122nd MPEG San Diego (CA), USA, 16 - 20 April 2018, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 Coded representation of Immersive Media

A first set of specifications is required in time for a market launch of products and services in 2018. It is highly likely that MPEG can deliver solutions that are more optimised in a longer time frame, which allows for more experiments and development. Since many believe that major market launch of VR 360 services will happen in 2020, a next set of specifications can be delivered 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.

MPEG-I phases
Given these results, MPEG is planning standards in support of Immersive Media, including those for 360º Audiovisual Media, to be developed in the following phases:

Phase 1a

- Timing is what guides this phase
- Goal: to deliver a Final Draft International Standard for up to 3 degrees of freedom 360 VR by end 2017.
- This phase aims to deliver a complete distribution system
- 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

Figure 1 - 3DoF

Phase 1b

- This phase is mainly motivated by desire by a significant part of the market to launch commercial services in 2020
- It is intended for systems and services that deploy in 2020; the specification must be ready in 2019 (which may match 5G deployments)
- Phase 1b will be published as an extension of the Phase 1a specification; it will focus on VR 360 with 3 DoF, with some additional depth clues, that would, for instance, allow moving the viewpoint in a limited space. (Like in phase 1a, including monoscopic and stereoscopic).
- In addition, this phase is thought to comprise elements like:
  - Optimization in projection mapping
  - Further motion-to-photon delay reductions
  - Optimizations for person-to-person communications
- Unlike phase 1a, phase 1b should have some quality definition and verification

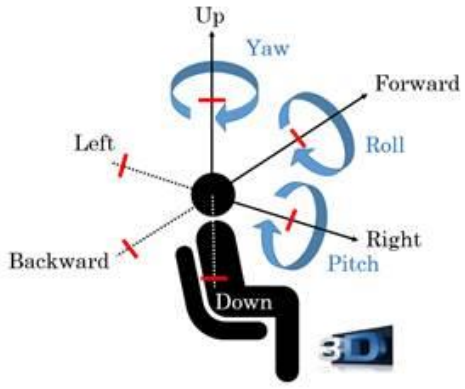


Figure 2 - 3DoF+

Phase 2

- A specification that is ready in 2021 or maybe 2022
- Goal is support for windowed 6DoF, omnidirectional 6DoF and 6DoF.
- Most important element probably new video codec with support for 6 DoF
- Audio support for 6 degrees of freedom
- Systems elements required in support of 6 DoF, as well as 3D graphics.
- Support for interaction with the virtual environment

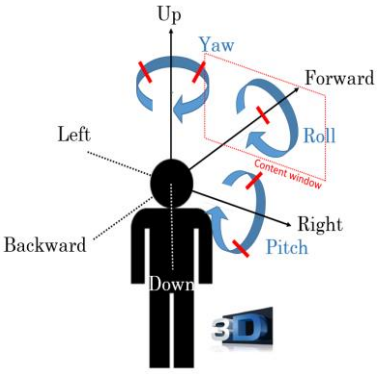


Figure 3 - Windowed 6DoF

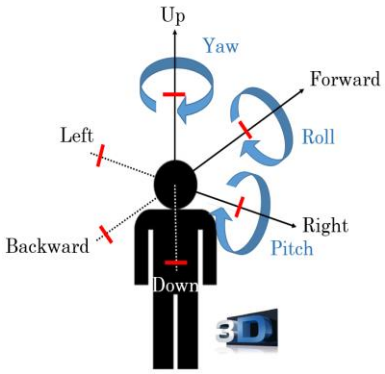


Figure 4 - Omnidirectional 6DoF

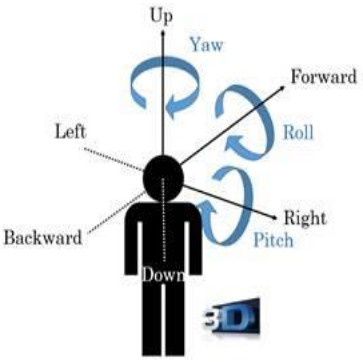


Figure 5 - 6DoF

Output Documents
N17504 - MPEG-I Use Cases phase 1
N17537 - Requirements on MPEG-I Scene Description
N17676 - Overview of Immersive media standards (v3)
N17686 - MPEG-I Project Plan

MPEG-I parts

Different technologies are used together to build immersive services. Consequently MPEG has structured MPEG-I as a suite of standards focusing on such specific technologies. The 5 parts of MPEG-I suite of standards are:

- Part 1 – Technical Report on Immersive Media (this present document)
- Part 2 – Application Format for Omnidirectional Media
- Part 3 – Versatile Video Coding
- Part 4 – Immersive Audio
- Part 5 – Point Cloud Compression
- Part 6 – Immersive Media Metrics
- Part 7 – Immersive Media Metadata
- Part 8 – Network-Based Media Processing (NBMP)

Part 1 – Technical Report on Immersive Media

This report forms the basis of the Project and investigate aspects of Immersive Media, which includes Virtual Reality, that are relevant to understand the needs for standardisation by WG11. This part is envisaged to be approved within one year.

The Technical Report will:

- Define a body of terminology - a vocabulary to be used across the Project.
- Define the elements of an immersive experience in the production and the consumption. The Technical Report will document minimum quality requirements as well as quality objectives for an uncompressed immersive media experience as a whole.
- Break down the elements that contribute to a full immersive media experience and assign quality requirements as well as quality and design objectives for those, preferably developing measurable metrics.
- Provide one or more integrated and architectural views on how these elements contribute to an overall immersive experience and how they are combined.

- Define an architectural view on the compression and coded representation of elements of immersive experiences as well as the coded representation and delivery of a full media experience, taking into account the individuality of the experience, while enabling scalable and efficient individual delivery as well as mass distribution. The representation and delivery will need to be mindful of rendering capabilities of consumption devices. This will ensure that the various technologies to be specified in the Project can integrate seamlessly. This Technical Report can possibly lead to a normative specification.
- Document standardization requirements to create interoperability in end-to-end systems. Such aspects are expected to include Audio, Video, Graphics and Systems with capture and rendering, as well as appropriate interfaces with sensors that record navigation in the immersive audiovisual space, as well as suitable formats for cost-conscious delivery to mass markets.

Output Documents
N17510 - MPEG-I Architecture
N17685 - Text of PDTR ISO/IEC 23090-1 Immersive Media Architecture

Part 2 – Application Format for Omnidirectional Media
This part defines a Media Application Format that enables storage and delivery of media content for an omnidirectional viewing experience. This part is envisaged to reach FDIS status at the end of 2017.

This part will reference technologies that include:
- Video coding profile(s), and a list of projection technologies that can be used for conversion of omnidirectional video content into a two-dimensional rectangular video;
- Immersive audio profile(s);
- Metadata for rendering and interaction;
- Encapsulation format that enable delivery using DASH and MMT

Output Documents
N17513 - Draft Requirements for MPEG-I Phase 2
N17683 - MPEG-I Phase 2 Use Cases
N17563 - Revised text of ISO/IEC FDIS 23090-2 Omnidirectional Media Format
N17584 - WD of ISO/IEC 23090-2 2nd edition OMAF
N17585 - Workplan on Reference Software and Conformance for OMAF
N17586 - Description and work plan of test framework for OMAF
N17684 - Defect report on ISO/IEC 23090-2 OMAF

Part 3 – Versatile Video Coding
This part represents the progression in visual coding technologies, to support the emerging and growing demand for higher efficiency video coding capabilities. In addition to the coding of traditional video content, this part will also address the new challenges presented by immersive video content by taking into account visual information that provides an increased sense of immersion beyond what existing video coding standards can provide. Immersion will be enabled by support of features that provide the viewer with the freedom to experience visual content with full parallax that is coherent to the movement of the user's viewing position and point of view, as well as to the motion of the objects in the scene – that is, the standard will provide up to 6 degrees of freedom of movement for the user. These features will be supported using coding technologies that facilitate the transmission and storage over networks whose deployment are envisaged to align with the completion of this standard.

- The JVET subgroup thanks Vittorio Baroncini for coordinating the test efforts, and the following test labs for conducting the subjective tests related to the CfP on Versatile Video Coding: BBC, DBW, EVATech, GBTech, Queen Mary University of London, RAI, Sisvel, University of Padova, and University of Western Scotland.
- The JVET subgroup thanks the following companies and organisations for responding to the CfP on Versatile Video Coding: DJI, Dolby, Ericsson, ETRI, Foxconn, Fraunhofer HHI, GoPro, Harbin Institute of Technology, HiSilicon, Huawei, InterDigital, KDDI, LG Electronics, MediaTek, NHK, Nokia, Panasonic, Peking University, Qualcomm, RWTH Aachen University, Samsung, Sejong University, Sharp, Sony, Technicolor, Tencent, TNO, University of Bristol, USTC, Wuhan University, Zhejiang University.

Output Documents
N17668 - Results from the Call for Proposals on Video Compression with Capability beyond HEVC
N17669 - Working Draft 1 of Versatile Video Coding
N17670 - Test Model 1 of Versatile Video Coding (VTM 1)

Part 4 – Immersive Audio
The goal of this Part is to provide audio coding for 6 degrees of freedom. Technologies may include ad-hoc sampled audio scenes 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 the capture, compressed representation and reproduction of audio sound fields is a means to providing a more realistic user experience of virtual audio presentations.

Output Documents
N17647 - Draft MPEG-I Architecture and Requirements
N17649 - Workplan for MPEG-I Evaluation Platform
N17650 - MPEG-I Audio Test Material

Part 5 – Point Cloud Compression
This Part will address compression of 3D visual media, in the form of point clouds. Point clouds can have attributes such as colors, material properties and/or other attributes. Point Clouds are typically captured using multiple cameras and depth sensors in various set-ups, however the acquisition is outside of the scope of this standard.
Point clouds typically have thousands up to billions of points to represent realistically reconstructed scenes.

The standard targets efficient geometry as well as attribute compression, scalable/progressive coding, and coding of sequences of point clouds captured over time. In addition, the compressed data format should support random access to subsets of the point cloud. The standard targets lossy compression useful for real-time communications, and lossless compression for, e.g., Geographical Information Systems, Computer-Aided Design, and cultural heritage applications. When combined with video, it enables immersive experiences that combine natural and synthetic content, and mixed reality applications.

Output Documents
N17675 - First ideas on Systems technologies for Point Cloud Coding
N17519 - PCC Test Model C13 v2
N17523 - Common Test Conditions for PCC
N17526 - PCC Test Model C2 v2
N17530 - PCC TMC13 performance evaluation and anchor results
N17531 - PCC TMC2 performance evaluation and anchor results
N17533 - PCC WD Categories 1, 3
N17534 - PCC WD Category 2
N17602 - CE 3.1 on kd-tree vs octree and plane encoding in octree
N17603 - CE 3.2 on point based prediction
N17613 - CE 3.4 on neighbour-dependent entropy coding
N17614 - CE 13.1 on lossy attributes coding
N17615 - CE 13.2 on point cloud tile and slice based coding
N17616 - CE 13.3 on lossless and near-lossless Compression
N17621 - CE 13.4 on entropy encoding symbols and contexts reduction
N17622 - CE 13.5 on inter-prediction of geometry coding
N17624 - CE 2.4 on lossless coding
N17625 - CE 2.7 on additional configurations of the video coding
N17630 - CE 2.8 on depth image coding
N17632 - CE 2.9 on occupancy map coding
N17633 - CE 2.10 on metadata coding
N17634 - CE 2.11 on in-loop vs out-of-loop texture and geometry interpolation
N17636 - CE 2.12 on visual quality improvement
N17637 - CE 2.13 on texture and depth padding
N17639 - CE 2.14 on auxiliary information coding

Part 6 – Immersive Media Metrics
This part specifies immersive media metrics and measurement framework to enhance the immersive media quality and experiences. This part also includes a client reference model with observation and measurement points to define the interfaces for the collection of the metrics.

Output Documents
N17564 - WD of ISO/IEC 23090-6 Immersive Media Metrics

Part 7 – Immersive Media Metadata
This part specifies immersive media metadata that can be consistently used in different application and system environments. The metadata includes definition of coordinate systems, projection formats, texture-to-sphere mappings, coverage definitions, or rotation parameters.

Output Documents
N17587 - WD of ISO/IEC 23090-7 Immersive Media Metadata
N17612 - Investigation of 3DoF+ Visual
N17617 - Call for Test Materials for 3DoF+ Visual
N17618 - Common Test Conditions on 3DoF+ and Windowed 6DoF

Part 8 – Network-Based Media Processing (NBMP)
With the development of immersive technologies in MPEG-I, such as 6DoF and AR/MR, the need for extensive media processing grows significantly. To address this challenge, part 8 will specify formats and interfaces to enable offloading of resource intensive immersive media processing tasks to the network. The Network-Based Media Processing framework will define external interfaces between the Media Source and the Media Processing Entities, that will allow users of the framework to access the framework, configure media processing tasks, upload/stream media data to the network for media processing, and access the processed media and the resulting metadata in real-time or in a deferred way. Workflow management that is used to configure media processing entities and to compose media processing services as a pipeline of media processing entities, is in scope of NBMP. The media and metadata formats that are used between Media Processing Entities in a media processing pipeline are also within scope.

Output Documents
N17502 - Use cases and requirements for NBMP (v4)
N17503 - Call for Proposals on Network-Based Media Processing
N17518 - Investigation on NBMP

2 Compressed representation of Neural Networks

NNR aims to define a compressed, interpretable and interoperable representation for trained neural networks. NNR shall be able to

- represent different artificial neural network types (e.g., feedforward networks such as CNN and autoencoder, recurrent networks such as LSTM, etc.)
- enable efficient incremental updates of compressed representations of NNs
- enable scalability, i.e. NNs of different performance can be obtained from subsets of the complete compressed representation
- inference with compressed network
- enable use under resource limitations (computation, memory, power, bandwidth)

The scope of existing exchange formats (NNEF, ONNX) is the interface between the framework used for training and the acceleration library/optimisation engine for a specific platform. If we consider a use case of deploying a trained network to a large range of target devices (e.g., mobile phones, signal processors in a vehicle), the process is likely to have these steps (see Figure 6):

- Training of the network with deep learning framework *L*, resulting in trained neural network *T*
- Export to an exchange format (neural network *T'*)
- Optimisation for 1,..,*n* target platforms, using acceleration/optimisation libraries A_1, …, A_n, resulting in optimised neural network O_1, …, O_n
- Distribute the networks O_k to the terminal devices, where they are executed with specific (software or hardware) inference engines I_k
- Alternatively, inference may be performed in an inference engine I* directly from the exported network, without using an accelerator library

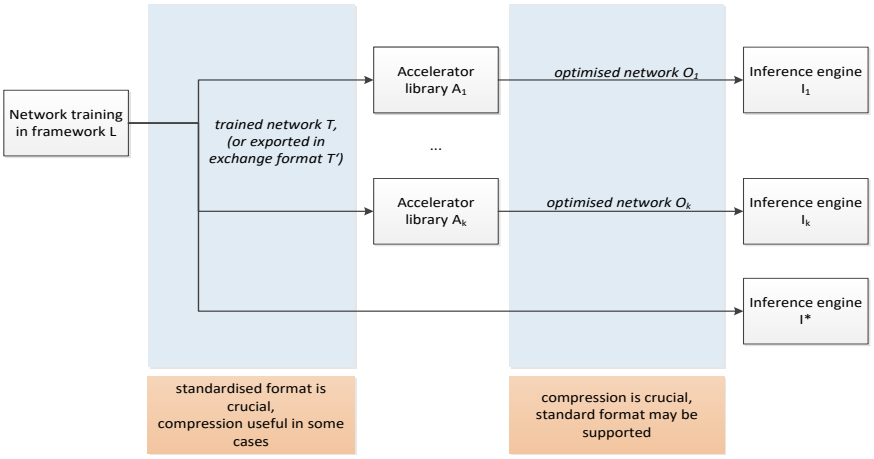


Figure 6 - Deployment/inference process.

We can make the following observations about the two interfaces:

- Compression based on a standard format would target *T'*, which is not where the volume of data is distributed
- Compression would be most beneficial for O_k, where networks are distributed to a large number of devices running the inference engine.
- O_k may already have a pruned network structure and quantised weights, so that compression is less (in the worst case not) effective
- O_k may not be easy to represent in a standard representation, and the format coming out of a specific acceleration library A_k may not be documented at all

Some of the NNR use cases would benefit from compression on the first interface, while for other only a compressed representation for the second interface would be beneficial.

A processing chain for neural network compression is shown in Figure 7. Some of the steps change (simplify) the model structure, while others affect only the weights. Several of these operations are also performed by the acceleration/optimisation libraries for specific target platforms.

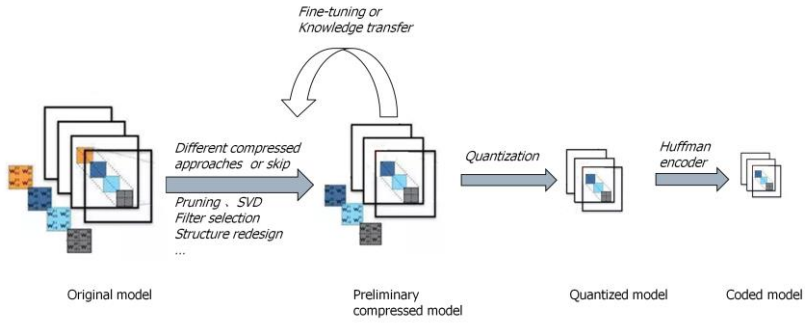


Figure 7 - Framework for neural network compression.

The following Table summarises the different requirements described, and the use cases that make reference to them, with respect to output documents.

	distribution									(re)training			processing			
	1	**2**	**3**	**4**	**5**	**6**	**9**	**14**	**15**	**7**	**8**	**10**	**11**	**12A**	**12B**	**13**
Compressed representation																
Exchange representation	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Efficient representation of the network	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Support lossless compression		X	X					X		X						
Support lossy compression	X	X	X					X		X			X			
Comparable performance of compressed network than original network	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Scalable compression	X							X	X				X			
Inference with compressed network		X	X		X						X		X			
Must support compression methods that do not require access to the original training data, may support others in addition	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

	distribution									(re)training			processing			
	1	**2**	**3**	**4**	**5**	**6**	**9**	**14**	**15**	**7**	**8**	**10**	**11**	**12A**	**12B**	**13**
Represent partial NN models	X		X	X	X	X	X		X	X	X	X				X
enable efficient incremental representation of NNs	X		X	X	X	X	X		X	X	X	X				X
Low computational complexity decoding		X		X	X	X		X	X	X	X		X	X	X	
Low memory consumption		X		X	X	X		X			X			X	X	
Support representation of different types of artificial neural networks											X	X		X	X	
Metadata																
Unique identification and time-stamping of an instance of a trained NN	X		X	X	X			X	X	X	X	X	X			
description of training metadata and reference to training data										X	X	X		X	X	X
Support validation of model updates	X					X						X				X
Metadata to support interpretability of trained models												X		X	X	
Metadata on required capabilities of inference engine	X		X		X	X			X			X	X	X	X	X
Extension mechanism to support application specific metadata												X		X	X	X
Robustness and security																
Error resilience of compressed representation	X			X				X				X		X	X	X
Ability to detect manipulation of the compressed network representation				X								X				
Other																
Interoperability of components using compressed and uncompressed versions of the network												X	X	X	X	
Signaling to negotiate relevant model updates						X						X				X
Mechanism to guide execution of NN models						X										X
Interface for providing training data										X	X	X		X	X	X
Signaling for negotiating the re-use of NN models							X					X				X
Future proof w.r.t delivery modes									X							
Robustness to adversarial attacks												X				

Output Documents
N17509 - Use cases and requirements for Compressed Representation of Neural Networks
N17623 - Draft Evaluation Framework for Compressed Representation of Neural Networks
N17680 - Call for Test Data for Compressed Representation of Neural Networks

3 MPEG-21 User Description (2nd Edition)

The use of multimedia is pervading more and more our daily life and getting evolved day by day into user-centric services. Although MPEG has developed standards related to User in MPEG-7, MPEG-21, and MPEG-M, they are probably superficial with respect to current and upcoming services like augmented reality and social network services. The goal is to develop an MPEG User Description (MPEG-UD) standard within the general scope of MPEG embedded within a framework that may be applicable to multiple application domains. In the following Table the use cases covered by the MPEG UD standard are shown.

Use Cases	
Use Case of Smart TV Recommendation Application	AnyAccess: Accessibility to the smart world
Scalable User Description	Damaged Vehicle
Providing User-Customized Contents	Incident report
Response to User's Emotion	Online shopping
Personalized Car Navigation	Car as a user
Casual Newsreaders	Scalable remote user interface
Professional News Editors	Visual communication
Infotainment	Advertisement for digital signage
MEdical Distributed Utilization of Services & Applications	Lossless-audio for users
Learning from User interaction with complex data for Recommendation	Companion Screen Service
Jini: An intelligent chatting partner in the smartphone	Ease and safe consumption of multimedia via loudness control on smart devices
Learn&Motiv: eLearning system based on the user’s motivation	Visual expression for user’s emotion and situation
TransTour: Automatic Translation system for tourists	Privacy protection

At the 114th MPEG San Diego meeting, MPEG successfully approved the 1st edition of MPEG-21 User Description (ISO/IEC 21000-22). Furthermore, MPEG-21 UD group published the Reference Software as AMD1 at the 117th MPEG Geneva meeting. The schemes for User Description, Context Description, Service Description and Recommendation Description in the standard were verified via the Reference Software for six applications proposed by the participants.

Since then, MPEG-21 UD group has worked for extending the scope of the standard so as to encompass new industrial applications. This resulted in the initiation of a 2nd edition of the MPEG-21 UD standard.

This 2nd edition of MPEG-21 UD standardizes the four data formats: User Description (UD), Context Description (CD), Service Description (SD), and Recommendation Description (RD).

- User Description (UD): a set of descriptions which may contain static and dynamic information about the user, including data, e.g., the history of the user's interactions, preferences and security settings.
- 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.) and security settings.
- Service Description (SD): a set of descriptions containing information about the service (or a set of sub-services), that is offered to the end-user application, e.g., video on demand and maps.
- Recommendation Description (RD): a set of descriptions containing information about recommended items, provided when a user 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.

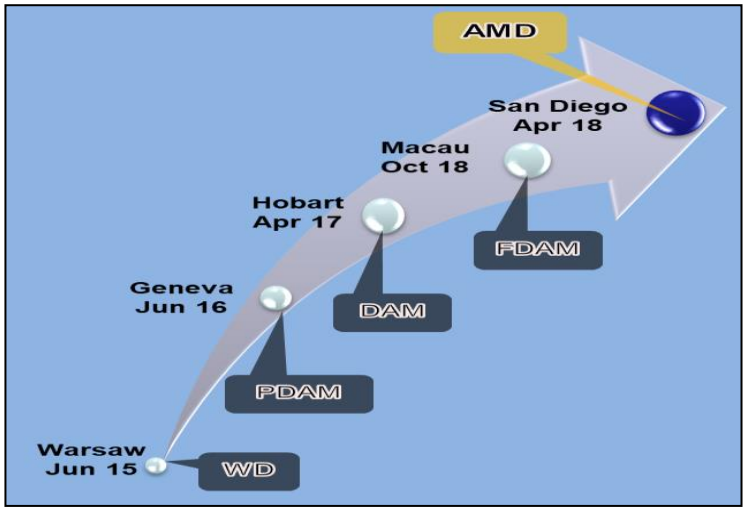
The DescriptionTools specified with respect to User Description, Context Description, Service Description and Recommendation Description are listed in the following Table. Their syntax and semantics are defined in the MPEG-21 UD standard.

DescriptionTools			
User Description	**Context Description**	**Sevice Description**	**Recommendation Description**
UserDescriptionType	ContextDescriptionType	BaseServiceType	RecommendationDescriptionType
UserProfileType	ContextIdentificationType	ServiceDescriptionType	compactUsageDescriptionType
PersonProfileType	DeviceCharacteristicsType	ServiceGeneralInformationType	QueryDescriptionType
OrganizationProfileType	NetworkInfoType	ServiceTargetInformationType	ProcessChainType
DeviceProfileType	WeatherType	ServiceTargetModelType	RecommendationInformationType
GroupedProfileType	OtherEnvironmentalInfo	VocabularySetType	RecommendableResourceType
UsageHistoryType	AudioEnvironmentType	ServiceInterfacesType	Resource
EventType	RecordingEnvironmentType	ServiceInterfaceType	resourceUsageType
InteractionAtomType	LoudnessEnvironmentTypeType	RequiredInputDataType	clusteringType
ArtefactType	VisualExpressionType	InternalServicesType	genericClusteringType
ObservableType		InternalServiceType	hierarchicalClusteringType
MultimediaExperienceType		AudioDBType	SequentialClusteringType
StateType		AudioDBDescriptorType	costFunctionMinimisationClusteringType
PreferenceType		VideoDBType	clusterStructureType
TextPresentationPreferencesType		VideoDBDescriptorType	genericAggregateType
WebLinkPreferenceType		ServiceObjectType	setMemberType
WebLinkUsageHistoryType		LoudnessInfoTypeType	orderedSetMemberType
ServicePreferenceType		VisualExpressionInfoType	genericSetType
GeneralAudioPreferenceType			labelledSetType
AudioPresentationPreferencesType			orderedSetType
AudioPresentationEnvironmentPreferenceType			equivalenceSetType
TranslationPreferencesType			linkageSetType
SpeechStylePreferenceType			Member
GenderType			OrderedMember
EmotionType			queryClauseType
ScheduleType			ORqueryClauseType
ScheduleEventType			ANDqueryClauseType
ActivityType			LoudnessControlTypeType
IntentionType			VisualExpressionType
LanguageType			
LanguageCompetenceReferenceType			
CompetenceLevelType			
AccessibilityType			
SocialInformationType			
KnowledgeType			
ObjectSharingType			
ObjectAccessibilityType			
UsagePatternType			
LoudnessPreferencesType			
VisualExpressionType			
BaseUserType			

Output Documents
N17567 - Text of ISO/IEC CD 21000-22 2nd edition User Description
N17568 - Use Cases for MPEG-21 User Description

4 Smart contracts for fair trade of music

At its 122nd meeting that took place in San Diego (CA), 16 – 20 April 2018, MPEG finalised its specification (ISO/IEC 21000-19 AMD 1) and associated reference software (ISO/IEC 21000-8 AMD 4) amendments on 'Media Value Chain Ontology (MVCO) Extensions to Time-Segments and Multi-Track Audio', also known, as 'Audio Value Chain Ontology (AVCO)'. Both of these amendments have completed successfully all standardisation stages and ballots, e.g., PDAM, DAM and FDAM and are currently in the process of publication as ISO/IEC international standards.

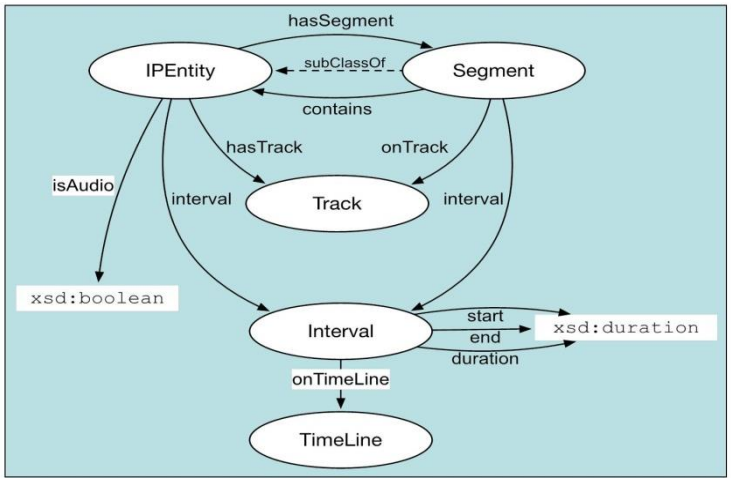


Audio Value Chain Ontology (AVCO) standardisation schedule.

MVCO facilitates IP rights tracking for fair and transparent royalties payment by capturing user roles and their permissible actions on a particular IP entity. However, widespread adoption of interactive music services, e.g., remixing, karaoke and collaborative music creation - thanks to MPEG-A: Interactive Music Application Format (ISO/IEC 23000-12) - raises the issue of rights monitoring when reuse of audio IP entities is involved, such as, individual tracks or even segments of them in new derivative works.

AVCO addresses this issue by extending MVCO functionality related to description of composite IP entities in the audio domain, whereby the components of a given IP entity can be located in time, and for the case of multi-track audio, associated with specific tracks. The introduction of an additional 'reuse' action enables querying and granting permissions for the reuse of existing IP entities in order to create new derivative composite IP entities.

Furthermore, MVCO/AVCO smart contracts by facilitating machine readable deontic expressions for permissions, obligations and prohibitions, with respect to particular users and IP entities, could be used in conjunction with distributed ledgers, e.g., blockchain, enabling both transparency and interoperability towards fair trade of music.



Audio Value Chain Ontology (AVCO) conceptualisation.

With qMedia::C4DM initiated MVCO Extensions on Time-Segments and Multi-Track Audio, 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 Extensions on Time-Segments and Multi-Track Audio Guidelines Document.

Output documents
Specification: ISO/IEC 21000-19 / AMD 1 MVCO Extensions on time-segments and multi-track audio
Reference Software: ISO/IEC 21000-8 / AMD 4 MVCO Extensions on time-segments and multi-track audio