1. **INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N17537**

**April 2018 – San Diego, CA, US**

|  |  |
| --- | --- |
| **Source** | **Requirements** |
| **Status** | **Output** |
| **Title** | **Requirements on MPEG-I Scene Descriptions** |
| **Author** | **Imed Bouazizi** |

# Introduction

ISO/IEC 14496-11 includes hybrid natural/synthetic scene technologies. After close to 20 years MPEG is interested in exploring newer and more functionality-rich solutions that may be available in the industry.

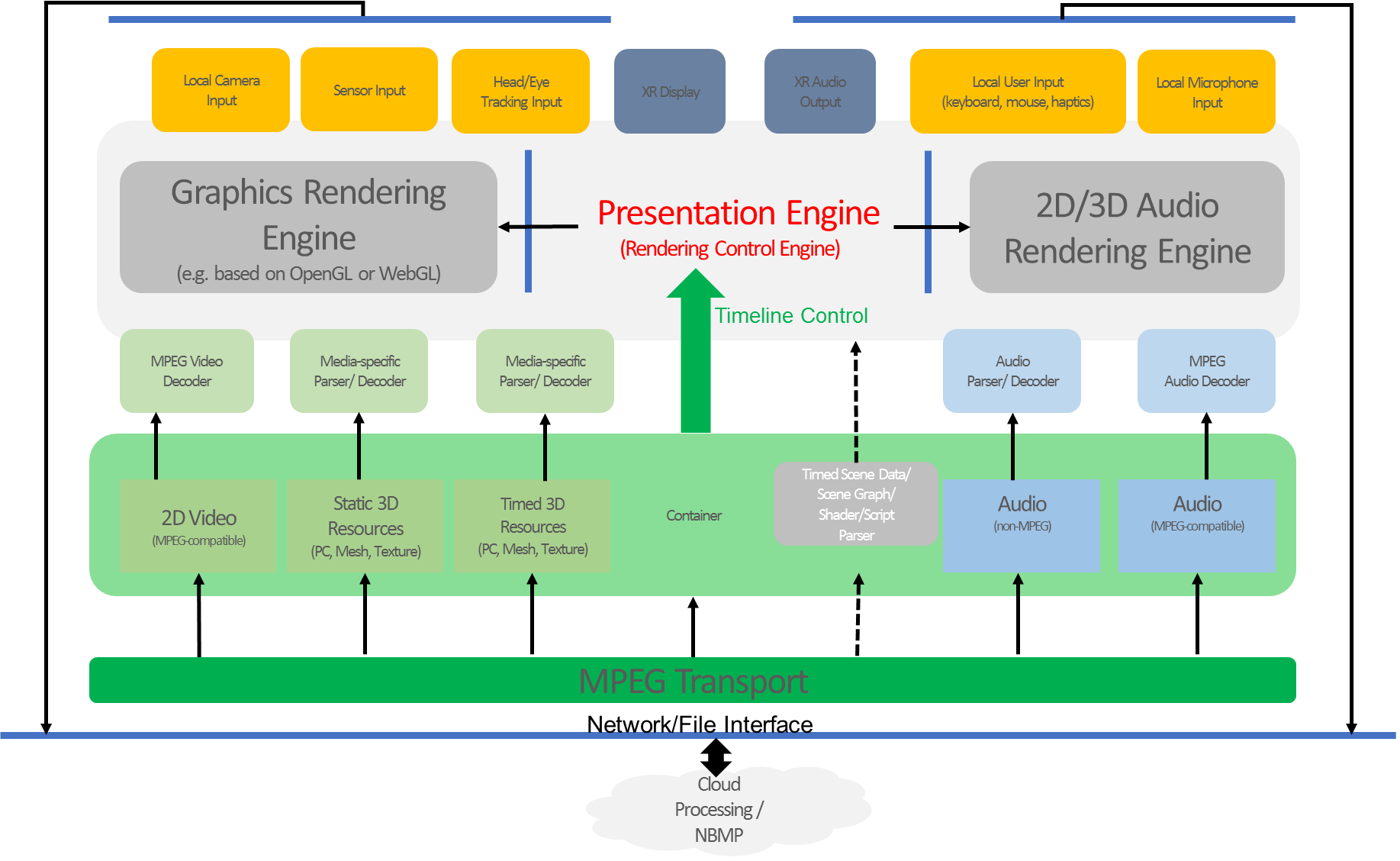
The scope of such a new work is the definition of a Format characterised by its ability to

1. Store 3D data that are
   1. *Synthetic*, i.e. created by a computer, and
   2. *Natural*, i.e. captured from the real world, and
2. Expose interfaces between components of the Format
3. To enable 3D interactive experiences.

In particular, this document assumes that MPEG will integrate existing and well-established scene description formats that fulfil MPEG-I’s requirements into an MPEG-I solution, starting from the architecture and definition of the interfaces. This document defines requirements for selection of the relevant scene description formats and the interfaces to the rest of the components of MPEG-I.

# Overview

MPEG-I is studying the definition of a new scene graph/description format as part of an effort to enable immersive applications and services based on 6DoF and AR/VR/MR technologies. The MPEG-I Architecture group has proposed the following architecture to be used as a reference for all MPEG-I technologies:



This document sets requirements on the interfaces and APIs that need to be developed to integrate existing scene description solutions as part of the overall MPEG-I architecture. For this purpose, a subset of the architecture is used as a reference.



# Definitions

***TBD***

# Requirements

1. General
   1. If possible, the solution shall define interfaces to integrate existing scene description formats rather than define a new scene description format
   2. The solution shall reuse existing interfaces/API definitions (also from other SDOs) whenever possible and appropriate
2. Scene Description Selection
   1. The scene description shall support audio, video and other media formats standardised by MPEG.
   2. The scene description shall enable the support of other visual or audio media formats.
   3. The scene description shall support definitions to indicate how sub-graphs and objects are related in terms of their temporal, spatial and logical relationships
   4. The scene description shall support composition of digital representations of natural and synthetic objects.
   5. The scene description shall support synchronisation between objects and attributes in the scene.
   6. The scene description shall support random access to spatial and temporal scene elements.
   7. The scene description should support information to enable a renderer to perform path tracing.
   8. The scene description shall support sub-graph representation that allows modular rendering e.g. leafs in the scene description tree can also be packaged and referenced individually from a parent scene description and container.
   9. The scene description should support nodes and attributes in order to implement natural laws of light, energy propagation and physical kinematic operations.
   10. The scene description should support nodes and attributes in order to implement natural laws of acoustic energy propagation and physical kinematic operations.
   11. The scene description should support description of ray-traced camera parameters for rendering
   12. The scene description shall support parametric models for use in rendering environmental acoustic behaviour (e.g. reverberation, occlusion and directivity).
   13. The scene description shall support references (e.g. URLs) to external media resources in place of embedded file references
   14. The scene description shall support a mechanism to safely customize behaviour for nodes like camera, texture, geometry, audio, and object placement nodes through sandboxed, validated domain specific shaders or scripts for these nodes without affecting the functionality or forcing changes to the root node graph or other node types; i.e. provide a mechanism to safely extend the scene description.
3. I-s Interface
   1. *It shall be possible to update the whole scene-graph, a sub-graph, or a node in the scene description*
   2. *It shall be possible to correctly render a 6DoF Presentation after a random access in time*
   3. *It shall be possible to perform timed scene description updates*
   4. *It shall be possible to associate a scene description update with the corresponding scene description*
   5. *It shall be possible to use a scene description as the entry point to a 6DoF presentation*
4. I-m Interface
   1. *It shall be possible to reference static and timed 2D media and 3D media (meshes, point clouds, …) stored locally or over the network*
   2. *It shall be possible to pre-fetch media that the presentation engine expects to be used in the presentation*
   3. *It shall be possible to retrieve and access referenced media partially in time and space*
   4. *It shall be possible to describe position, orientation, and visual/acoustic characteristics when rendering referenced media*
5. I-l Interface
   1. *It shall be possible to discover and configure local capture modalities*
   2. *It shall be possible to adjust the presentation based on local capture modality availability*
   3. *It shall be possible to reference media objects that are captured locally using different capture modalities*
6. I-i Interface
   1. It shall be possible to discover user interactivity modules
   2. it shall be possible to define custom interactivity procedures based on input from the user or from the user’s devices and sensors
7. Export
   1. *The scene description shall support information to enable a renderer to output raster data (image, and video), volumetric data (point clouds, meshes, arrays of voxels, and reflectance fields) and audio.*
   2. The scene description shall support a scriptable export output node for asynchronously exporting (as a file stream or buffer) any or all parts of the scene description connected to a node into a simpler or flattened representation.

# References

[1] “ITU-T/ITU-R/ISO/IEC Common Patent Policy”, March 2007#