More Instant Messaging Interoperability
Internet-Draft
Intended status: Informational
Expires: 5 January 2026

R. Mahy

4 July 2025

Audio, Video, and Image Metadata extensions for the More Instant Messaging Interoperability (MIMI) Content format
draft-mahy-mimi-av-metadata-00

Abstract

The More Instant Messaging Interoperability (MIMI) content format is a container for rich content, which can reference image, video, and audio files. This document describes metadata for these files to allow for more pleasant rendering.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at https://rohanmahy.github.io/mimi-av-metadata/draft-mahy-mimi-av-metadata.html. Status information for this document may be found at https://datatracker.ietf.org/doc/draft-mahy-mimi-av-metadata/.

Discussion of this document takes place on the More Instant Messaging Interoperability Working Group mailing list (mailto:mimi@ietf.org), which is archived at https://mailarchive.ietf.org/arch/browse/mimi/. Subscribe at https://www.ietf.org/mailman/listinfo/mimi/.

Source for this draft and an issue tracker can be found at https://github.com/rohanmahy/mimi-av-metadata.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 5 January 2026.

Table of Contents

1. Introduction

The MIMI content format [I-D.ietf-mimi-content] can convey a variety of media types, as either inline or referenced external content. In messaging applications it is common to display audio, video, and static image content, collectively audio/video (AV). The layout for messaging applications often reserves a placeholder for the AV content. While it is common for static images to be immediately displayed, audio and video content is often not immediately downloaded and rendered. Even if image data is downloaded immediately, if there is a network or server delay there can be time when the aspect ratio or dimensions of the image are not yet know. It is therefore useful to have some rendering hints about the media for more pleasant rendering. This document defines extensions to the MIMI content format to provide these hints.

2. Conventions and Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

This document uses a variety of terms from the MIMI content format definition, especially NestedPart, SinglePart, ExternalPart, and MultiPart.

3. AV Metadata Extensions

The AV Metadata MIMI content extension is an array of AV metadata entries. Each AV metadata entry is a CBOR map of AV metadata properties, all of which are optional except the part_index and type. The semantics of the individual property fields is as follows:

* part_index: refers to the order of MIMI parts from the relevant part inside the NestedPart structure in a MIMI content message. It can refer to a SinglePart or ExternalPart.

* type: an integer enumeration representing the media type (not-including the subtype). audio is 1, image is 2, and video is 3. An extension socket is defined, although its need is not anticipated.

* width: the width of the image or video in pixels

* height: the height of the image or video in pixels

* duration: the duration of the audio or video in seconds. It can be expressed as an unsigned integer or a positive floating point number

* preview_index: for a video part, the partIndex of another related part that represents its image preview. It can refer to a SinglePart or ExternalPart, or a MultiPart with chooseOne partSemantics which contains only SinglePart or ExternalPart types, all of which must be an image with a disposition value of preview.

* accessibility_text: this text could be rendered instead of the audio, image, or video when various accessiblity settings are enabled, or during no or slow network access when a cached or preview image is not available.

* rotation: one of four values: 0, 90, 180, or 270. This integer refers to the number of degrees of clockwise rotation (in 90 degree increments) needed to correctly view the image.

  Note that an orientation field is not necessary. Any image and video with a width field which is larger than its height is assumed to have a landscape mode orientation, while one with a height larger than its width is assumed to have a portrait mode orientation.

The following snippet of Concise Data Definition Language (CDDL) [RFC8160] is used to formally define the structure of the extension.

```
av_metadata_array = (
    "av_metadata" : [ * metadata_entry ]
)

metadata_entry = {
    &(part_index: 1) : uint16,
    &(type: 2)       : audio / image / video / $ext_media,
    ? &(width: 3)              : uint,
    ? &(height: 4)             : uint,
    ? &(duration: 5)           : nonnegative_number,
    ? &(preview_index: 6)      : uint16,
    ? &(accessibility_text: 7) : tstr,
    ? &(rotation: 8)           : 0 / 90 / 180 / 270
    $ext_av_metadata
}

nonnegative_number = uint / float .gt 0.0
uint16 = uint .size 2

audio = 1
image = 2
video = 3
```

4. Example

Below is an example of a video of puppies, a preview image, and an audio clip.

```
   "av_metadata" : [
     {
        /partIndex /         1: 2,
        /type      /         2: 3, /video/
        /width     /         3: 1920,
        /height    /         4: 1080,
        /duration  /         5: 37, / in seconds. can be uint or float /
        /preview_index /     6: 4,
        /accessibility_text/ 7: "two golden retriever puppies playing in" +
                                "overgrown grass lit with low sunlight"
     },
     {
        /partIndex /         1: 4,
        /type      /         2: 2, /image/
        /width     /         3: 1920,
        /height    /         4: 1080
     },
     {
        /partIndex /         1: 7,
        /type      /         2: 1, /audio/
        /duration  /         5: 9.45, / in seconds. can be uint or float /
        /accessibility_text/ 7: "uproarious laughter"
     }
   ]
```

5. Security Considerations

TODO Security

6. IANA Considerations

TODO register the extension with IANA.

7. Normative References

[I-D.ietf-mimi-content]
Mahy, R., "More Instant Messaging Interoperability (MIMI) message content", Work in Progress, Internet-Draft, draft-ietf-mimi-content-06, 28 February 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-mimi-content-06>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/rfc/rfc2119>.

[RFC8160] Tatham, S. and D. Tucker, "IUTF8 Terminal Mode in Secure Shell (SSH)", RFC 8160, DOI 10.17487/RFC8160, April 2017, <https://www.rfc-editor.org/rfc/rfc8160>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

Acknowledgments

TODO acknowledge.

Author’s Address

Rohan Mahy
Email: rohan.mahy@gmail.com