

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: 18 October 2026

ssw. Whited, Ed.
Independent
16 April 2026

Matroska Stem Files
draft-swhited-mka-stems-10

Abstract

This document defines a multi-track profile of the Matroska container format for distributing stems. It is intended to be used by DJ applications and Digital Audio Workstations while remaining backwards compatible with existing media players.

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 18 October 2026.

Copyright Notice

Copyright (c) 2026 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

1. Introduction	2
1.1. Requirements Language	3
2. Versioning	3
3. Track Layout	3
3.1. Audio Streams	3
3.2. Stem Metadata	4
4. Mastering	5
5. Format Support	6
6. IANA Considerations	6
6.1. Matroska Tag Names Registry	6
6.2. Stem Types Registry	6
7. Security Considerations	7
8. Normative References	7
9. Informative References	8
Acknowledgements	8
Author's Address	9

1. Introduction

Stems are recordings of individual instruments, or clusters of instruments, used by DJs and music producers for live mixing of music. Historically stems have been stored as individual audio files, or using patent-encumbered or vendor specific, proprietary container formats.

A common feature of modern software used by DJs is "dynamic" or "live" stem separation where the DJ software attempts to algorithmically separate the audio signals in a track to allow the DJ to mute, solo, or apply effects to individual instruments. The results of such dynamic separation vary but are, generally speaking, noticeably different from the original stems used by the producer and frequently contain distortions and other artifacts that sound undesirable. A better model is to have the producer release the original stems along with the original track. This allows the final mix to sound closer to the producers original vision for the track, even while it is being remixed and re-interpreted by a DJ or another artist.

This specification documents a profile for the Matroska container format [RFC9559] that allows it to store the final mix for a track alongside the lossless or lossy stems used to mix the track in a single file. The target consumer of these stem files are DJ applications meant for live remixing and performance, as well as Digital Audio Workstations (DAWs) used by producers who want their music to be remixed.

1.1. Requirements Language

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.

2. Versioning

Files that meet the criteria of this document MUST have a global UTF-8 tag with the name "STEM_VERSION" with the value "1.0". The presence of this tag indicates that this is a stem file and the value indicates compatibility with this document. The value may or may not be changed in the future if new versions of this format are created.

3. Track Layout

3.1. Audio Streams

Each stem file may contain an arbitrary number of tracks containing audio and MUST include at least three audio tracks (the mixed audio and at least two stems). For stem files meant for live DJ use, it is RECOMMENDED that four or fewer stem tracks be used (as opposed to stem files meant for music production or non-live remixing where a DAW may utilize a significantly larger number of tracks).

For ease of decoding each track SHOULD be encoded using the same codec with the same parameters including bitrate, and sample rate. Stems are often recorded with a single channel and only the final mix is in stereo. For stem files that are meant to be re-mixed by a DAW this is fine, but DJs may want to maintain a similar balance and channel layout to the original track. Stems MAY have a different channel count or layout than the main audio track, however it is RECOMMENDED that all stem tracks maintain the same channel count and layout as the main track and have the same channel balance as their component parts in the final mix. For example, if the final mix is a stereo track that contains a fiddle that is 75% in the right channel and only 25% in the left channel, the stem track for the fiddle would also be in stereo with the stem mostly appearing from the right channel as in the final mix.

The first track containing audio data MUST be the final post-mix audio in the default language (the mixdown track). All mixdown tracks regardless of language MUST have the Matroska "Default" flag set to "1" ([RFC9559], Section 18.1, 5.1.4.1.5). This helps preserve backwards compatibility in media players which do not support this format which typically play the first audio stream found or may

select based on the default flag. In addition, the "Enabled" flag for any mixdown tracks MUST be set to "1" ([RFC9559], Section 5.1.4.1.4).

The remaining audio tracks will be individual stems and MUST have the same effective length as the mixdown track such that playing each stem track from the beginning would result in roughly the same audio (excluding the final stages of mastering) as the final mix present in the mixdown track. For example, if the original track is three minutes long and the stem file includes a percussion track but the percussion does not start until minute two, the percussion stem would still be three minutes long but would contain a minute of silence at the start of the track, or would have a block timestamp ([RFC9559], Section 10) that sets the effective start time to one minute.

Each stem track MUST have the Matroska "Default" flag set to "0" and MUST have the "Enabled" flag set to "0".

When creating the file the stem tracks SHOULD NOT have any intra-stem gain normalization applied to bring the stems up to the same perceived volume. Instead they should retain the same levels as they would have in the final mix present in the mixdown track so that if all stems were played at unity gain the overall level would be equivalent to the level of the final mix.

On playback, DJ software MAY choose to normalize the gain on any combination of stems currently being played to make it equivalent to the mixdown track or any other tracks being mixed in even if some stems are muted or have their individual gains adjusted. However, the exact mixing behavior of DJ applications is outside the scope of this specification.

3.2. Stem Metadata

For each stem track a tag ([RFC9559], Section 5.1.8) SHOULD also be set with its target set to the stem track and a tag name of "STEM_COLOR". The tag value must be a string in RGB hex format set to a color representing the stem (ie. #145374).

Each stem track MUST set the value of the track "Name" element ([RFC9559], Section 5.1.4.1.18) to a short, human-meaningful, track name for the stem that describes its contents, for example "Percussion" or "Vocals". These names are intended for display in playback applications and therefore should remain concise, but no specific format is defined. The track Name element MAY also be duplicated or overridden as a tag, in which case the order of precedence from Section 24.1 of [RFC9559] SHOULD be respected. The maximum length of the name string (whether in the "Name" field or a tag) MUST NOT be longer than 20 bytes.

The free-text "Name" field provides a label that is backwards compatible with existing software that supports Matroska, but does not necessarily provide enough information for DJ software implementing this format. Some software uses pre-defined stems (ie. always has "Vocal", "Melody", "Bass", and "Drum" stems), but the name is free-text and may not exactly match the name of the specific stem in that software. Or the software may want to display icons that represent each stem, or display translations in a language that is not already embedded in the files "Name" tags. For example, in most tracks with words there might be a stem with its name set to "Vocals" but another track may call it "Singer" or a traditional called folk dance track might name it "Caller". Software for stem playback might want to show a Microphone icon next to each of these tracks. To help map stems to individual icons or controller buttons, a binary tag with the name "STEM_TYPE" and a target of the stem track MAY be added to the file. The value of this tag is an unsigned 64-bit integer from the IANA Stem Types Registry, the initial contents of which are defined in Section 6.2.

4. Mastering

Because the final stages of mastering happens post-mix and the stems are pre-mix audio the stem tracks SHOULD NOT have the final post-mix mastering steps applied (including any final compression or limiter steps). This means that a DJ playing the track using the stem tracks instead of the mixdown track will result in different audio from the final mix. This is deemed an acceptable trade off since the final sound of the DJs version of a track is likely to be significantly different from what the original track producer had created either way. Even without the final mastering steps this method still gives the producer more control over the final sound than if the DJ were to use an auto stem separation algorithm.

Other mastering steps MAY be applied to the stems as well as to the final mix at the mastering engineers discretion.

5. Format Support

The Matroska container format can store many types of audio, not all of which are suitable for DJing or music production. To ensure compatibility between playback and encoding applications the following formats SHOULD be supported depending on the use case of the software as shown in the following table. Formats with the use case "Live remixing" are intended largely for playback applications meant for live performance (ie. DJ software). Formats with the use case "Music production" are intended to be distributed for remixing in a non-live setting (ie. with a DAW or music tracker).

Codec	Use Case	Codec ID
FLAC [RFC9639]	Live remixing, Music production	A_FLAC [RFC9639], Section 10.2
Opus [RFC6716]	Live remixing	A_OPUS [I-D.ietf-cellar-codec], Section 3.4.32

Table 1: Audio codec support

6. IANA Considerations

6.1. Matroska Tag Names Registry

This memo modifies the "Matroska Tag Names" registry defined in Section 6.1 of [I-D.ietf-cellar-tags] to add the following values:

Tag Name	Tag Type	Reference
STEM_VERSION	UTF-8	This document, Section 2
STEM_COLOR	UTF-8	This document, Section 3.2
STEM_TYPE	binary	This document, Section 3.2

Table 2: Additions to the "Matroska Tag Names" Registry

6.2. Stem Types Registry

IANA will create a new registry called the "Stem Types Registry".

To register a new Stem Type in the registry a unique unsigned 64-bit integer ID, Description, and an optional reference to a document describing the stem type are required.

The initial values of the Stem Types Registry are as follows:

ID	Description	Reference
0x01	Stems that contain sung vocals.	This document
0x02	Stems that contain spoken vocals including traditional dance calls, or spoken words over a tune.	This document
0x03	Stem that contains the melody (regardless of instrument)	This document
0x04	Stems that include bass elements (regardless of instrument)	This document
0x05	General percussion	This document
0x06	Synths	This document

Table 3: Initial Contents of the "Stem Types" Registry

7. Security Considerations

This document inherits security considerations from both [RFC8794] and [RFC9559]. It does not have additional security considerations.

8. Normative References

- [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/info/rfc2119>>.
- [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/info/rfc8174>>.

- [RFC9559] Lhomme, S., Bunkus, M., and D. Rice, "Matroska Media Container Format Specification", RFC 9559, DOI 10.17487/RFC9559, October 2024, <<https://www.rfc-editor.org/info/rfc9559>>.

9. Informative References

- [RFC6716] Valin, JM., Vos, K., and T. Terriberry, "Definition of the Opus Audio Codec", RFC 6716, DOI 10.17487/RFC6716, September 2012, <<https://www.rfc-editor.org/info/rfc6716>>.
- [RFC8794] Lhomme, S., Rice, D., and M. Bunkus, "Extensible Binary Meta Language", RFC 8794, DOI 10.17487/RFC8794, July 2020, <<https://www.rfc-editor.org/info/rfc8794>>.
- [RFC9639] van Beurden, M.Q.C. and A. Weaver, "Free Lossless Audio Codec (FLAC)", RFC 9639, DOI 10.17487/RFC9639, December 2024, <<https://www.rfc-editor.org/info/rfc9639>>.
- [I-D.ietf-cellar-codec]
Lhomme, S., Bunkus, M., and D. Rice, "Matroska Media Container Codec Specifications", Work in Progress, Internet-Draft, draft-ietf-cellar-codec-18, 12 April 2026, <<https://datatracker.ietf.org/doc/html/draft-ietf-cellar-codec-18>>.
- [I-D.ietf-cellar-tags]
Lhomme, S., Bunkus, M., and D. Rice, "Matroska Media Container Tag Specifications", Work in Progress, Internet-Draft, draft-ietf-cellar-tags-23, 26 February 2026, <<https://datatracker.ietf.org/doc/html/draft-ietf-cellar-tags-23>>.

Acknowledgements

Thanks to the members of #matroska on the libera.chat IRC network, and to mosu and JanC in particular, for patiently explaining the basics of the format to me and for all their feedback.

Thanks also to the members of the Ardour forums for their feedback on DAWs and mastering.

Thanks to the developers of Mixxx and the members of the Mixxx forums for pointing out the potential pain points for DJ software.

Finally, thanks to the members of the IETF CELLAR working group, especially Steve Lhomme, for their feedback.

Author's Address

Sam Whited (editor)
Independent
Email: sam@samwhited.com
URI: <https://blog.samwhited.com>