Open source tools

Some examples of open source projects
used (present, future) by archivists

Jérôme Martinez
MediaArea

FIAF, Prague, April 2018

The four freedoms

- 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

FFmpeg

Tool for manipulating A/V files

Very versatile

Lot of formats supported

Lot of possibilities to manipulate A/V files
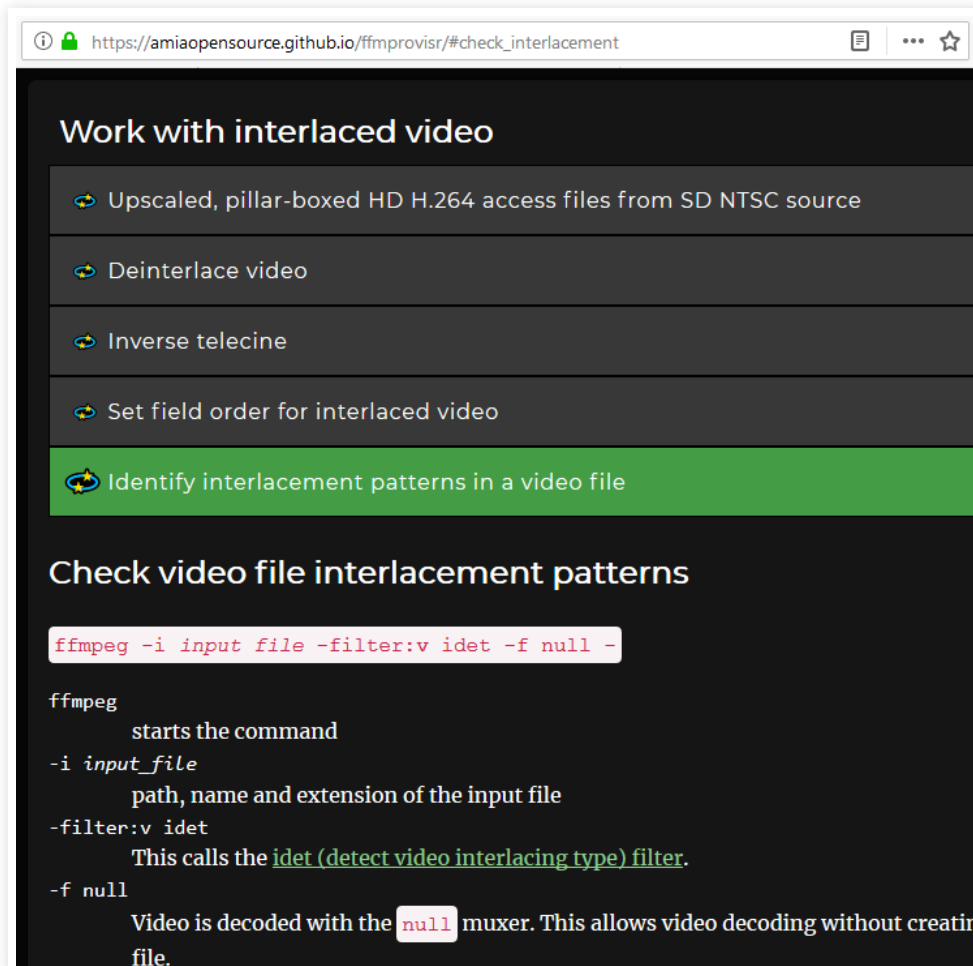
Command line tool

ffmprovisr

FFmpeg has a steep learning curve

ffmprovisr helps users through the command generation process so that more people can reap the benefits of FFmpeg

Maintained by archivists like you

https://amiaopensource.github.io/ffmprovisr/

ffmprovisr

QCTools

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

Based on FFmpeg

Graphical interface or command line

https://www.bavc.org/preserve-media/preservation-tools/qctools

(Development snapshots on https://MediaArea.net/QCTools)

QCTools

QCTools

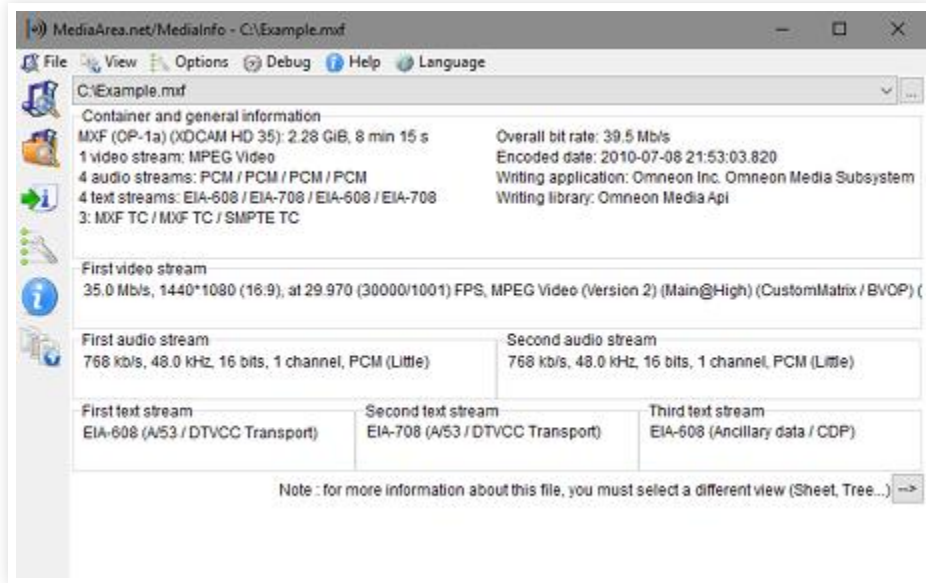Colors of digitilzed video are weird? Let's check...

MediaInfo

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

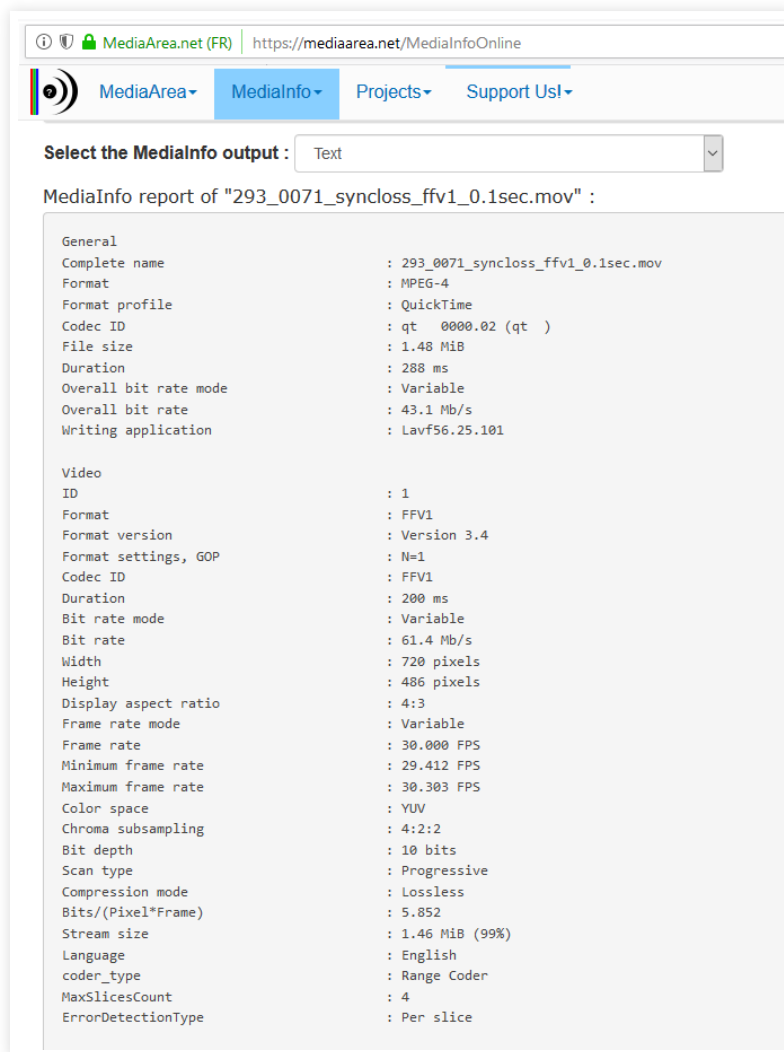Graphical interface or command line or software library

https://MediaArea.net/MediaInfo

MediaInfo

MediaInfoOnline

)))   MediaArea▾   MediaInfo▾   Projects▾   Support Us!▾

Select the MediaInfo output :   Text

MediaInfo report of "293_0071_syncloss_ffv1_0.1sec.mov" :

```
General
Complete name          : 293_0071_syncloss_ffv1_0.1sec.mov
Format                 : MPEG-4
Format profile         : QuickTime
Codec ID               : qt   0000.02 (qt  )
File size              : 1.48 MiB
Duration               : 288 ms
Overall bit rate mode  : Variable
Overall bit rate       : 43.1 Mb/s
Writing application    : Lavf56.25.101

Video
ID                     : 1
Format                 : FFV1
Format version         : Version 3.4
Format settings, GOP   : N=1
Codec ID               : FFV1
Duration               : 200 ms
Bit rate mode          : Variable
Bit rate               : 61.4 Mb/s
Width                  : 720 pixels
Height                 : 486 pixels
Display aspect ratio   : 4:3
Frame rate mode        : Variable
Frame rate             : 30.000 FPS
Minimum frame rate     : 29.412 FPS
Maximum frame rate     : 30.303 FPS
Color space            : YUV
Chroma subsampling     : 4:2:2
Bit depth              : 10 bits
Scan type              : Progressive
Compression mode       : Lossless
Bits/(Pixel*Frame)     : 5.852
Stream size            : 1.46 MiB (99%)
Language               : English
coder_type             : Range Coder
MaxSlicesCount         : 4
ErrorDetectionType     : Per slice
```

MediaConch

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

Based on MediaInfo

Graphical interface or command line

https://MediaArea.net/MediaConch

MediaConch

Implementation and Policy reporter

Check by file upload	Check online files	Check server files

Policy [Choose a policy ▾] **Display** [MediaConch Html ▾] **Verbosity** [Default level ▾]

[Check files]

Results

× Close all results

Apply a policy to all results [Choose a new policy to apply ▾]

Show [10 ▾] entries

Search: [   ]

Files ⇅	Implem ⇅	Policy ⇅	MediaInfo	MediaTrace	Status ⇅
ffv1_test_pixfmt-yuv444p10le…	✔ Valid	✖ Matroska is well described?	👁 ⊕	👁 ⊕	✔
ffv1_test_pixfmt-yuva422p_co…	✔ Valid	✖ Matroska is well described?	👁 ⊕	👁 ⊕	✔
ffv1_test_pixfmt-yuva444p_co…	👁 ⊕	👁 ⊕	👁 ⊕	👁 ⊕	↻ 🗑
veraPDF test suite 6-1-10-t0…	✖ Not valid	✖ Matroska is well described?	👁 ⊕	👁 ⊕	✔
train1.tif	✖ Not valid	✖ Matroska is well described?	👁 ⊕	👁 ⊕	✔
buggy_header.pdf	✖ Not valid	✖ Matroska is well described?	👁 ⊕	👁 ⊕	✔

Showing 11 to 16 of 16 entries

Previous	1	2	Next

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-DOCVER-COH** Tests 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-IDLENGTH-LIMIT** Tests run: 1 | Results: ✓
- **ELEMENTS-WITHIN-MAXIDLENGTH** 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-SLICE-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

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

MediaConch

Inspect your files

Offset	Key	Value
0x00000000	📁 EBML (30 bytes)	
0x0000001e	📁 Segment (35726 bytes)	
0x0000001e	📁 Header (12 bytes)	
0x0000002a	📁 SeekHead (115 bytes)	
0x0000009d	📁 Void (88 bytes)	
0x000000f5	📁 Info (139 bytes)	
0x00000180	📁 Tracks (112 bytes)	
0x00000180	📁 Header (5 bytes)	
0x00000185	📁 CRC-32 (6 bytes)	
0x0000018b	📁 TrackEntry (101 bytes)	
0x0000018b	📁 Header (2 bytes)	
0x0000018d	📁 CRC-32 (6 bytes)	
0x00000193	📁 TrackNumber - 1 (3 bytes)	
0x00000196	📁 TrackType - 1 (3 bytes)	
0x00000199	📁 CodecID - V_FFV1 (8 bytes)	
0x000001a1	📁 TrackUID - 1 (4 bytes)	
0x000001a5	📁 FlagLacing - 0 (3 bytes)	
0x000001a8	📁 Language - und (7 bytes)	
0x000001af	📁 DefaultDuration - 40000000 (8 bytes)	
0x000001b7	📁 Video (12 bytes)	
0x000001c3	📁 CodecPrivate (45 bytes)	
0x000001c3	📁 Header (3 bytes)	
0x000001c6	📄 version	3 (0x3)
0x000001c6	📄 micro_version	4 (0x4)
0x000001c6	📄 coder_type	0 (0x0)

MediaConch

Policy editor

Policy list:

🔍 Search

- 📁 User policies
  - 📁 Video file is MKV + FFV1-Intra + PCM or FLAC with CRC32 everywhere (or)
    - 📁 MKV, FFV1 Intra, PCM/FLAC, error detection (and)
      - 📄 Container is MKV
      - 📄 Video is FFV1
      - 📄 GOP size of 1
      - 📄 Container uses error detection
      - 📄 Video uses error detection
    - 📁 Audio is PCM or FLAC (or)
    - 📄 Has no video track
  - 📁 matrix_coefficients not same (and)
- 📁 System policies
  - 📁 Is this NTSC or PAL SD? (and)
  - 📁 Example MKV FFV1 digitization policy (and)
  - 📁 Matroska is well described? (and)
  - 📁 CAVPP Preservation Master (and)
  - 📁 Memoriav Video files Recommendations (or)

Rule type [MediaInfo] [MediaTrace]

Rule name ⓘ

Container is MKV

Track type ⓘ *

General

Field ⓘ *

Format

Occurrence ⓘ

Validator ⓘ

Is equal (=)

Content *

Matroska

[Save] [Duplicate] [Delete]

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

BWF MetaEdit

Embedding, validating, and exporting of metadata in Broadcast WAVE Format (BWF) files

Supports the FADGI Broadcast WAVE Metadata Embedding Guidelines

Graphical interface or command line

https://MediaArea.net/BWFMetaEdit

BWF MetaEdit

| Technical Metadata | Core Metadata | Rules | File management |

Select which standards and rule sets to follow during use of BWF MetaEdit.
Selection of rule sets will constrained the allowed data entry and may add a
See documentation on BWF MetaEdit Rules within the Help documentation.

- ☑ BWF (EBU Tech 3285) requirements
- ☐ BWF (EBU Tech 3285) recommendations
- ☐ BWF CodingHistory (EBU Tech R98-1999) recommendations
- ☐ BWF OriginatorReference (EBU R99-1999) recommendations
- ☑ INFO (Microsoft definition) requirements
- ☐ INFO (Microsoft definition) recommendations
- ☐ Federal Agencies Digitization Guidelines Initative recommendations

BWF MetaEdit

unks	bext	INFO	XMP	aXML	iXML	MD5Stored	MD5Generated	Errors	Information
1	Version 0	No	No	No	No		4D532065CED4C7523C28EAAA6C61E6B6		MD5, no existing MD5 chunk
2	Version 0	No	No	No	No	8B20F853F7D355973E00E0056EB06CDF	4D532065CED4C7523C28EAAA6C61E6B6	MD5, failed verification	
3	Version 0	No	No	No	No		BE1B34CD10ACB56B3990CE78AF00595E		MD5, no existing MD5 chunk
4	Version 0	No	No	No	No		7767E5E97F2CD9096DCE8F501CAC90BC		MD5, no existing MD5 chunk
5	Version 0	No	No	No	No		3000B27C7A4DA205F96A8D7622FB8AD6		MD5, no existing MD5 chunk

AVI MetaEdit

Embedding, validating, and exporting of metadata in AVI (Standard and OpenDML) files

Supports the U.S. National Archives Guidelines recommandations

Graphical interface or command line

Is similar to BWF MetaEdit? Right, code from BWF MetaEdit was reused, limiting development cost

https://MediaArea.net/AVIMetaEdit

AVI MetaEdit

Technical Metadata	Core Metadata	Rules

Select which standards and rule sets to follow during u
Selection of rule sets will constrained the allowed data
See documentation on AVI MetaEdit Rules within the H

- ☑ AVI (New Multimedia Data Types and Data Technic
- ☐ AVI (New Multimedia Data Types and Data Technic
- ☑ INFO (Microsoft definition) requirements
- ☐ INFO (Microsoft definition) recommendations
- ☐ U.S. National Archives recommandations

MOV MetaEdit

Embedding and editing of metadata in MOV (Apple QuickTime) or MP4 (ISO/IEC 14496-14 a.k.a. MPEG-4 Part 14) files.

It is currently focused on Universal Ad ID metadata and Pixel Aspect Ratio edition

Sponsors were not interested in more. Could be expanded on request

Graphical interface or command line

https://MediaArea.net/MOVMetaEdit

MOV MetaEdit

vrecord
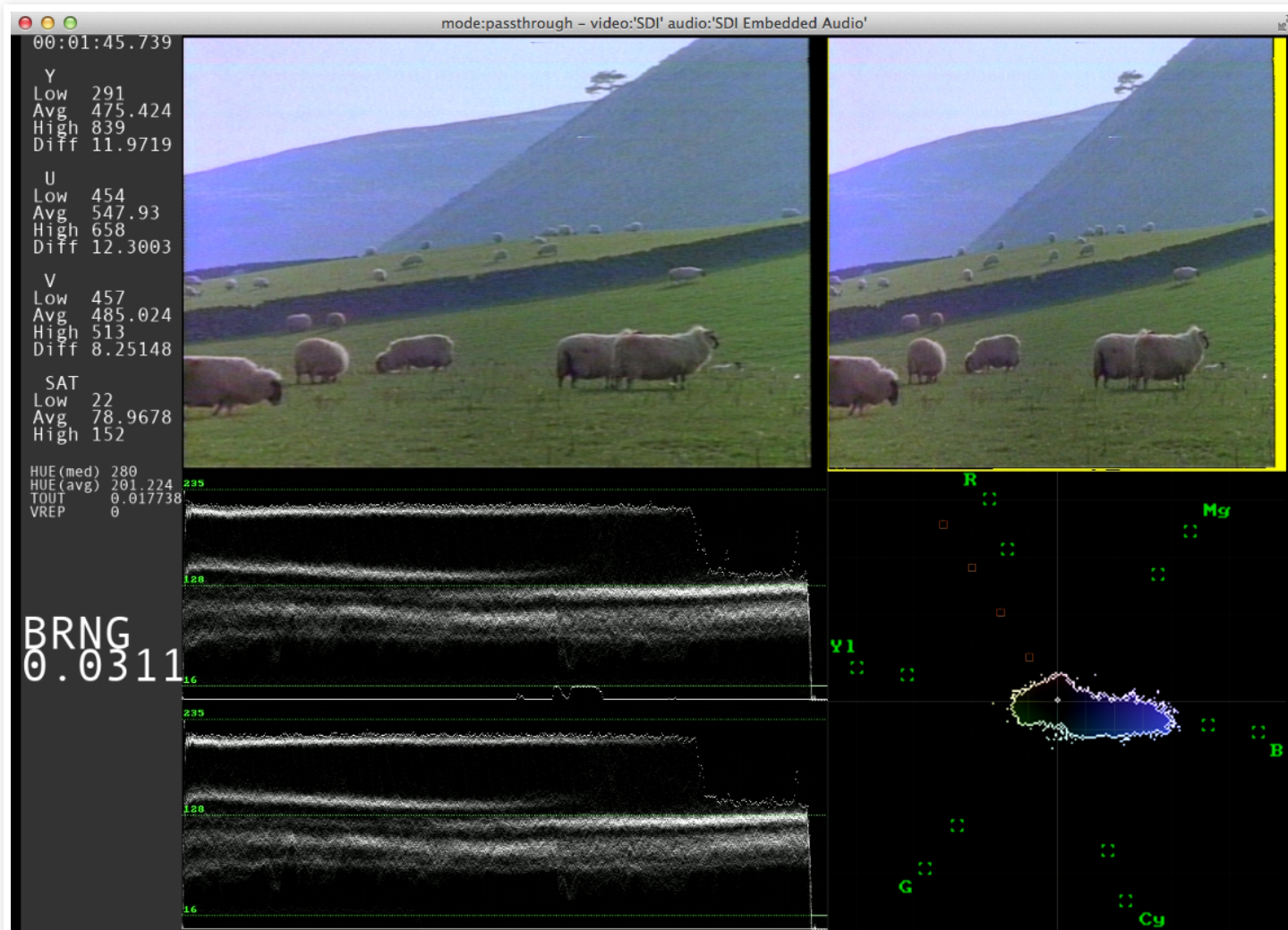
Tool to digitise analogue video

Make videotape digitization or transfer easier.

Blackmagic Design capture cards currently supported

https://github.com/amiaopensource/vrecord

vrecord

VirtualDub FFV1

VirtualDub is a Windows transcoding tool with a GUI

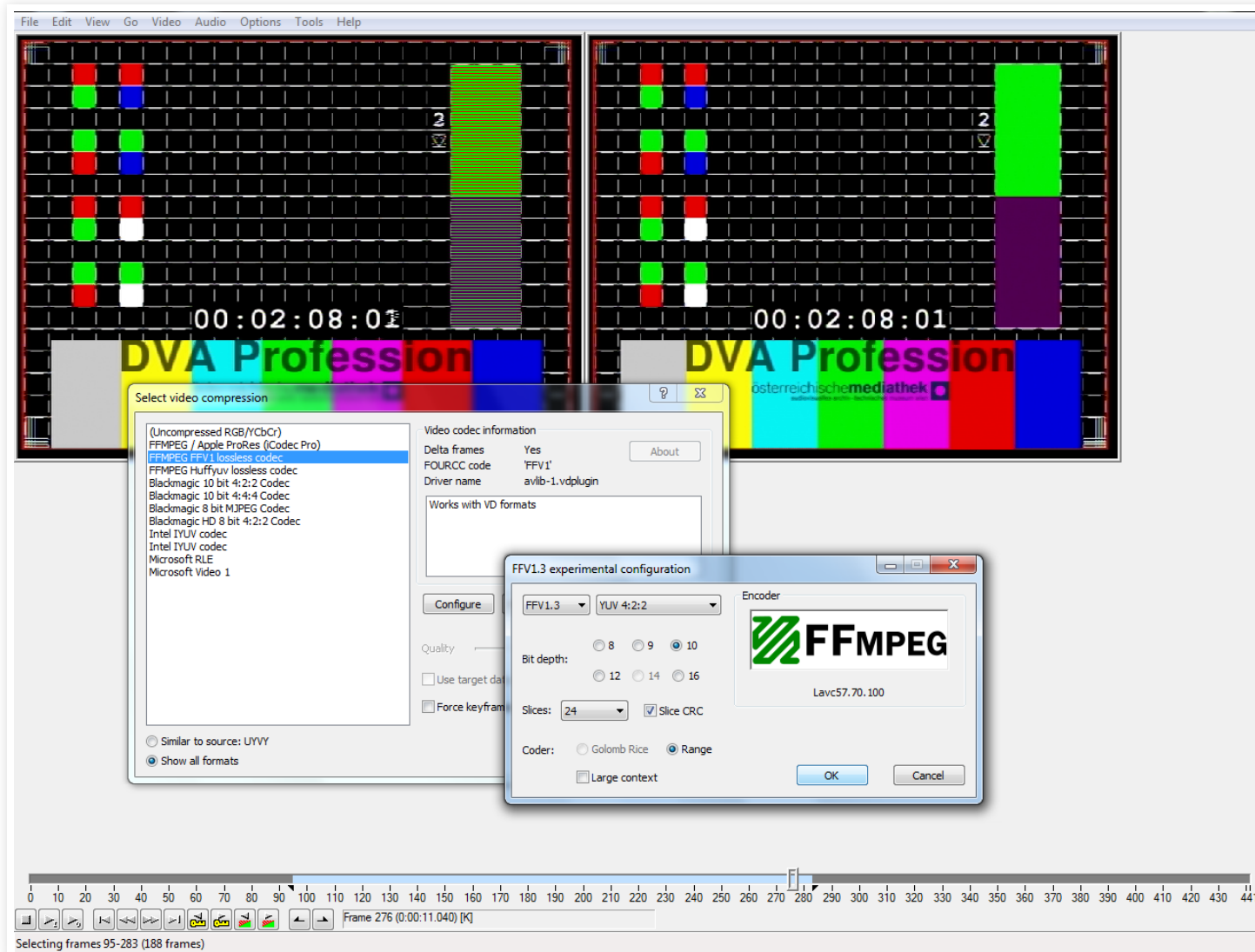No FFV1 supported by current version

Developers abandonned the tool (no more development)

But it is open source

"Nobody wants to update the tool? Let's add FFV1 support ourselves"

http://www.av-rd.com/projects/2017-virtualdub_ffv1.html

VirtualDub FFV1

RAWcooked

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

Metadata accompanying the RAW data are preserved (reversiblility)

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

https://MediaArea.net/RAWcooked

RAWcooked

- 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...)

RAWcooked

- Storage
  Save HDD/LTO space: either ~2x less cost for same redundacy or 2x more redundancy for ~ same cost
- Transport
  Encode, transport, decode; you save bandwidth (€... and transfer speed) without changing something else in your workflow (same files after revert to DPX/TIFF)

RAWcooked

In development

Stable release planned next month

~20 DPX flavors (RGB/RGBA * bitdepth * filled/packed * Big/Little endian) supported

Uses FFmpeg FFV1 encoder (internal encoder planned), uses internal FFV1 decoder

Transparent development, all at
https://github.com/MediaArea/RAWcooked

RAWcooked

Developed by MediaArea
https://MediaArea.net

Main sponsorship by AV Preservation by reto.ch
https://reto.ch

With additional financial support from some other archives:

- CNA (National Audiovisual Centre of Luxembourg)
- Nasjonalbiblioteket (National Library of Norway)
- IFI (Irish Film Institute)
- Northwestern University Libraries
- You?

Projects have similar patterns

This is open source

One can peek code from another project

Lot of libraries are shared

A communuity (e.g. archives from different countries) can build an ecosystem

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

Users have full control

Actors

Funding: European Commission, MoMA, National Endowment for the Humanities, the Knight Foundation, Library of Congress and FADGI, NARA (National Archives and Records Administration), CNA (National Audiovisual Centre of Luxembourg), Nasjonalbiblioteket (National Library of Norway), IFI (Irish Film Institute), Northwestern University Libraries and lot of small sponsoring from tens of other small to large entities

Actors

Management: MediaArea, AVP, Bay Area Video Coalition...

Development: MediaArea, lot of individual developers, and lot of archivists!

Funding, management, development: not always same people. You decide.

Stay in touch

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

Jérôme Martinez: jerome@MediaArea.net

Slides: https://MediaArea.net/Events

License: CC BY