114th MPEG San Diego (CA), USA, 22 - 26 February 2016, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 Media-centric IoT and Wearable (M-IoTW)

The proposed Media-centric IoT and Wearable (M-IoTW) root elements are shown in Figure 1 followed by their definitions.

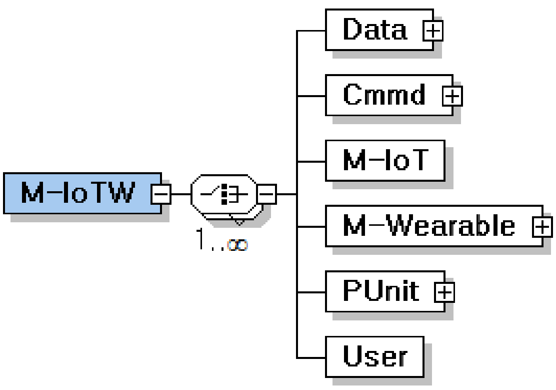

Figure 1 - Root elements of M-IoTW.

Name	*Definition*
`M-IoTW`	Serves as the root element of M-IoTW description. `M-IoTW` shall be used as the topmost element in a description.
`M-IoT`	Describes an instance of one M-IoT description. An M-IoT description can be used to represent information of the M-IoT device itself, which includes the type and equipped I/O devices information.
`M-Wearable`	Describes an instance of one M-wearable description. An M-Wearable description can be used to represent information of the M-Wearable device itself, which includes the type and equipped I/O devices information.
`Data`	Describes an instance of data description. A data description can be used to represent information of data, which includes processing data incoming from input devices or intermediate data generated during the processing, and media data to be presented to a user.
`Cmmd`	Describes an instance of command description. A command description can be used to represent commands, which includes interaction commands for interaction between a user and IoT/Wearable devices, and acting commands for controlling sensors equipped in/connected to IoT/Wearable devices.

Name	Definition
PUnit	Describes an instance of information of processing unit description. A processing unit description can be used to represent information of processing to be done in a processing unit.
User	Describes an instance of user information description. A user description can be used to represent information related to a user.

Output Documents
N16120 - State of Discussions on M-IoTW technology
N16140 - Overview, context and objectives of Media-centric IoTs and Wearables
N16141 - Conceptual model, architecture and use cases for Media-centric IoTs and Wearables
N16139 - Draft Requirements for Media-centric IoTs and Wearables

2 Visual Identity Management AF (VIM-AF)

Citizens are voluntarily sharing significant amounts of information (data, pictures and video) about themselves and their activities through social networking applications, which are being commercially exploited and traded for marketing purposes, but also used for malicious activity.

In those social networks, access control is provided and done by the social networks themselves without any control by the user. To restore citizens' confidence in online data collection practices, submitted media should be encrypted by the user and can only be viewed according to user wishes such as: by a group of friends, purpose of sharing, time, date or metadata.

In order to provide user control over processing and sharing of multimedia content, a flexible, effective and scalable mechanism is to provide users a way to express their control desires in a form that can be processed and monitored systematically, consistently and persistently throughout the lifecycle of the multimedia content. These control desires can be considered as multimedia usage-control policies, much like data usage-control policies many websites (e.g., Facebook) use to protect user identity and data usage-control. *The difference here is that, in the user control case, it is the users, not the websites, who specifies how their multimedia content should be processed and shared.*

In order to express users' desire on controlling usage of their multimedia content, the following should be supported in a usage-control policy expression language such as a profile of the MPEG-21 CEL:

- Specification of ownership of the multimedia content and their components, with optional authentication checking mechanisms.
- What kinds of processing, or types of operations, can be applied to the multimedia content, and their deontic characteristics such as permissions, obligations and prohibitions (i.e., rights, duties and bans in the MPEG-21 CEL terms).
- How to share, with whom, for how long and under what conditions, also along with their deontic characteristics such as permissions, obligations and prohibitions (i.e., rights, duties and bans in the MPEG-21 CEL terms).

System for image identity privacy

Figure 2 illustrates a conceptual framework for managing privacy of users when pictures are taken and being shared among users.

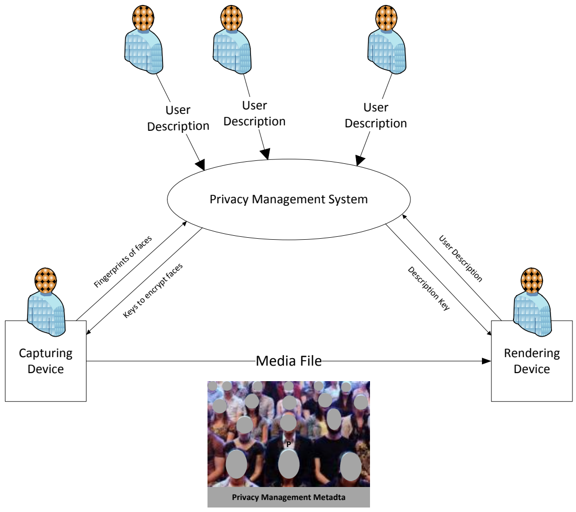

Figure 2 - Framework for privacy management of pictures.

- Each user registers his privacy management metadata to a Privacy Management System (PMS), containing the following information
    - Unique ID of the user
    - A fingerprint of the face of the user

    A list of the IDs of the users who are allowed to see his face
- The PMS automatically generates encryption keys for each registered user's face and shares the decryption keys with the users who are allowed to see his face
- When a picture is taken by a capturing device, a unique ID of the picture is generated and it is provided to PMS with the list of fingerprints of the faces in that picture with sequential numbers associated for every face in the picture
- The PMS identifies the list of users in the picture by matching the fingerprints received and the fingerprints of the registered users and provides appropriate encryption keys for each face to the capturing device
- The capturing device creates a file of the picture according to the user's privacy management metadata capturing the picture, including the following information
    - Unique ID of the picture
    - List of sequential numbers of encrypted faces and their location and size
- The capturing device shares the file of the picture to a rendering device
- A rendering device tries to decrypt the faces by using the privacy management metadata of the user trying to see the picture and the keys received from the PMS; in doing so
- The rendering device sends the following information to the PMS
    - The ID of the picture the user is trying to see
    - The description of the user trying to see the picture
    - The context of the user
- The PMS evaluates whether the user trying to see the picture is listed as one of the users who can see the faces of the users in the picture. If the user trying to see the picture is not listed as the one who can see some of the faces of the users in the picture, then the PMS sends to those users a description about the user trying to see the picture. If those users agree to show their faces to the user trying to see the picture then the PMS sends to the rendering device the keys to decrypt their faces.

Output Documents

N16142 - Use cases and requirements on Visual Identity Management AF

3 Common Media Application Format (CMAF)

Most video, such as movies and TV shows, that has been delivered in the past only to TV sets by broadcast or disc, is now available over the internet. In the last decade, several billion new internet video devices such as phones, tablets, game consoles, and connected TVs have been bought and used by consumers to play video in a wide range of bitrates, resolutions, languages, etc. that reflect the variety of devices, networks, and the global reach of the internet. As a result, the industry has evolved an “adaptive” media format that allows each device to select and combine media objects, such as a particular audio and video tracks, and segments of video at a bitrate and resolution that is optimized for the device and the current network throughput. This object-oriented, late binding media format that has evolved allows one set of media objects to be encoded that devices can combine into thousands of customized media presentations without having to encode and deliver thousands of different audio/video files or streams.

The segmented media format that has been widely adopted for internet delivery using DASH, Web browsers, commercial services such as Netflix and YouTube, etc. 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 (ARIB, ATSC, DECE, DASH-IF, DVB, HbbTV, SCTE, 3GPP, DTG, DLNA, etc.), 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 would further benefit from a common format, embodied in an MPEG standard, to improve interoperability and distribution efficiency.

MPEG Technologies in the Common Media Application Format

Some of the MPEG technologies that have been widely adopted for internet video and are expected to be referenced by the Common Media Application Format are:

- ISO Base Media File Format (MPEG-4 Part 12)
- MPEG codecs for audio and video
- MPEG-4 file specifications for the delivery of audio, video, and subtitles/captions derived from the Base format (MPEG-4 Part 15, Part 30, etc.)
- Common Encryption (ISO/IEC 23001-7)

Use Cases

1. Adaptive bitrate streaming of Media Segments using HTTP(S) and any Presentation Description, such as DASH MPD, Smooth Streaming Manifest, Apple HTTP Live Streaming Manifest (m3u8), etc. Note: Specific CMAF bindings to MPEG DASH and other presentation descriptions are expected to be defined separately.
2. Broadcast/multicast streaming of Media Segments over one-way networks such as terrestrial broadcast, satellite, or cellular network.
3. Hybrid network streaming of live content via broadcast/multicast or cellular, and unicast for time shifted or individualized content, and Segments lost during broadcast delivery.
4. Download of streaming files for local playback, or local playback of downloaded content combined with streamed Segments.
5. Server-side and Client-side ad insertion with media segment and manifest-level signaling of messages for ad insertion and other purposes, such as SCTE-35, VAST, VMAP, etc.

Requirements

1. The Common Media Application Format (CMAF) shall be based on existing MPEG standards and industry practices; and shall support the following requirements:
    a. Shall define specific profiles based on the ISO Base Media File Format (ISOBMFF), each defining a conformance point that provides interoperability between CMAF conformant devices (CMAF players) and CMAF presentations that support that profile. The highest priority is to specify the minimum requirements both CMAF presentations and CMAF players must support for interoperability, and to specify an easily adopted baseline profile.
    b. Shall be capable of carrying encoded audio, video, and subtitles, including signaling to support track selection and common accessibility use cases.
2. Shall be delivery mechanism agnostic, but support at least the following delivery mechanisms:

a. Adaptive streaming with seamless adaptive switching of CMAF tracks encoded with different bitrates, frame rates, and video resolutions.
b. Late binding of independently created and/or delivered tracks for combined playback.
   Note: In late binding, media components' segments are separately delivered and combined at the playback time, while in early binding, the media component segments are added together as part of delivery and fed combined into the player.
c. Early binding of independently created and/or delivered tracks for combined delivery, i.e. playable delivery segments that contain media segments of differing media types but the same approximate time-range.
d. Multicast and broadcast delivery.
e. Hybrid network delivery (broadcast + unicast).
f. Physical media delivery (memory sticks, discs, etc.).
g. Low latency live streaming.
h. Multitrack file progressive download and playback.
i. Efficient CDN delivery with uniform identification of each media segment by URI (and possible byte range) for efficient caching and reuse.
j. Shall support random access of the content.
k. Shall support signaling of various random access points and their types.
l. Shall support delivery and playback of selected tracks of the content.

3. Each CMAF profile shall define:
   a. A video codec, its profile and level, associated color space, EOTF and other rendering constraints.
   b. An audio codec, its profile, features, and channel configuration.
   c. Optional closed caption and subtitle formats.
   d. An optional encryption scheme and key management constraints for Common Encryption (CENC) of media samples.
   e. Track and segment encryption metadata that enables the use of any DRM systems that conform to MPEG Common Encryption for key management and decryption.
   f. A set of constraints for ISOBMFF files, tracks, movie fragments, samples, and elementary streams bindings.
   g. CMAF profile identifier(s) for the conformance points supported by a presentation or device.

Output Documents
N16144 - Requirements for the Common Media Application Format
N15947 - WD v.1 of Common Media Application Format

4 MPEG-M: Multimedia Service Platform Technologies (3rd Edition)

The MPEG-M Architecture has been enriched with new Technology Engines as in Table 1, as well as, support of the Pub/Sub protocols as described in the following section.

Table 1 - Technology Engines and classification of their APIs

No.	Engine	Creation API	Access API	Engine-specific API
1.	CDVS	Y	Y	
2.	CEL	Y	Y	• Search • Check With
3.	Digital Item	Y	Y	
4.	Event Reporting	Y	Y	• Register Events • Transmit Event Reports
5.	Green Metadata	Y	Y	
6.	IPMP	Y	Y	• IPMP Tool Instantiation • IPMP Tool Initialisation
7.	Media Framework	Y	Y	
8.	Metadata	Y	Y	
9.	MPEG-21 File Format	Y	Y	

10.	Overlay	Y	Y	• Store/Retrieve messages • Propagate message
11.	REL	Y	Y	• Validation • Authorisation
12.	Search	Y	Y	
13.	Security	Y	Y	• Authentication • Integrity
14.	Sensory Effects	Y	Y	• Manipulate a Sensory Effect Metadata (SEM) structure; • Retrieve by identifier or object; • Adapt to get device capabilities and user preferences; • Actuate to set device commands of the actuator

Orchestration

The case analysed in this section shows the operation of the Orchestrator Technology Engine in a particularly articulated case. If PubSub is used to access resources, when Subscriber has completed a PubSub workflow, i.e. has received the notification that a match of his subscription with a publication has been found, the user clicks on the notification. The workflow unfolds based on Figure 3.

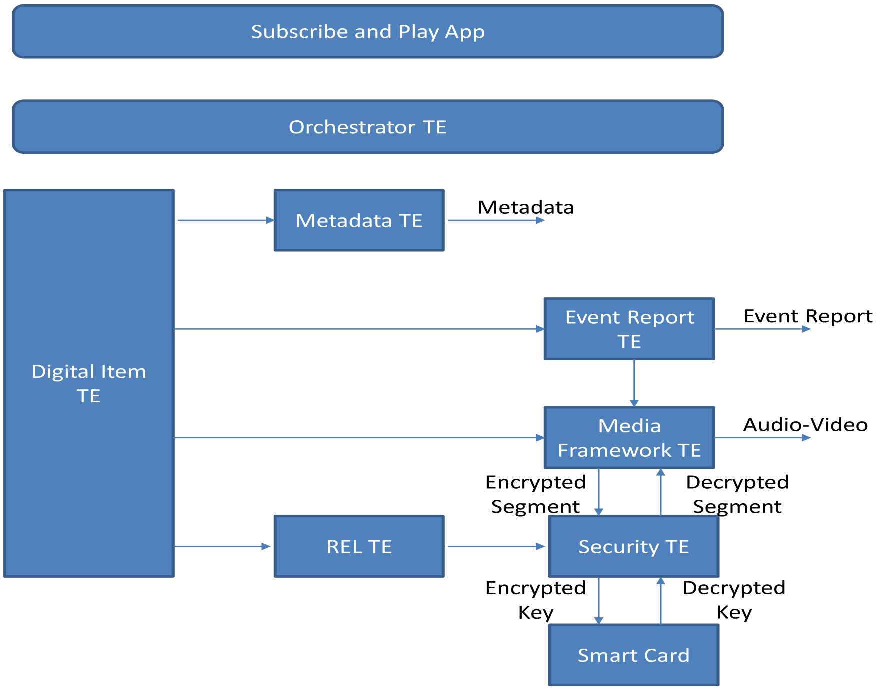

Figure 3: Operation of Orchestrator in resource playing

1. Subscriber user
   a. Parses notification
   b. Requests Network TE to get Resource Information
   c. Receives Resource Information
   d. Decides to see what the video is about
   e. Presses Preview button on Subscribe and Play App
2. Subscribe and Play App requests video metadata
3. DI TE
   a. Parses DI
   b. Passes metadata to Metadata TE
4. Metadata TE
   a. Parses metadata
   b. Passes metadata to Subscribe and Play App
5. If user decides to view video Subscribe and Play App requests MXM to play video
6. DI TE
   a. Continues parsing DI

    b. Passes REL licence to REL TE
7. REL TE
    a. Parses REL licence
    b. Passes encrypted decryption key to Security TE
8. Security TE passes Encrypted decryption key to Smart Card
9. Smart Card
    a. Tries to decrypt encrypted decryption key
    b. If successful
        i. Passes key to Security engine
        ii. Informs Orchestrator TE
    c. Otherwise informs Orchestrator TE of failure
10. If decryption key is successfully decrypted, DI TE
    a. Parses DI
    b. Sends Resource ID to Media Framework TE
    c. Parses Event Report Request
    d. Sends Event Report Request to Event Report TE
11. Media Framework TE
    a. Requests Network TE to get MPD
    b. Receives MPD
    c. Parses MPD
    d. Requests Network TE to get first segment
    e. Receives first segment (encrypted)
    f. Passes first segment to Security TE
12. Security TE
    a. Decrypts first segment using decryption key (in clear)
    b. Passes decrypted first segment to Media Framework TE
13. Event Report TE
    a. Parses Event Report Request
    b. Dispatches Event Report to listed users
14. Media Framework TE
    a. Decodes first segment
    b. Writes decoded first segment to decoder buffer

The process continues. However, Event Reports are sent only after the first segment has been successfully decrypted (of course other criteria are possible, e.g. only when the entire video has been played).

Output Documents
N15953 - Text of ISO/IEC DIS 23006-1 3rd edition MXM Architecture
N15954 - Text of ISO/IEC IS 23006-2 3rd edition MXM API
N15955 - Text of ISO/IEC IS 23006-3 3rd edition MXM Conformance

5 Workshop on 5G / Beyond UHD Media, Wed 24 Feb 2016

The media distribution industry is moving at a rapid pace, driven by a very competitive environment to provide the best quality of experience to the end customer. 4K content distribution over the Internet has already been started and 4K broadcasting is coming up soon. UHD content offering high frame rate, HDR, and wide colour gamut, is forecasted to drive a large scale upgrade of displays and receivers, capable of handling the new formats and content capabilities. Additionally, 360 video, virtual reality (VR) and 8K content are promise to be the next wave of disruptive technologies, paving the way for a wide range of use cases through freeing the user from production constraints and providing an immersive user experience controlled by the end user.

At the same time, advances are being made to accommodate for the exponential growth of bandwidth needs. Recent studies show that 70% of the traffic on the Internet is generated by video traffic, the majority of which is delivered over mobile and wireless networks. 5G is making steady steps towards providing Gigabit bitrates to the end user through advances in path diversity, modulation, higher cell density, and the usage of new frequency bands. 5G is also aiming at reducing delays over the air interface down to 1 millisecond.

These new developments raise questions about the requirements to be fulfilled by MPEG in bridging the codec layer and the system layer to enable the emergence of unprecedented media experiences. Considering the above, MPEG held a public workshop aiming to assess the potential of these new developments and their impact on MPEG technologies as well as to initiate an industry wide activity to address the needs of emerging media services and networks enabling them.

Program
14:00 – *Cloud Computing and the Internet of Things*, Dr. James Kempf **(**Principal Research Engineer, Ericsson)
14:25 – *Virtual Reality*, Ian Harvey (Senior Vice President, FOX)
14:50 – *Unlocking Opportunities: Unbundling Technologies and Creating Relevant Enablers*, Kent Walker (Vice President, Qualcomm)
15:15 – *Challenges and opportunities in Next Generation Mobile Multimedia Services*, Sungmin Cho (Team Leader, SK Telecom)
15:35 Break
16:00 – *5G and Its Challenges in Support of UHD Media: A WWRF Outlook*, Dr. Sudhir Dixit (Vice Chair - Americas, Wireless World Research Forum)
16:25 – *5G and future media consumption*, Emmanuel Thomas (TNO)
17:00 – *5G Key Technologies and Recent Advances*, Charlie Zhang (Vice President, Samsung Research America)

17:30 – Panel Discussion
What will 5G bring to Multimedia Services and what role can MPEG play?
Panel Participants:

- James Kempf (Ericsson)
- Ian Harvey (FOX)
- Sungmin Cho (SK Telecom)
- Sudhir Dixit (WWRF)
- Charlie Zhang (Samsung Research America)
- Thomas Stockhammer (Qualcomm)
- Ozgur Oyman (Intel)

Output Documents
N16138 - Requirements for a Future Video Coding Standard v2
N16125 - Presentations of the Workshop on 5G and beyond UHD Media

6 Workshop on JPEG Privacy & Security, Tue 23 Feb 2016

Privacy and security support for image data is becoming steadily more important seen the fact that image collections are increasingly more stored in distributed and cloud repositories rather than in private repositories. Moreover, social media and online photo repositories, for example, are currently offering insufficient means to secure privacy-sensitive information carried by the picture or to signal associated IPR metadata. Observing that on a daily basis billions of pictures are shared in JPEG legacy formats on these media, it is evident that embedding additional functionality that would safeguard this type of information and functionality would benefit a significant user base.

Hence, the JPEG Committee has launched a new activity called JPEG Privacy & Security. This activity aims at developing a standard for realizing secure image information sharing which is capable of ensuring privacy, maintaining data integrity, and protecting intellectual property rights. This activity is not only intended to protect private information carried by images - in the image itself or the associated metadata - but also to provide degrees of trust *while sharing image content and metadata based on individual preferences*. It is necessary to extend the existing coding standards by adding such preferences. JPEG Privacy & Security will explore ways on how to design and implement the necessary functionality without significantly impacting on coding performance while ensuring scalability, interoperability, and forward and backward compatibility with current JPEG standard frameworks.

Program
13:30 – *JPEG Privacy and Security - Introduction and Scope*, Touradj Ebrahimi (JPEG Convenor, EPFL)
13:40 – *Summary of Brussels workhop*, Ambarish Natu
13:50 – *Who Can you Trust? (When you're trusting trust),* Jeremy Malcolm (EFF)
14:15 – *The PLUS Coalition Connecting JPEG Files to Rights Holders and Rights Information*, Jeff Sedlik (PLUS Coalition)
14:40 – *Media-JSON for Creating Visual Web*, Ramesh Jain (University of California, Irvine)
15:05 – *SC27 Standards*, Gregg Brown (Microsoft)
15:30 – Break
16:00 – *Legacy JPEG compliant transmorphing to preserve privacy in social networks*, Touradj Ebrahimi (EPFL)
16:25 – *Embedded metadata in the cultural heritage community*, Greg Reser (Univerisity of California, San Diego)
16:50 – *JPSearch metadata and its use for Security & Privacy in JPEG images*, Jaime Delgado (Professor of Distributed Multimedia, UPC)
16:10 – Panel Discussion

Output Documents
N71026 (WG1) - Proceedings of 2nd JPEG Privacy and Security Workshop

7 Call for Proposals on Media Orchestration (MORE)

MPEG is now working on "Media Orchestration". 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.

With this Call for Proposals ("CfP"), MPEG invites all parties that have technologies that fulfil all or some of the requirements in the Media Orchestration Requirements Document (N16132) to make a proposal to MPEG for the inclusion of such technologies in the upcoming Media Orchestration Standard. MPEG also invites parties holding datasets relevant for the development of the Media Orchestration standard to contribute such data sets to MPEG, so that these can be used in the development of this standard including possibly the evaluation of technology proposals.

The Context and Objectives for Media Orchestration (N16131) provides background on the subject matter for the Call for Proposals, as well as proposal for a Functional Architecture (see Figure 4). The Requirements for Media Orchestration (N16132) provide the requirements that those wishing to respond to the CfP are asked to address.

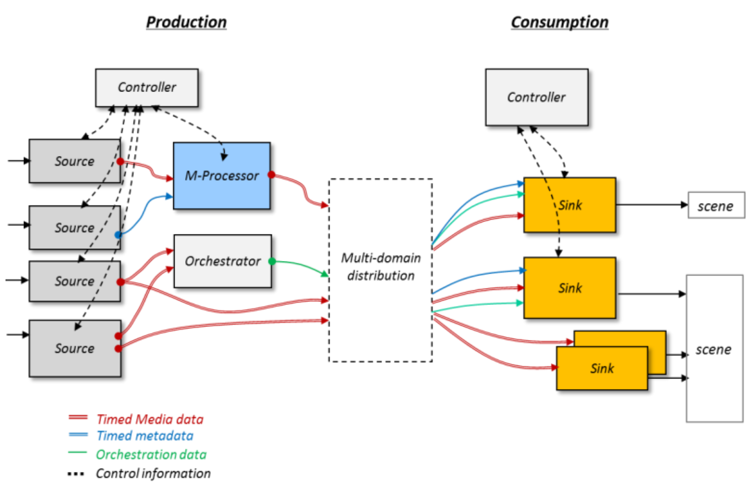

Figure 4 - Examples of Timed Data in the context of media orchestration

qMedia::MMV group contributed a use case on media orchestration in the security domain: the orchestration of media streams captured across a variety of sources including CCTV, professional quality content captured by broadcasters and user-generated content from social networks forms a critical source of information which can be used towards emergency relief, crime prevention, crime solving or better provision of security. Linking and merging heterogeneous data retrieved from multiple sources can be used to guide investigation processes and perform inference based on evidence extracted from that media. A pictorial description of the use-case is presented in Figure 5.

Figure 5 - The heterogeneous content sources available for security, emergency and relief agencies

Output Documents
N16131 - Context and Objectives for Media Orchestration v.3
N16132 - Requirements for Media Orchestration v.3
N16133 - Call for Proposals on Media Orchestration Technologies

8 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). In the following section use cases are described related to the following two new requirements introduced by qMedia::C4DM in MVCO:

- Audio Segments: 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.
- Multi-track Audio: 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.

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 3: 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 4: 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 5: 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.

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.

Output documents
N15936 - WD of ISO/IEC 21000-19 AMD 1 Extensions on Time Segments and Multi-Track Audio