Une image contenant texte

Description générée automatiquement ISO/IEC JTC 1/SC 29/AG 3 N197

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

**MPEG Liaison and Communication Convenorship: KATS (Korea, Republic of)**

**Document type:** Output document

**Title:** Convenor Report, the 19th SC 29/AG 3 Meeting, 2025-03-31/04-04, Geneve, CH

**Date of document:** 2025-04-06

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

**No. of pages:** 14 (without cover pages)

**Email of Convenor:** kyuheonkim@khu.ac.kr

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29ag3>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 2**

**MPEG TECHNICAL REQUIREMENTS**

**ISO/IEC JTC 1/SC 29/AG 3 N197**

**Online – April 2025**

|  |  |
| --- | --- |
| **Title** | **White paper on MPEG technologies for Metaverse** |
| **Source** | **AG 3, MPEG Liaison and Communications** |
| **Status** | **Approved** |
| **Serial Number** | **25193** |

# 1. Introduction to Metaverse in MPEG technologies

The Metaverse is an evolving concept, representing a vast, shared virtual space that merges real-world and digital experiences. As the Metaverse grows, the role of media standards such as developed by ISO/IEC JTC1 / SC 29 - MPEG (Moving Picture Experts Group) becomes critical for ensuring immersive, seamless, high-quality, and interoperable experiences. This white paper explores MPEG’s contributions to the Metaverse, spanning from foundational 3D graphics standards like MPEG-4 and MPEG-V to the cutting-edge MPEG-I technologies driving today’s immersive experiences. By tracing this evolution, we highlight how MPEG’s long-standing expertise in media standards supports the seamless, interoperable, and high-quality virtual environments envisioned for the Metaverse.

# 2. The challenge of defining the Metaverse

## 2.1 Introduction

Rather than attempting to provide a generic definition of the Metaverse, MPEG has selected a limited set of key parameters representative of what could be a Metaverse experience. Those parameters are listed below.

## 2.2 Real-time

The Metaverse functions in real time, meaning it enables users to interact instantly within a shared virtual space, similar to how communication occurs in physical reality. This immediacy is essential for creating a seamless, immersive experience, allowing users to engage with others, respond to events, and make decisions that affect their surroundings instantaneously. Real-time interactions build a sense of presence and connectivity, fostering environments where live collaboration, gaming, and social interactions feel as dynamic as real life.

## 2.3 3D experience

The Metaverse enables the augmentation of our 3D physical world or to replace it with a virtual environment which improves productivity and engagement. Unlike flat, 2D interfaces, its 3D environments allow for a richer, more immersive experience where users can explore depth, spatial relationships, and varied perspectives. This 3D framework helps in building realistic worlds, making interactions and environments more engaging, especially when accessed via VR (Virtual Reality) or AR (Augmented Reality) technologies.

## 2.4 Interactivity

Interactivity is a key feature of the Metaverse, where users can actively engage with both the environment and other participants. This interactivity ranges from giving the user freedom to change their pose, to enabling users to make choices, manipulate objects, create content, inject media into the virtual space, and influence the outcomes of experiences. The interactivity fosters personalized experiences, giving users the agency to shape their virtual surroundings and engage meaningfully with the digital landscape.

## 2.5 Social impact

The Metaverse is largely social, designed to bring people together across geographic barriers. It encourages socialization and community-building, where users can collaborate, play, or simply spend time together in shared spaces. Social elements are foundational to the Metaverse, enabling users to express themselves, form friendships, and build communities that might not be possible in the physical world, thereby enhancing the sense of connection and belonging.

## 2.6 Persistence

Persistence in the Metaverse means that its environments and interactions continue to exist and evolve even when a user logs off. Much like the physical world, the Metaverse is always "on," with user-created changes and interactions remaining in place. This continuity ensures that users can return to a stable, evolving environment, making their digital contributions lasting and giving a sense of ongoing reality within the virtual world.

## 2.7 Mapping to MPEG activities

Each of the above parameters were assessed when considering the applicability of MPEG activities to Metaverse experiences. The resulting technologies listed in this white paper are therefore considered as technical enablers suitable for the implementation of Metaverse services.

# 3. Key scenarios in the Metaverse supported by MPEG

Various scenarios in the Metaverse require robust support from MPEG technologies to ensure optimal performance. These scenarios highlight the essential elements of user interactivity, persistence, and immersive experiences.

## 3.1 Virtual environments for digital interactions

In virtual environments, users interact with digital assets such as avatars, objects, and spaces. These environments could be anything from social meeting spaces to virtual exhibitions or workspaces. Users can personalize avatars and environments, engaging in activities that mirror real-world interactions, such as meetings, presentations, or collaborative projects.

Une image contenant texte, diagramme, reçu, Plan

Le contenu généré par l’IA peut être incorrect.

Figure – High level service architecture example of virtual environment for digital interactions (Virtual Museum case)

**Common requirements:** Real-time rendering of digital assets, interaction with 3D environments, and seamless integration of different 3D formats are necessary for a fully immersive user experience.

## 3.2 Immersive entertainment experiences

In the entertainment domain, users might engage in immersive experiences like virtual concerts, theater performances, or multiplayer games. These scenarios involve real-time interaction with virtual worlds, where users can navigate through environments, interact with other participants, and engage with dynamic digital content. The media content may vary significantly depending on the devices being used, such as VR headsets, holographic displays, or augmented reality glasses.

Une image contenant texte, diagramme, Plan, Parallèle

Le contenu généré par l’IA peut être incorrect.

Figure – High level service architecture example of immersive entertainment experiences (Live performance case)

**Common requirements:** Real-time streaming and synchronization of high-resolution 3D environments across various devices, support for multiple immersive displays, and low-latency communication. Additionally, functionalities offering accessibility and personalization features of streamed content is of particular importance.

## 3.3 Virtual commerce and digital asset interaction

Virtual commerce involves the ability to engage with digital products, spaces, and services within the Metaverse. For example, a user could browse and purchase digital assets, ranging from virtual goods to real-world items represented virtually. Users could also interact with complex 3D representations of products, making detailed evaluations in digital showrooms or virtual marketplaces.

Une image contenant texte, diagramme, Parallèle, Plan

Le contenu généré par l’IA peut être incorrect.

Figure – High level service architecture example of virtual commerce and digital asset interaction (Virtual dressing room case)

**Common requirements:** High-fidelity 3D asset rendering, secure and reliable streaming of digital products, and low-latency interaction for seamless commercial transactions.

## 3.4 Remote collaboration and mixed reality

Virtual workspaces enable collaboration across distances by replicating real-world office environments or creating entirely new digital spaces. In this scenario, users interact with virtual tools, documents, and colleagues in real-time, as if they were in the same physical location. Collaboration tools might include virtual whiteboards, 3D models, or interactive presentations.

Une image contenant texte, diagramme, Plan, reçu

Le contenu généré par l’IA peut être incorrect.

Figure – High level service architecture example of remote collaboration and mixed reality (AR two-party call case)

**Common requirements:** Real-time synchronization of virtual assets, low-latency communication between users, and support for a variety of interactive media formats.

## 3.5 Digital twin systems for monitoring and control

Digital twin systems involve creating virtual representations of physical assets, processes, or systems, with numerous applications in industry, engineering, mobility and smart cities, among others. For example, in a manufacturing scenario, a digital twin could represent a production line, enabling real-time monitoring, simulation, and predictive maintenance. Users can interact with these digital models, observe updates from sensors in the physical environment, manipulate parameters, visualize the impact of changes in real-time, and perform simulations.

Une image contenant texte, diagramme, Plan, ligne

Le contenu généré par l’IA peut être incorrect.

Figure – High level service architecture example of Digital twin systems for monitoring and control (B2B digital twin case)

**Common requirements:** Real-time data integration, high-fidelity rendering of 3D objects, and the ability to simulate and interact with complex systems.

# 4. MPEG technologies supporting the Metaverse

MPEG standards play a key role in enabling many aspects of the Metaverse experiences. Below is a breakdown of some relevant technologies that support various Metaverse scenarios. Notably MPEG’s ISO/IEC 23090 standards (also known as MPEG-I) are essential for the representation of assets and scenes in the Metaverse.

## 4.1 Evolution of MPEG 3D Graphics Standards

Over the past 25 years, MPEG has developed a robust suite of 3D graphics standards that have progressively built the foundation for Metaverse-like experiences. This journey began with MPEG-4, introduced in 1998, which extended beyond video and audio to include advanced 3D graphics capabilities. A key component, the Binary Format for Scenes (BIFS, ISO/IEC 14496-11), enabled structured descriptions of interactive 3D scenes, allowing objects, animations, and user interactions to be encoded efficiently. BIFS provided an early framework for virtual environments, supporting features like real-time rendering and scene compositing that are essential to the Metaverse’s interactivity and persistence.

Further advancing 3D capabilities, MPEG-4’s Animation Framework eXtension (AFX, ISO/IEC 14496-16) introduced tools for representing and animating complex 3D objects, including meshes and skeletal structures. AFX offered efficient representation and compression techniques, laying groundwork for later standards like dynamic mesh coding. Additionally, MPEG-4 Part 25 (ISO/IEC 14496-25) defined a 3D Graphics Compression Model, offering a standardized method for compressing diverse 3D data types and serving as a precursor to modern MPEG techniques for compressing arbitrary scene graph formats. These early efforts established MPEG as a pioneer in 3D graphics, directly influencing the immersive and interoperable systems required by the Metaverse today.

## 4.2 MPEG-V: Sensory Effects and Virtual-Real Integration

Published in 2011, MPEG-V (ISO/IEC 23005) marked one of MPEG’s earliest efforts to standardize interactions between virtual and physical worlds, a foundational concept for the Metaverse. Also known as 'Media Context and Control,' MPEG-V defined formats for sensory effects—such as light, wind, temperature, and vibration—synchronized with multimedia content to enhance immersion. This standard enabled devices like 4D cinema systems or haptic interfaces to respond in real time to virtual scenes, bridging the gap between digital experiences and physical sensations.

MPEG-V’s focus on real-time sensory feedback and device interoperability aligns with the Metaverse’s goals of interactivity, presence, and social engagement. For example, its Sensory Effect Description Language (SEDL) allowed creators to script multi-sensory experiences, while its architecture supported integration with virtual environments—a precursor to the multi-sensory immersion targeted by modern Metaverse platforms. For more details, see: ISO/IEC, MPEG-V Standard Overview, ISO/IEC JTC1/SC29/WG11, available from: https://www.mpeg.org/standards/MPEG-V/. While MPEG-V emphasized sensory integration, subsequent MPEG-I standards build on this legacy by enhancing 3D scene representation and real-time streaming, creating a more comprehensive Metaverse framework.

## 4.3 Dynamic point cloud compression

Building on MPEG-4’s early 3D compression efforts, such as Part 16’s Animation Framework eXtension, V-PCC and G-PCC represent a significant advancement in handling dynamic, high-density point clouds for real-time Metaverse applications.

V-PCC (Video-based Point Cloud Compression - part of ISO/IEC 23090-5) is designed to efficiently compress assets represented as dense dynamic point cloud data by projecting 3D points onto 2D video frames. This method leverages existing video compression technologies to handle the geometry and texture information of point clouds, making it suitable for applications requiring real-time transmission and rendering.

G-PCC (Geometry-base Point Cloud Compression - ISO/IEC 23090-5) focuses on compressing static point cloud data by exploiting the geometric properties of 3D objects. It employs techniques such as octree decomposition and predictive coding to achieve high compression efficiency. Although designed for applications like automotive lidar, 3D objects archival and cultural heritage preservation, it can also be used for compression of dense dynamic point cloud assets.

*For more details on G-PCC see:*

* *ISO/IEC, White Paper on MPEG-I Geometry-based Point Cloud Compression, 2023, ISO/IEC JTC1/SC29/WG6, MPEG 3D Graphics. Available from:* [*https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/142\_Antalya/w22804.zip*](https://www.mpeg.org/wp-content/uploads/mpeg_meetings/142_Antalya/w22804.zip)

Une image contenant croquis, dessin humoristique, squelette

Le contenu généré par l’IA peut être incorrect.

Figure – Samples of point cloud assets

## 4.4 Dynamic mesh compression

V-DMC extends the animation and mesh capabilities introduced in MPEG-4’s Animation Framework eXtension (AFX), adapting them for the bandwidth and temporal demands of Metaverse environments. V-DMC (Video-based Dynamic Mesh Coding - ISO/IEC 23090-29) is designed to efficiently compress assets represented as dynamic 3D mesh supporting time varying connectivity information and time varying attribute maps. By utilizing video-based techniques, V-DMC achieves efficient compression while maintaining the temporal coherence of dynamic meshes. MPEG’s V-DMC standard is meant to allow for real-time transmission and manipulation of 3D objects, reducing bandwidth requirements without compromising visual quality.

Une image contenant texte, chaussures, croquis, diagramme

Le contenu généré par l’IA peut être incorrect.

Figure – MPEG V-DMC encoder workflow

*For more details see:*

* *ISO/IEC, Video-based dynamic mesh coding web page, ISO/IEC JTC1/SC29/WG7, MPEG 3D Graphics Coding. Available from:* [*https://www.mpeg.org/standards/MPEG-I/29/*](https://www.mpeg.org/standards/MPEG-I/29/)

## 4.5 Immersive video

Immersive video helps to create a sensation of presence that is difficult to achieve with graphics alone. MPEG immersive video (MIV, MPEG-I Part 12) bridges the gap between video and 3D graphics by rendering video natively in 3D space. The video can be captured by a modest number of (range-sensing, industrial or professional) cameras, and it can be combined with graphics and other elements through MPEG-I SD.

Une image contenant texte, capture d’écran, conception

Description générée automatiquement

Figure – MPEG MIV workflow

*For more details see:*

* *ISO/IEC, White Paper on MPEG-I Immersive Video, 2024, ISO/IEC JTC1/SC29/WG6, MPEG Video Coding. Available from:* [*https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/137\_OnLine/w21274.zip*](https://www.mpeg.org/wp-content/uploads/mpeg_meetings/137_OnLine/w21274.zip)

## 4.6 Immersive audio

Audio is critical for creating a fully immersive environment. MPEG’s immersive audio standards, particularly those outlined in MPEG-I Part 4 and MPEG-H Part 3, provide realistic immersive audio experiences. These standards ensure that users experience high quality spatial audio that changes dynamically based on their environment and user interactions, adding depth and realism to the virtual world. Specifically, MPEG-I processes sound objects, channels and HOA and renders 6DoF source localization and realistic room effects of a virtual scene. It additionally supports real-time inputs for social VR. It also offers an interface for accessibility settings that are a requirement for the successful inclusion of hearing-impaired persons into an enjoyable audio experience.

Ein Bild, das Text, Screenshot, Diagramm, Design enthält.

Automatisch generierte Beschreibung

Figure - MPEG-I immersive audio architecture

*For more details see:*

* *ISO/IEC, White Paper on MPEG-I Immersive Audio, 2024, ISO/IEC JTC1/SC29/WG6, MPEG Audio Coding. Available from:* [*https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/146\_Rennes/w23975.zip*](https://www.mpeg.org/wp-content/uploads/mpeg_meetings/146_Rennes/w23975.zip)

## 4.7 MPEG Scene description and scene representation

MPEG-I Scene Description (ISO/IEC 23090-14) succeeds MPEG-4’s Binary Format for Scenes (BIFS), offering enhanced support for dynamic 3D scenes and media orchestration in the Metaverse. BIFS, rooted in the Virtual Reality Modeling Language (VRML), provided a text-based foundation for describing interactive 3D scenes, which MPEG adapted into a binary format for efficiency and real-time applications. In contrast, MPEG-I SD builds on top of glTF (GL Transmission Format), a modern, widely adopted standard optimized for compact and efficient transmission of 3D assets across platforms. This shift from VRML-based BIFS to glTF-based SD reflects MPEG’s adaptation to the demands of today’s Metaverse, prioritizing interoperability and performance in complex virtual environments.

MPEG-I SD (Scene Description - ISO/IEC 23090-14) adds the support of MPEG media and ensures its localization in space as well as its orchestration in a streaming context.

Une image contenant texte, capture d’écran, conception

Le contenu généré par l’IA peut être incorrect.

Figure - MPEG Extensions defined in MPEG-I Scene Description

*For more details see:*

* *ISO/IEC, White Paper on MPEG-I Scene Description, 2022, ISO/IEC JTC1/SC29/WG3, MPEG System. Available from:* [*https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/140\_Mainz/w22138.zip*](https://www.mpeg.org/wp-content/uploads/mpeg_meetings/140_Mainz/w22138.zip)

Further, MPEG-I Part 28 focuses on the interchange and compression of 3D graphics, ensuring that digital assets can move fluidly across different systems.

## 4.8 Avatar Representation Format

MPEG’s work on avatar technologies spans over two decades, beginning with foundational standards in MPEG-4 that addressed the compression and animation of avatars and avatar-like dynamic objects. MPEG-4 Part 2 (ISO/IEC 14496-2) introduced Face and Body Animation (FBA), a pioneering standard for encoding facial expressions and body movements using parametric models. This allowed efficient transmission and rendering of animated humanoid characters, supporting early virtual environments. Later, MPEG-4 Part 16 (ISO/IEC 14496-16), through the Animation Framework eXtension (AFX), added Bone-Based Animation, enabling skeletal structures to drive realistic deformations of 3D models. These early standards established MPEG’s expertise in avatar representation, compression, and animation, directly influencing modern Metaverse applications.

Building on this legacy, MPEG is now developing the Avatar Representation Format (ARF) to address interoperability needs for various immersive applications. The specification is currently in Committee Draft stage as ISO/IEC 23090-39. The format defines an ARF container, which is used to store all user’s avatar assets, as well as a set of animation stream formats, addressing facial, body, hand, landmark, and other body part animations. ARF advances beyond FBA and Bone-Based Animation by offering a unified, interoperable framework tailored to the diverse requirements of today’s Metaverse, ensuring efficient streaming and rendering across platforms.

The following figure depicts a screen shot of an animated avatar in a 3D scene.

A person standing in a room with blue chairs

AI-generated content may be incorrect.

Figure - MPEG animated avatar in a 3D scene

*For more details see:*

* *ISO/IEC, Avatar representation formats web page, ISO/IEC JTC1/SC29/WG3, MPEG Systems. Available from:* [*https://www.mpeg.org/standards/MPEG-I/39/*](https://www.mpeg.org/standards/MPEG-I/39/)

## 4.9 Real-time streaming and synchronization

MPEG technologies like DASH (Dynamic Adaptive Streaming over HTTP) allow for real-time, adaptive streaming of multimedia content. These technologies are crucial for ensuring that users experience synchronized, high-quality media across different devices and network conditions. In scenarios involving multiple users or real-time collaboration, synchronized media streaming is essential to maintain a consistent user experience.

## 4.10 Neural-network compression for scene representations

Recently proposed scene representations such as neural radiance fields (NeRFs) represent visual content as learned parameters of neural networks. MPEG has developed a standard Neural Network Coding (NNC) for compression of neural networks, likely applicable to these parameters, as well as updates of them.

*For more details see:*

* *ISO/IEC, White Paper on MPEG-7 Neural Network Coding, 2024, ISO/IEC JTC1/SC29/WG4, MPEG Video Coding. Available from:* [*https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/145\_OnLine/w23564.zip*](https://www.mpeg.org/wp-content/uploads/mpeg_meetings/145_OnLine/w23564.zip)

## 4.11 Haptics Representation & Coding Format

As part of the immersive experience to be expected within a Metaverse application, tactile communication is a significant feature to enable. In this context, MPEG developed a standard for the coded representation of Haptics. The phase 1 of the standard will be published by H1 2025 and aims to standardize the representation and transmission of haptic data across the vibrotactile and kinesthetic modalities as a new media, enabling interoperable and high-quality haptic interactions across various devices and applications.

In this standard Haptic is considered as an additional media synchronized with audio and video (in the figure below). MPEG specified the representation and coding formats allowing interchange and distribution of haptics and considering both parametric and sampled-based haptic signals.

A diagram of a network

Description automatically generated

Figure - Haptic media pipeline. It is similar to traditional A/V pipelines. MPEG standardizes the encoded representation and the decoder. The encoder and renderer are proprietary.

Proper extensions of the MPEG-I Scene Description have also been done to include such haptic data into current 3D scenes, to enable interactions with objects and avatars (see section 4.5 MPEG Scene description and scene representation).

A phase 2 is started to extend this standard for spatial and interactive Haptics. A complete framework for supporting localized and dynamic changing haptics based on the environment and user interactions is currently defined.

Carriage over DASH and ISOBMFF has been also specified making it usable for all MPEG transport protocols.

*For more details see:*

* *ISO/IEC 23090-31 Haptics Coding - Committee Draft text available from:* [*https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/140\_Mainz/w22111.zip*](https://www.mpeg.org/wp-content/uploads/mpeg_meetings/140_Mainz/w22111.zip)*.*
* *ISO/IEC 23090-32 Carriage of Haptic Data - Final Draft International Standard text available from* [*https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/143\_Geneva/w22941.zip*](https://www.mpeg.org/wp-content/uploads/mpeg_meetings/143_Geneva/w22941.zip)

# 5. Challenges and future directions

Although MPEG technologies provide a robust foundation for the Metaverse, certain challenges remain:

**Interoperability:** As the Metaverse is meant to be applicable to multiple platforms and devices, ensuring seamless interaction between these systems is a significant challenge. MPEG standards continue to evolve to support broader interoperability across different 3D formats and platforms.

**Advanced compression:** The increasing complexity of 3D environments and assets requires efficient compression algorithms to manage bandwidth and maintain real-time performance. Developing standards for AI-assisted compression is one of the future directions that MPEG is undertaking to define best-in-class solutions.

**Security and privacy:** With the proliferation of digital assets and virtual spaces, security and privacy concerns will become paramount. Ensuring that MPEG technologies can support secure interactions, particularly in commercial and personal environments, will be a growing area of focus.

# 6. Conclusion

From MPEG-4’s BIFS and AFX to MPEG-V’s sensory integration and the advanced MPEG-I framework, MPEG standards have formed the backbone of virtual world technologies for over 25 years, enabling real-time interaction, immersive audio-visual content, and efficient distribution of complex 3D scenes and digital assets vital to the Metaverse. As the Metaverse continues to evolve, MPEG technologies remain essential to ensuring high-quality, scalable, and interoperable virtual experiences. By continuing to innovate in areas such as dynamic mesh compression, immersive audio, and real-time streaming, MPEG is well-positioned to support the converged immersive experiences aimed to become THE Metaverse.

NOTE: The attached file contains a detailed list of MPEG standards foreseen applicable to Metaverse applications and experiences.