersive Media, including those for 360º Audiovisual Media, will be developed in the following phases:

Phase 1a

- Timing is what guides this phase
- Goal: to deliver a standard for 3DoF 360 VR in the given timeframe (end 2017 or maybe early 2018)
- Aim for a complete distribution system
- Based on OMAF activity; using OMAF timelines;
- Audio: a 3D Audio profile of MPEG-H geared to a 360 Audiovisual experience with 3 DoF,
- Transport: Basic 360 streaming, and if possible optimizations (e.g., Tiled Streaming)
- Video: Adequate tiling support in HEVC (may already exist) and projection, monoscopic and stereoscopic
- MPEG should be careful not to call this MPEG VR, as the quality that can be delivered in the given timeframe may not be enough.

Phase 1b

- Mainly motivated by desire by a significant part of respondents to launch commercial services in 2020
- deploy in 2020; spec ready in 2019, (which may match 5G deployments)
- Extension of 1a; focus very likely still on VR 360 with 3 DoF (again monoscopic and stereoscopic)
- If there are elements that could not be included in phase 1a, improving quality – it is not a foregone conclusion that there will be a phase 1b, and if there is such a phase, it is to be further defined what this would comprise
    - E.g., optimization in projection mapping
    - E.g., further motion-to-photon delay reductions
    - Optimizations for person-to-person communications
- Phase 1b should have some quality definition and verification

Phase 2

- A specification that is ready in 2021 or maybe 2022
- This would be a "native" VR spec ("MPEG VR")
- Goal is support for 6 DoF
- Most important element probably new video codec with support for 6 DoF; to be decided by Video Group what tools are most suitable
- Audio support for 6 degrees of freedom
- Systems elements perhaps required too in support of 6 DoF, as well as 3D graphics.

It is the Requirements Group's understanding that these phases will all be accommodated in the newly proposed Work Item for *Coded Representation of Immersive Media*.

The project is currently thought to comprise at least the following parts:

- Part 1: Technical Report on Immersive Media
- Part 2: Application Format for Omnidirectional Media
- Part 3: Immersive Video
- Part 4: Immersive Audio
- Part 5: Point Cloud Compression

Output Documents
N16541 - New Work Item Proposal on Coded Representation of Immersive Media
N16542 - Summary of the "Survey on Virtual Reality"
N16543 - Planned Phases for Standardization of Immersive Media
N16562 - Planned work in the area of coded representation of immersive Media

7 Media Orchestration (MORE)

With the abundance of capture and display devices, and with applications and services moving towards a more immersive experience, we need the tools to be able to manage multiple, heterogeneous devices over multiple, heterogeneous networks, to create a single experience. 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, although this is not yet represented in the use cases. This may have a relation with the notion of "Media Internet of Things" that is also discussed in MPEG.

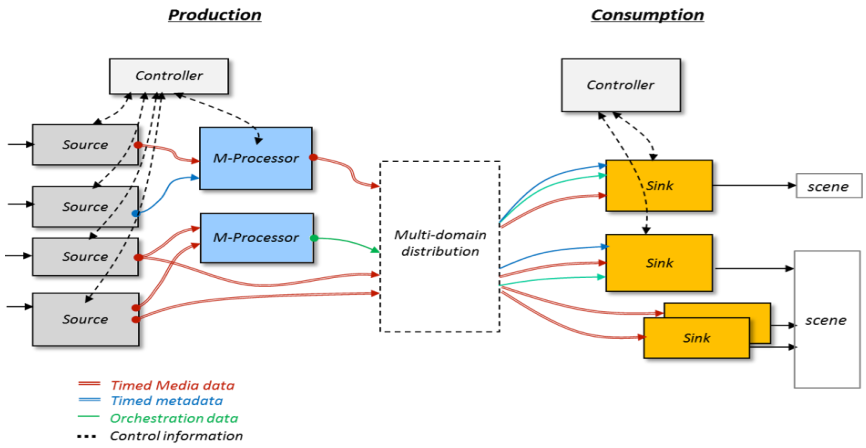

Figure 5 - Media Orchestration Model.

The Media Orchestration Model, as in Figure 5, shows a complex scenario where many sources contribute to an experience with many, orchestrated sinks. This is broadcast-oriented example, with a clear distinction between the production side and the consumption side. The Media Orchestration Model also supports mixed cases, for example shared experiences, where sinks are used for broadcast media consumption in orchestration with sources and sinks used for communication between users.

MPEG issued a second Call for Proposals (CfP) on Media Orchestration at its 115th meeting. The CfP resulted in 12 responses at the 116th meeting, which were reviewed and evaluated. The conclusions of the meeting are summarised, as:

- Adding technology from a number of proposals to the *Working Draft* and the *Technologies Under Consideration*.
- Developing an *Object Model* for Media Orchestration that includes a text-based approach, the exact format to be selected (JSON or XML).

- Determining which *categories of metadata* MORE will focus on, and to consider CEs that can be used to evaluate proposed solutions and existing standards in MPEG.
- Developing a Section or Annex that specifies '*Source Sync*' and to keep discussing with DVB how to best harmonize this with the DVB CSS specification.
- Checking its mapping onto *MPEG-2* with individual MPEG-2 experts.
- Recommends that the MORE activity meets with *CICP* experts and with *quality metrics experts* to harmonize its approach with other activities in MPEG, such as *MPEG-B parts 8 and 10.*

qMedia::MMV contributed the following data set:

m39214	MORE	LASIE Dataset for Media Orchestration	Krishna Chandramouli, Panos Kudumakis, Ebroul Izquierdo	

- The contribution offers a test set of content to those willing to sign an agreement. The data set can be used for experiments with orchestration of surveillance media data.
- Times and locations are available with the data, but not orientation of the Sources.
- The contribution was for information for those that want to have a data set, and does not propose technologies for the specification.

The Systems subgroup thanks the European FP7 LASIE project for making its Surveillance Dataset available for use in developing Media Orchestration technologies. Interested parties can find details on obtaining a license to this dataset in contribution m39214.

Output Documents
N16443 - WD of ISO/IEC 23001-13 Media Orchestration
N16444 - Technologies under Consideration for ISO/IEC 23001-13 Media Orchestration

8 MPEG-21 Media Value Chain Ontology (MVCO) - Audio Extensions

Widespread adoption of multi-track formats such as the MPEG-A: Interactive Music Application Format (ISO/IEC 23000-12:2010) raises the issue of rights monitoring for fair and transparent royalties payment with respect to reusable tracks or even segments of them in derivative new works. This MVCO Amendment 1 for IP Entities in the audio domain addresses this issue by facilitating complex matrix based rights monitoring on time vs tracks throughout the media value chain. It defines concepts for the representation of time segments and tracks of multi-track audio IP entities. Segments and tracks may contain IP entities that can be treated as conventional IP entities as defined in MVCO. The introduction of an additional Action 'ReuseIPEntity' in MVCO enables granting permissions for the reuse of existing IP Entities in order to create new derivative composite IP Entities.

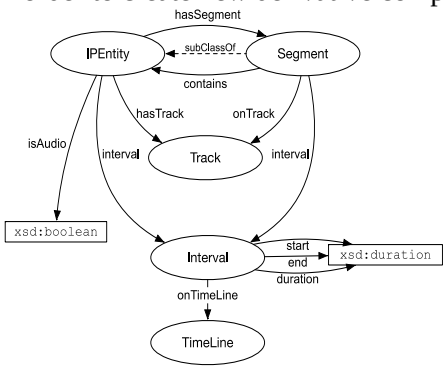

Figure 6 - Classes and relationships for the representation of IPEntity that contain other existing IP Entities. Segments may also be associated with individual Tracks of a Multi-track IP Entity.

With qMedia::C4DM proposed MVCO Audio Extensions, it is now possible to query for information about user collectives and the components of the composite IP Entities.

List members of a user collective:

```
$ java -jar rvdac.jar -r -lcu Performers
RVD Administration Console
Guitarist
Vocalist
```

List components of a composite IP Entity (including locations specified by segment and tracks where applicable):

```
$ java -jar rvdac.jar -r -lic MusicInstance
RVD Administration Console
LyricsInstance | segment: 30s to 150s | track: 2
GuitarInstance | track: 1
```

Further information on this particular use case can be found in the MVCO - Audio Extensions guidelines document. At the 116th MPEG meeting MVCO - Audio Extensions reached DAM status with 16 countries voted in favor and 10 abstentions. It is expected to reach FDAM in April 2017 and full AMD status in June 2017.

Output documents
N16432 - Text of ISO/IEC 21000-19:2010 DAM 1 Extensions on Time Segments and Multi-Track Audio

9 Dr Panos Kudumakis awarded by ISO/IEC for MPEG-M: Architecture

The MPEG-M technology specifications (ISO/IEC 23006) have reached the status of International Standard at the Incheon (KR), 104th MPEG meeting. MPEG-M specifies the means to access individual MPEG tools through standardized APIs and is expected to help the creation of a global market of MPEG-M applications that can run on devices supporting MPEG-M APIs in addition to the other MPEG technologies. The MPEG-M standard should also help the deployment of innovative business models because it will enable the easy design and implementation of media-handling value chains. The standard also provides reference software as open source with a business friendly license.

MPEG-M developed through a number of Calls for Proposals for technologies (i.e., Interoperable DRM Platform, Open Media Market, MPEG Extensible Middleware, Advanced IPTV Terminal, Open Connected TV & Multimedia Service Platform Technologies) under the auspices of Digital Media Project (2003-2014).

Last but not least, at the Chengdu (CN), 116th MPEG meeting, Dr Panos Kudumakis presented on behalf of ISO/IEC with a *'Certificate of Appreciation*' as project editor in the development of the international standard ISO/IEC 23006-1, Information Technology - Multimedia Service Platform Technologies - Part 1: Architecture.

In the following links further info is provided wrt our work on MPEG-M: Past, Current and Future. You could also join us at 'Interactive Music Summit', 25 Nov. 2016, Google Campus, London, UK.