MediaConch

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

Jérôme Martinez, MediaArea

Open Source Preservation Workshop - April 2016

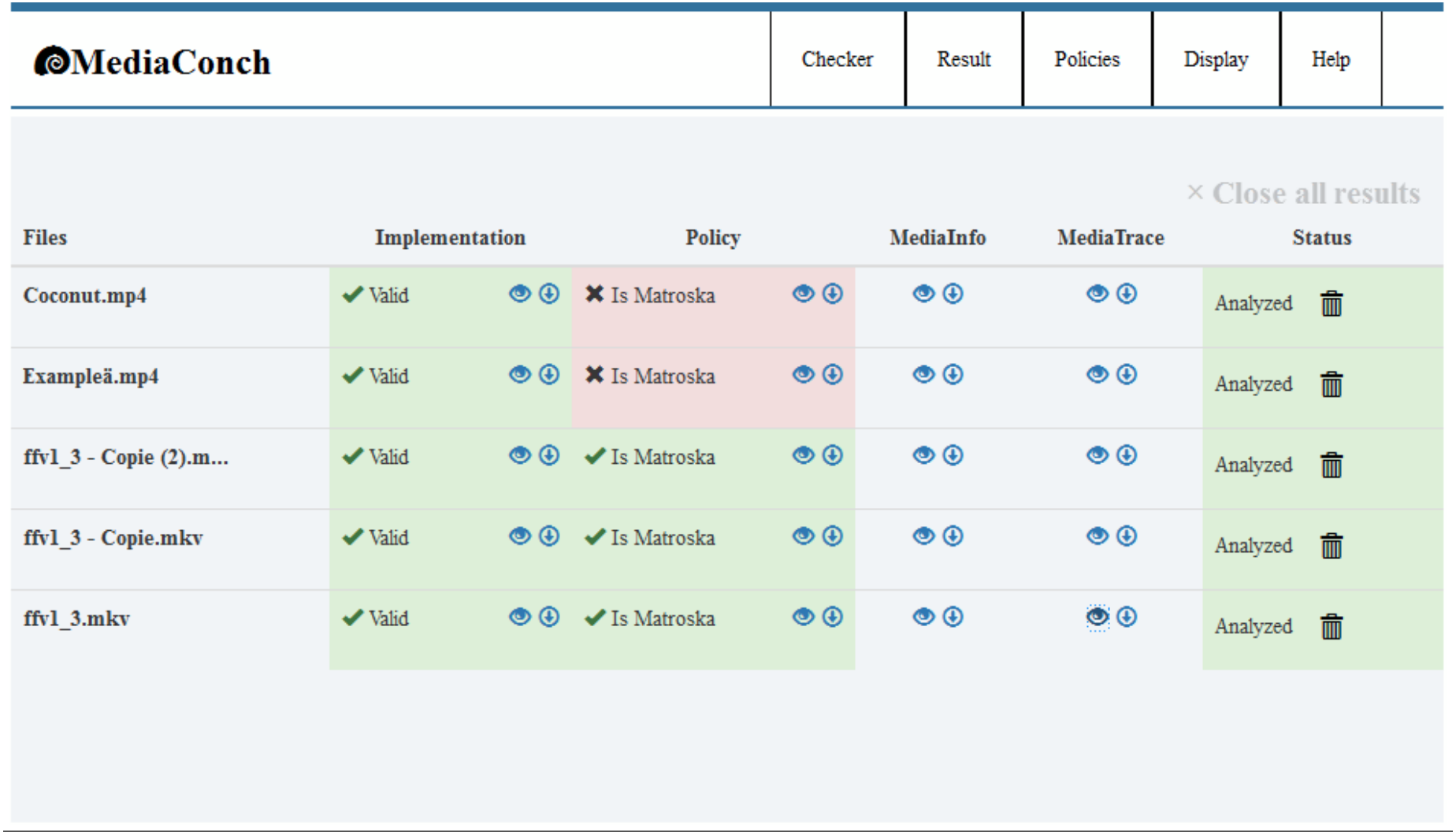
What is MediaConch?

MediaConch is a conformance checker

- Implementation checker
- Policy checker
- Reporter
- Fixer

What is MediaConch?

Implementation and Policy reporter



What is MediaConch?

Example of report

MKV/FFV1 (reformatted QT/v210)

Policy check for Matroska-wrapped FFV1 and LPCM file, transcoded from QuickTime-wrapped v210 and LPCM file.

General Format equals Matroska

Context (field): Format

Context (value): Matroska

tracktype	tracktypeorder	trackid	actual	outcome	reason
General			Matroska	✓ pass	

Video Width is equal to 720 (pixels)

Context (field): Width

Context (value): 720

tracktype	tracktypeorder	trackid	actual	outcome	reason
Video		1	640	✗ fail	is not equal

What is MediaConch?

General information about your files

Key	Value
C:/Programmation/PreFormaMediaInfo/SampleTestFiles/FFV1/ffv1_3.mkv	
General	
UniqueID	88323790047680325859674626238128084708
Format	Matroska
Format_Version	4
FileSize	126167
Duration	1.000
OverallBitRate	1009336
FrameRate	25.000
FrameCount	25
StreamSize	2511
Video	
StreamOrder	0
ID	1
UniqueID	1
Format	FFV1
Format_Version	3.4
CodecID	V_MS/VFW/FOURCC / FFV1
Duration	1.000
BitRate	989250
Width	320

What is MediaConch?

Inspect your files

Offset	Key	Value
0x00000000	EBML (47 bytes)	
0x0000002f	Segment (126120 bytes)	
0x0000002f	Header (12 bytes)	
0x0000003b	SeekHead (66 bytes)	
0x0000007d	Void (157 bytes)	
0x0000011a	Info (81 bytes)	
0x0000016b	Tracks (167 bytes)	
0x0000016b	Header (12 bytes)	
0x00000177	TrackEntry (155 bytes)	
0x00000177	Header (9 bytes)	
0x00000180	TrackNumber - 1 (3 bytes)	
0x00000183	TrackUID - 1 (4 bytes)	
0x00000187	FlagLacing - 0 (3 bytes)	
0x0000018a	Language - eng (7 bytes)	
0x00000191	TrackType - 1 (3 bytes)	
0x00000194	DefaultDuration - 40000000 (8 bytes)	
0x0000019c	CodecID - V_MS/VFW/FOURCC (17 bytes)	
0x000001ad	Video (16 bytes)	
0x000001bd	CodecPrivate - Copy of vids (85 bytes)	
0x000001bd	Header (3 bytes)	
0x000001c0	Size	81 (0x51)

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

CAVPP Preservation Video Files rules

- Quicktime wrapper
- 10 bit uncompressed 4:2:2 video
- Total Bitrate
- Frame Rate 29.970
- Aspect Ratio 4/3
- Is Interlaced
- Audio is 48kHz
- Audio is stereo
- Audio is 24 bit

Add a new rule

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)

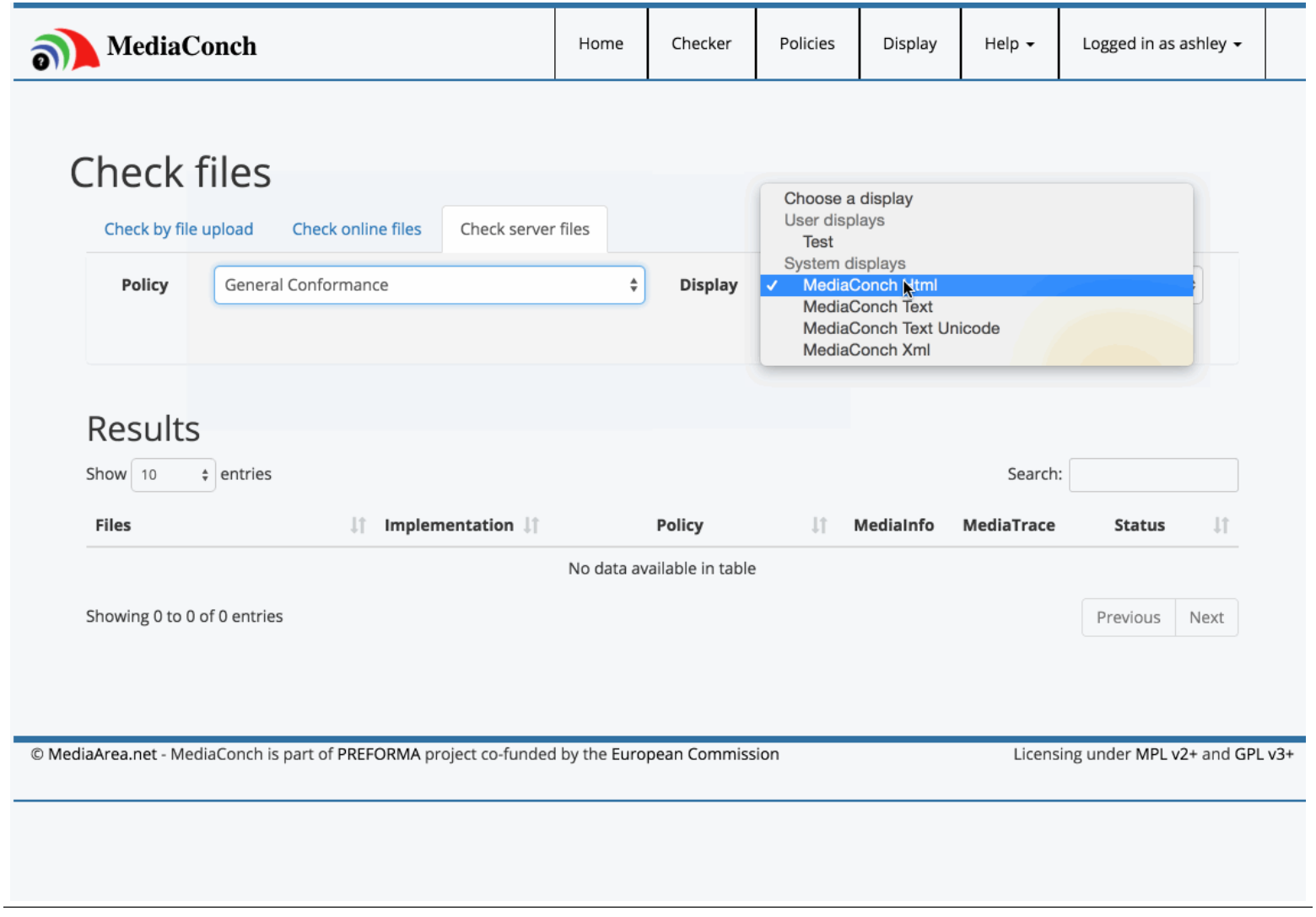
- Better handling of huge collections
- Better user interface
- Statistics
- Standardize Matroska and FFV1
- More conformance tests
- Integration in Archivematica
- Fixer

And after PREFORMA sponsorship?

It depends on you!

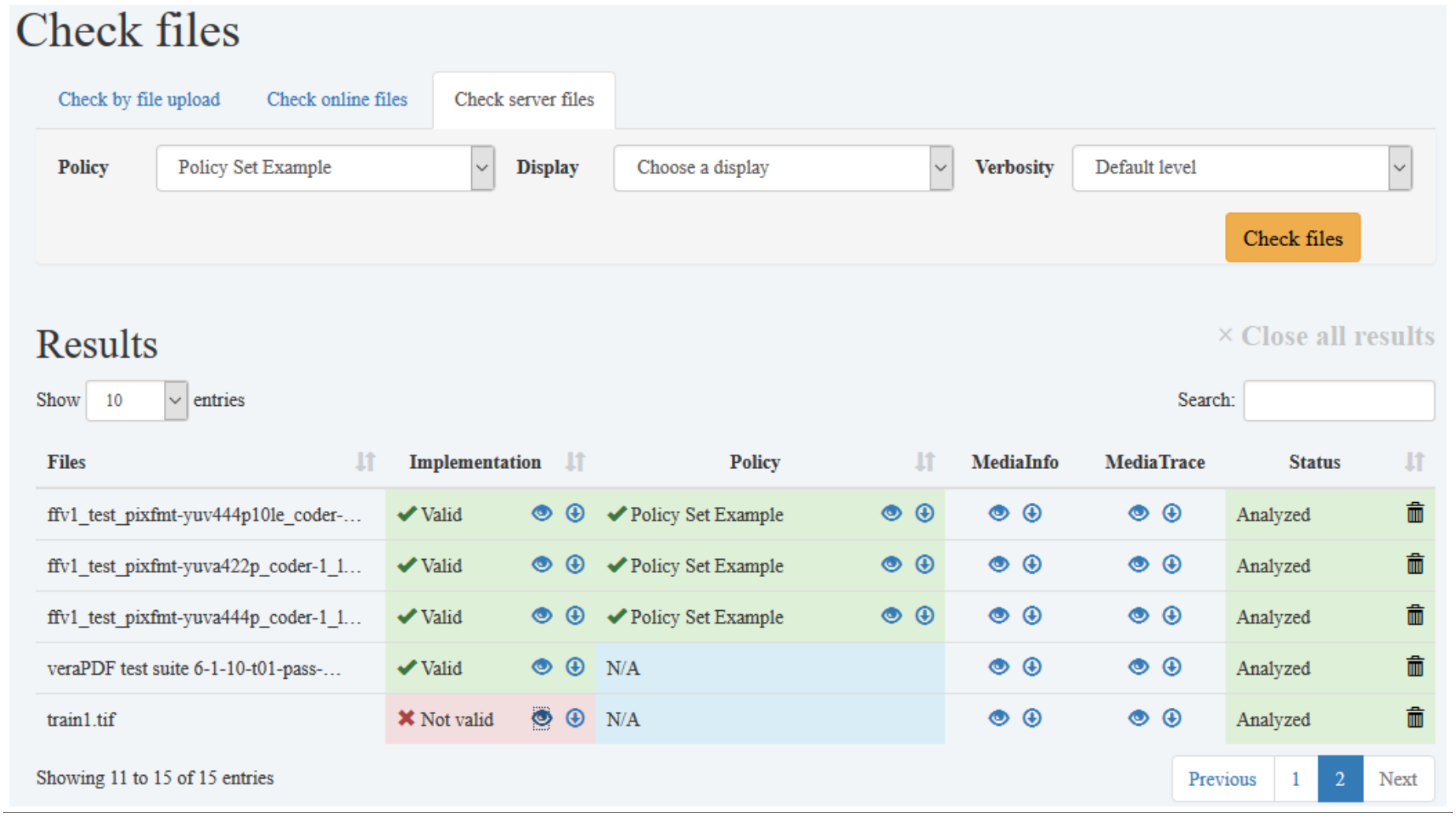
- This is open source
- Driven by user requests
- Everyone can develop or sponsor a development
- Potential features:
  - Integration of test of your prefered format (MXF? doable. JP2k? doable. WAV? doable...)
  - Integration of other checkers (BWF MetaEdit? QCTools?)
  - Better integration in your workflow
  - ...

Example (Online)

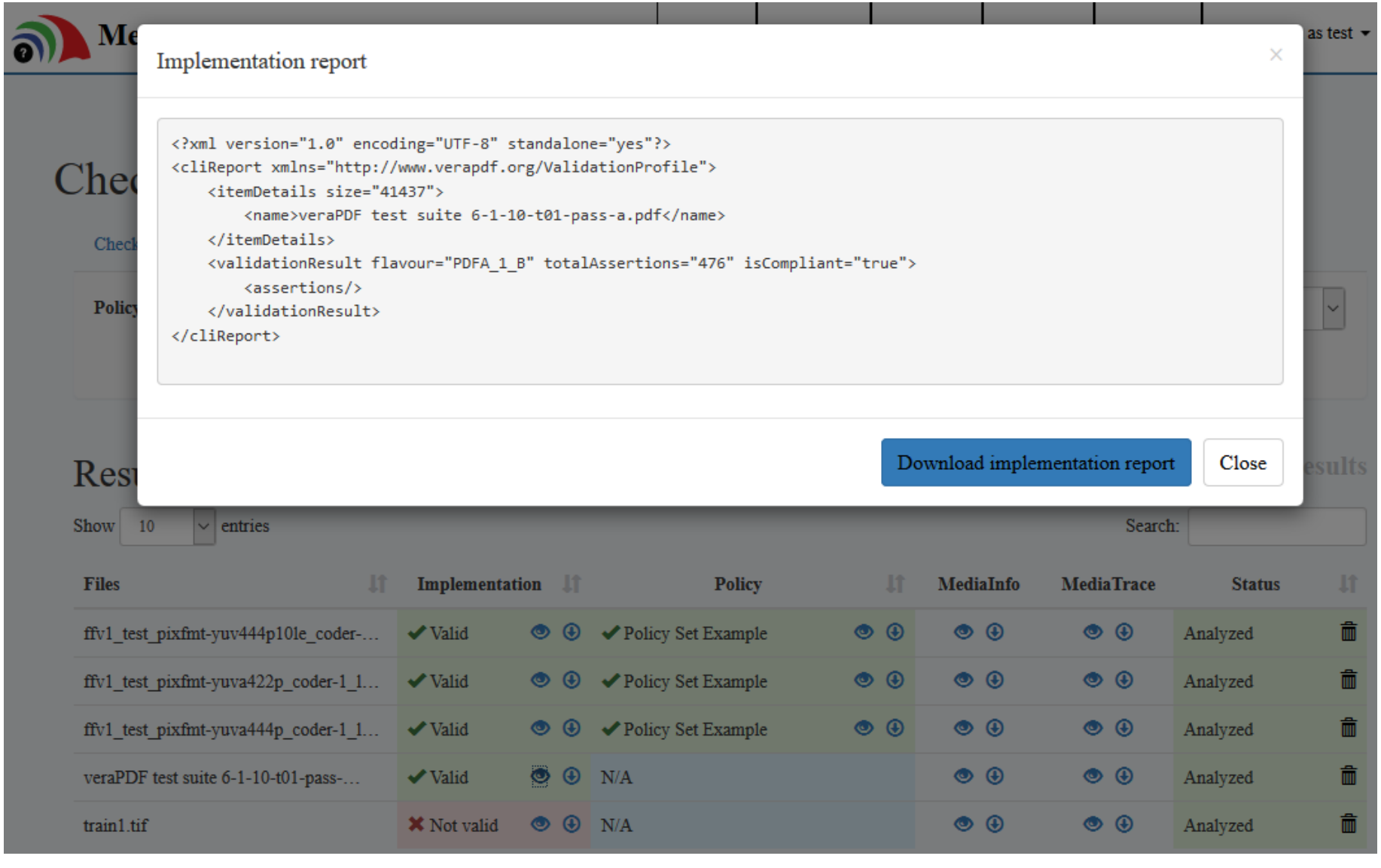


Example (Command line)

Example (Plugins)



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