FRAME

Preservation and digitization of audiovisual media

June 18-22

ina

FRAME

Open source quality control and quality management

MediaArea

Open source software company focused on digital media analysis.

FRAME

Open source quality control
and quality management

Why quality control is important?

Why quality control is important?

If you don't handle that now, content is definitely partially lost.

artifact: Vacuum Guide Adjustment Error
source: 2" Quadruplex

Why quality control is important?

Your digital copy provider is available now
> In the future, your content provider may have disappeared, trashed content…

The original tape is still there, but for how long?
> The original tape is still there, but for how long?
> Or your tape becomes no more readable.
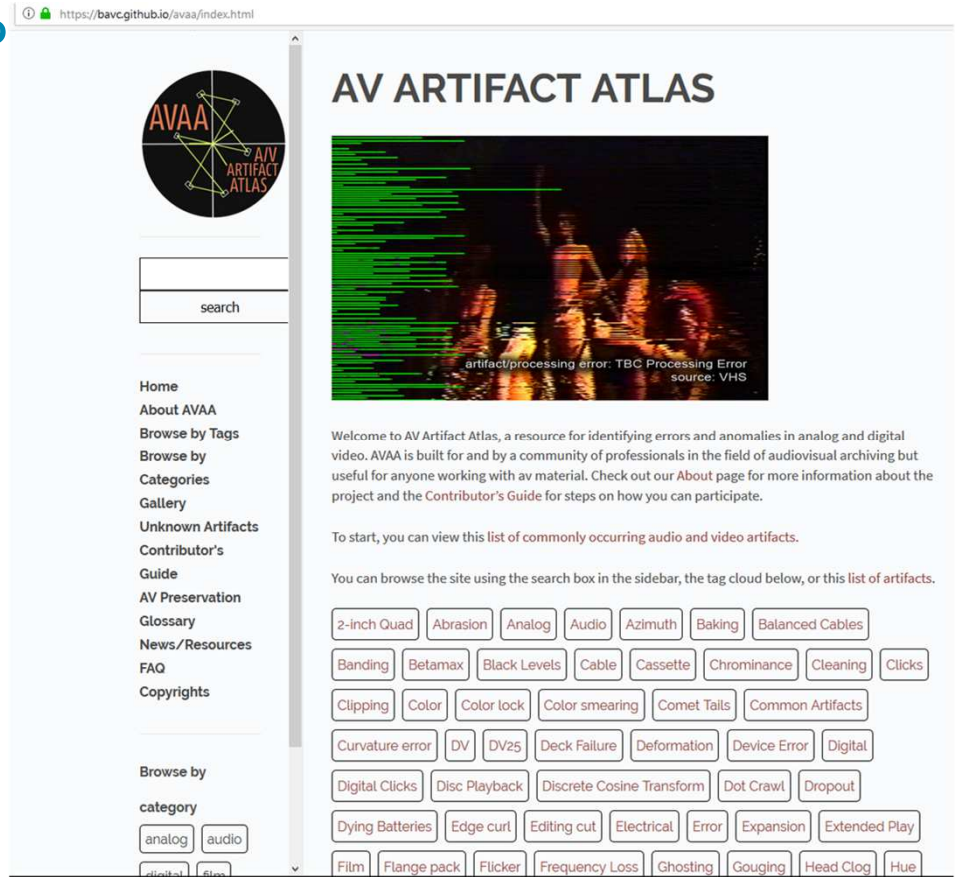
Why quality control is important?

Easier to catch quality issues and fix them now than later
> When someone else discovers quality issues, it may be too late.

Why quality control is important?

You are not alone

Example of online support:

AV Artifact Atlas, a resource for identifying errors and anomalies in analog and digital video. AVAA is built for and by a community of professionals in the field of audiovisual archiving but useful for anyone working with av material.

Why quality control is important?

You are not alone

Example of symposiums:

"No Time To Wait" symposiums are free events focused on open media, open standards, and digital audiovisual preservation.
Symposium program is made by archivists, who present how they work and the associated (open source) tools.

Organization of symposiums has a cost.
- Hosting granted by a main sponsor

- Other archives or vendors may provide additional financial support

"No Time To Wait" 2018 will be at British Film Institute on October 25-26.
Other sponsors are AV Preservation by reto.ch, Centre national de l'audiovisuel Luxembourg (CNA) and the International Federation of Film Archives (FIAT/IFTA)

Overview of open source

Overview of open source

The archival domain
>Relatively small market (niche)
>
>Off-the-shelf products often focus on different use cases
>
>Highly specialized demands
>
>High potential for vendor dependence...

Overview of open source

Proprietary tools and formats are linked to one, and only one, vendor

If vendor disappears, it is impossible to get bug fixes

Without source code, you can not adapt yourself to new platforms: are you sure you'll be able to run the software in 100 years?

Vendor decides on the new features
Are you big enough for making the vendor develop your prefered feature?

Overview of open source

The four freedoms of open source (a.k.a. free software)

The freedom to run the program as you wish, for any purpose

The freedom to study how the program works, and change it so it does your computing as you wish

The freedom to redistribute copies so you can help your neighbor

The freedom to distribute copies of your modified versions to others

Overview of open source

You are no more tied to only one vendor

YOU decide of the new features
If you don't like anymore what your supplier does, just change, YOU decide

For small project, usually one developer, but YOU can fork if it does not go as you want.

For bigger projects, YOU may decide between different competitors for new development or support.

Overview of open source

Reusability of existing solutions
- Free choice of (local) support/suppliers
- No black-box
- Common tools/codebase = larger userbase
- Less "forced" upgrades

Overview of open source

Benefits of paying for open source software

      Better support/updates

      Pooling resources

      Designed for your use-cases

      Overall better cost-effectiveness

      Public money = public solution?

Overview of open source

Open source does not mean lack of professional support:
    Support contracts
    Paid installation/integration
    Hire developers

Non-financial contributions:
    Documentation
    Testing / bug-reports
    Helping others

Overview of open source

A real-world example: The lossless video codec FFV1

    Austrian National Archive (Mediathek) wanted to do lossless digital video archiving

    Not satisfied with existing products (Interoperability issues)

    Found FFV1 in FFmpeg: Excellent codec, but we wanted/needed more...

    Contacted and hired a FFV1 developer

    Other parties involved for advices (pooled resources)

    Budget calculated in reference to costs of proprietary alternatives

    Now FFV1 version 3: faster and integrity-aware

    Important: published our experiences with FFV1

    So: other archives using FFV1 now profit from improvements, too!

Overview of open source

Quality of tools

    Chicken and egg: if everyone waits, nothing happens

    We start with needs easy to handle and we create dedicated open source tools

    Tools become bigger, step by step, when more people join

    YOU can decide about participating in lowering the overall cost for archives.

Overview of open source

Funding:

      Driven by user requests

Most of tools were funded after a need is detected by users

Why paying for something you don't control? (is your current choice future proof?)

Everyone (you included) can develop or sponsor a development

You can fork if you think you have a better idea than others

Overview of QCTools

Overview of QCTools

Helps users analyze and understand their digitized video files
through use of audiovisual analytics and filtering

Overview of QCTools

Overview of QCTools
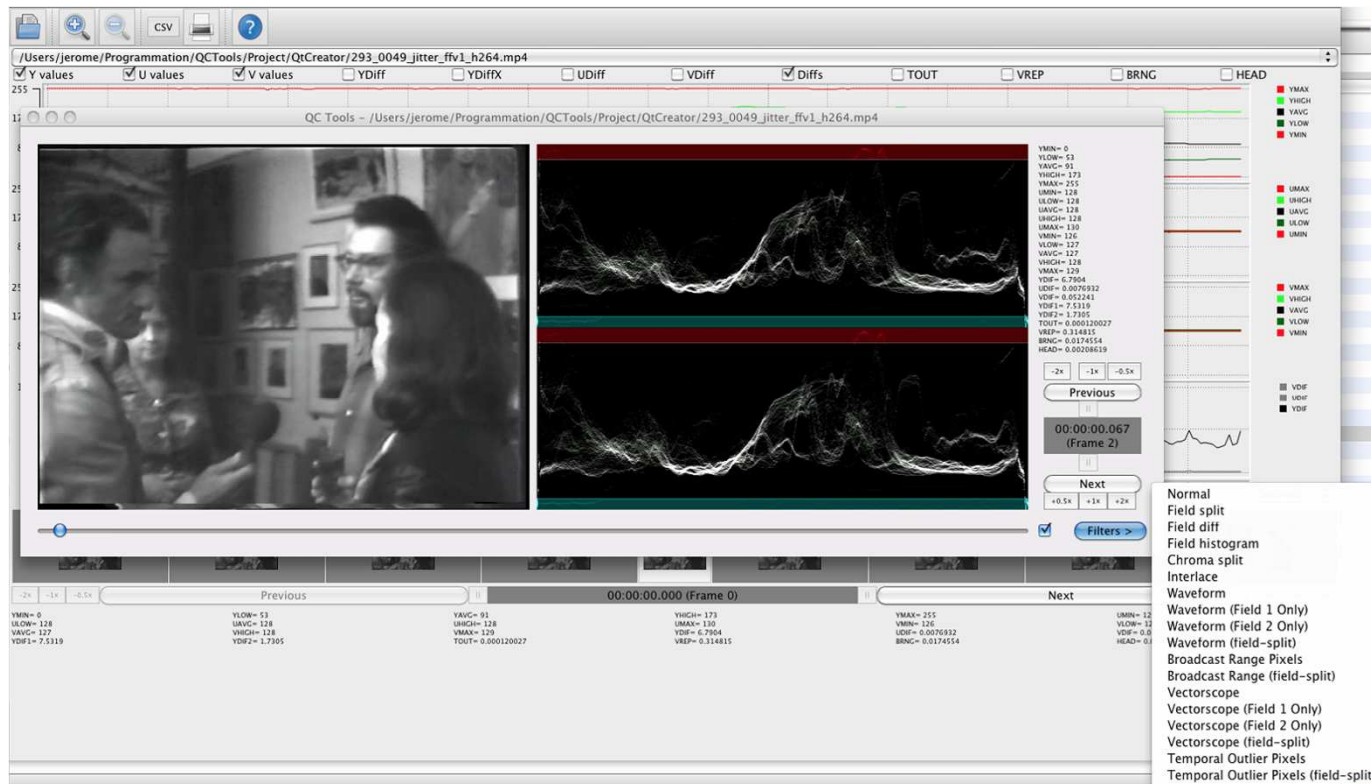
Overview of QCTools

Help in detecting issues. Example with field 2 missing color.

Overview of QCTools

Only 1 field has issues.

Overview of QCTools

Example of drop out.

Dropout is e.g. a brief loss of
RF caused by a media defect
such as a scratch on the
media surface.

Can be detected through
"TOUT" graph.
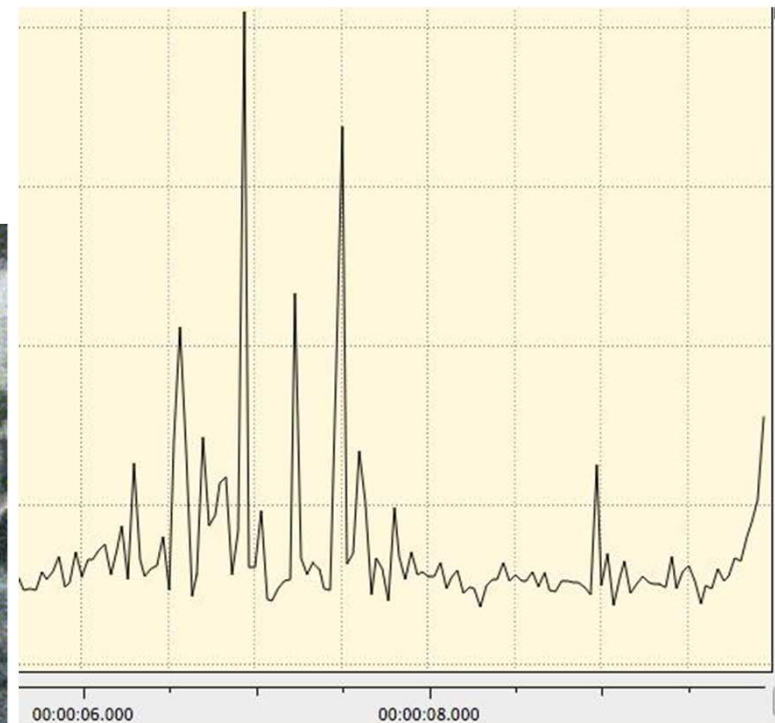
Overview of QCTools

Graph Domain	Graph(s) Name	8-bit values	10-bit values	Ideal Range (8/10)
Y Channel	Y MIN, Y LoW, Y AVG, Y HIGH, Y MAX	0-255	0-1023	16/64 (Y Low) 128/512 (Y AVG) 235/940 (Y HIGH)
UV Channels	MIN, LoW, AVG, HIGH, MAX	0-255	0-1023	16/64 (Low) 85/341-170/682 (AVG) 240/960 (HIGH)
Saturation	SAT MIN, SAT LoW, SAT AVG, SAT HIGH, SAT MAX	0-88.7 (75 %) 88.7-118.2 (100%) 118.2-181.02 (Illegal YUV)	0-354.8 (75%) 354.8-472.8 (100%) 472.8-724.08 (Illegal YUV)	75% values
Hue	Hue	0-360°	0-360°	120-147°
Temporal outliers	ToUT	0-1		0-0.009
Vertical Line Repetitions	VREP	0-1		0
Mean Square Error per Field	MSEf	0-1		As close to 0 as possible
Peak Signal to Noise Ratio	PNSR	0-60 dB		30-50 dB

Overview of MediaConch

Overview of MediaConch

MediaConch is a conformance checker
- Implementation checker
- Policy checker
- Reporter
- Fixer

Overview of MediaConch

How it was developed:

Initial development sponsored by European Union (PREFORMA project)

Focused on FFV1 and Matroska.

We added all formats supported by MediaInfo to the policy checker.

Overview of MediaConch

Implementation and Policy reporter

Overview of MediaConch

Implementation report:

Policy report:

Compared to a specification
or
Compared to your policy

Overview of 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

Overview of MediaConch

Inspect your files

Overview of MediaConch

Policy editor

Overview of MediaConch

Public policies

Public policies page lists policies our users would like to share with you.
If you want to share yours, go to policy editor page (don't forget to login in order to associate your policy to your account), select the policy you want so share and set the "policy visibility" field to "public".

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 can not contain error detection) or FLAC (it has CRC by design)

Maintainer: Jérôme Martinez (MediaArea)
License: CC-BY-SA-4.0+

[Add to my policies]   [Export]

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+

[Add to my policies]   [Export]

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+

[Add to my policies]   [Export]

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+

[Add to my policies]   [Export]

Overview of MediaConch

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

Overview of MediaConch

Supported formats:

      Priorities for the implementation checker:

      Matroska

      FFV1

      Can accept any format supported by MediaInfo for the policy checker

      MXF + JP2k

      QuickTime/MOV

      Audio files (WAV, BWF, AIFF…)

      …

Overview of MediaConch

Support of plugins

Overview of MediaConch

Versatile
Input can be from:
     Files (local/network)
     FTP/FTPS/SFTP
     HTTP/HTTPS
     Amazon S3

Binaries are provided for:
     Windows
     Mac (Homebrew users:"brew install mediaconch", that's all!)
     Linux (Ubuntu, Debian, Fedora, OpenSUSE...)

Overview of MediaConch

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?)
  - Integration in your workflow
  - ...

Other tools

MediaInfo

Convenient unified display of the most relevant technical and tag data for video and audio files.

**Output**: text, XML, HTML, PBCore, EBUCore, NISO Z39.87…

Container: format, profile, commercial name of the format, duration, overall bit rate, writing application and library, title, author, director, album, track number, date, duration…

Video: format, codec id, aspect, frame rate, bitrate, color space, chroma subsampling, bit depth, scan type, scan order…

Audio: format, codec id, sample rate, channels, bit depth, language, bit rate…

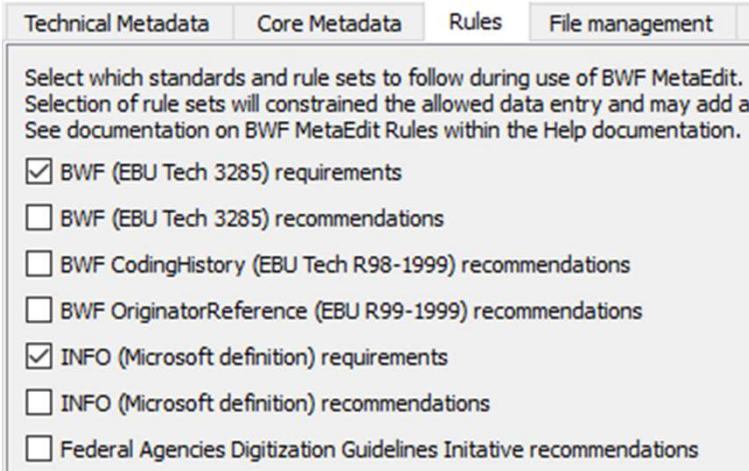**Text**: format, codec id, language of subtitle…

Chapters: count of chapters, list of chapters…
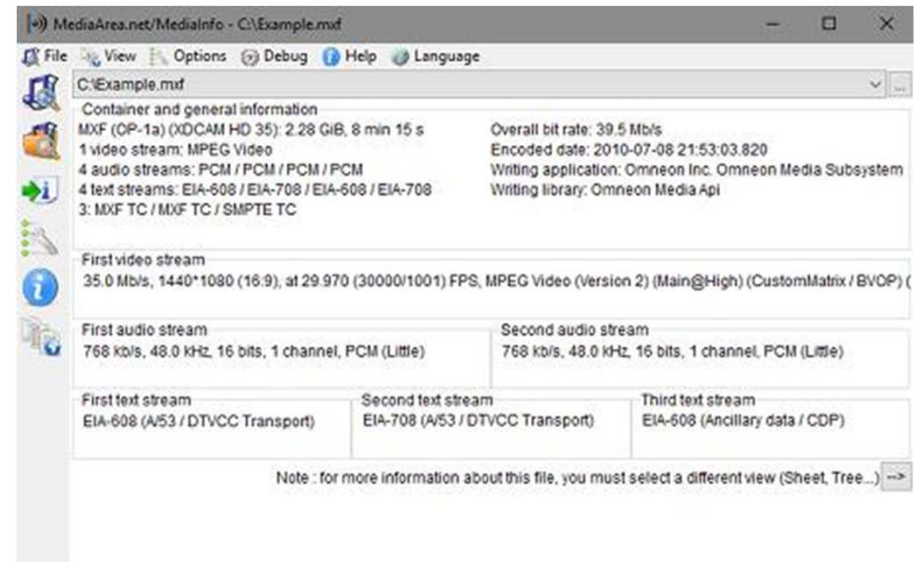
BWF MetaEdit

Supports embedding, validating, and exporting of metadata in Broadcast WAVE Format (BWF) files.

It supports the FADGI Broadcast WAVE Metadata Embedding Guidelines.

DV Analyzer

Technical quality control and reporting tool that examines DV streams in order to report errors in the tape-to-file transfer process, such as video error concealment information, invalid audio samples, timecode inconsistency, inconsistent use of arbitrary bits in video DIF blocks, and DIF structural problems. Also reports on patterns within DV streams such as changes in DV time code, changes in recording date and time markers, first and last frame markers within individual recordings, and more.

	Absolute time	Arb bits	Video error concealment
0	00:00:00.000	F	
398	00:00:15.920	X	
399	00:00:15.960	1	0.20% (1 4 errors, 1 5 errors, 1 C errors)
400	00:00:16.000	2	0.20% (2 A errors, 1 E errors)
401	00:00:16.040	X	
402	00:00:16.080	9	0.07% (1 8 errors)
403	00:00:16.120	3	0.27% (1 2 errors, 1 9 errors, 1 D errors, 1 E errors)
404	00:00:16.160	X	
405	00:00:16.200	B	0.07% (1 B errors)
406	00:00:16.240	X	
407	00:00:16.280	6	0.13% (1 6 errors, 1 8 errors)
408	00:00:16.320	6	0.20% (1 3 errors, 1 B errors, 1 C errors)
409	00:00:16.360	E	0.13% (1 1 errors, 1 9 errors)

RAWcooked

Encodes RAW audio-visual data (DPX/TIFF) into a losslessly compressed file (no more unplayable TAR!)

Metadata accompanying the RAW data are preserved (reversibility)

Sidecar files, like MD5, LUT or XML, are in container attachments

Final package is 1.5-3x (usually 2x) smaller than DPX/TIFF

Checksum by "Cluster" (usually 1 second) at container level

Checksum by "Slice" (you choose how many per frame) at video level

Files are natively playable by lot of tools (FFmpeg, VLC...)

Save HDD/LTO space: either ~2x less cost for same redundancy or 2x more redundancy for ~ same cost

https://MediaArea.net/RAWcooked

Practical examples

Stay in touch

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

Jérôme Martinez: jerome@MediaArea.net

Slides: https://MediaArea.net/Events

License: CC BY