 ISO/IEC JTC 1/SC 29/WG 3 N0383

**ISO/IEC JTC 1/SC 29/WG 3**

**MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title:** Exploration Experiments for MPEG-I Scene Description

**Status:** Approved

**Date of document:** 2020-10-15

**Source:** ISO/IEC JTC 1/SC 29/WG 3

**No. of pages:** 17 (with cover page)

**Email of Convenor:** young.L@samsung.com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 3**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 3 N** **0383**

**October 2021, Virtual**

|  |  |
| --- | --- |
| **Title** | **Exploration Experiments for MPEG-I Scene Description** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | **20860** |

[1 Carriage of Random Access Support in Scene Description 3](#_Toc85228361)

[2 Dynamic Scene Update 8](#_Toc85228367)

[3 Codec Support in MPEG-I SD 14](#_Toc85228384)

1. **Carriage of Random Access Support in Scene Description** 
   1. **Introduction**

The goal of EE1 is to evaluate the proposed solutions on carriage of information, more specifically glTF objects and JSON Patches to enable temporal random access of MPEG scene description and the efficient storage of those elements. This will enable the Scene Description adhoc to validate random access capability, and to evaluate the efficiency and behavior of the proposed solutions. EE1 is expected to provide considerable insights in terms of the cumulative size of temporal random access data, and also the processing steps required, from the Presentation Engine perspective.

* 1. **Mandates**

The mandates for this EE are as follows:

* To study the types of samples (including their data structure and storage format) which enable temporal random access
* To study the types of samples (including their data structure and storage format) which enable efficient storage of samples for a sequence of scene descriptions documents
* To study possible coding structures to provide efficient access of required data, in terms of
  + The required number of samples to construct an independent version of a scene description
  + The total size of sample payloads to deliver the randomly accessible dynamic scene descriptions
  + The benefit of storing the required samples in either one track or two (or more) tracks
  + Processing steps to construct an independent version of a scene description for random access operation
  + Processing steps to construct an independent version of a scene description for plain playback operation (i.e., normal scene update with no trick play)
* To provide recommendations on the best practices for the composition of the various types of samples
* To mandate text to be incorporated into the Second DIS of ISO/IEC 23090-14
  1. **Participants**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participant | Contact | Email | Type | - |
| Samsung Electronics | Sungryeul Rhyu | [suzz.rhyu@samsung.com](mailto:suzz.rhyu@samsung.com) | P |  |
| Nokia Corp. | Lukasz Kondrad | [lukasz.kondrad@nokia.com](mailto:lukasz.kondrad@nokia.com) | P |  |
| Xiaomi | Emmanuel Thomas | [thomase@xiaomi.com](mailto:thomase@xiaomi.com) | P |  |

(P = proponent, C = cross checker)

* 1. **Information about proposed technologies**
     1. **Anchor: M56439**

See [1] for recommendation on the anchor.

See [2] and [3] for details and proposed specification text for the Second DIS accordingly.

* + 1. **Method M57079**

See [4] for details.

* + 1. **Method M57542**

See [5] for details.

* 1. **Evaluation**
     1. **Test sequence**
* ABC\_v2 is used as the test sequence #1.
  + Use tag: [release “v1.0”] of EE1 test framework [6]
  + Use coding structure that provides random access point on 3rd, 5th, 7th samples
* A sequence generator that automatically generate random sequence is expected to be available by following the timeline.
  + The generator will generate a sequence of glTF/JSON by randomly adding/removing nodes.
  + The number of Scene Description versions and frequency of the random access point can be configured by config file.
  + The sequence generator, one specific sequence from the execution, and the configuration were discussed as following bullet points:
  + Five parameters were discussed to generate test sequences.
    - The number of sequences
    - The number of scenes
    - The number of objects
    - The number of updated objects in a scene update
    - Use of repeated or unique scenes
  + The first and every fifth samples thereafter are RAP samples (e.g., 1st, 6th, 11th, 16th, and so on).
  + Parameter sets for the sequence generator generated test sequences is shown in Table 1.

Table 1. Test sequences for EE1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sequence number | The number of scenes | Initial number of objects | The number of updated operations  (CRUD) | Unique scenes | Description |
| SC#1 | 10 | 10 | 5 (50%) | All | Short sequence of small scenes |
| SC#2 | 100 | 100 | 50 (50%) | All | Medium sequence of medium scenes |
| SC#3 | 1,000 | 1,000 | 500 (50%) | All | Long sequence of large scenes |
| SC#4 | 10 | 100 | 5 (5%) | All | Few updates in small scene |
| SC#5 | 10 | 1,000 | 50 (5%) | All | Few updates in medium scene |
| SC#6 | 10 | 10,000 | 500 (5%) | All | Few updates in large scene |
| SC#7 | 10 | 100 | 100 (100%) | All | Lots of updates in small scene |
| SC#8 | 10 | 1,000 | 1,000 (100%) | All | Lots of updates in medium scene |
| SC#9 | 10 | 10,000 | 10,000 (100%) | All | Lots of updates in large scene |
| SC#10 | 10 | 100 | 50 (50%) | 3 unique scenes  A 🡪 B 🡪 C 🡪 B 🡪 B 🡪 A 🡪 **C** 🡪 B 🡪 C 🡪 A | Repeating scenes |

* + 1. **Evaluation criteria**
       1. **Evaluation framework**

For the evaluation of the technologies, the proponents are invited to use and add their proposal to the evaluation framework located at [7].

The proposed methods should be added to the ‘/methods’ folder of the framework.

* + - 1. **For random access**

In this test case, the file is expected to have a first track providing the sequence of scene description documents over time. On top of this, the file provides efficient access to a subset of those scene description documents for enabling random access, i.e., the media player can start at those positions with a minimum of processing. The following criteria are considered for evaluating such technologies:

* Test No. A-1) Total size of sample payloads (from all tracks).
* Test No. A-2) Random access: A simple python code using the evaluation framework to construct an independent version of v3, v5, v7 for test sequence ABC\_v2 and v1 and every fifth samples thereafter (v1, v6, …) for test sequence sc#1 to sc#10.

For this test case, the anchor to compare against is a track containing one sync sample for each complete scene description document at those positions.

* + - 1. **For efficient storage of glTF document sequence as samples**

In this test case, the file is expected to have one track providing the sequence of scene description documents over time. This track may have one or more sync samples for reducing the total size of the track but random access to arbitrary positions is not the purpose. The following criteria are considered for evaluating such technologies:

* Test No. B-1) The required number of samples to construct each independent version (v1, v2, v3, and so on).
* Test No. B-2) Total size of sample payloads (from all tracks if more than one although not expected).
* Test No. B-3) Plain playback: A simple python code or processing steps to construct an independent version of v2 from v1, v3 from v2, v4 from v3, and so on.

For this test case, the anchor to compare against is a track containing one sync sample for each complete scene description document.

* + 1. **Evaluation**

Evaluation experiments are to be carried out under the same evaluation criteria compared with other proposals. Option A should be included to assure participant’s common understandings on the evaluation.

* + - 1. **For random access**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Test no |  | Proposed method | Analysis on method A | Analysis on method B | Analysis on method C | Analysis on method D |
| A-1 | all versions | 000 bytes |  |  |  |  |
| A-2 | RAP | Code or report |  |  |  |  |

* + - 1. **For efficient storage of glTF document sequence as samples**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Test no |  | Proposed method | Analysis on method A | Analysis on method B | Analysis on method C | Analysis on method D |
| B-1 | v1 | 0 samples |  |  |  |  |
|  | v2 | 0 samples |  |  |  |  |
|  | v3 | 0 samples |  |  |  |  |
|  | v4 | 0 samples |  |  |  |  |
|  | v5 | 0 samples |  |  |  |  |
|  | … | 0 samples |  |  |  |  |
|  | vlast | 0 samples |  |  |  |  |
| B-2 | all versions | 000 bytes |  |  |  |  |
| B-3 | all versions | Code or report |  |  |  |  |

* 1. **Timeline**
* 2022-01-12: MPEG document upload deadline
* 2022-01-17: MPEG #137(online) meeting starts
  1. **References**

1. M57550, Report of Exploration Experiments 1
2. M56439, Carriage of glTF JSON documents and JSON patch documents
3. M57647, Specification text for selected method from EE1
4. M57079, Response to EE on Scene Description
5. M57542, Optimised glTF target file patching
6. <https://gitlab.com/mpeg-i/scene-description/ee1-framework/-/releases#v1.0>
7. <https://gitlab.com/mpeg-i/scene-description/ee1-framework>
8. **Dynamic Scene Update**
   1. **Introduction**

The goal of EE2 is to identify suitable enhancements for advanced dynamic scene updates in phase 2 of MPEG scene description. Based on the list of advanced dynamic scene update scenarios identified, and listed in this description, the potential solutions listed are expected to be used as a starting point for discussions and further contributions.

* 1. **Mandates**

The mandates for this EE are as follows:

* To study the different potential solutions for the list of advanced dynamic scene update scenarios, in particular the necessary metadata and their usage to enable such dynamic scene update scenarios.
* To provide recommendations on dynamic scene update scenarios and solutions to support such scenarios.
* To mandate text to be incorporated into the phase 2 of ISO/IEC 23090-14.
  1. **Participants**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participant | Contact | Email | Type | - |
| Samsung Electronics | Eric Yip | [eric.yip@samsung.com](mailto:eric.yip@samsung.com) | P |  |
| Nokia Corp. | Lukasz Kondrad | [lukasz.kondrad@nokia.com](mailto:lukasz.kondrad@nokia.com) | P |  |
| InterDigital | Ahmed Hamza | [ahmed.hamza@interdigital.com](mailto:ahmed.hamza@interdigital.com) | P |  |

(P = proponent, C = cross checker)

* 1. **Advanced dynamic scene update scenarios**

According to requirements 13, 21, 22, 23, 24, 85 and 86 as listed in the Annex of this document, the dynamic scene update scenarios gathered satisfy a combination of these requirements related to time and interactivity related scene updates,

The advanced dynamic scene update scenarios to be supported are:

* Timed dynamic scene updates
  + Wallclock time triggered dynamic scene updates
  + Presentation timed triggered dynamic scene updates (supported in DIS)
  + Conditional timed dynamic scene updates
* Event triggered dynamic scene updates
  + Trigger update to event scene and go to new presentation timeline
  + Trigger update to event scene and return to original scene
  + Trigger update to event scene and skip to different scene in same presentation timeline
  + Trigger update to skip to different scene in same presentation timeline
  1. **Potential solutions**

**Timed dynamic scene update scenarios**

**Wallclock time triggered dynamic scene updates [CD text]**

The current DIS text does not contain any support for the indication of a wallclock time to be used for triggering dynamic scene updates.

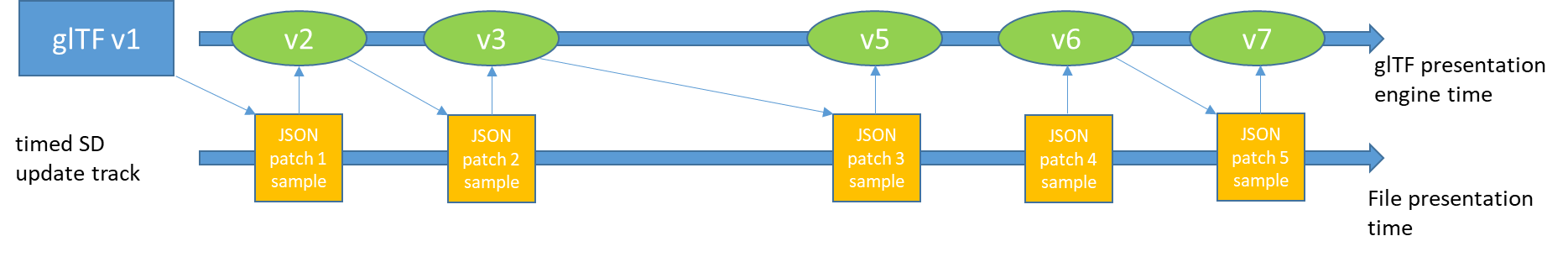
A potential solution is to include dynamic scene update attributes related to wallclock time triggers, as shown in table 1.

**Table 1 – Definitions of wallclock time attributes for dynamic scene updates**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Type** | **Default** | **Description** |
| absolute\_time\_UTC | DateTime | n/a | Wall clock time identifying the execution time of the scene update transaction on the glTF object. The value is denoted in UTC. |
| absolute\_time\_TAI | DateTime | n/a | Wall clock time identifying the execution time of the scene update transaction on the glTF object. The value is denoted in TAI. |

**Presentation time triggered dynamic scene updates [DIS text]**

The current DIS text supports timed dynamic scene updates which are triggered by presentation time. The presentation timestamp of the track samples containing JSON patch documents (as defined in clause 8.4 of the DIS) can be used to trigger the dynamic scene update as shown in figure 4.



**Fig. 4 Presentation time triggered dynamic scene updates**

**Conditional timed dynamic scene updates [CD text, m56736]**

Dynamic scene updates may be restricted by conditions such as the version of a scene description which is being maintained in the Presentation Engine memory at a given time. In such a scenario, the scene update sample is only applied when the version of the current scene description matches to that specified by the update sample.

One solution is to include version related dynamic scene update attributes; namely a target\_version\_id, which specifies the version of the target scene description for which the dynamic scene update is applicable, as well as a result\_version\_id, which specifies the resulting scene description when the dynamic scene update is applied. These attributes are shown in table 2.

**Table 2 – Definitions of version attributes for dynamic scene updates**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Type** | **Default** | **Description** |
| target\_version\_id | String | n/a | Identifier for the version of the target scene description for which the dynamic scene update is applicable |
| result\_version\_id | String | n/a | Identifier for the version of the resulting scene description when the dynamic scene update is applied |

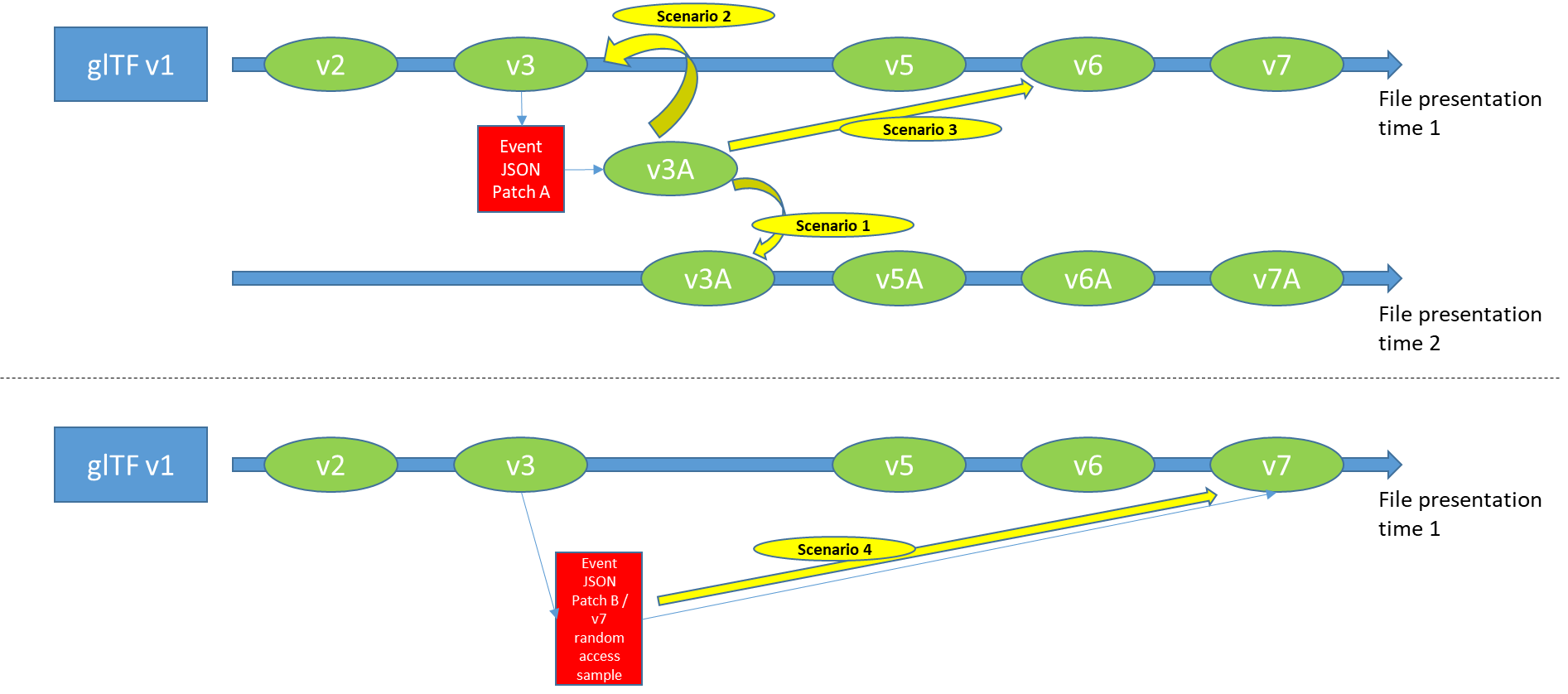
**Non-timed dynamic scene update scenarios**

**Event triggered dynamic scene updates [m56736]**

The current DIS text does not contain any support for the indication of dynamic updates which are triggered by interaction events.

The triggering of a scene update by an interaction event is highly content dependent, and can be expressed as content metadata through event scene update attributes. Unlike a timed scene update which typically progresses the timeline of the content, an event triggered scene update may allow for deviations from the main content timeline, hence allowing for loops or skips in the content timeline.

Depending on the content author’s intent, there may be four different event update scenarios as shown in figure 5.



**Fig. 5 Event triggered dynamic scene update scenarios**

Scenario 1

* An event A triggers the dynamic scene update of glTF v3 to glTF v3A. This update utilises the event update sample containing JSON Patch document A.
* The updated scene description (v3A) is executed by the presentation engine, after which further updates may be enabled through timed or non-timed scene updates.

Scenario 2

* An event A triggers the dynamic scene update of glTF v3 to glTF v3A. This update utilises the event update sample containing JSON Patch document A.
* The updated scene description (v3A), as a result of, and through the nature of the event, has a certain playout period (a certain presentation time period).
* After the playout of scene description v3A, the presentation engine returns the scene to the scene description version before the occurrence of the event (v3).

Scenario 3

* An event A triggers the dynamic scene update of glTF v3 to glTF v3A. This update utilises the event update sample containing JSON Patch document A.
* The updated scene description (v3A), as a result of, and through the nature of the event, has a certain playout period (a certain presentation time period).
* After the playout of scene description v3A, the presentation engine skips to a specified scene description version (v6).

Scenario 4 (a special case of scenario 1, mixed with scenario 3)

* An event B triggers the dynamic scene update of glTF v3 to glTF v7.
* glTF v7 is a scene description version which exists within the same presentation timeline of glTF v3, and so the update can also be considered a direct scene skip.
* This update may utilize an event update sample containing JSON Patch document B, or may be realised by the execution of a random access scene description sample containing glTF v7.

The support of these scenarios can be specified by event scene update attributes shown in table 3.

**Table 3 – Definitions of event related attributes for dynamic scene updates**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Type** | **Default** | **Description** |
| event\_id | String | n/a | Identifier for the event triggering the dynamic scene update |
| skip\_version\_id | String | n/a | Identifier for the version of the skip to scene description which is executed by the presentation engine, after the playout of the event updated scene graph |
| skip\_time | integer | n/a | Specifies the skip to time (referenced to the presentation time) of the scene graph version to skip to by the presentation engine, after the playout of the event updated scene graph |
| return\_event | integer | n/a | Flag to indicate whether the scene graph version is returned after the event scene playout. When set to 1, this flag indicates that the scene graph version is returned to the version previous to the event, after the playout of the scene graph version triggered by the event scene update. |
| playout\_time | integer | n/a | Specifies the playout time of the updated scene graph version triggered by the event scene update (in seconds, or any other time unit) |

* 1. **Timeline**
* 2022-01-12: MPEG document upload deadline
* 2022-01-17: MPEG #137(online) meeting starts

**Annex**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number | Requirements N19511 | Fulfillment gltf2.0, including N18869 | Status MPEG Extensions | Suitability as Test Scenario |
| Phase 2a Requirements | | | | |
| General | | |  | 1 – must have  2 – quite important  3 – nice to have  4 – if there is lots of time  5 – unnecessary |
| 13 | The scene description shall support definitions to indicate how sub-graphs and objects are related in terms of their temporal, spatial and logical relationships | This requirement is partially supported, since with exception of animations, all nodes of a scene graph are assumed to be active at time 0 and there is no concept of scene updates in glTF2. Regarding the logical relationships between media elements within a scene, means of interactivity and possible constrains to it should be defined. | Partially supported in gltf by nodes (spatially and logically).  Scene description updates are under development and partially addressed in DIS. More work in EEs and TuC. | 4 |
| 21 | It shall be possible to update the whole scene-graph, a sub-graph, or a node in the scene description | glTF2 does not come with a scene update mechanism.  MPEG needs to provide appropriate extensions to address this issue | Partially supported in the DIS, work ongoing in EEs. | 1  Need to define a scenario. This is part of the current discussion |
| 22 | It shall be possible to correctly render a 6DoF Presentation after a random access in time | glTF2 does neither have support for a timing model nor scene updates through time and as such every glTF2 is considered a random access point in time.  MPEG needs to provide appropriate extensions to address this issue | Partially supported in the DIS, work ongoing in EEs. | 2  Likely covered by extension from above |
| 23 | It shall be possible to perform timed scene description updates | glTF2 does not come with a scene update mechanism.  MPEG needs to provide appropriate extensions to address this issue | Partially supported in the DIS, work ongoing in EEs. | 1  See 21 |
| 24 | It shall be possible to associate a scene description update with the corresponding scene description | glTF2 does not come with a scene update mechanism.  MPEG needs to provide appropriate extensions to address this issue | Partially supported in the DIS, work ongoing in EEs. | 2  See 21 |
| 85 | It shall be possible to discover user interactivity modules | glTF2.0 supports this, but interaction with timed media still needs to be defined.  MPEG needs to provide appropriate extensions to address this issue | Left to phase 2b. Needs work |  |
| 86 | it shall be possible to define custom interactivity procedures based on input from the user or from the user’s devices and sensors | glTF2.0 supports this, but interaction with timed media still needs to be defined.  MPEG needs to provide appropriate extensions to address this issue | Left to phase 2b. Needs work |  |

1. **Codec Support in MPEG-I SD**
   1. **Background**

At the 136th MPEG meeting, WG7, WG4, and WG3 agreed jointly to establish a new EE as part of the MPEG-I Scene Description AHG to study and specify the necessary extensions to add support for the V3C codecs (V-PCC and MIV) in particular, and all immersive MPEG codecs in general.

The background of the discussion can be found in [1].

* 1. **Current Understanding**

It was established that when adding codec support, the following options are possible:

* Codec independent: in this option, all the necessary decoding and post-processing is performed to produce a primitive format that is natively supported by the Presentation Engine.
* Codec dependent: in this option, the Presentation Engine needs to have some level of support for the codec, in order to be able to render the object.
  + Variant a: in this variant, an intermediate uncompressed format is passed to the Presentation Engine for rendering. A Presentation Engine that supports this format may then load the appropriate shader programs to perform post-processing (e.g. 3D reconstruction) and rendering of the object. The Presentation Engine must support the intermediate format.
  + Variant b: in this variant, samples of the compressed stream are passed to the Presentation Engine for decompression, post-processing, and rendering. The Presentation Engine must support the compressed format.

These variants vary in the split of tasks between the media pipeline in the MAF and the Presentation Engine. The following diagram depicts example pipelines for these different options.



**Figure 1 Example Pipeline Options**

The following example show how the different options could be described in the MPEG-I SD glTF document:

**Table 1 Example glTF for Option 1**

|  |
| --- |
| *.*  *.*  *.*  {  *"name"*: ”vpcc\_longdress",  *"primitives"*: [  {  *"attributes"*: {  *"POSITION"*: 15,  *"COLOR\_0"*: 16  },  *"mode"*: 0  }  ]  }  *.*  *.*  *.*  *"extensions"*: {  *"MPEG\_media"*: {  *"media"*: [  {  *"name"*: "longdress",  *"timeOffset"*: 0.0,  *"alternatives"*: [  {  *"mimeType"*: "video/mp4;codec=v3e1.L2.0.0.1, avc1.4D401E",  *"uri"*: "https://example.com/vpcc\_longdress.mp4"  }  ]  }  ]  }  }, |

**Table 2 Example glTF for Option 2a**

|  |
| --- |
| {  *"name"*: "vpcc\_longdress",  *"primitives"*: [  {  *"attributes"*: {  *"\_MPEG\_V3C\_ATLAS\_0"*: 1136,  *"\_MPEG\_V3C\_GEOMETRY\_0"*: 1134,  *"\_MPEG\_V3C\_OCCUPANCY\_0"*: 1135,  *"\_MPEG\_V3C\_COLOR\_0"*: 1137  },  *"mode"*: 0  }  ]  }  *.*  *.*  *.*  *"extensions"*: {  *"MPEG\_media"*: {  *"media"*: [  {  *"alternatives"*: [  {  *"mimeType"*: "video/mp4;codec=v3e1.L2.0.0.1, avc1.4D401E",  *"tracks"*: [  {  *"track"*: "#track\_ID=1"  }  ],  *"uri"*: "https://example.com/vpcc\_longdress.mp4"  }  ],  *"loop"*: true,  *"timeOffset"*: 0  }  ]  }  }, |

**Table 3 Example glTF for Option 2b**

|  |
| --- |
| {  *"name"*: "vpcc\_longdress",  *"primitives"*: [  {  *"attributes"*: {  *"\_MPEG\_V3C\_POINTCLOUD"*: 165,  },  *"mode"*: 0  }  ]  }  *.*  *.*  *.*  *"extensions"*: {  *"MPEG\_media"*: {  *"media"*: [  {  *"alternatives"*: [  {  *"mimeType"*: "video/mp4;codec=v3e1.L2.0.0.1, avc1.4D401E ",  *"tracks"*: [  {  *"track"*: "#track\_ID=1"  }  ],  *"uri"*: " https://example.com/vpcc\_longdress.mp4"  }  ],  *"loop"*: true,  *"timeOffset"*: 0  }  ]  } |

* 1. **Mandates**

This EE will have the following mandates:

* Identify the MPEG codecs to be supported in MPEG-I SD
* Define the MIME type and any necessary signaling and extensions to enable options 1 and 2
* For codec dependent support:
  + Evaluate the codec-dependent options and decide on which ones to enable
  + Define the exact buffer formats and any necessary restrictions on the formats
  + Define any necessary glTF extensions and register any new \_MPEG attributes
  + Provide guidelines on how to implement the variant
* Provide test scenarios, assets, and implementation in the reference software
  1. **Support for V-PCC**
  2. **Support for MIV**
  3. **Support for G-PCC**
  4. **Participants**

The following EE participants are identified:

|  |  |  |
| --- | --- | --- |
| **Participant** | **Affiliation** | **NB** |
| Imed Bouazizi | Qualcomm | US |
| Basel Salahieh | Vimmerse | US |
| Lauri Illola | Nokia | FI |
| Lukasz Kondrad | Nokia | FI |
| Ahmed Hamza | Interdigital | CA |
| Gurdeep Bhullar | Interdigital | CA |
| Bart Kroon | Philips | NL |

* 1. **Timeline**
* 2022-01-12: MPEG document upload deadline
* 2022-01-17: MPEG #137(online) meeting starts
  1. **References**

[1] m58329, Codec Support in SD

[2] MIV Test Model, <https://dms.mpeg.expert/doc_end_user/documents/135_OnLine/wg11/MDS20596_WG04_N00112.zip>