MediaConch

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

Jérôme Martinez
MediaArea.net SARL

Open Source Preservation Workshop - April 2016
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>Files</th>
<th>Implementation</th>
<th>Policy</th>
<th>MediaInfo</th>
<th>MediaTrace</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coconut.mp4</td>
<td>✅ Valid</td>
<td>✖️ Is Matroska</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>Example.mp4</td>
<td>✅ Valid</td>
<td>✖️ Is Matroska</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_3 - Copie (2).m...</td>
<td>✅ Valid</td>
<td>✅ Is Matroska</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_3 - Copie.mkv</td>
<td>✅ Valid</td>
<td>✅ Is Matroska</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_3.mkv</td>
<td>✅ Valid</td>
<td>✅ Is Matroska</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
</tbody>
</table>
What is MediaConch?

Example of report

<table>
<thead>
<tr>
<th>Tracktype</th>
<th>Tracktypeorder</th>
<th>Trackid</th>
<th>Actual</th>
<th>Outcome</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>Matroska</td>
<td></td>
<td>✔ pass</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Video Width is equal to 720 (pixels)

<table>
<thead>
<tr>
<th>Tracktype</th>
<th>Tracktypeorder</th>
<th>Trackid</th>
<th>Actual</th>
<th>Outcome</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video</td>
<td></td>
<td>1</td>
<td>640</td>
<td>❌ fail</td>
<td>is not equal</td>
</tr>
</tbody>
</table>
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

To create policies, users may select fields from MediaInfo's stream types (General, Video, Audio, etc.) and apply validators (Equal, Not Equal, Greater Than, etc.) to the associated field values. For example, a policy item may state that the "Width" field from the file's Video stream equal 640 pixels. When adding the field's value, however, be sure to leave off any associated strings, such as "pixels" "Kbps" "fps", etc.

**Rule name** *

Quicktime wrapper

**Editor type** :
- Editor
- Free text

**Track type** *

General

**Field** *

Format_Profile

**Occurrence** *

*

**Validator** *

Is equal (==)

**Value**

QuickTime
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 attachement 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...)
  Ubuntu 16.04 and Debian Testing/9 users:
  "apt-get install mediaconch", 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...)

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/
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?

Still under development but already usable
(PREFORMA prototyping phase up to end 2016)

- Better handling of huge collections
- Better user interface
- Statistics
- Standardize Matroska and FFV1
- More conformance tests
- Integration in Archivematica
- Fixer
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:
  - Integration of test of your preferred format
    (MXF? doable. JP2k? doable. WAV? doable...)
  - Integration of other checkers
    (BWF MetaEdit? QCTools?)
  - Better integration in your workflow
  - ...
### Check files

**Policy**: General Conformance  
**Display**: MediaConch Html  

![Check files](https://example.com/checkfiles.png)

### 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_0.mov</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-yuva444p_coder-1_1...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-bgr0_coder-0_level...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-gbrp14le_coder-1_1...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_0.mkv</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-yuv422p_coder-0_le...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffv1_test_pixfmt-yuv420p_coder-...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>In queue</td>
</tr>
</tbody>
</table>
Example (Command line)
Example (Plugins)

Check files

[Check by file upload] [Check online files] [Check server files]

Policy: Policy Set Example
Display: Choose a display
Verbosity: Default level

Check files

Results

Show: 10 entries

<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>ffmpeg_test_pixfmt-yuv444p10le_coder-1165664662</td>
<td>Valid</td>
<td>Policy Set Example</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffmpeg_test_pixfmt-yuv444p10le_coder-1165664662</td>
<td>Valid</td>
<td>Policy Set Example</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>ffmpeg_test_pixfmt-yuv444p10le_coder-1165664662</td>
<td>Valid</td>
<td>Policy Set Example</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>vcrasPDF test suite 6-1-10-401-pass-1165664662</td>
<td>Valid</td>
<td>N/A</td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
<tr>
<td>train1.tif</td>
<td>Not valid</td>
<td></td>
<td></td>
<td></td>
<td>Analyzed</td>
</tr>
</tbody>
</table>

Showing 11 to 15 of 15 entries

[Previous] [1] [2] [Next]
Example (Plugins)
Example (Plugins)
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