Data Migration, Our Frenemy

Reto Kromer • AV Preservation by reto.ch

No Time to Wait!
Open Media, Open Formats, Open Archives
Österreichisches Filmmuseum, Vienna, Austria
9–10 November 2017

The Reel Thing XLII
New Orleans, Louisiana, USA
29 November 2017

Table of Contents

Lessons Learned

Data Migrations

2014

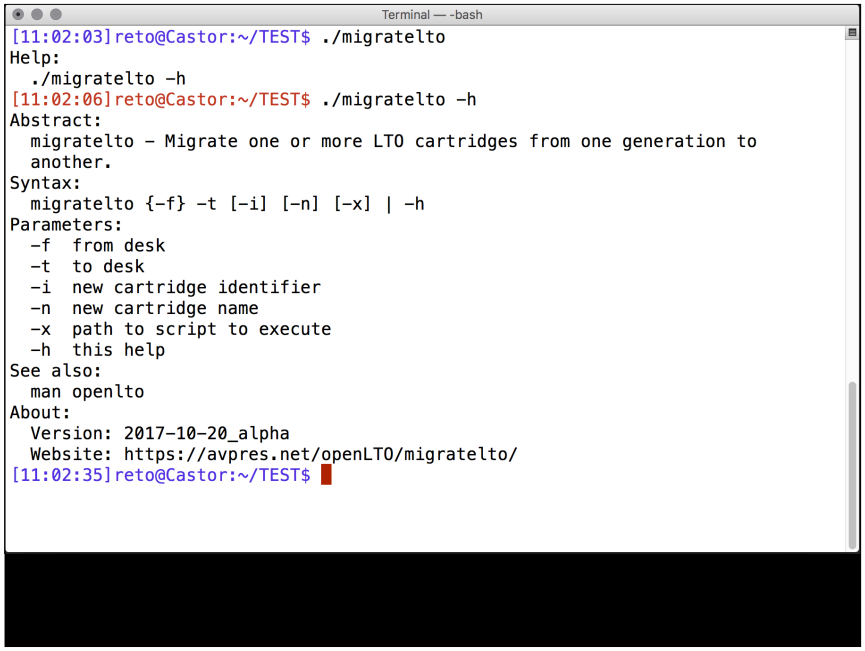
- our internal archive from LTO-4 to LTO-6 (5.7 PB)

2014–2017

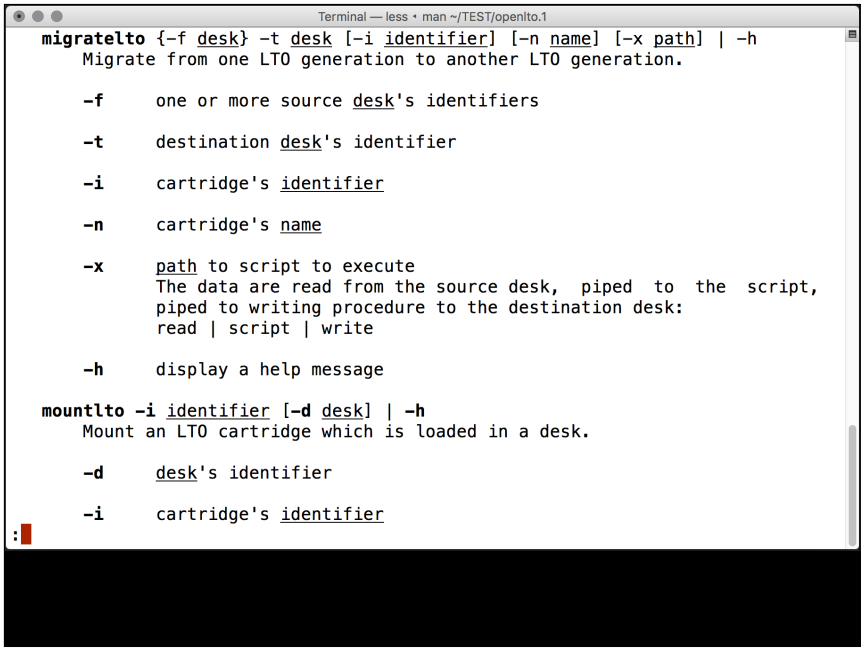
- […]

2017

- a client's archive from LTO-5 to LTO-7 (270 GB)



5



6

read | script | write

script to modify

- container
- codec
- both container and codec

7

Examples

8

Video

from:

- AVI / 8-bit and 10-bit uncompressed
- MOV / 8-bit and 10-bit uncompressed
- MP4 / 8-bit and 10-bit uncompressed

to:

- Matroska / FFV1

Container and Codec

→ read file from source LTO
→ demultiplex file
→ decode file

- $Y'C_BC_R$, 4:2:2, 8 bit, «raw» [uyvy422]

→ encode file
→ multiplex file
→ write file to destination LTO

Container and Codec

→ read file from source LTO
→ demultiplex file
→ decode file

- $Y'C_BC_R$, 4:2:2, 10 bit, «raw» [yuv422p10le]

→ encode file
→ multiplex file
→ write file to destination LTO

Film

from:

- TIFF
- DPX

to:

- Matroska / FFV1 using RGB48 and additional metadata

Reading

Reto Kromer: **Matroska and FFV1: One File Format for Film and Video Archiving?**, in «Journal of Film Preservation», n. 96 (April 2017), FIAF, Brussels, Belgium, p. 41–45

➔ https://retokromer.ch/publications/JFP_96.html

Old Experimentations

from:

- AVI / HuffYUV
- AVI / FFV1 version 1

to:

- Matroska / FFV1 [version 3]

Mid_{old} Experimentations

from:

- AVI / CineForm (VC-5) with Bayer
- MOV / CineForm (VC-5) with Bayer
- MOV / ProRes

to:

- Matroska / FFV1 after de-mosaicking
- Matroska / ProRes

Container, Codec and Data

➔ read file from source LTO
➔ demultiplex file
➔ decode data
  - 12 bit [bayer_bggr16le]
➔ encode data
➔ multiplex file
➔ write file to destination LTO

Container, Codec and Data

- → read file from source LTO
- → demultiplex file
- → decode data
  - R′G′B′, 4:4:4, 16 bit, «raw» [rgb48le]
- → encode data
- → multiplex file
- → write file to destination LTO

Container, Codec and Data

- → read file from source LTO
- → demultiplex file
- → decode data
- → **filter data (de-mosaic)**
- → encode data
- → multiplex file
- → write file to destination LTO

Container

- → read file from source LTO
- → demultiplex file
  - ProRes 422, 10 bit [yuv422p10le]
  - ProRes 4444, 10 bit [yuv444p10le or yuva444p10le] or 12 bit [yuv444p12le]
- → multiplex file
- → write file to destination LTO

Mid_{new} Experimentations

from:

- AVI / CineForm (VC-5) with Bayer
- MOV / CineForm (VC-5) with Bayer
- MOV / ProRes

to:

- Matroska / FFV1 version 4 [with Bayer]
- Matroska / ProRes with 12-bit support

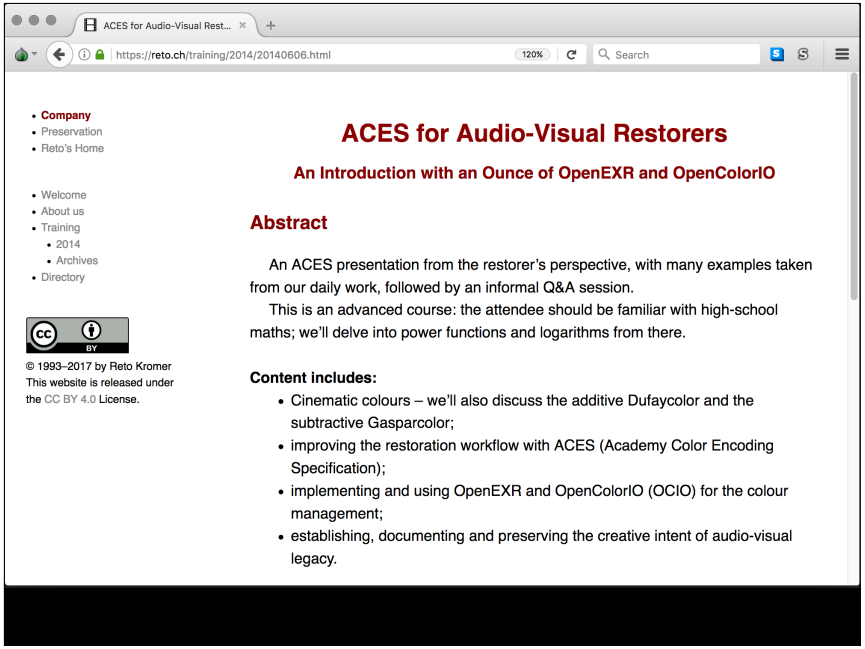
New Experimentations

from:

- MXF / OpenEXR
- flavour of NUT / multiple RGB48
- flavour of NUT / RGB72 or $Y'C_BC_R$ 24-bit
- flavour of NUT / expanded OpenEXR

to:

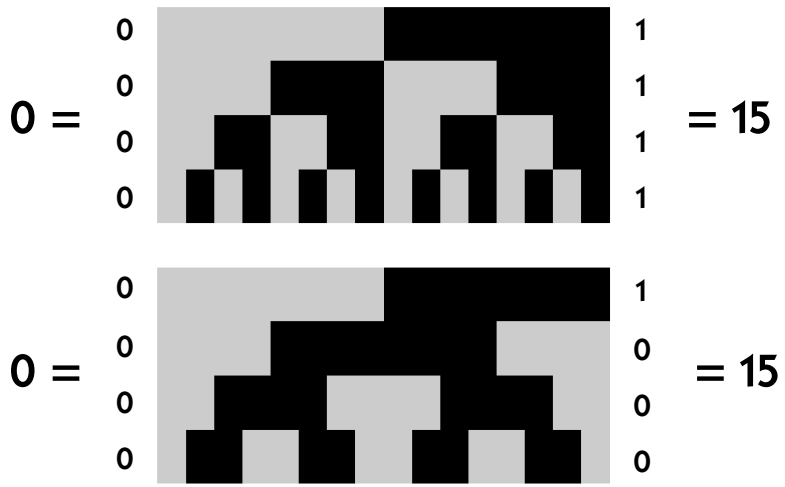
- Matroska / FFV1 version 4 using RGB48, RGB72, floats and additional metadata



Wishes

EBML

- additional support:
  - ➔ Gray codes

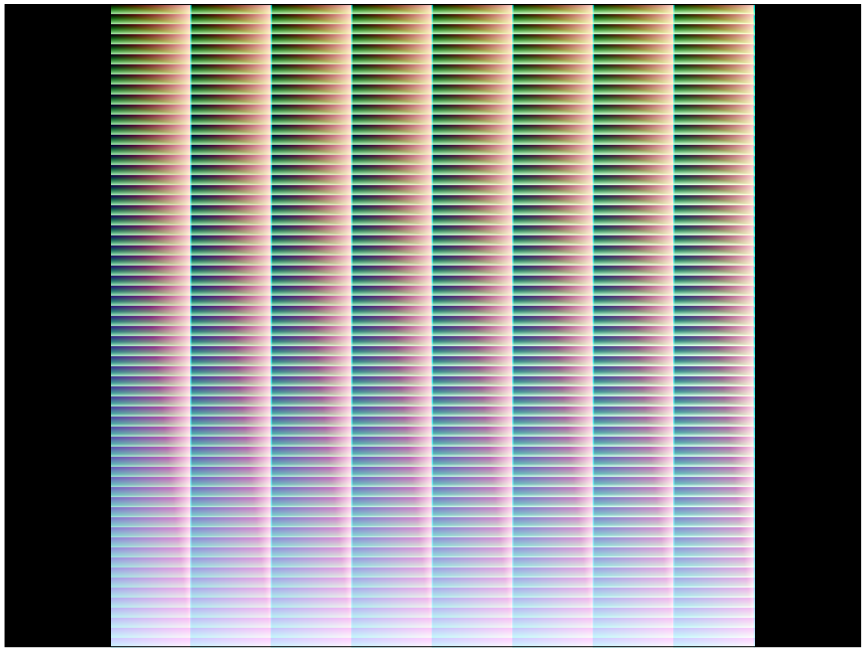


25

Matroska

- additional metadata
- additional integration:
  - → 1D LUT
  - → 3D LUT
- additional support:
  - → Gray codes

26

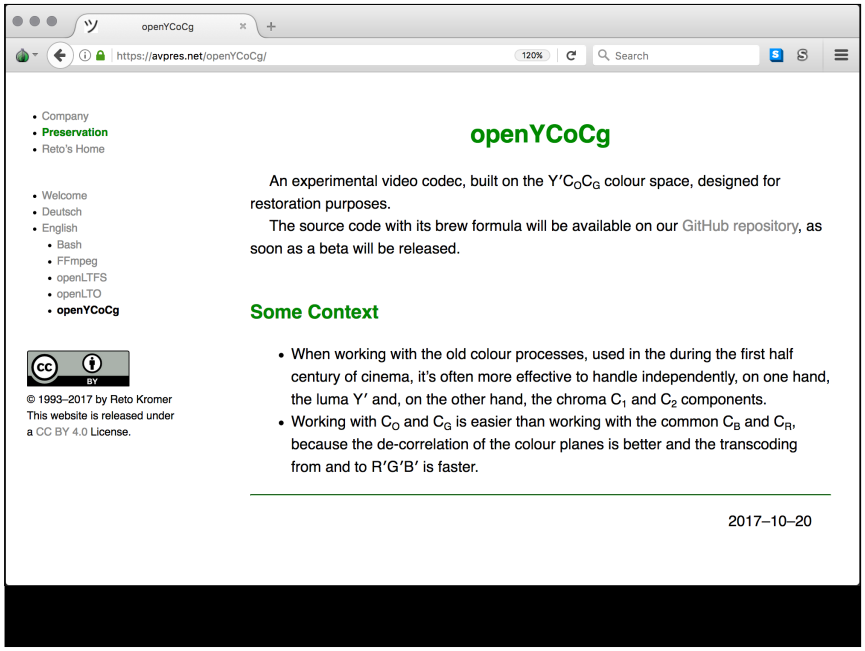


27

FFV1

- additional metadata
- additional encoding:
  - → RGB72, $Y'C_BC_R$ 24-bit and floats
  - → Bayer
- additional support:
  - → Gray codes
  - → $Y'C_OC_G$

28



29

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1.396523 \\ 1 & -0.342793 & -0.711348 \\ 1 & 1.765078 & 0 \end{pmatrix} \begin{pmatrix} Y' \\ C_B \\ C_R \end{pmatrix}$$

$$\begin{pmatrix} Y' \\ C_B \\ C_R \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.168074 & -0.329965 & 0.498039 \\ 0.498039 & -0.417947 & -0.080992 \end{pmatrix} \begin{pmatrix} R' \\ G' \\ B' \end{pmatrix}$$

30

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} Y' \\ C_O \\ C_G \end{pmatrix}$$

$$\begin{pmatrix} Y' \\ C_O \\ C_G \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{pmatrix} \begin{pmatrix} R' \\ G' \\ B' \end{pmatrix}$$

31

FLAC

- additional support:
  → Gray codes
- additional encoding:
  → 24-bit sample format

32

Kate Murray:

«More adoption means better sustainability.»

[while proudly speaking about MXF/JPEG 2000 at the Matroska/FFV1 symposium in Vienna]

Matroska Reader

- full transparency with DPX [et similia]
  - → RAWcooked
- create, alter and inspect Matroska files
  - → MKVToolNix
- verify Matroska and WebM files
  - → mkvalidator

Matroska Player

- apply 1D and 3D LUT to the video stream
  - → VLC
  - → mpv
  - → ffplay
  - → FFmpv

AV Preservation by reto.ch

chemin du Suchet 5
1024 Écublens
Switzerland

Web: reto.ch
Twitter: @retoch
Email: info@reto.ch