

Reversibly Normalise Film Scans & Optimize Storage

# The RAWcooked project

*Data is never RAW.*

*Data is always cooked in a way or another.*

Jérôme Martinez

No Time to Wait 4, December 2019

# MediaArea

Open source software company focused on digital media analysis. We work (different levels of involvement) on:

- **MediaInfo**

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

- **MediaConch**

Implementation checker, policy checker, & reporter

- **QCTools**

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

- **BWF MetaEdit, AVI MetaEdit, MOV MetaEdit**

Embedding, validating, and exporting of metadata

- **DV Analyzer**

Checking presence of technical errors in DV captures

# Raw A/V files

Huge size

(4K+ is there, can be 100 MB/frame, several TB per hour)

1 file per video frame (thousands of files in a directory)

Not playable as is by several players (VLC...)

So many DPX or TIFF format flavors  
(interoperability issues)

# FFV1

Lossless video compression format

Open source, patent free

Adopted by several archives

Being standardized (IETF)

Frames are divided by slices, with checksums

# Compression

Example with 1 second at 24 fps 10-bit HD film on a 6-core (12-thread) Skylake-X CPU:

- 24 DPX files (or in ZIP/TAR uncompressed): 189 MB
- 1 compressed ZIP file: 175 MB in 10 seconds
- 1 compressed LZMA2 file: 154 MB in 30 seconds
- 1 FFV1/MKV Intra 16-slice file: 105 MB in 1.5 seconds

# Disadvantages of FFV1 alone

- You lose some metadata  
(DPX/TIFF header: scan software, some colorimetry info, film type, DPX time code, shutter angle, gamma...)
- Not like ZIP or TAR (exact same files)
- Complicated command line
- But...

# RAWcooked

- Easy: just a short command line  
"rawcooked YourDirectoryName"
- Store DPX/TIFF headers/footers in a specific Matroska attachment
- Store other sidecar files as Matroska attachments
- Output is a single Matroska/FFV1/FLAC file
- Encoding is reversible (bit-by-bit to original files)  
"rawcooked YourMatroskaFileName.mkv"

# Easy check of integrity

- Check if the file is healthy  
`"rawcooked --check YourMatroskaFileName.mkv"`
- Check if DPX headers are conform to specs  
`"rawcooked --conch YourMatroskaFileName.mkv"`
- Add error correction codes while encoding the file  
e.g. with overhead of 1.5%, you can lose 4 blocks every 252 blocks without losing any content  
`"rawcooked --ecc YourDirectoryName"`
- Fix the corrupted file  
`"rawcooked --fix YourMatroskaFileName.mkv"`



# Use case

- Archive asks a digitilization to their supplier  
Classic workflow with the scanner  
+ "rawcooked --all YourDirectoryName"
- Transport... (2x less file sizes, less costly)
- Archive receives content & checks the integrity  
(file health, DPX conformance...)  
"rawcooked --check YourMatroskaFileName.mkv"
- Archive can visually check the content with  
e.g. VLC Media Player
- Storage (cost divided by 2 due to compression)
- Revert to exact original DPX if someone needs it  
"rawcooked YourMatroskaFileName.mkv"

# Supported input formats

- DPX/Raw: 8/10/12/16 bit, RGB/RGBA
- TIFF/Raw: 16 bit, RGB
- WAV/PCM: 16/24 bit, 1/2/6 channel, 44/48/96 kHz
- AIFF/PCM: 16/24 bit, 1/2/6 channel, 44/48/96 kHz
- Based on files from our sponsors
- More formats or format flavors on request

# Our sponsors



# Our sponsors

- AV Preservation by reto.ch (main sponsor)
- National Audiovisual Centre Luxembourg (CNA)
- National Library of Norway
- Irish Film Institute (IFI)
- Northwest University Library
- National Library of Wales
- Walter J. Brown Media Archives
- The MediaPreserve
- British Film Institute
- New York Public Library

# Financial sustainability

- Open source code provided without lock to sponsors
- Deliveries on our website are with a lock
- DPX 8/10 bit RGB & WAV 2ch 48kHz flavors are usable by default
- We provide a key for other format flavors and features (temporary key possible)
- 1000 € for first flavor/feature  
+ 500 € per additional flavor/feature
- 500 €/year for maintenance (priority support)
- To be compared with storage cost saving (storage cost divided by 2)

# Current developments (with sponsors)

- DPX conformance checker
- Integrated auto-check
- Erasure code
- Speed improvement through CPU (SSE/AVX)  
(looking for additional sponsors)
- Graphical interface

# Potential improvements (no sponsors yet)

- Support of reels?
- Speed improvement through GPU?
- CFA/Bayer/RGGB support?
- Creation of an access file at the same time?
- Better support of audio?
- More input formats?

# Stay in touch

MediaArea: <https://mediaarea.net>, @MediaArea\_net

RAWcooked: <https://MediaArea.net/RAWcooked>

Jérôme Martinez: [jerome@mediaarea.net](mailto:jerome@mediaarea.net)

Slides: <https://MediaArea.net/Events>

License (except images): CC BY