

# Datenspeicherung und -migration

Reto Kromer • AV Preservation by reto.ch

**Open Source im Archivalltag**  
Staatsarchiv Aargau, 20. August 2024

1

# Data Migrations

## **2014**

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

## **2014–2021**

- many migrations for clients

## **2021**

- our internal archive from LTO-6 to LTO-8 (25.2 PB)

2

# Magnetic Tape

- in use since the 1950s by IT
- cartridges are always on polyester base (old open reels can be on triacetate base)

3

# Packaging

- open reel
- cassette
- cartridge

4

## Recording

- linear or diagonally
- analogue or digital

5

## LTO

- Linear Tape-Open
- answer from the IT industry to the bank and insurance sector
- in 2000 LTO-1
- currently LTO-9
- currently the LTO Consortium consists in: Hewlett Packard Enterprise, IBM and Quantum

6

## LTO-8

- only one-generation backward reading capabilities
- format M8 = LTO-7 cartridges formatted as LTO-8
- M8 can be used on LTO-8 drives only

7

## LTO-9

- LTO-9 drives manufactured by IBM only
- LTO-9 cartridges manufactured by Fujifilm and Sony Group only
- only one-generation backward reading capabilities
- only 50% capacity increase
- backward reading capabilities for regular LTO-8 (L8), but not M8

8

## LTO-10

- Will there be two-generation backward reading capabilities?
- Is LTFS strong enough?
- Release possibly end of 2024, probably beginning of 2025.

9



10

## Formatting

### TAR

- from LTO-1 to LTO-4 only possibility
- still possible possible today

### LTFS

- possible (and recommended) since LTO-5

11

## TAR

- standard TAR
  - bloc size
  - number of archives per cartridge
  - archives needing more than one cartridge
- TAR with a proprietary data encoding (e.g. BRU, Retrospect)

12

# LTFS

- different versions
- almost one implementation per vendor, but...  
... "ltfs" and "mklts" common commands
- lossless compression (default) or uncompressed data
- unencrypted (default) or encrypted data

13

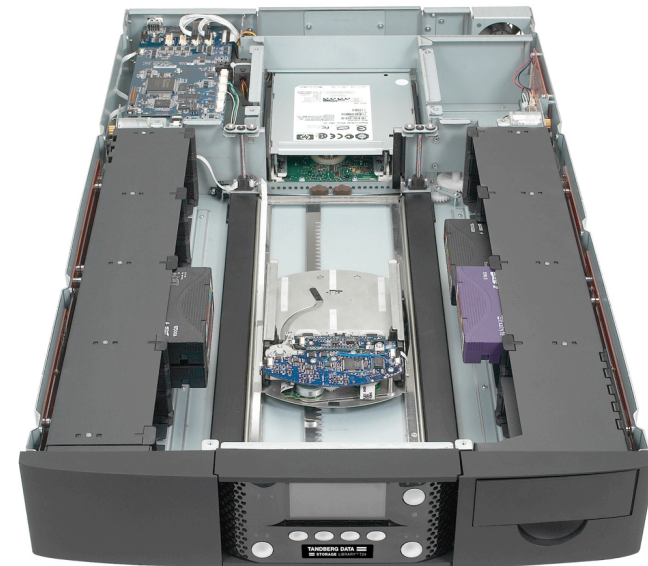
# Drive

- internal or external unit
- library

14



15



16

## Storage of the Tapes

- in a tape library
- on a shelf
- in a fire-proved cabinet

17

## Software

- proprietary or open source
- graphical user interface (GUI) and/or command-line interface (CLI)

18

## Plan the Next Migration

- file naming
- barcodes
- checksums
- write the full index to the cartridge
- technical metadata
- code to retrieve the files

19

## #1: Film

### FILM

- FILM\_DPX/Film\_nnnnnn.dpx
- Film\_PCM.wav
- Film\_ProRes.mov
- Film\_H264.mp4

20

## #2: Video

### VIDEO

- Video\_YCbCr422.mkv
- Video\_ProRes.mov
- Video\_H264.mp4

21

## File Naming (Example)

- title\_codec.container
- title\_codec\_container\_algorithm.txt
- film\_H264.mp4
- film\_H264\_mp4\_md5.txt

22

## Checksums

### **cryptographic**

- MD5
- SHA-1
- SHA-256
- SHA-512

### **non-cryptographic**

- CRC-32
- xxHash 32
- xxHash 64
- xxHash 128

23

## Longterm

- storage of the cartridges
- three copies...  
... in geographically distant locations
- data integrity check
- data migration
- availability of LTO desks

24

## Reading

Reto Kromer: **On the Bright Side of Data Migrations**, in «IASA Journal», n. 49 (December 2018), IASA, p. 18–22

→ [retokromer.ch/publications/IASA\\_49](http://retokromer.ch/publications/IASA_49)

25

## read | script | write

script to modify

- container
- codec
- both container and codec
- metadata
- filename

26

## #1: ProRes-born Content

**from:**

- ProRes stored in a QuickTime (.mov) container

**to:**

- ProRes stored in a Matroska (.mkv) container

27

## Update the 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

28

## #2: 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

29

## Container and Codec

→ read file from source LTO

→ demultiplex file

→ decode file

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

→ encode file

→ multiplex file

→ write file to destination LTO

30

## Container and Codec

→ read file from source LTO

→ demultiplex file

→ decode file

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

→ encode file

→ multiplex file

→ write file to destination LTO

31

## #3: Filename

### from:

- Title\_YUV422.mkv

### to:

- Title\_YCbCr422\_9d5084b5b0a08d5022b39e0e75241d12.mkv

32



## Equipment

- servers
- network
- storage
- ventilation (cooling)

33



34

## Working Place

- software
- computer with at least one good monitor
- storage
- illumination
- chair

35

## Common Equipment

- LTO desk or library
- software
- network

36

## Consumables

- LTO cartridges

37

## Sustainability

### from 2004 to 2012

- all power consumption from the grid

### from 2013 to 2019

- no electricity from the grid is used for the IT

### from 2020 to 2024

- no electricity at all comes from the grid

38

## No Time to Wait

- Berlin 2016
- Wien 2017
- London 2018
- Budapest 2019
- [Hilversum] 2021
- Den Haag 2022
- Prag 2023
- Karlsruhe 2024

39

AV Preservation by  
**reto.ch**

Sandrainstrasse 3  
3007 Bern  
Schweiz

reto.ch  
info@reto.ch



40