113th MPEG Geneva, Switzerland, 17 - 23 October 2015, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 MPEG-21 User Description Reference Software

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 MPEG-21 UD standard is composed of four Descriptions: User Description (UD), Context Description (CD), Service Description (SD), and Recommendation Description (RD). These Descriptions will be encoded using XML and decoded (parsed) by the reference software decoder. Although the MPEG policy is usually to offer only the reference software decoder, the User Description group decided to provide in addition a reference software encoder with aim to help the widespread use of the standard. Furthermore, a number of applications will be provided as examples for the verification of the MPEG-21 UD reference software encoder/decoder.

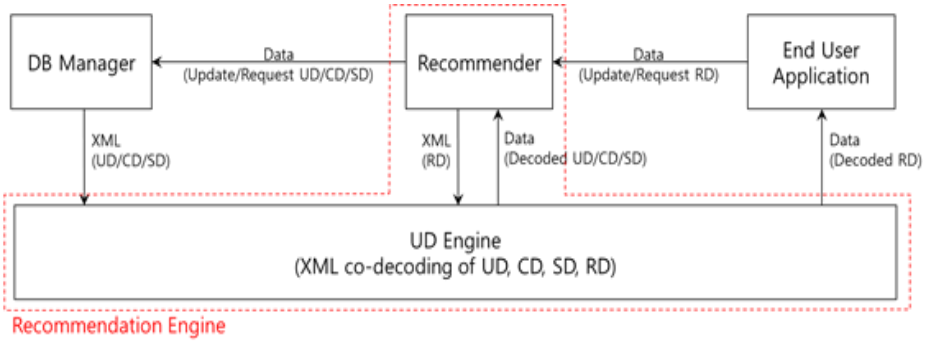


Figure 1: MPEG-21 User Description reference software block diagram.

Output documents
N15649 - Study of ISO/IEC DIS 21000-22 User Description
N15650 - WD of Reference Software and Implementation Guidelines of User Description
N15648 - Work plan for Reference Software on ISO/IEC 21000-22 User Description

2 Publish/Subscribe Application Format (PSAF)

Publish/Subscribe (PubSub) is an established communication paradigm where senders do not communicate information directly to the intended receivers but rely instead on a service that mediates the relationship between senders and receivers. In PubSub senders (called Publishers) post information to and receivers (called Subscribers) declare their interest in information to a service.

The typical steps of a Publish/Subscribe workflow in a media distribution context are given by Table 1, where "Resource" is the media item that is announced (i.e. published) by Publisher.

Table 1 – Steps in multimedia Publish/Subscribe

Step	Step	Information type	Acronyms
1.	Publisher stores information on resource	Resource Information	RI
2.	Publisher publishes information on resource	Publication Information	PI
3.	Subscriber subscribes to a class of resources	Subscription Information	SI
4.	Service matches subscription with publication		
5.	Service issues notification to targets	Notification Information	NI
6.	Target opens notification		
7.	Subscriber requests/plays resource		

NB1 It is possible for Steps 3 and 4 to also happen in reverse order, i.e. a Subscriber can look for information before the information is actually published. In other words there is no specified order for Publications and Subscriptions to be made.

NB2 A target of Step 6 is called Consumer, a user who receives a notification from an MSP that a match has been found based on instructions received from Publisher and Subscriber. A Consumer can be one of them but also a third party.

Requirements

1. Publishers and Subscribers shall be able to define Users or groups of Users to be/not to be notified
2. Publishers shall be able to define a standard information package (Resource Information) containing at least the following information elements
    a. Identifiers of Resource
    b. Descriptions of Resource
    c. Information related to the multimedia service provided, the target users and the recommended context of use
    d. Permissions, Obligations and Prohibitions for the use of the Resource by users of the Resource
    e. Request to notify a list of users that a specific use of a Resource has been made
3. Publishers shall be able to define a standard information package (Publication Information) containing at least the following information elements
    a. Metadata related to Resource Information
    b. Information related to the multimedia service provided, the target users and the recommended context of use
    c. Permissions, Obligations and Prohibitions for the use of Metadata related to Resource Information by the Match Service Provider
    d. Request to notify/not to notify a list of users of a Match between this Publication and Subscriptions
4. Subscribers shall be able to define a standard information package (Subscription Information) containing at least the following information elements
    a. Query related to Metadata of Resource Information
    b. Metadata related to Consumer and the Context in which the Resource sought will be consumed
    c. Permissions, Obligations and Prohibitions for the use of the Query, the Consumer-related metadata and the Context of use by the Match Service Provider
    d. Request to notify/not to notify a list of users of a Match between this Subscription and Publications
5. Match Service Providers shall be able to define a standard information package (Notification Information) containing at least Recommendation of a Resource Information ID
6. Publishers and Subscribers shall be able to
    a. Select the Match Service Provider(s) that shall perform match of their Publications and Subscriptions

b. Define the period of time during which the Match Service Provider shall notify Matches of their Publications or Subscriptions
c. Update their respective Publications and Subscriptions

7. Publishers, Subscribers, Consumers and Match Service Providers shall be able to guarantee the authenticity of their Resource, Publication; Subscription; Metadata related to Consumer and Context and date of issuance; and Notification Information

Relationships between Users

The relationships between the Users of multimedia Publish/Subscribe are depicted in Figure 1. Note that UD = User Description, RD = Recommendation Description, ERR = Event Report Request

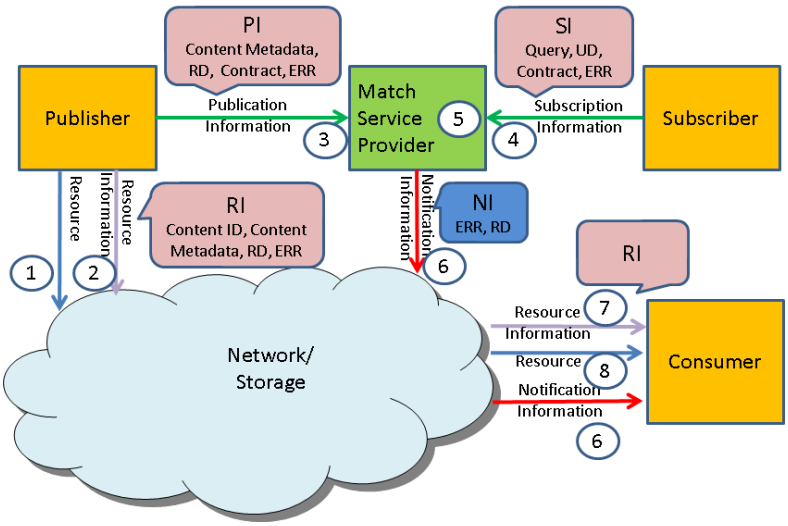


Figure 1: Relationships between users of Publish/Subscribe

Table 2 lists some of the acts performed by Publishers, Subscribers, MSPs and Consumers as a result of the relationships between them.

Table 2 – Possible acts in multimedia Publish/Subscribe

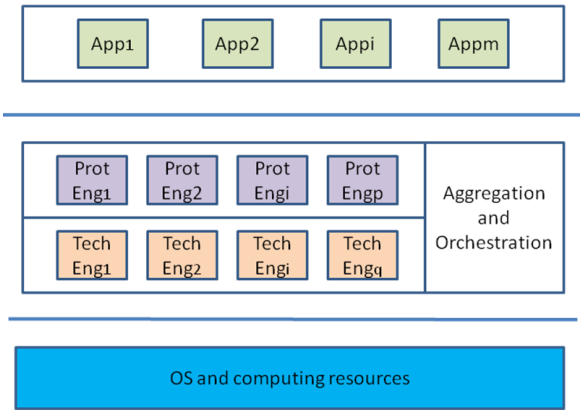
Party 1	Act	Party 2	Act
Publisher	Requests match of pub with subs	MSP	E.g computes statistics
Subscriber	Requests match of sub with pubs	MSP	E.g computes statistics
Publisher	Grants right to use (e.g. play) resource	Consumer	E.g. pays for play

Output documents

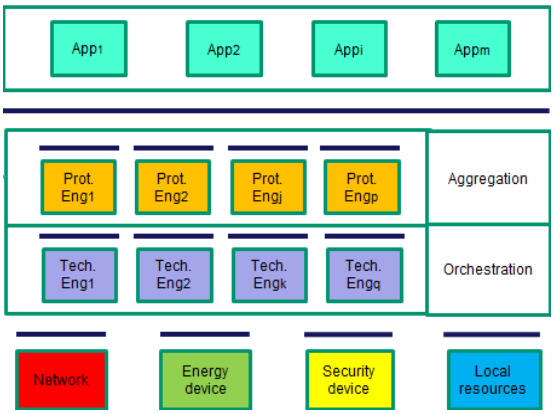
N15743 - Requirements for Publish and Subscribe Application Format (PSAF) v.2

3 MPEG-M Architecture (3rd Edition)

MPEG-M APIs are further refined in its 3rd Edition with respect to its 2nd Edition, as shown in the following figures. In particular, in its 3nd Edition, MPEG-M APIs are grouped in terms of: Network, Energy, Security and Local Resources.



MPEG-M device architecture (2nd Edition)



MPEG-M device architecture (3rd Edition)

Output documents
N15665 - Text of ISO/IEC CD 23006-1 3rd edition

4 Media Linking Application Format (MLAF)

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.

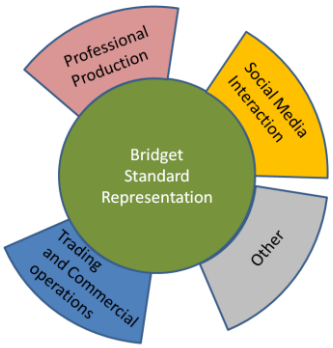


Figure 3: Bridget creation workflow.

Functional Requirements

1. Source and Destination Content of a bridget can be
    a. A file or a stream
    b. A static medium (e.g. an image) or a dynamic medium (e.g. a video)
2. A bridget shall include a unique and persistent identifier
3. A bridget may include authoring information (e.g. title, date, copyright, etc...)
4. A bridget shall include identification of its Source Content
5. A bridget shall include information about the link between Source and Destination Content
    a. A bridget shall include reference to
        i. A specific time instant of a dynamic medium, or
        ii. A specific time period of a dynamic medium, or
        iii. The entire duration of a dynamic medium
    b. A bridget may include reference to
        i. A specific spatial region of a static medium
        ii. A specific spatiotemporal region (e.g. a 2D or 3D object) of a dynamic medium
    c. A bridget may include metadata describing the context of the link between source and destination content (e.g., the semantics of the relation between source and destination)
6. A bridget may include metadata describing the Source Content
    a. Title, date, copyright, etc...
    b. Content type, media duration, MIME type, file size, etc...
7. A bridget may include textual descriptions of the Source Content
8. A bridget shall include identification of Destination Content
9. A bridget may include metadata describing the Destination Content
    a. Title, date, copyright, etc...
    b. Content type, media duration, MIME type, file size, etc...
10. A bridget may include textual descriptions of the Destination Content
11. A bridget may include information on how a bridget should be presented
12. A bridget may include an icon or image to visually present the bridget to the user
13. A bridget may include information on visual/acoustic presentation of the Destination Content
14. A bridget may include consumption-related information, e.g.
    a. Parental rating
    b. Accessibility
    c. Device requirements (e.g. stereo headphones)
    d. Target users
15. A bridget may include rights and payment information for destination content consumption
16. A bridget may include information about the service providing the destination content

Enabling Technologies

- MPEG-4 Binary Format for Scenes (BIFS)
- MPEG-21 Digital Item Declaration (DID)
- MPEG-7 Audio Visual Description Profile (AVDP)
- EBU Core Metadata Set

Two examples are currently provided for the MLAF standard verification:

- Example 1 - This example illustrates how bridget production metadata are represented using MLAF.
- Example 2 - This example shows a simple usage of the MLAF XMTA Restricted Scene. The presentation shows a logo (media1.jpg) on a background (media0.jpg). The logo starts zooming in and out once clicked and it stops when clicked again.

Output documents

N15656 - Text of ISO/IEC DIS 23000-18 Media Linking Application Format

N15657 - Work plan for reference software of ISO/IEC 23000-18 Media Linking Application Format

5 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 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
N15352 - Requirements for Media Value Chain Ontology
N15646 - WD of ISO/IEC 21000-19 AMD 1 Extensions on Time Segments and Multi-Track Audio

6 Exploration of multi-stream DASH delivery of audio, including MPEG-H 3D Audio

Next generation audio coding technologies enable personalization use cases which draw benefits from components delivered in individual streams and potentially over different networks. Such use cases require the audio components being delivered individually to be combined in the audio decoder at the receiver.

New audio codecs enable content creation in many ways. The term audio scene describes the composition of a particular audio experience presented to the user. Audio scenes can be encoded to form single self-contained streams. Alternatively the audio components (like dialogue or music and effects) that form one audio scene can be encoded separately so that they can be delivered in multiple individual elementary streams. Delivering components of an audio scene in separate streams enables service providers to deliver their content more efficiently for the case that not all components are of interest for the whole audience.

The idea is to deliver only essential audio components via the available resources on bit rate constrained broadcast emission paths and to provide additional components that belong to the same scene through alternative routes and other networks. This very flexible concept enables richer and more personalized user experiences compared to options available when using just the quite limited capacity of a broadcast emission path. Another reason for providing audio as components is that in a DASH on-demand environment the user is enabled to request only those audio components that are of interest for him, e.g. one dialogue only in one language or a number of music stems.

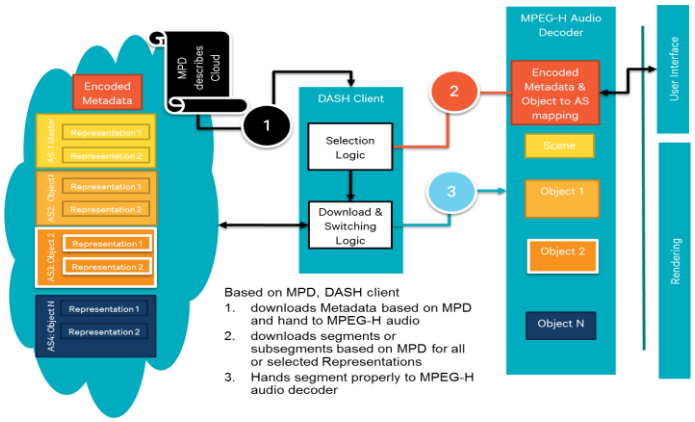


Figure 4: Basic setup for MPEG-H audio based DASH delivery

Output documents
N15687 - Exploration of multi-stream DASH delivery of audio, including MPEG-H 3D Audio

7 Exploration on playing multiple MPDs using playlist

In over the top (OTT) delivery, the content may come from difference origin servers and a content aggregator composes the linear program. Different content servers may provide live or on-demand content. The aggregator may be considered being a server on the network or an application at the device. The following figure demonstrates the general frame work.

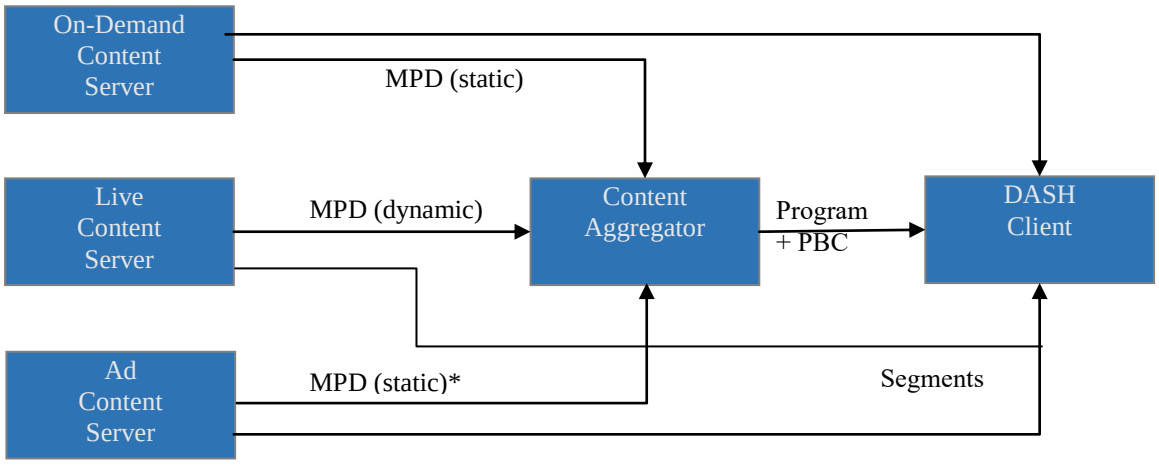


Figure 5: Live Content Aggregation

In this Figure:

1. The live content server provides the live content. The initial content offering is delivered through an initial MPD and any updates will be delivered using MPD updates.

2. The on-demand content server hosts on-demand content and their manifest.

3. The ad content server is also an on-demand content server which provides the ad content and their manifest. Note that the ad-server content might be in DASH format or other streaming formats in some use-cases.

4. The program aggregator server combines the live, on-demand and ad content together into a "program" or "channel" and provides it to the DASH client. Therefore, in the context of this CE, program/channel is a composed media presentation that its various parts may come from different on-demand and live origin server and its media time-line is not necessarily linear. The media segments are delivered from the original servers.

5. Along with the program/channel, the program aggregator may provide playback control (PBC) behavior to the client. The behavior may require the client to apply different rules for different parts of the program (on different periods) or require certain order of playback (for instance play back of ads before the content).

Output documents
N15811 - Exploration on playing multiple MPDs using playlist

8 Draft Call for Proposals on Media Orchestration Technologies

MPEG is currently 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 Preliminary Call for Proposals MPEG would like to make industry aware that it currently intends to issue a Call for Proposals ("CfP") for Media Orchestration technologies at its 114th meeting in San Diego, February 2016.

The Context and Objectives for Media Orchestration (N15734) provides background on the subject matter for the Call for Proposals, as well as a draft for a Functional Architecture. The Requirements for Media Orchestration (N15735) provide the requirements that those wishing to respond to the CfP will be asked to address.

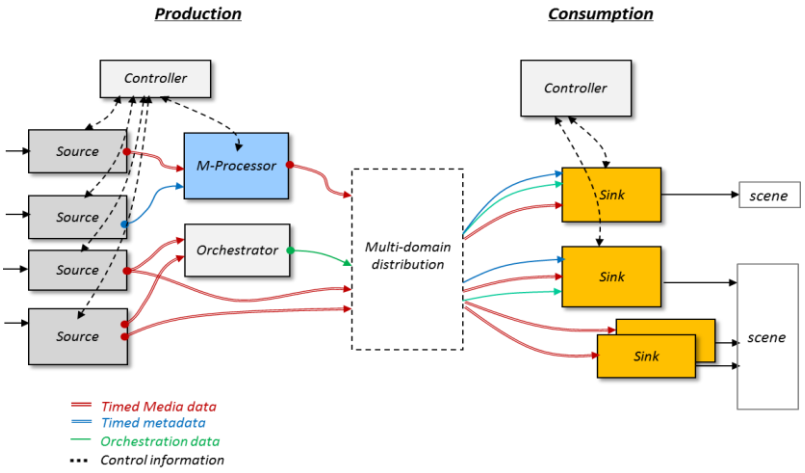


Figure 6: Example of Timed Data in the context of media orchestration

Output documents
N15736 - Draft Call for Proposals on Media Orchestration Technologies
N15735 - Requirements for Media Orchestration v.2
N15734 - Context and Objectives of Media Orchestration v.2

9 3D Audio-Visual formats

The figure below shows a general block diagram of an audio/visual system where six interfaces are defined.

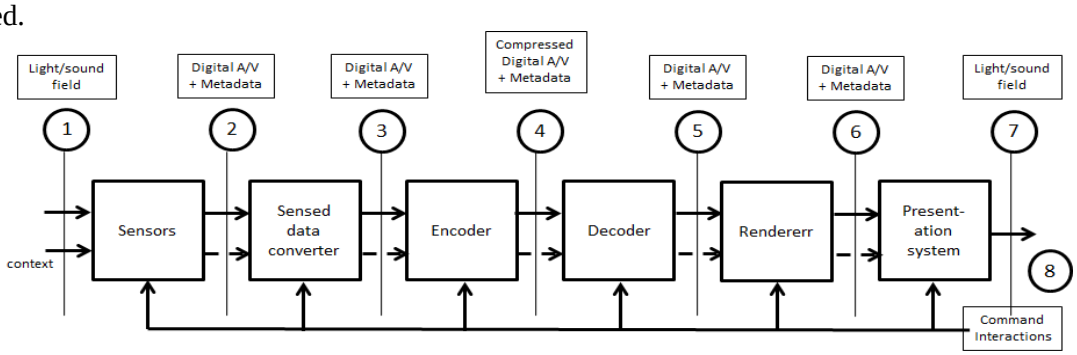


Figure 7: Components and interfaces of a 3D A/V system

Legend
a. **Sensors:** a set of devices capable of sensing the light/sound field and the context in which the field is captured
b. **Sensed data converter:** converts the sensed data to a format that can be used to feed the encoder
c. **Encoder:** compresses audio-visual information
d. **Decoder:** processes the data from Encoder to produce a version of the data input to Encoder
e. **Renderer:** converts the data from Decoder for presentation
f. **Presentation System:** generates audio and visual fields from the data generated by Decoder
g. **Command interactions**: converts human actions or information from environment to input to the 3D AV system

Output documents
N15731 - Requirements und Use Cases for Omni-directional Media Format
N15737 - Formats for 3DAV