FILE FORMATS FOR AUDIOVISUAL PRESERVATION: HOW TO CHOOSE?

by Peter Bubestinger

10. November 2016

MAIN INTENTION

- We want our data to be "Futureproof™".
- Digital video preservation is *different* than audio...

THE QUEST FOR THE HOLY GRAIL

STEPS ON THE QUEST

- Know your material
- Know your use cases
- Know your limits
- Make a choice

INTRODUCTION

USE-CASES

Different institutions, different use-cases

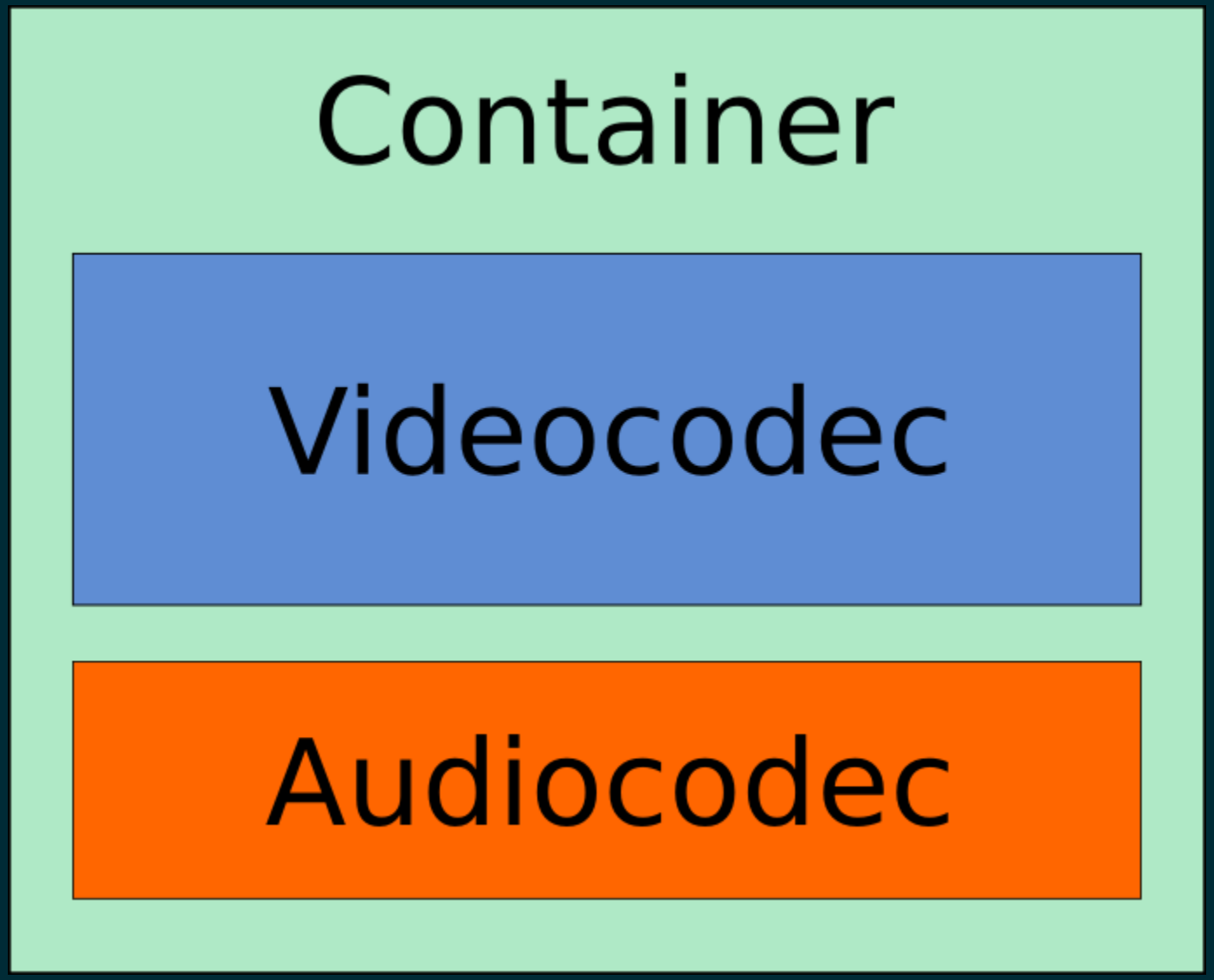
- Preservation
  (Archive master)
- Production
  (Mezzanine Format for edits)
- Access
  (Access copies / often web)

THE DIGITAL VIDEO TRINITY

FILE ENDING = CONTAINER

Saying "*The videos are in Flash/AVI/MOV format*" or "*The camera creates MP4*", usually **only** tell you which container is used.

But: What *is* a "container"?


Container
Videocodec
Audiocodec

"THE VIDEOFORMAT"

...actually consists of ≥ 3 formats.
Correct would be to name all 3 component formats:

Format = "V-Codec / A-Codec in Container"

Examples:

- H.264 / AAC in MP4
- XviD / MP3 in AVI
- FFV1 / PCM in MKV

DESIRED PROPERTIES FOR LTP

PROPERTY CATEGORIES

1. Significant properties
2. Preservation improving properties

1. SIGNIFICANT PROPERTIES

- Resolution independent
- Aspect ratio preserving
- Color as native as possible
- No digital loss

2. PRESERVATION-IMPROVING PROPERTIES

- Handleable data-amount
- Non-proprietary
- Hardware independent
- Avoiding unnecessary complexity

1. SIGNIFICANT PROPERTIES

RESOLUTION INDEPENDENT

- Preserve original resolution "as-is"
- Some implementations limit resolution
- Some formats define resolution (DVD,BluRay,etc)
- Sometimes "multiple of X" required (XviD, H.264)

ASPECT RATIO PRESERVING

- DAR: Display Aspect Ratio
- SAR: Storage Aspect Ratio
- PAR: Pixel Aspect Ratio
- Metadata required
- Unconventional ratios (consumer/web)

COLOR AS NATIVE AS POSSIBLE

- Colorspaces: RGB, YUV, XYZ, etc
- Conversion might be lossy
- Formats often support only a few
- Chroma subsampling (4:2:2, 4:2:0, ...)
- Interlaced / Progressive

NO DIGITAL LOSS

- Popular formats currently "lossy"
- Cameras produce lossy
- "*looking good now*" might not be so good later...
- To consider: generation loss by editing
- Future format migrations
- Best = lossless (e.g. FFV1) or uncompressed

2. PRESERVATION-IMPROVING PROPERTIES

HANDLEABLE DATA-AMOUNT

- No sense in sizes one cannot handle
- Storage size
- Network speed
- Disk speed
- Data bus speed (RAM, etc)
- ≥ 1 Backups

HANDLEABLE DATA-AMOUNT

- How much storage for 1 copy?
- + backup?
- How much MB/s for real-time playback?
- Required network speed for this?
- Enough network bandwidth for daily ingest + backups?
- How many concurrent users (web, in-house)?

NON-PROPRIETARY

- Open specification / standard
- Open format + closed implementation = black-box
- Proprietary: possible vendor lock-in
- Archive = market niche
- Open Source implementation available?
- Try (with other tools) before you buy!

NON-PROPRIETARY

Benefits of Open Source / Free Software implementation

- **YOU** are in control
- No vendor lock-in
- No black-box debugging
- Ability to use/study/share/improve tools
- Independence of market interests
- No format obsolescence (virtual immortality)
- Future proof by archiving source code

HARDWARE INDEPENDENT

Hardware dependency is bad.

- Similar to vendor lock-in
- Encoding/decoding card required?
- Rare/proprietary drive (remember "JAZ"?)
- Try decoding (with other software) before you buy!

AVOIDING UNNECESSARY COMPLEXITY

Let's reconsider the "All-in-one" approach...

RECONSIDERING "ALL-IN-ONE"

More features = more likely that:

- Implementation only covers parts
- Interoperability problems between different implementations

MINIMALISTIC DATA FORMAT

- As simple as possible
- As complicated as necessary

Source: Bernhard Reiter: "Minimalgebot für Datenformate - Offener Standard sein reicht nicht"

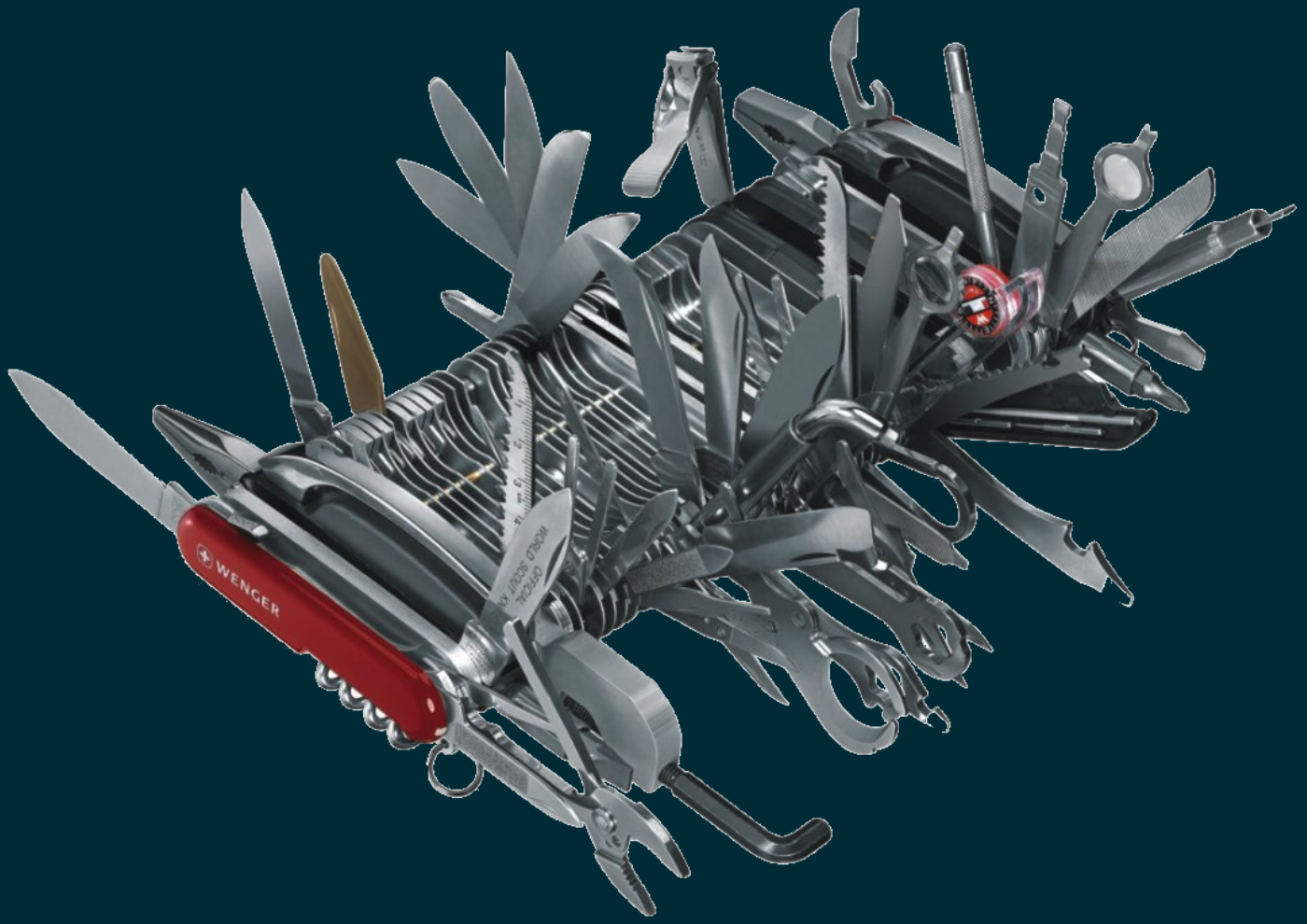
MINIMALISTIC DATA FORMAT

- How well does the data-format solve the problem?
- Is there a simpler format that could solve the problem just as well?

Source: Bernhard Reiter: "Minimalgebot für Datenformate - Offener Standard sein reicht nicht"

DECIDE WHAT YOU **REALLY** NEED

OTHERWISE...?



THAT'S IT.

IN A NUTSHELL :)

QUESTIONS?

Peter Bubestinger-Steindl
(pb@av-rd.com)

ABOUT MYSELF

PETER BUBESTINGER-STEINDL

(pb@av-rd.com)

- Studied media computer-science at the TU Vienna
- Developer, trainer and tech-consultant since 2000
- 8 years working with broadcast audio archives around the globe (NOA)
- 5 years R&D at the "Austrian Mediathek"
- Dec. 2015: Founding of company "AV-RD"

LICENSE AND CREDITS

Peter Bubestinger-Steindl
(pb@av-rd.com)