MediaConch

Implementation and policy checking on FFV1, Matroska, LPCM (and more)

Jérôme Martinez, MediaArea

Innovation Workshop - March 2017
What is MediaConch?

MediaConch is a conformance checker

- Implementation checker
- Policy checker
- Reporter
- Fixer
What is MediaConch?

Implementation and Policy reporter

<table>
<thead>
<tr>
<th></th>
<th>Check by file upload</th>
<th>Check online files</th>
<th>Check server files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy</td>
<td>Choose a policy</td>
<td>Display MediaConch Html</td>
<td>Verbosity Default level</td>
</tr>
</tbody>
</table>

Check files

Results

Apply a policy to all results: Choose a new policy to apply

Show 10 entries

<table>
<thead>
<tr>
<th>Files</th>
<th>Implem</th>
<th>Policy</th>
<th>MediaInfo</th>
<th>MediaTrace</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>ffv1_test_pixfmt-yuv444p10le...</td>
<td>✓ Valid</td>
<td>✗ Matroska is well described?</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-yuva422p_co...</td>
<td>✓ Valid</td>
<td>✗ Matroska is well described?</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-yuva444p_co...</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✗</td>
</tr>
<tr>
<td>veraPDF test suite 6-1-10-t0...</td>
<td>✗ Not valid</td>
<td>✗ Matroska is well described?</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>train1.tif</td>
<td>✗ Not valid</td>
<td>✗ Matroska is well described?</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>buggy_header.pdf</td>
<td>✗ Not valid</td>
<td>✗ Matroska is well described?</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Showing 11 to 16 of 16 entries
What is MediaConch?

Implementation report:

MediaConch Report
File: C:\temp\FFV1+PCM_WithChecksum_Untouched.mkv
MediaConch EBML Implementation Checker
Toggle all verbosity: ☑

- EBML-ELEM-START Tests run: 1 | Results: ✓
- EBML-VER-COH Tests run: 1 | Results: ✓
- EBML-DOCV-CH Test run: 1 | Results: ✓
- EBML-ELEMENT-VALID-PARENT Tests run: 87 | Results: ✓
- EBML-ELEMENT-NONMULTIPLES Tests run: 70 | Results: ✓
- EBML-ELEMENT-CONTAINS-MANDATES Tests run: 43 | Results: ✓
- EBML-ELEMENT-IN-SIZE-RANGE Tests run: 43 | Results: ✓
- EBML-VALID-MAXID Tests run: 1 | Results: ✓
- EBML-VALID-MAXSIZE Tests run: 1 | Results: ✓
- HEADER-ELEMENTS-WITHIN-LENGTH LIMIT Tests run: 1 | Results: ✓
- ELEMENTS-WITHIN-MAXLENGTH Tests run: 1 | Results: ✓
- HEADER-ELEMENTS-WITHIN-MAXSIZELENGTH Tests run: 1 | Results: ✓
- ELEMENTS-WITHIN-MAXSIZELENGTH Tests run: 1 | Results: ✓
- MKV-SEEK-RESOLVE Tests run: 4 | Results: ✓
- EBML-CRC-FIRST Tests run: 6 | Results: ✓
- EBML-CRC-VALID Tests run: 6 | Results: ✓
- MKV-VALID-TRACKTYPE-VALUE Tests run: 2 | Results: ✓
- MKV-VALID-BOOLEANS Tests run: 3 | Results: ✓
MediaConch FFV1 Implementation Checker
- FFV1-LOCAL-CRC-VALID Tests run: 4 | Results: ✓
MediaConch PCM Implementation Checker

Policy report:

MediaConch Report
File: C:\temp\FFV1+PCM_WithChecksum_Untouched.mkv

Example MKV FFV1 digitization policy ✗ fail
Example of a digitization specification of analog SD video to FFV1 and Matroska.
Type: and Rules run: 17 | Fail count: 5 | Pass count: 12

- Is it Matroska? ✓ pass
- Matroska version 4 or greater? ✓ pass
- Unique ID is present? ✓ pass
- Is the video FFV1? ✓ pass
- FFV1 is version 3.4 or later? ✓ pass
- FFV1 is encoded in GOP size of 1? ✗ fail
- FFV1 uses slice crcs? ✓ pass
- Display Aspect Ratio is 4/3? ✗ fail (Actual: 1.222)
- Frame Rate is Constant? ✓ pass
- ColorSpace is YUV? ✗ fail (Actual: RGB)
- Chroma Subsampling is 4:2:2? ✗ fail
- Audio is PCM? ✓ pass
- Audio is 48000 Hz? ✓ pass
- Is this NTSC or PAL SD? ✗ fail
- Bit Depth is 8 or 10? ✓ pass
- Audio is Stereo or Mono? ✓ pass
- Bit Depth is 16 or 24? ✓ pass
What is MediaConch?

General information about your files

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C:\Programmation\PreFormaMediaInfo\SampleTestFiles\FFV1\ffv1_3.mkv</td>
<td></td>
</tr>
<tr>
<td>General</td>
<td></td>
</tr>
<tr>
<td>UniqueID</td>
<td>88323790047680325859674626238128084708</td>
</tr>
<tr>
<td>Format</td>
<td>Matroska</td>
</tr>
<tr>
<td>Format_Version</td>
<td>4</td>
</tr>
<tr>
<td>FileSize</td>
<td>126167</td>
</tr>
<tr>
<td>Duration</td>
<td>1.000</td>
</tr>
<tr>
<td>OverallBitRate</td>
<td>1009336</td>
</tr>
<tr>
<td>FrameRate</td>
<td>25.000</td>
</tr>
<tr>
<td>FrameCount</td>
<td>25</td>
</tr>
<tr>
<td>StreamSize</td>
<td>2511</td>
</tr>
<tr>
<td>Video</td>
<td></td>
</tr>
<tr>
<td>StreamOrder</td>
<td>0</td>
</tr>
<tr>
<td>ID</td>
<td>1</td>
</tr>
<tr>
<td>UniqueID</td>
<td>1</td>
</tr>
<tr>
<td>Format</td>
<td>FFV1</td>
</tr>
<tr>
<td>Format_Version</td>
<td>3.4</td>
</tr>
<tr>
<td>CodecID</td>
<td>V_MS/VFW/FOURCC / FFV1</td>
</tr>
<tr>
<td>Duration</td>
<td>1.000</td>
</tr>
<tr>
<td>BitRate</td>
<td>989250</td>
</tr>
<tr>
<td>Width</td>
<td>320</td>
</tr>
</tbody>
</table>
What is MediaConch?

Inspect your files
What is MediaConch?

Policy editor
What is MediaConch?

Public policies

- **Video file is MKV + FFV1-Intra + PCM or FLAC with CRC32 everywhere**
  - Test that the video file is suitable for archiving purposes from my point of view ;).
  - Container format is Matroska with error detection (CRC)
  - Video format is FFV1 with error detection (CRC) and with Intra mode (each frame is independent)
  - Audio format is PCM (unfortunately it cannot contain error detection) or FLAC (it has CRC by design)
  - Maintainer: Jérôme Martinez (MediaArea)
  - License: CC-BY-SA-4.0+

- **PDF is PDF/A**
  - Test that a PDF is suitable for archives.
  - Note: for the moment, test that it is marked as PDF/A. Other ideas?
  - Maintainer: Jérôme Martinez (MediaArea)
  - License: CC-BY-SA-4.0+

- **TIFF is Raw**
  - Test that a TIFF file is suitable for archive.
  - Note: for the moment, test that it is raw. Other ideas?
  - Maintainer: Jérôme Martinez (MediaArea)
  - License: CC-BY-SA-4.0+

- **Austrian Mediathek: Preservation Master (Video)**
  - PAL/NTSC, FFV1 version 0/1, PCM 44.1/48kHz in AVI
  - Maintainer: Peter B.
  - License: CC-BY-4.0+
What is MediaConch?

Fixer

- Segment sizes in Matroska
- Matroska “bit flip” correction
- FFV1 “bit flip” correction
Integration

Archivematica is an integrated suite of open-source software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model.
MediaConch interfaces

- Graphical interface
- Web interface
- Command line
- Server (REST API)
- (Work in progress) a library (.dll/.so/.dylib)
MediaConch output formats

- XML (native format)
- Text
- HTML
- (Work in progress) PDF
- Tweakable! (with XSL)
Open source

- GPLv3+ and MPLv2+
- Relies on MediaInfo (metadata extraction tool)
- Use well-known open source libraries: Qt, sqlite, libevent, libxml2, libxslt, libexslt...
Supported formats

• Priorities for the implementation checker
  - Matroska
  - FFV1
  - PCM

• Can accept any format supported by MediaInfo for the policy checker
  - MXF + JP2k
  - QuickTime/MOV
  - Audio files (WAV, BWF, AIFF...)
  - ...
Supported formats

Can be expanded

- By plugins
  - Support of PDF checker: VeraPDF plugin
  - Support of TIFF checker: DPF Manager plugin
  - You use another checker? Let us know
- By internal development
  - More tests on your preferred format is possible
  - It depends on you!
Versatile

Several input formats are accepted

- FFV1 from MOV or AVI
- Matroska with other video formats
- (Work in progress) Extraction of a PDF or TIFF attachment from a Matroska container and analyze with a plugin (e.g. VeraPDF and DPF Manager)
- ...
Versatile

Input can be from:

- Files (local/network)
- FTP/FTPS/SFTP
- HTTP/HTTPS
- Amazon S3
Versatile

Binaries are provided for:

- Windows
- Mac
  Homebrew users: "brew install mediaconch", that's all!
- Linux (Ubuntu, Debian, Fedora, OpenSUSE...)
  Since Ubuntu 16.04 and Debian Testing/9 users:
  "apt-get install mediaconch" or in Ubuntu Store, that's all!
  (it is in the official distros repository)
- Embedded devices? Doable
  (we tested it on a Raspberry Pi )
- Can be ported on other distros (BSD...)

(it is in the official distros repository)
Standardization

- Matroska is widely used but not (yet) standardized
- FFV1 is gaining increasing usage in preservation contexts but is not (yet) standardized
CELLAR: IETF workgroup

- Open standards group
- Goal to IETF-standardize Matroska/FFV1/FLAC
- A lot of progress, especially with Matroska/EBML specs
- [https://datatracker.ietf.org/wg/cellar/charter/](https://datatracker.ietf.org/wg/cellar/charter/)
**FFV1 performance**

- **NOA tested on SD 8-bit content:**
  - i7-2600 (4 cores+HT, 3.4-3.8 GHz)
  - 3-4x real time
  - 4-5x decoding speed increase compared to JP2k

- **VIAA tested on SD 10-bit content (FFmpeg 3.2):**
  - E5-2698V3 (16 cores+HT, 2.3-3.6 GHz)
  - 0.7x real time/thread, 11-12x real time/all cores+HT
  - 3-4x decoding speed increase compared to JP2k
  - Better compression ratio by 2% compared to JP2k
    (9% with additional FLAC audio compression)
FFV1 performance

- This is an average, results varies depending on the content of files
  - From 0.4x to 2.4x (average 0.7x) real time/thread (encoding/decoding)
  - From 0.7x to 16x (average 3.5x) the speed of JP2k (FFmpeg)
- Not convinced?
  - Test on your own files
  - MediaArea will provide test scripts
  - We can perform tests for you
Worldwide

- 2 project leaders
  - Jerôme Martinez (Digital Media Analysis Specialist, France)
  - Dave Rice (Archivist, USA)

- Presentations worldwide
  - IASA, France
  - FIAT/IFTA, Austria
  - FOSDEM, Belgium
  - AMIA, USA
  - Code4Lib, USA
  - JTS, Singapore
  - (3-6 October 2016) IPRES, Switzerland
  - (25-29 September 2016) IASA, USA
Matroska research corpus

- We analyze all Matroska files from archive.org
- Interface with some statistics of Matroska elements usage (e.g. files with CRC-32 elements...)

https://mediaarea.net/MediaConchCorpus/
What's next?

- Continue to improve handling of huge collections
- Continue to improve user interface
- Support of embedded attachments
- Statistics
- Finish standardization of Matroska and FFV1
- More conformance tests
- More fixing cases
And after PREFORMA sponsorship?

It depends on you!

- This is open source
- Driven by user requests
- Everyone can develop or sponsor a development
- Potential features:
  - Support of tests for your prefered format (MOV? MXF? JP2k? WAV?)
  - Support of other checkers (BWF MetaEdit? QCTools?)
  - Integration in your workflow
  - ...
### Example (Plugins)

#### Results

<table>
<thead>
<tr>
<th>Files</th>
<th>Implementation</th>
<th>Policy</th>
<th>MediaInfo</th>
<th>MediaTrace</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>ffv1_test_pixfmt-yuv44p10le...</td>
<td>✔️ Valid</td>
<td>✔️ PDF is PDF/A</td>
<td>📋</td>
<td>📋</td>
<td>📊 Analyzed</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-yuva422p_co...</td>
<td>✔️ Valid</td>
<td>✔️ PDF is PDF/A</td>
<td>📋</td>
<td>📋</td>
<td>📊 Analyzed</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-yuva44p_co...</td>
<td>✔️ Valid</td>
<td>✔️ PDF is PDF/A</td>
<td>📋</td>
<td>📋</td>
<td>📊 Analyzed</td>
</tr>
<tr>
<td>veraPDF test suite 6-1-10-40...</td>
<td>✗ Not valid</td>
<td>✔️ PDF is PDF/A</td>
<td>📋</td>
<td>📋</td>
<td>📊 Analyzed</td>
</tr>
<tr>
<td>train1.tif</td>
<td>✗ Not valid</td>
<td>✔️ PDF is PDF/A</td>
<td>📋</td>
<td>📋</td>
<td>📊 Analyzed</td>
</tr>
<tr>
<td>buggy_header.pdf</td>
<td>✗ Not valid</td>
<td>✗ PDF is PDF/A</td>
<td>📋</td>
<td>📋</td>
<td>📊 Analyzed</td>
</tr>
</tbody>
</table>

Showing 11 to 16 of 16 entries
Example (Plugins)

MediaConch Report

File: buggy_header.pdf
PDF/A-1B validation profile
PDF file is not compliant with Validation Profile requirements.
  Toggle all verbosity: 

- ISO 19005-1:2005/6.3.7(3) Tests run: 1 | Results: ✗ Fail count: 1
  Name: isSymbolic == false || nrCmaps == 1
  Results: fail ✗
  specification: ISO 19005-1:2005
  clause: 6.3.7
  testNumber: 3
  description: Font programs' "cmap" tables for all symbolic TrueType fonts shall contain exactly one encoding
  object: TrueTypeFontProgram
  Value context: root/document[0]/pages[0](4 0 obj PDPage)/contentStream[0](5 0 obj PDContentStream)/operators[9]/font[0](NFXXYB+Calibri)
  /fontFile[0]

- ISO 19005-1:2005/6.2.3(2) Tests run: 1 | Results: ✗ Fail count: 1
- ISO 19005-1:2005/6.1.8(1) Tests run: 14 | Results: ✗ Fail count: 2
- ISO 19005-1:2005/6.1.7(2) Tests run: 5 | Results: ✗ Fail count: 1
- ISO 19005-1:2005/6.7.11(1) Tests run: 1 | Results: ✗ Fail count: 1
Example (Plugins)

MediaConch Report

File: train1.tif
dpfmanager: Baseline 6.0

{count[tags.tag[name=SubIFDs]] == 1} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=ImageLength] > tags.tag[name=ImageLength]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=ImageWidth] > tags.tag[name=ImageWidth]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=NewSubfileType]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=NewSubfileType.cardinality == 1]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=NewSubfileType] == 0} | {tags.tag[name=SubIFDs].ifd.tags.tag[name=NewSubfileType] == 1} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=ImageDescription]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=ImageDescription.cardinality == 1]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=ImageWidth.cardinality == 1]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=Compression.cardinality == 1]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=XResolution]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=YResolution]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=XResolution.cardinality == 1]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=YResolution.cardinality == 1]} Tests run: 1 | Results: * Fail count: 1
{tags.tag[name=SubIFDs].ifd.tags.tag[name=Make]} Tests run: 1 | Results: * Fail count: 1
Stay in touch

MediaArea: https://mediaarea.net, @MediaArea_net

MediaConch: https://mediaarea.net/MediaConch, @MediaConch

Jérôme Martinez: jerome@mediaarea.net

Slides: https://mediaarea.net/Events

License: CC BY