



# Outline

- **Draft History**
- **Update section review**
- **Next Step**

# Draft History

- ***Individual draft : 00 - 02 : 2025 July - 2025 Dec***
- ***WG draft adoption call : Accepted : 2026 .02***
  - ***Changed draft title***
    - RTP Payload Format for Avatar Representation Format (ARF) Animations
    - RTP Payload Format for Avatar Representation Format (ARF) Animation **Stream**

# Updated section from 04(individual)

- **Revise the Abstract and Introduction to align certain terms with the MPEG specification.**

## Abstract

This memo outlines RTP payload formats for the animation stream format as defined in the ISO/IEC 23090-39 specification (MPEG-I Avatar Representation Format). An animation stream format is composed of Avatar Animation Units (AAU) including an AAU header and zero or more AAU packets. The RTP payload header format allows for packetization of an AAU unit in an RTP packet payload as well as fragmentation of an AAU into multiple RTP packets.

## Abstract

This memo outlines RTP payload formats for the animation stream format as defined in the ISO/IEC 23090-39 standard (MPEG-I Avatar Representation Format), in the following referred to as ARF. ARF is composed of Avatar Animation Units (AAU) including an AAU header and zero or more AAU packets. The RTP payload header format allows for packetization of an AAU unit in an RTP packet payload as well as fragmentation of an AAU into multiple RTP packets.

## Introduction

[ISO.IEC.23090-39] specifies the Avatar Representation Format (ARF) to offer an interoperable exchange format for the storage, carriage and animation of 3D avatars. It defines the "Avatar Animation Unit"(AAU) as a unit of packetization suitable for Avatar animation streaming, and similar in essence to the NAL unit defined in some video specifications. This document describes how AAUs can be transmitted using the RTP protocol. This document followed recommendations in [RFC8088] and [RFC2736] for RTP payload format writers.

[ISO.IEC.23090-39] specifies the Avatar Representation Format (ARF) to offer an interoperable exchange format for the storage, carriage and animation of avatars. It defines the "Avatar Animation Unit"(AAU) as a unit of packetization suitable for Avatar animation streams, and similar in essence to the NAL unit defined in some video specifications. This document describes how AAUs can be transmitted using the RTP protocol. This document follows recommendations in [RFC8088] and [RFC2736] for RTP payload format writers.

# Updated section from 04(individual)

- ***Added MPEG specification as a reference to clarify original source***

Avatar animation data may be stored as samples in an avatar container, such as the MPEG Avatar Representation Format container [ISO.IEC.23090-39], along with the avatar model representation. This data may also be generated on-the-fly as cameras and sensor capture a person's motion and generate corresponding commands to mimic this movement for an avatar that represent the user. Avatar animation samples may be structured into a bitstream comprising a sequence of Avatar Animation Units (AAUs), whose general structure is provided in Figure 2.

Each AAU is associated with an Avatar ID that indicates the target avatar to which the animation data applies. In addition, it is also associated with a Level of Detail (LoD), which indicates the quality of the avatar animation. Different LoDs may, for example, correspond to different numbers of animation joints and thus different animation sample sizes. The animation data within an AAU is generated by an tracking/animation framework (e.g., OpenXR or ARKit) based on a schema identified using a URN. An avatar container corresponds to a single avatar ID, and each asset within the container holds data for one or more LoDs.

The animation streams may be read from a file, or generated on-the-fly as cameras and/or sensors capture a person's motion and generate corresponding commands to mimic this movement for an avatar that represents the user. Avatar animation samples are structured into a bitstream comprising a sequence of Avatar Animation Units (AAUs), defined in [ISO.IEC.23090-39], and whose general structure is provided in Figure 2.

An avatar animation is associated with a Base Avatar, using an avatar ID. Each AAU is associated with an Avatar ID that indicates the target avatar to which the animation data applies. In addition, it is also associated with a Level of Detail (LoD), which indicates the level of detail of the asset to which the animation data is associated. The animation data within an AAU can for example be generated by a tracking and animation framework (e.g., OpenXR or ARKit) . [ISO.IEC.23090-39] defines this identified using a URN.

# Updated section from 04(individual)

- ***Avatar Animation Unit – need to be updated based on next MPEG meeting result***

D (Dependency, 1 bit): this field indicates whether an AAU included in the avatar animation packet payload is an independent AAU (D=0) or dependent (D=1). If D=1, the AAU is dependent on other AAUs for decoding. If D=0, the AAU can be decoded independently.

UT (Unit Type, 4 bits): this field indicates the type of the payload, which can be the type of the AAU for single unit payload, or the type of the payload otherwise, as shown in Figure 5.

L (Level of Detail, 3 bits): this field indicates the level of detail to which the AAU(s) within the RTP packet applies. If the RTP packet includes multiple AAUs, L MUST indicate the lowest LoD.

AvID (Avatar ID, 8 bits): this field identifies the avatar to which the animation data in the payload of the packet applies. The avatar corresponds to the digital assets to be animated.

D (Dependency, 1 bit): this field indicates whether an AAU included in the avatar animation packet payload is an independent AAU (D=0) or dependent (D=1). If D=1, the AAU is dependent on other AAUs for decoding. If D=0, the AAU can be decoded independently. Editor's Note: in the current version of [ISO.IEC.23090-39] all AAUs are independent AAUs.

UT (Unit Type, 4 bits): this field indicates the type of the payload, which can be the type of the AAU [ISO.IEC.23090-39] for single unit payload, or the type of the payload otherwise, as shown in Figure 5.

L (Level of Detail, 3 bits): this field indicates the level of detail to which the AAU(s) within the RTP packet applies. If the RTP packet includes multiple AAUs, L MUST indicate the lowest LoD.

AvID (Avatar ID, 8 bits): this field identifies the avatar to which the animation data in the payload of the packet applies. The avatar corresponds to the digital assets to be animated.

# Next Step

- *Suggestions and feedback are welcome*
- *We are planning to update draft and request WGLC when MPEG starts FDIS process.*