

Status of **CELLAR**

Codec **E**ncoding for **L**oss**L**ess **A**rchiving
and **R**ead-time transmission

2016-09-27 @dericed @tessaAfallon #iasa_web

PREFORMA Challenge

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

LOC's Sustainability Factors

- Disclosure
- Adoption
- Transparency
- Self-documentation
- External dependencies
- Impact of patents
- Technical protection mechanisms

Navigation

[Main page](#)
[Recent changes](#)
[Random page](#)

Toolbox

[What links here](#)
[Related changes](#)
[Special pages](#)
[Printable version](#)
[Permanent link](#)

[Page](#)
[Discussion](#)

[Read](#)
[Edit](#)
[View history](#)

[Go](#)

[Search](#)

Video

[Main Page](#) > [Documentation](#) > [Format policies](#) > [Video](#)

Significant characteristics of video files [edit]

Preservation Format [edit]

- [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 [edit]

MPEG-1/MP2

Normalization tool [edit]

FFmpeg

Comments [edit]

FFV1/MKV [edit]

- [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 [edit]

- 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 \(H.262\)](#) [↗](#)
- For preservation of audio streams, [WAV PCM](#) or [WAV BWF](#) are preferred formats and [AIFF](#) is acceptable. See [Guidelines for the Creation of Digital Collections: Digitization Best Practices for Audio, Consortium of Academic and Research Libraries in Illinois, 2009](#) [↗](#), p. 2.
- [More information](#)
 - [More information on the Material Exchange Format \(MXF\) is available at Library of Congress Sustainability of Digital Formats: MXF](#) [↗](#)
 - [Digital Video Preservation Reformatting Project](#), Prepared by Media Matters, LLC for the Dance Heritage Coalition, Presented to The Andrew W. Mellon Foundation, June 2004 [↗](#)
 - [Arts and Humanities Data Service Preservation Handbook: Moving Image](#), Gareth Night, 2005 [↗](#)
 - [WP2 - Preservation Strategies](#), Richard Wright, PrestoPRIME WP2 BBC Research & Development, UK, 2009 [↗](#)
 - [A Primer on Codecs for Moving Image and Sound Archives: 10 Recommendations for Codec Selection and Management](#). Chris Lacinak, AudioVisual Preservation Solutions, 2010 [↗](#)

Motion JPEG 2000 [edit]

- [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 Dávila, System Architect and Technology Consultant, Media Matters LLC, date unknown [↗](#)

Media type preservation plans

[\[edit\]](#)

Media type	File formats	Preservation format(s)	Access format(s)	Normalization tool
Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Email	PST	MBOX	MBOX	readpst
Email	Maildir**	Original format	MBOX	md2mb.py
Office Open XML	DOCX, PPTX, XLSX	Original format	PDF for PPTX	OpenOffice
Plain text	TXT	Original format	Original format	None
Portable Document Format	PDF	PDF/A	Original format	Ghostscript
Presentation files	PPT	Original format	PDF	OpenOffice
Raster images	BMP, GIF, JPG, JP2*, PCT, PNG*, PSD, TIFF, TGA	Uncompressed TIFF	JPEG	ImageMagick
Raw camera files/Digital Negative format**	3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F	Original format	JPEG	ImageMagick/UFRaw
Spreadsheets	XLS	Original format	Original format	None
Vector images	AI, EPS, SVG	SVG	PDF	Inkscape
Video	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV	FFV1/LPCM in MKV	MPEG-1	FFmpeg

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
- 2016: Standardization in progress
- 2016: Added 16bit RGB

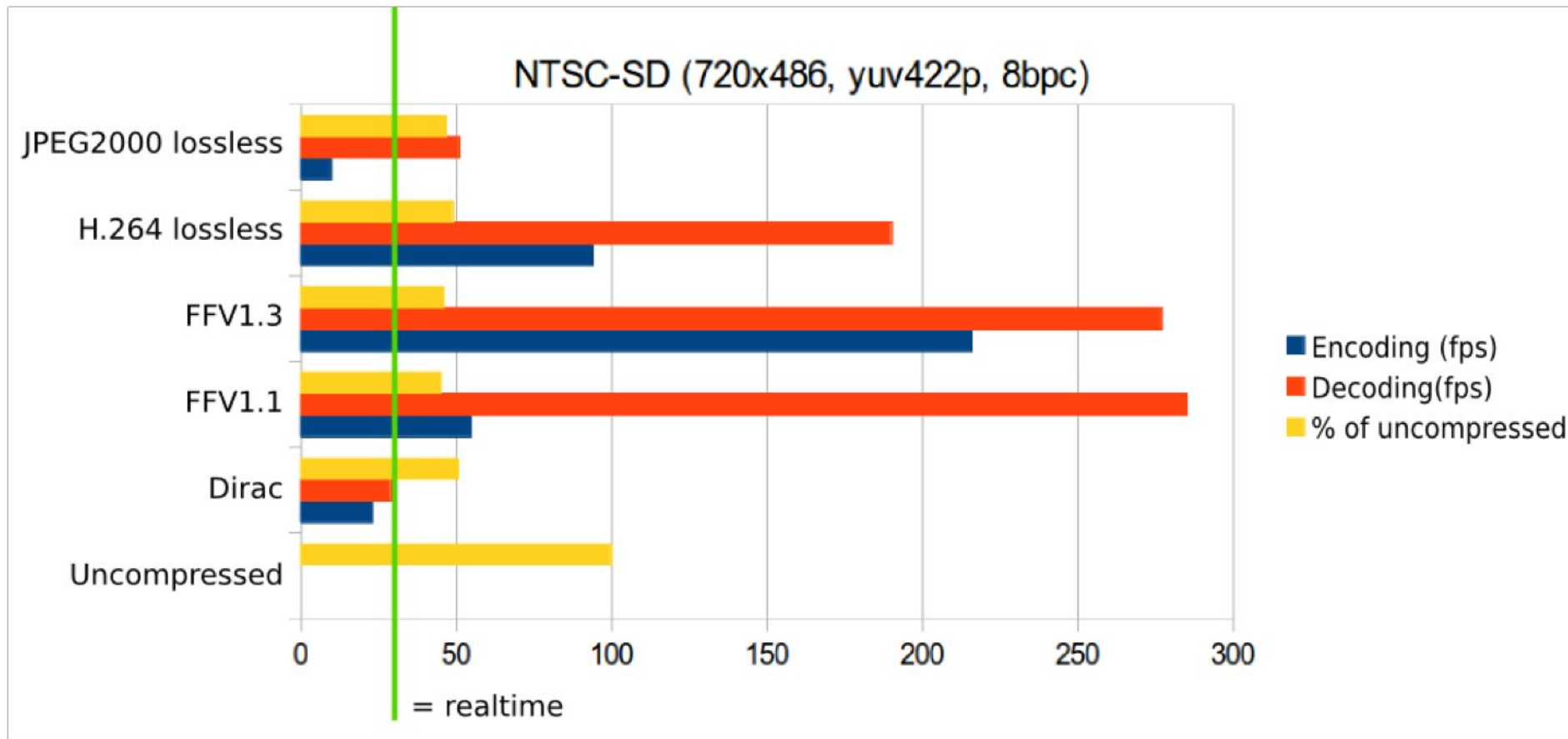


Figure 1: Speed / size comparison chart

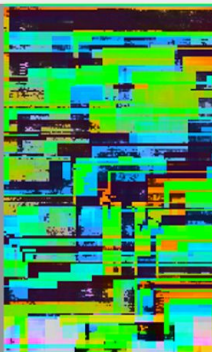
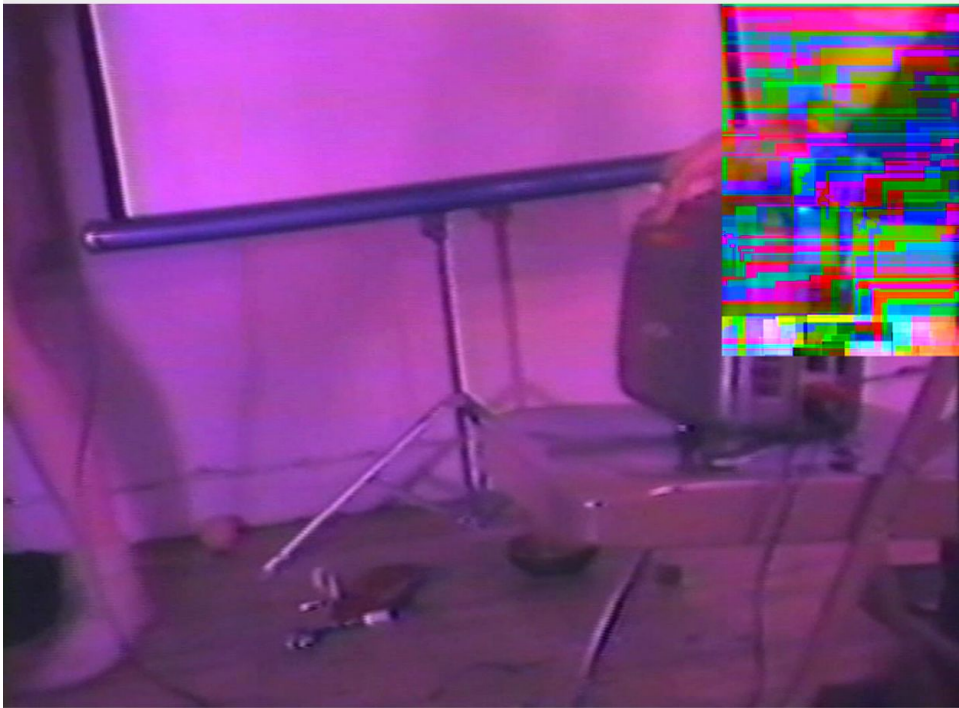
Losslessness

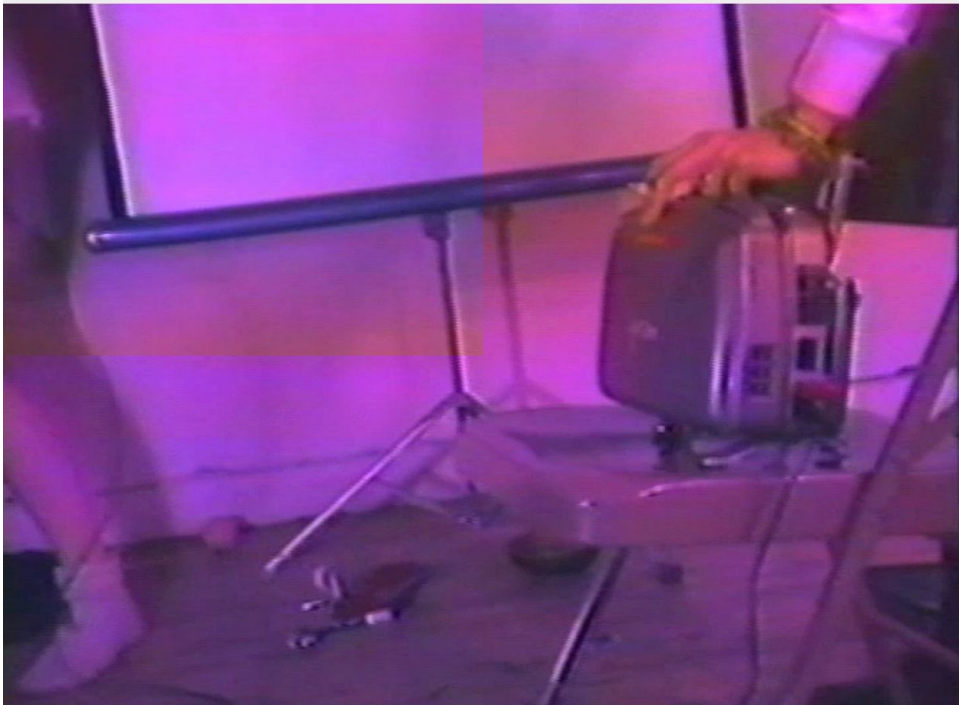
- Fixity
- Self-description
- Size



“[ffv1 @ 0x7f9855046e00] CRC mismatch FC686A4F! frame 215”









MATROŠKA

EBML & Matroska

- Extensible Binary Meta Language (EBML is a Binary XML format)
- An EBML Schema defines an EBML Document like an XML Schema defines an XML Document
- 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

EBML & Variable Sized Integer (VINT)

```
0100000010000001010001000011
011010110000101110100011100
100110111101110011011010110
1100001
```

EBML & Variable Sized Integer (VINT)

```
01000010  10000010  10001000
01101101  01100001  01110100
01110010  01101111  01110011
01101011  01100001
```


EBML & Variable Sized Integer (VINT)

01000010 10000010 10001000
01101101 01100001 01110100
01110010 01101111 01110011
01101011 01100001

EBML & Variable Sized Integer (VINT)

0 1000010	10000010	1 0001000
01101101	01100001	01110100
01110010	01101111	01110011
01101011	01100001	

EBML & Variable Sized Integer (VINT)

Element ID: **01** 000010 10000010

Element Data Size: **1** 0001000

Element Data: 01101101 01100001

01110100 01110010 01101111 01110011

01101011 01100001

EBML & Variable Sized Integer (VINT)

Element ID: 0x4282

Element Data Size: 8 bytes

Element Data: 01101101 01100001
01110100 01110010 01101111 01110011
01101011 01100001

EBML Schema Element Definitions

name: DocType

path: /EBML/DocType (1,1)

id 0x4282

type: String

description: A string that describes and identifies the content of the EBML Body that follows this EBML Header.

EBML & Variable Sized Integer (VINT)

Element ID: 0x4282

Element Data Size: 8 bytes

Element Data: "matroska"

EBML & Variable Sized Integer (VINT)

```
010000001000001010001000011  
011010110000101110100011100  
100110111101110011011010110  
1100001
```

<DocType>matroska</DocType>

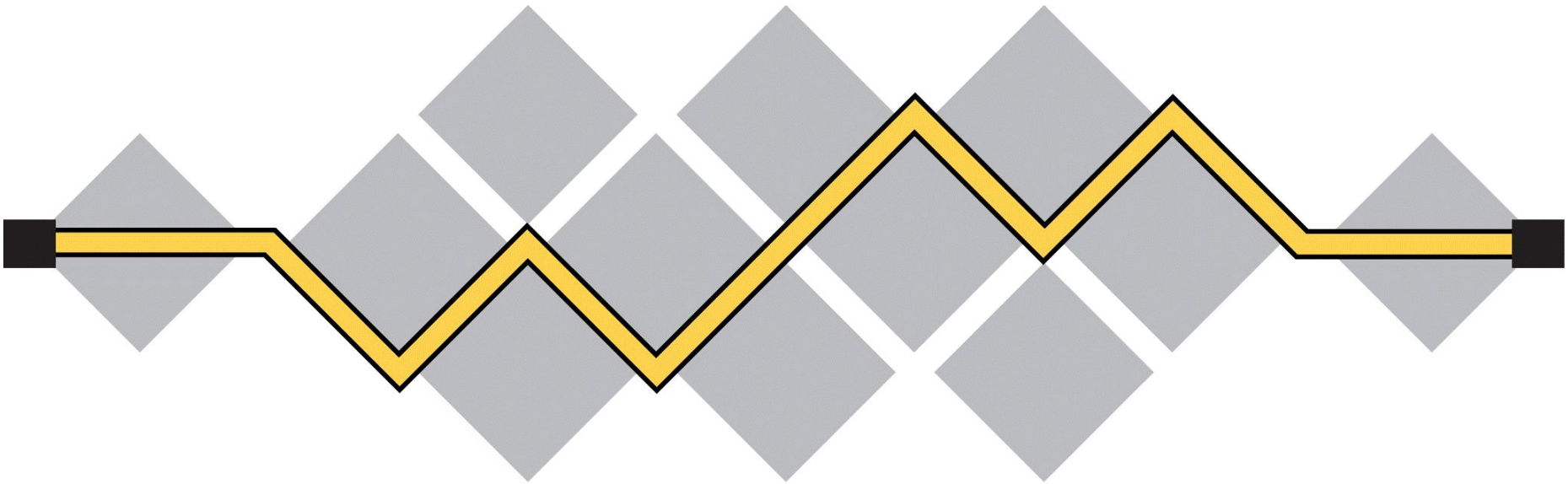
EBML & Variable Sized Integer (VINT)

```
<EBML>
  <EBMLVersion>1</EBMLVersion>
  <EBMLReadVersion>1</EBMLReadVersion>
  <EBMLMaxIDLength>4</EBMLMaxIDLength>
  <EBMLMaxSizeLength>8</EBMLMaxSizeLength>
  <DocType>matroska</DocType>
  <DocTypeVersion>4</DocTypeVersion>
  <DocTypeReadVersion>2</DocTypeReadVersion>
</EBML>
<Segment>
  <Info>
    <TimecodeScale>1000000</TimecodeScale>
    <MuxingApp>Lavf57.50.100</MuxingApp>
    <WritingApp>Lavf57.50.100</WritingApp>
    <SegmentUID>efd46d3ed630381ef9021d1d4ed5a81a</SegmentUID>
    <Duration>40.0</Duration>
  </Info>
  <Tracks>
    <TrackEntry>
```


mission



TL;DR: 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.



I E T F[®]

IETF Mission Statement

The mission of the IETF is to make the Internet work better by producing high quality, relevant technical documents that influence the way people design, use, and manage the Internet. [...] cardinal principles:

Open process - any interested person can participate in the work, know what is being decided, and make his or her voice heard on the issue. [...]

Technical competence - the issues on which the IETF produces its documents are issues where the IETF has the competence needed to speak to them, and that the IETF is willing to listen to technically competent input from any source. [...]

Volunteer Core - our participants and our leadership are people who come to the IETF because they want to do work that furthers the IETF's mission of "making the Internet work better". [...]

Rough consensus and running code - We make standards based on the combined engineering judgement of our participants and our real-world experience in implementing and deploying our specifications. [...]

Protocol ownership - when the IETF takes ownership of a protocol or function, it accepts the responsibility for all aspects of the protocol, even though some aspects may rarely or never be seen on the Internet. [...]

User

[Sign in](#)
[New account](#)
[Preferences](#)

Groups

[Active WGs](#)
[Active RGs](#)
[Other](#)

By area/parent

[Applications and Real-Time](#)
[General](#)
[Internet](#)
[Ops & Mgmt](#)
[Routing](#)
[Security](#)
[Transport](#)
[IRTF](#)

New work

[Chartering groups](#)
[BOFs](#)

Other groups

[Concluded groups](#)
[Non-WG lists](#)

Documents

[Draft submissions](#)
[Sign in to track docs](#)

RFC streams

[IAB](#)
[IRTF](#)
[ISE](#)

Meetings

[Agenda](#)
[Materials](#)
[Past proceedings](#)
[Upcoming](#)
[Request a session](#)
[Session requests](#)

Codec Encoding for LossLess Archiving and Realtime transmission (cellar)

[Documents](#)[Charter](#)[Meetings](#)[History](#)[Photos](#)[Email expansions](#)[List archive](#)[Tools »](#)

WG **Name** Codec Encoding for LossLess Archiving and Realtime transmission

Acronym cellar

Area Applications and Real-Time Area (art)

State Active

Charter charter-ietf-cellar-01 Approved

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>

Archive <https://mailarchive.ietf.org/arch/browse/cellar/>

Jabber chat

Room address <xmpp:cellar@jabber.ietf.org?join>

Logs <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.

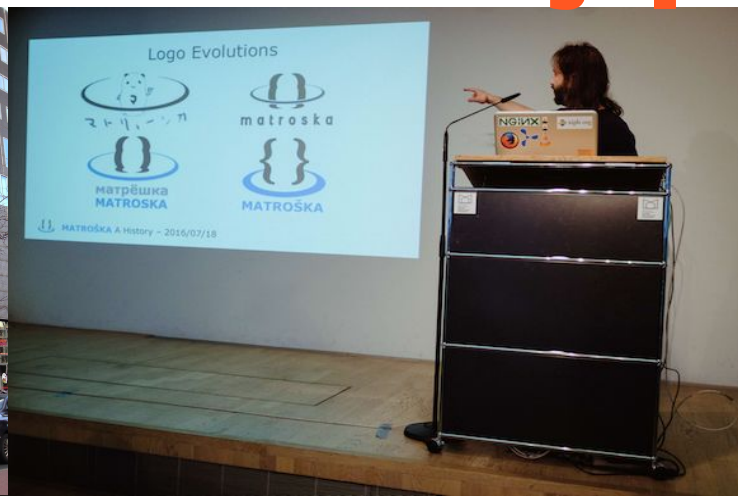
There is also a need for a lossless audio format to complement the lossless video codec and container format. FLAC is a lossless audio codec that has seen widespread adoption in a number of different applications including archival applications. While there are open source implementations of the codec, 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>

No Time To Wait - An Matroska & FFV1 Symposium



Compiling Specification

Prior to CELLAR

- EBML and Matroska documentation was in HTML
- 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 fraction of data).

The goal of this document is to define a generic, binary, space-efficient format 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 and content on the same structural layer to be used with multiple, possibly different semantic layers. Except for the `EBML Header` and a few global specifications, the `EBML` format specification does not define particular `EBML` format semantics. This specification is intended to define how other `EBML`-based formats are defined.

`EBML` uses a simple approach of building `Elements` upon their length, and value) as this approach is well known, easy to parse, and selective data parsing. The `EBML` structure additionally allows for an arrangement to support complex structural formats in an efficient manner.

Notation and Conventions

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

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

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

EBML: Extensible Binary Meta Language

Element Data: The value(s) of the `EBML Element` which is identified by the `Element ID` and `Element Data Size`. The form of the `Element Data` in this document and the corresponding `EBML Schema` of the `Element` is defined by the `Element Type`.

Element Data Size: An expression, encoded as a `Variable String`, that defines the length in octets of `Element Data`.

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

[[Docs](#)] [[txt](#)] [[pdf](#)] [[xml](#)] [[html](#)] [[Tracker](#)] [[WG](#)] [[Email](#)] [[Diff1](#)] [[Diff2](#)] [[Nits](#)]

Versions: ([draft-lhomme-cellar-ebml](#)) 00

cellar

S. Lhomme

Internet-Draft

Intended status: Standards Track

D. Rice

Expires: March 27, 2017

M. Bunkus

September 23, 2016

Extensible Binary Meta Language draft-ietf-cellar-ebml-00

Abstract

This document defines the Extensible Binary Meta Language (EBML) format as a generalized 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/broadcastn range)
- Definition of EBML Schema to express the structure of EBML Document Types
- 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

Matroska Colour Management Metadata

Colour

MatrixCoefficients
BitsPerChannel
ChromaSubsamplingHorz
ChromaSubsamplingVert
CbSubsamplingHorz
CbSubsamplingVert
ChromaSitingHorz
ChromaSitingVert
Range
TransferCharacteristics
Primaries
MaxCLL
MaxFALL

MasteringMetadata

PrimaryRChromaticityX
PrimaryRChromaticityY
PrimaryGChromaticityX
PrimaryGChromaticityY
PrimaryBChromaticityX
PrimaryBChromaticityY
WhitePointChromaticityX
WhitePointChromaticityY
LuminanceMax
LuminanceMin

Defining how to define support for encodings within Matroska.

Many encodings requires private data to contextualize the encoding.

Codec ID	Name	Description	
Video			
V_MS/VFW/FOURCC	Microsoft (TM) Video Codec Manager (VCM)	V_MS/VFW/FOURCC - Microsoft (TM) Video Codec Manager (VCM) The private data contains the VCM structure BITMAPINFOHEADER including the extra private bytes, as defined by Microsoft . The data are stored in little endian format (like on IA32 machines). Where is the Huffman table stored in HuffYUV, not AVISTREAMINFO ??? And the FourCC, not in AVISTREAMINFO.fccHandler ???	
V_UNCOMPRESSED	Video, raw uncompressed video frames	The private data is void, all details about the used colour specs and bit depth are to be put/read from the KaxCodecColourSpace elements.	
V_MPEG4/ISO/???	MPEG4 ISO Profile Video	The stream complies with, and uses the CodecID for, one of the MPEG-4 profiles listed below.	
	V_MPEG4/ISO/SP	MPEG4 ISO simple profile (DivX4)	stream was created via improved codec API (UCI) or even transmuxed from AVI (no b-frames in Simple Profile), frame order is coding order
	V_MPEG4/ISO/ASP	MPEG4 ISO advanced simple profile (DivX5, XviD, FFMPEG)	stream was created via improved codec API (UCI) or transmuxed from MP4, not simply transmuxed from AVI! Note there are differences how b-frames are handled in these native streams, when being compared to a VFW created stream, as here there are no dummy frames inserted, the frame order is exactly the same as the coding order, same as in MP4 streams!
	V_MPEG4/ISO/AP	MPEG4 ISO advanced profile	stream was created ... (see above)
V_MPEG4/MS/V3	Microsoft (TM) MPEG4 V3	and derivatives, means DivX3, Angelpotion, SMR, etc.; stream was created using Vfw codec or transmuxed from AVI; note that V1/V2 are covered in Vfw compatibility mode	
V_MPEG1	MPEG 1	The matroska video stream will contain a demuxed Elementary Stream (ES), where block boundaries are still to be defined. Its recommended to use MPEG2MKV.exe for creating those files, and to compare the results with selfmade	

User

- [Sign in](#)
- [New account](#)
- [Preferences](#)

Groups

- [Active WGs](#)
- [Active RGs](#)
- [Other](#)

By area/parent

- [Applications and Real-Time](#)
- [General](#)
- [Internet](#)
- [Ops & Mgmt](#)
- [Routing](#)
- [Security](#)
- [Transport](#)
- [IRTF](#)

New work

- [Chartering groups](#)
- [BOFs](#)

Other groups

Codec Encoding for LossLess Archiving and Realtime transmission (cellar)

Documents

[Charter](#)[Meetings](#)[History](#)[Photos](#)[Email expansions](#)[List archive](#)[Tools »](#)

Document	Date	Status	IPR	AD / Shepherd
Active Internet-Draft				
draft-ietf-cellar-ebml-00 Extensible Binary Meta Language	2016-09-23 27 pages New	I-D Exists WG Document		

Document	Date	Status	IPR	AD / Shepherd
Related Internet-Drafts				
draft-lhomme-cellar-ebml-00 Extensible Binary Meta Language	2016-07-06 27 pages	I-D Exists		
draft-lhomme-cellar-matroska-00 Matroska	2016-07-08 220 pages	I-D Exists		
draft-niedermayer-cellar-ffv1-00 FF Video Codec 1	2016-07-06 28 pages	I-D Exists		

Atom feed:

[All changes](#)[Significant](#)[Subscribe to changes](#)[Export as CSV](#)

**Participation /
onlooking welcome via
GitHub repos and
mailing list**

Thanks

2016-09-27 @dericed @tessaAfallon #iasa_web