112th MPEG Warsaw, Poland, 20 - 26 June 2015, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 MPEG-21 Media Contract Ontology (MCO) - 2nd Edition

MPEG-21 aims at defining an open framework for multimedia applications, where users distribute, consume, operate on and transact with content represented as Digital Items. These transactions can be governed by the Media Value Chain Ontology (ISO/IEC 21000-19). Furthermore, its complement, the Media Contact Ontology (ISO/IEC 21000-21) expresses business agreements and contracts between parties in a semantic representation, thus facilitating trade of media rights. MCO features include the identification of the contract itself and of its parties, and an unambiguous expression of the agreed permissions, obligations and prohibitions, in a machine readable way, so that their verification can be implemented in software. In particular the MCO deontic expressions address the rights for the exploitation of intellectual property entities, including the specification of the associated conditions, together with other contractual aspects, such as payments, notifications and material delivery.

It is worth to be noticed, in this MCO 2nd Edition, the addition of the following elements:

- *mco-core:Event* – A real life event subject to intellectual property. Examples include sport events, concerts, and musical contexts. The particularity of this class is that its individuals occur at a given time and thus their status can vary in time with respect to be started, in progress, suspended or finished.

- *mco-core:Match* – models the action of executing a comparison between two (or more) IP-Entities and evaluating the degree of matching of the respective properties. The action may result in either a simple boolean value or more complex structured information including ranking, confidence, or statistical data, depending on the application context. The input IP-Entities may play a different role in the action execution, e.g. one may be used as query or sample and another one as candidate under matching evaluation.

- *mco-ipre:MakeRadioProduct* – for modeling the transformation into an IP-Entity mainly made up of sound components.

- *mco-ipre:Remix* – for modeling the recombination of audio tracks or channels from a recording to produce a new or modified audio recording.

- *mco-ipre:CreativeTransform* – for modeling any transformation of the IP-Entity into a new one, containing new creative editorial elements. With the following subclasses:

    - *mco-ipre:Novelization* – for modeling the transformation of the original IP-Entity into a new literary work.

- *mco-ipre:Prequel* – for modeling the transformation into a new IP-Entity, having same or related characters and/or same or related context elements than the original one, for which the fictionary story is consistently conceived to happen before the original story.

- *mco-ipre:Remake* – for modeling the creation of a new IP-Entity with the same theme or concept or plot than the original one.

- *mco-ipre:Sequel* – for modeling the transformation into a new IP-Entity, having same or related characters and/or same or related context elements than the original one, for which the fictionary story is consistently conceived to be the continuation of the original story.

- *mco-ipre:Spinoff* – for modeling the transformation into a new IP-Entity, having some common elements with the original one.

The Contract Expression Language (ISO/IEC 21000-20) which is using XML instead of OWL which is used in Media Contract Ontology (ISO/IEC 21000-21), updated accordingly.

Output documents
N15486 - DoC on ISO/IEC CD 21000-20 2nd edition Contract Expression Language
N15487 - Text of ISO/IEC DIS 21000-20 2nd edition Contract Expression Language
N15488 - DoC on ISO/IEC CD 21000-21 2nd edition Media Contract Ontology
N15489 - Text of ISO/IEC DIS 21000-21 2nd edition Media Contract Ontology

2 MPEG-21 Media Value Chain Ontology (MVCO) - Extensions

There are several use cases, where rights tracking of composite audio IP entities is beneficial, for instance for the delivery of DJ Mixes, and multi-track audio material, as well as in the description of works which creation involved the reuse of other existing works (derivative works). This section describes use cases, in order to illustrate the range of scenarios, where the proposed - by C4DM QMUL - MVCO Extensions (WD of ISO/IEC 21000-19/AMD1) can be employed.

Use Case 1: Podcast
Consider podcast, a program of music or talk made available in digital format that consists several music pieces, each with its own rights holders. A podcast may be defined as a single IP entity. However, in many cases a podcast consists of a number of media items, such as songs with individual property rights. Using MVCO it is possible to identify the rights and permissions for the podcast as a whole, as well as for specific segments of the podcast, improving transparency to underlying rights holders.

Use Case 2: Mashup
A mashup is a song or composition created by blending two or more pre-recorded songs, usually by overlaying the vocal track of one song seamlessly over the instrumental track of another. Although such works are often considered "transformative" of original content, and thus may find protection from copyright claims under the fair use doctrine of copyright law, the rights management of such creative works remains complex. The proposed MVCO extension enables the description of the components of such a production including the definition of overlapping segments. Moreover, the individual components can be described in the same fashion as other IP entities, thus it is possible to describe the full media value chain from the inception of the original work to its reuse in the mashup. Information about the potential transformations the original audio material has undergone in the process of its reuse may additionally be described using future extensions to MVCO or other existing ontologies. The description of individual components of a musical work also applied to hip-hop remixes, where the remixing producer produces a new instrumental track for an existing vocal track.

Use Case 4: Creative Sampling
Sampling of existing music is an established technique in many music genres, especially in hip-hop production. However, as opposed to the "mashup" example above, in this example only segments of a song are used. A music production may consist of a large number of music samples from different sources. The samples may be of different length, may be used repetitively as loops or occur only sporadically. In order to track the media value chain including all IP components, MVCO can be used to describe the sources, permissions, and rights holders of all reused work that is part of the music production. For instance, a record label may grant the permission to reuse a given IP entity to another record label in order to produce a new recording. The detailed description of audio segments facilitates the identification of a reused IP entity in the music production.

Use Case 5: Multi-track Audio Player
A multi-track player implemented in a Web site or mobile app can improve the user experience by providing detailed information about the music stream. For instance, in addition to the multiple audio tracks, it may display information about a given segment or track depending on the cursor position. Further functionality may include the assisted navigation within the audio stream depending on MVCO descriptors, as well as the highlighting and identification of tracks associated with specific users or rights.

Use Case 6: Collaborative Music Production
Another growing field of innovation is collaborative music production. A collaborative music production tool supporting MVCO keeps track of the media value chain. Each user that is registered for the project is associated with his/her contributions and can grant permissions for actions such as making copies or adaptations. Existing IP entities that have been reused in the production are associated with individual information as well. The system makes use of the proposed segment and track concepts, which aid in the management of the complexities of rights associated with collaborative composite content. This model can also be used for the case of remixing existing music, where components of the original production undergo transformations, or additional external IP entities are reused.

Requirements related to aforementioned use cases on time segments and multi-track audio introduced during the 112th Warsaw MPEG meeting as well as related terminology.

Terminology

Interval	A temporal entity with specified duration
ReuseIPEntity	The action of using one IP Entity in the creation of a composite IP Entity.
Segment	An identifiable part of an IP Entity
Track	A single track of a multi-track audio IP Entity
Timeline	Represents the passage of time in relation to time-based IP Entities

Audio Segments

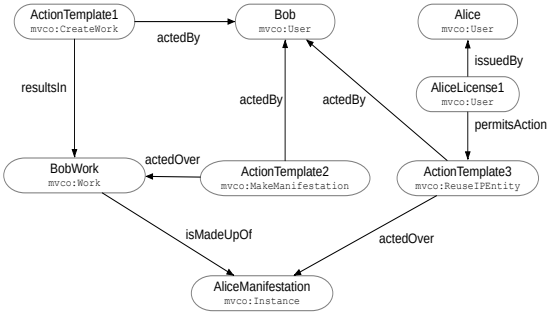
Requirement 20	**The Ontology shall support the description of IP Entities that appear in defined segments on the timeline of a composite IP Entity.**
Description	MVCO shall be able to represent the content of individual segments of an audio IP Entity, defined by a start and end point. A Segment may contain an individual IP Entity and is defined by an interval with a start and end point on the Timeline of a composite IP Entity.
Rationale	Composite IP Entities may consist of existing IP Entities that appear within a specific time segment, associated with individual rights.
Benefit Example	Flexibility in rights association for individual parts of composite audio IP Entities. A broadcaster produces a podcast consisting of multiple speech and music segments. The rights holders of the songs in the podcast issue licenses for the use of the IP Entities. The rights holders for specified time segments can be identified.

Multi-Track Audio

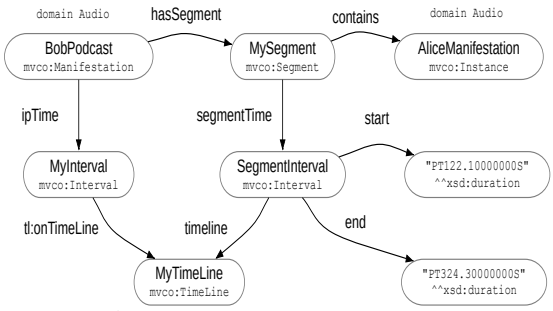
Requirement 21	**The Ontology shall support the description of individual tracks of an audio multi-track entity.**
Description	MVCO shall be able to represent the content of individual tracks of an audio multi-track IP Entity. The content of audio tracks may be treated as individual IP Entities as part of a composite IP Entity.
Rationale	Audio IP Entities may consist of multiple tracks. The tracks of a multi-track audio IP Entity may contain existing IP Entities associated with individual rights.
Benefit Example	Flexibility in rights association for individual tracks of multi-track audio IP Entities. Music files may be distributed in a multi-track format. Different rights holders and permissions can be represented for the IP Entities that appear on individual tracks.

Specifying tracks and segments that contain IP Entities enables the user to:

- Answer queries about the components of composite IP Entities.
- Answer queries about which kind of Role a User plays with respect to a certain IP Entity of a particular track or time segment.
- Answer queries about provenance, rights and permissions concerning individual parts of a composite IP Entity.



Reusing an IP Entity for the creation of a new work.



Segment of a media resource, holding an existing IP Entity.

References

[1] Victor Rodriguez-Doncel and Jaime Delgado , 'A Media Value Chain Ontology for MPEG-21', IEEE Multimedia, Issue 99, 2009.

[2] Inseon Jang, Panos Kudumakis, Mark Sandler, Kyeongok Kang, 'The MPEG Interactive Music Application Format Standard', IEEE Signal Processing Magazine, pp. 150-154, Vol. 28, Issue 1, Jan. 2011.

Output documents

N15352 - Requirements for Media Value Chain Ontology

N15485 - WD of ISO/IEC 21000-19 AMD 1 Extensions on Time Segments and Multi-Track Audio

3 MPEG-21 User Description

The use of multimedia is pervading more and more our daily life and services are becoming more and more customized to user needs. Although MPEG has already developed other standards such as MPEG-7, MPEG-21, and MPEG-M that address User Description, the level of specification in those standards cannot cope with the needs of current and upcoming services such as augmented reality and social networks.

In a context of a large number of competing service providers, a user typically relies on a Recommendation Service that suggest choices to Users. Conceivably the use of more than one Recommendation Service in combination could provide better choices to users. However, comparing different recommendations can be difficult if the users seeking recommendations, the contexts in which they operate and the services they are using are described in incompatible fashions.

The aim of the MPEG-21 User Description standard, referred to as MPEG-21 UD, is to enable Recommendation Services that provide standard, i.e. compatible, recommendations.

The scope of MPEG-21 UD can be exemplified by the following use scenario. A Smart TV manufacturer has included a new type of recommendation functionalities in the TV set to help customers select programmes. The TV set supports standard User Description (UD), Context Description (CD) and Service Description (SD). With these data User can access different Recommendation Services each of which provides a standard Recommendation Description (RD). The TV set has an application, which mashes up the different recommendations and provides customized recommendations to User.

Therefore MPEG-21 UD standardizes the four data formats represented in Figure 1: User Description (UD), Context Description (CD), Service Description (SD), and Recommendation Description (RD).

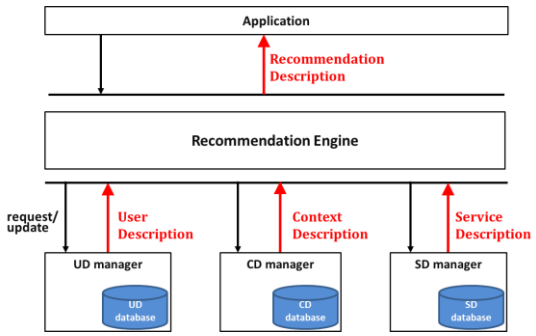


Figure 1: Conceptual Model of MPEG-21 User Description

- User Description (UD): a set of descriptions which may contain static and dynamic information about the user, including some other data like the history of the user's interactions, preferences, security settings regarding these information, etc.
- Context Description (CD): a set of descriptions of the environmental situation in which the user operates, e.g., user's device in use, physical position, environmental variables (temperature, humidity, sound level, etc.), security settings regarding these information, etc.
- Service Description (SD): a set of descriptions containing information (including security settings) about the service (or a set of sub-services), that is offered to the end-user application, e.g. video on demand, maps, etc.
- Recommendation Description (RD): a set of descriptions containing information about recommended items, provided when a customer requests a service in a certain context and in a certain environment. RD may include 1) the recommended content, 2) information extracted from UD, CD, SD; 3) additional logical relations among UD/CD/SD (or their subsets) and 4) metadata from UD/CD/SD.

Output documents
N15491 - Text of ISO/IEC DIS 21000-22 User Description
N15492 - WD of Implementation Guidelines of User Description

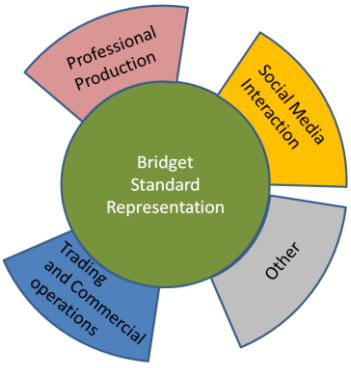
4 Media Linking Application Format (MLAF)

ISO/IEC 23000 (also known as "MPEG-A") is an MPEG standard that supports a fast track to standardization by selecting readily tested and verified technologies taken from the MPEG corpus of standards and combining them to form a MAF (Multimedia Application Format). If a needed piece of technology is not available from the said corpus, then additional technologies originating from other organizations may also included by reference in order to facilitate the envisioned MAF.

The development of the MAF called "Media Linking Application Format" (MLAF ISO/IEC 23000-18) has been prompted by existing many examples of services where media transmitted for consumption on a primary device give hints to users to consume related media on a secondary or companion device. To facilitate interoperability of such services it is beneficial to define a data structure (a "format") that codifies the relationship between the two information sources.

Bridgets are links which exist because of some inherent semantic relationship between content items. As such, they can be products of an editorial decision, taken by someone as the result of the inspection (which can be manual or automatic) of content items, and can be objects of a workflow which involves different roles taking care of finding, organising and finally crafting the data that constitute them. The nature of a Bridget is however quite different than traditional linearity of media content, and as such it induces a different, more "distributed" workflow. In fact, whether a piece of media content is a candidate source or destination for a Bridget can be the result of an editorial decision taken at any moment and by quite different kind of users.

What is foreseeable is a sort of "layered" approach at producing Bridgets in which actors with different roles defines Bridgets under different perspectives and possibly concurring at the same time. Authors of programmes will define Bridget end points (i.e. sources and destination content items) following criteria matching with the editorial intention, main distribution channel or target audience of the programme. At the same time marketing and commercial operators (e.g., advertisement agents) will define such end points following their own mind-setting, which may be independent from the authorial perspective. Last, but definitely not least, final users can define their own ways for 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.



Bridget creation workflow.

Therefore, a standard format for representing and exchanging Bridget-related information fosters integration of all those systems having a role in generating Bridget information in the different and heterogeneous aforementioned domains. The following MPEG-4 BIFS nodes are used to describe the Bridget Presentation Information. The description of AudioClip node is given as an example.

Bridget MPEG-4 BIFS nodes	As an example, the AudioClip node is described.
1. Appearance	**AudioClip {** **exposedField SFString description** "" **exposedField SFBool loop FALSE** **exposedField SFFloat pitch** 1.0 **exposedField SFTime startTime** 0 **exposedField SFTime stopTime** 0 **exposedField MFString url** [] **eventOut SFTime duration_changed** **eventOut SFBool isActive** **}** An AudioClip node specifies audio data that can be referenced by other nodes that require an audio source. The description field is a textual description of the audio source. A browser is not required to display the description field but may choose to do so in addition to or in place of playing the sound. The url field specifies the URL from which the sound is loaded. Browsers shall support at least the wavefile format in uncompressed PCM format [WAVE]. It is recommended that browsers also support the MIDI file type 1 sound format [MIDI]. MIDI files are presumed to use the General MIDI patch set. Results are not defined when the URL references unsupported data types. The loop, startTime, and stopTime exposedFields and the isActive eventOut, and their affects on the AudioClip node, are discussed in detail in the "Concepts - Time Dependent Nodes" section (ISO/IEC 14772-1:1997). The "cycle" of an AudioClip is the length of time in seconds for one playing of the audio at the specified pitch. The pitch field specifies a multiplier for the rate at which sampled sound is played. Only positive values are valid for pitch (a value of zero or less will produce undefined results). Changing the pitch field affects both the pitch and playback speed of a sound. A set_pitch event to an active AudioClip is ignored (and no pitch_changed eventOut is generated). If pitch is set to 2.0, the sound should be played one octave higher than normal and played twice as fast. For a sampled sound, the pitch field alters the sampling rate at which the sound is played. The proper implementation of the pitch control for MIDI (or other note sequence sound clip) is to multiply the tempo of the playback by the pitch value and adjust the MIDI Coarse Tune and Fine Tune controls to achieve the proper pitch change. The pitch field must be > 0.0. A duration_changed event is sent whenever there is a new value for the "normal" duration of the clip. Typically this will only occur when the current url in use changes and the sound data has been loaded, indicating that the clip is playing a different sound source. The duration is the length of time in seconds for one cycle of the audio for a pitch set to 1.0. Changing the pitch field will not trigger a duration_changed event. A duration value of -1 implies the sound data has not yet loaded or the value is unavailable for some reason. The isActive eventOut can be used by other nodes to determine if the clip is currently active. If an AudioClip is active, then it should be playing the sound corresponding to the sound time (i.e., in the sound's local time system with sample 0 at time 0): fmod (now - startTime, *duration* / pitch).
2. AudioClip	
3. BitWrapper	
4. Circle	
5. Color	
6. Coordinate	
7. FontStyle	
8. Group	
9. ImageTexture	
10. IndexedFaceSet	
11. IndexedLineSet	
12. Inline	
13. Layout	
14. LineProperties	
15. Material2D	
16. Normal	
17. OrderedGroup	
18. PointSet	
19. PositionInterpolator2D	
20. Rectangle	
21. ScalarInterpolator	
22. Script	
23. Shape	
24. Sound2D	
25. Switch	
26. Text	
27. TextureCoordinate	
28. TimeSensor	
29. TouchSensor	
30. Transform	
31. Transform2D	

Output documents
N15498 - Text of ISO/IEC CD 23000-18 Media Linking Application Format

5 Exploration on Media Orchestration

With so many 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. This may have a relation with the notion of "Media Internet of Things" that is also discussed in MPEG.

Timed Content

Timed content has an intrinsic timeline. It may have a start and/or end (e.g. Content on-Demand), or it may be continuous (broadcast). It may be atomic (video-only) or it may be composite (A/V, 3D-audio, multiplex). Classic examples of timed content are video, audio and timed text. In the context of media orchestration, also streams of location and orientation, as well as other sensor outputs are timed content, see Figure 2. Note that the notion of time may have various applications in Media Orchestration. It may pertain to the *delivery* or on the media itself, or the *presentation.* It may also pertain to the *capture* of media.

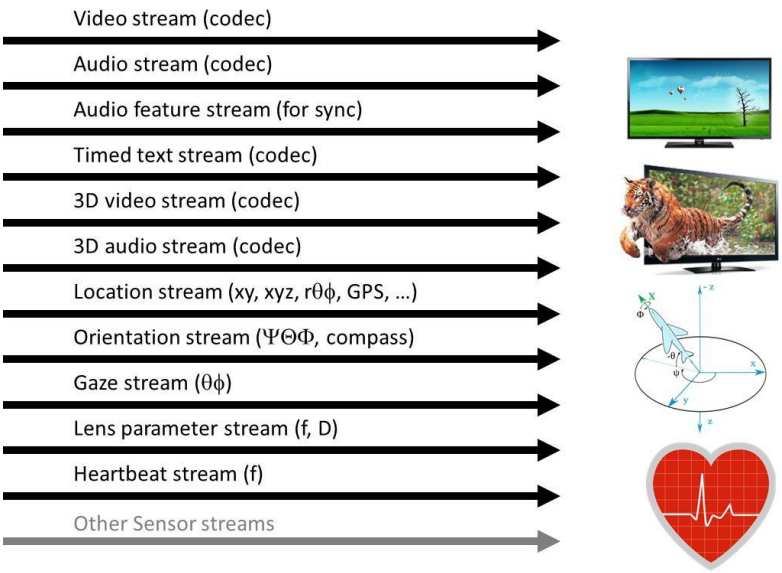


Figure 2: Examples of timed content in the context of media orchestration

A couple of representative media orchestration use cases are following:

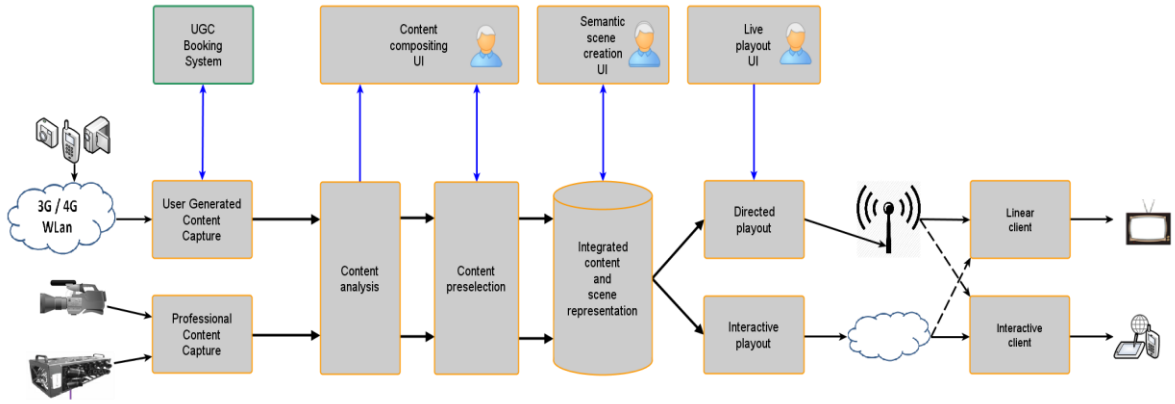
Video Street View for an Event

A lot of people visiting an event, e.g. a carnival/festival, use their smart phones to create recordings. Furthermore, the event has own cameras to create other recordings, on the stage, on the event objects, performer, on drones, etc. Their recorded video are uploaded to the cloud, and they are used to create/playback video street view. When a certain time and a certain position in the event are specified, video corresponding to the time and position starts to be played back. The specified positions are changed through the playback period, and the videos are changed. In order to playback videos continuously, i.e. with no-gap, during the playback period, videos near the specified positions can be selected to be played back as well as the ones on the strict positions. The videos near the positions can be also processed, e.g. zooming, panning, synthesis, in order to be more seamlessly.

Video Street View for an Event

Immersive Coverage of Spatially Outspread Live Events (ICoSOLE)

ICoSOLE aims at supporting use cases that enables users to experience live events which are spatially spread out, such as festivals (e.g. Gentse feesten in Belgium, Glastonbury in the UK), parades, marathons or bike races, in an immersive way by combining high-quality spatial video and audio and user generated content. Therefore, a platform shall provide support for a context-adapted hybrid broadcast-Internet service, providing efficient tools for capture, production and distribution of audiovisual content captured by a heterogeneous set of devices (professional and user-generate content) spread over the event site. An overview of such a generic example system is shown in the figure below where content is coming both from professional content capture devices and user-generated content capture devices. After the content is being analyzed and preselected – possibly supported by an editor through an appropriate user interface (or done automatically) – the content is integrated using an appropriate scene representation and prepared for dissemination (live and on-demand enabling interactivity) to a plethora of heterogeneous devices.



In particular, the capturing app (incl. Moments app) allows for audio-visual content capturing including various metadata (e.g., location, orientation, A/V recording parameters, sensor information, manual annotations by the user, user profile data incl. psychophysiological data, obviously subject to privacy precautions) and live streaming or (later) uploading to a central storage and processing unit on site (typically in an outside broadcast van). These user-generated media assets may be referred to as 'Moments'.

On the consumption side, a feature called 'The Wall of Moments' allows a remote user to have a unique festival experience by using 'Moments' captured by peers attending the live event. The remote user pitches in to the live event via a Moment and can enjoy a high quality A/V stream from there on, as Moments are precisely synchronized with the professional production feed. In this way, the user still has an immersive experience by having direct social contact with the crowd (more importantly, his/her friends/peers). The Wall of Moments could be realized as a mosaic representing the most popular moments users shared (e.g., via Facebook, Twitter), moments the production highlighted, live production feeds and a virtual map. It acts as an online interactive portal to the live event for users not attending the event, yet providing an immersive experience bringing them closer to the event. Various interaction possibilities for professional editors and users of the system can be envisioned for this use case.

Output documents
N15343 - Requirements for Media Orchestration v.1
N15342 - Draft of Context and Objectives for Media Orchestration

6 Exploration on Wearable MPEG

The exploratory work carried out by the wearable MPEG group during the MPEG 111th and 112th meetings brought to light the synergies and complementarities among wearable MPEG and several standard families, being inside or outside MPEG (MPEG V, MPEG-UD, MPEG-U, MIoT, W3C). This will be further jointly investigated and assessed from the standardization point of view during the next 113rd MPEG meeting.

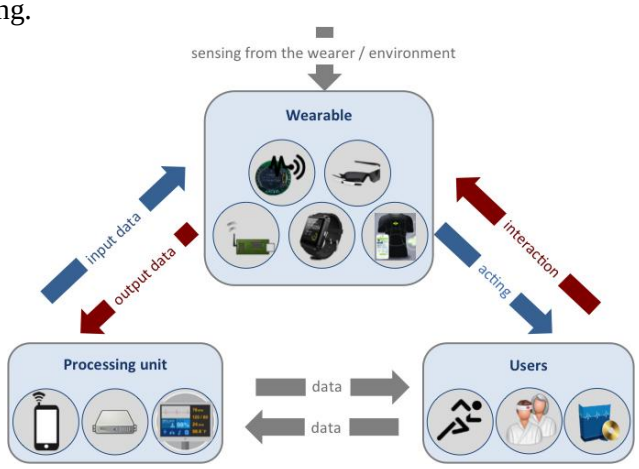


Figure 3: Conceptual model for Wearable MPEG

The scope of the Wearable MPEG is to standardize:

1. The interaction commands from User to Wearable
2. The format of the aggregated and synchronized data sent from the Wearable to the Processing unit (represented by red arrows in Figure 3)
3. A focused list of sensors that the Wearable may integrate.

By investigating a large number of use cases, the following formats of data / commands are identified for standardization as shown in the following Table.

Use case		Exchange types of data (formats) / commands			
		Wearable → User	**User → Wearable**	**Wearable → Processing Unit**	**Processing Unit → Wearable**
1	Multimedia communication	• Mono image • Stereoscopic Image • Audio and Voice • Useful data of User	• Gesture of Hand Motion of head • Voice • Touch • Skin keyboard	• Mono image • Stereoscopic image • Depth image • Data of Physical sensors	• Mono image • Stereoscopic Image • Audio and Voice • Useful data of User • Gesture command • Voice command
2	Gesture Command	not defined	• Gesture of Hand • Motion of head	• Mono image • Stereoscopic image • Depth image • Data of Physical sensors	• Mono image • Stereoscopic Image • Useful data of User • Gesture command
3	Speech translation	• audio • text	• audio • text • gender • location • speech style	• audio • text • gender • location • speech style	• audio • text
4	Natural language communication	• audio • text • interaction	• audio • video • text • interaction • location • gender	• audio • video • text • interaction • location • gender	• audio • video • text
5	Accessibility/ protection functionality	• audio • video • text • interaction • vibration • light	• audio • video • text • interaction • location • accessibility features	• audio • video • text • interaction • location • accessibility features	• audio • video • text • interaction • signal for vibration and light
6	Wearable device in cloud	• render visual • notifications	• single touch • multi touch • voice control	• audio • video • interaction	• audio • video • graphics
7	Visual Communication	• Visual Object (Recommendation)	• Gesture Command • Touch Command • Voice Command (Intention)	• Profile of Wearable Device • Gesture Command • Touch Command • Voice Command (Intention) • Location • Weather • Traffic (Sensed Data)	• Visual Object (Recommendation)
8	Digital shirt	• audio • video • temperature • vibration	• single touch • multi touch • surface touch • hands gesture • arms gestures • voice control	• audio • video • temperature • pulse	• run • jump • stand • lay down • walk
9	Artificial heart	- -	- -	• heart rate • core temperature • timing	• heart rate • compression type
10	In-body	- -	- -	• heart rate • body temperature • blood pressure • heart disorder • blood pressure too high	- -
11	QA	• audio • image • text	• speech • image • text • UD	• speech • image • text • UD	• answer in audio • image • text
12	Multimodal	•	• Speech data • Gesture data • Facial expressions (emotion and intention) • Other modality data. Ex) text, touch input	• Multimodal data in predefined representation • User preferences	•
13	Self-adaptive application	• Audio • Video •	• Setup data • Policy management data •	• Status information • Application's static/dynamic information	• Control commands •