Status of FFV1 & Matroska

Update on Matroska and FFV1

2017-03-07 @MediaConch #preforma
PREFORMA Challenge

Empower memory institutions to gain full control over the technical properties of digital content intended for long-term preservation.
FFV1 - A lossless video encoding

- 2003: Created in Open Source project "FFmpeg"
- 2006: Bitstream frozen (version 1)
- 2009: Picked up for preservation
- 2010: Funding improvements
- 2012: Added 14bit RGB, Multithreading, SliceCRC
- 2013: Official release of "FFV1.3"
- 2014: PREFORMA Project
- 2015: Standardization in progress
- 2016: Added 16bit RGB
Video

Main Page > Documentation > Format policies > Video

Significant characteristics of video files

Preservation Format

- FFV1/PCM in Matroska wrapper (MKV) (Archivematica 0.7.1 and later)
- MPEG-2/PCM in Material eXchange Format wrapper (MXF) (Archivematica 0.7 and earlier)

Access Format

MPEG-1/MP2

Normalization tool

FFmpeg

Comments

FFV1/MKV

- FFV1 is a completely lossless video codec. For a comparison of lossless codecs, see Video Codecs Comparison 2007.
- Matroska (pronounced maTROShka) is an open standard free video container format which can hold a large number of video and audio codecs. See http://www.matroska.org/.

Other containers and codecs

- According to Library of Congress, "For file-based compressed video, conform to or approximate MPEG-2_422 (4:2:2 Profile) at Main Level (aka MPEG-2 422@ML) or MPEG-2_MP (Main Profile) at Main Level (aka MPEG-2 MP@ML). Uncompressed or losslessly compressed copies are preferred to compressed (for future development)." Library of Congress Sustainability of Digital Formats: MPEG-2 Video Encoding (R.202)

More information

- More information on the Material Exchange Format (MXF) is available at Library of Congress Sustainability of Digital Formats: MXF.

Motion JPEG 2000

- Motion JPEG 2000 (MJPEG2K) is emerging as a preferred format for video files. See for example:
  - UKOLN Multimedia standards: MJ2
  - Lossless Video Compression for Archives: Motion JPEG2k and Other Options, Ian Gilmour, National Film and Sound Archive, Australia, R. Justin Davila, System Architect and Technology Consultant, Media Matters LLC, date unknown.
FFV1 perks

- Losslessness
- Fixity
- Self-description
- Size
Figure 1: Speed / size comparison chart
"[ffv1 @ 0x7f9855046e00] CRC mismatch FC686A4F! frame 215"
Matroska - a metadata-infused wrapper

- Active use since 2002
- Widespread adoption as internet video format
- Foundation of Google's webm (web-streaming video)
- Subtitle management, chaptering abilities
- Extensible structured metadata
- File attachment capabilities (mostly used for subtitles)
- Broad support of audiovisual encodings
EBML & Matroska

- Extensible Binary Meta Language (EBML is a Binary XML format)
- Matroska and webm are EBML Document Type
- Storage is based on a structure of Element ID, Element Data Size, and Element Data
- Unlike XML, an EBML Document requires an EBML Schema to be interpreted semantically
No Time To Wait - An Matroska & FFV1 Symposium
## Codec Encoding for LossLess Archiving and Realtime transmission (cellar)

<table>
<thead>
<tr>
<th>WG</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Codec Encoding for LossLess Archiving and Realtime transmission</td>
</tr>
<tr>
<td><strong>Acronym</strong></td>
<td>cellar</td>
</tr>
<tr>
<td><strong>Area</strong></td>
<td>Applications and Real-Time Area (art)</td>
</tr>
<tr>
<td><strong>State</strong></td>
<td>Active</td>
</tr>
<tr>
<td><strong>Charter</strong></td>
<td>charter-ietf-cellar-01 Approved</td>
</tr>
</tbody>
</table>

### Dependencies

Document dependency graph (SVG)

### Personnel

- **Chairs**: Tessa Fallon, Tim Terriberry
- **Area Director**: Ben Campbell

### Mailing list

- **Address**: cellar@ietf.org
- **To subscribe**: [https://www.ietf.org/mailman/listinfo/cellar](https://www.ietf.org/mailman/listinfo/cellar)
- **Archive**: [https://mailarchive.ietf.org/arch/browse/cellar/](https://mailarchive.ietf.org/arch/browse/cellar/)

### Jabber chat

- **Room address**: xmpp:cellar@jabber.ietf.org@join
- **Logs**: [https://jabber.ietf.org/logs/cellar/](https://jabber.ietf.org/logs/cellar/)

## Charter for Working Group

The preservation of audiovisual materials faces challenges from technological obsolescence, analog media deterioration, and the use of proprietary formats that lack formal open standards. While obsolescence and material degradation are widely addressed, the standardization of open, transparent, self-descriptive, lossless formats remains an important mission to be undertaken by the open source community.

FFV1 is a lossless video codec and Matroska is an extensible media container based on EBML (Extensible Binary Meta Language), a binary XML format. There are open source implementations of both formats, and an increasing interest in and support for use of FFV1 and Matroska. However, there are concerns about the sustainability and credibility of existing specifications for the long-term use of these formats. These existing specifications require broader review and formalization in order to encourage widespread adoption.

There is also a need for a lossless audio format to complement the lossless video codec and container format. FLAC is a lossless audio codec that has seen widespread adoption in a number of different applications including archival applications. While there are open source implementations of the codec, no formal standards for either the codec itself or its use in container formats currently exist. Review and formalization of the FLAC codec standard and its use in Matroska container formats is needed for wider adoption.

Using existing work done by the development communities of Matroska, FFV1, and FLAC, the Working Group will formalize specifications for these open and lossless formats. In order to provide authoritative, standardized specifications for users and developers, the Working Group will seek consensus throughout the process of refining and formalizing these standards. Initial specifications can be accessed here:
Specification Development via GitHub

- https://github.com/Matroska-Org/ebml-specification
- https://github.com/Matroska-Org/matroska-specification
- https://github.com/FFmpeg/FFV1/
- https://github.com/xiph/flac

- https://www.ietf.org/mailman/listinfo/cellar
What does a specification look like?

Prior to CELLAR

- EBML and Matroska documentation was in Drupal site
- FFV1 documentation was in LyX

Within CELLAR

- EBML, Matroska, and FFV1 is managed in Markdown, converted to HTML and RFC formats
Introduction

`EBML`, short for Extensible Binary Meta Language, specifies (byte) aligned format inspired by the principle of XML (a fra data).

The goal of this document is to define a generic, binary, spa that can be used to define more complex formats (such as content) using an `EBML Schema`. The definition of the `EBML` idea behind HTML and XML as a good one: separate structure an the same structural layer to be used with multiple, possibly semantic layers. Except for the `EBML Header` and a few glbob specification does not define particular `EBML` format massnt specification is intended to define how other `EBML`-based fo

`EBML` uses a simple approach of building `Elements` upon thr length, and value) as this approach is well known, easy to pa selective data parsing. The `EBML` structure additionally all arrangement to support complex structural formats in an effic

Notation and Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this do interpreted as described in [#RFC2119].

This document defines specific terms in order to define the f of `EBML`. Specific terms are defined below:

`Child Element`: A `Child Element` is a relative term to desc `Elements` immediately contained within a `Master Element`.

`EBML`: Extensible Binary Meta Language

`Element Data`: The value(s) of the `EBML Element` which is i `Element ID` and `Element Data Size`. The form of the `Element` this document and the corresponding `EBML Schema` of the Elem

`Element Data Size`: An expression, encoded as a `Variable Si length in octets of `Element Data`.

`EBML Body`: All data of an `EBML Document` following the `EB considered the `EBML Body`.

Abstract

This document defines the Extensible Binary Meta Language (EBML) format as a geneeralized file format for any type of data in a hierarchical form. EBML is designed as a binary equivalent to XML and utilizes a storage-efficient approach to building nested Elements with identifiers, lengths, and values. Similar to how an XML Schema defines the structure and semantics of an XML Document, this document defines an EBML Schema to convey the semantics of an EBML Document.

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 http://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 March 27, 2017.
Recent CELLAR work

- EBML focus
- Adopting of underlying standards and references to existing standards
- Defining color and display metadata (colorspace, HDR, full/broadcast range)
- Definition of EBML Schema to express the structure of EBML Document
- Clarified interlacement support
- Support for unknown display aspect ratios
- FFV1 slice structure clarification (more explicit definition of each slice element)
- Security considerations section
- Test file libraries in GitHub (examples of logical errors and extent of what is permitted)
CELLAR work in progress

- Matroska reference timecode support
- Updating how encoding support is defined
- Review of Matroska’s metadata registry
- 360 degree / VR video metadata
- Language authority updates
- Rationale numbers as timestamps
- FFV1 version 1.4, context of color and range
- Attachment updates
- Extend subtitle support for other data forms of temporal data (captions, etc)
- Recommendations of practices for use of Matroska and FFV1 in preservation
Thanks

2017-03-07 @MediaConch #preforma