

PREFERRED, OBSOLETE OR IN-BETWEEN? DEVELOPING A CRITERIA CATALOGUE FOR AV-MATERIAL

Preservation planning at the German National Library of Science and Technology (tib)

Merle Friedrich

*German National Library of Science and Technology,
Germany*

merle.friedrich@tib.eu

0000-0001-7158-8583

Abstract - The born-digital audio-visual (AV) holdings of the German National Library of Science and Technology are analyzed regarding the present file formats. The most frequent AV file formats are examined in terms of suitability as preservation format based on a catalogue of criteria. Furthermore their risk of obsolescence is evaluated using view paths. The examined file formats are not preferred as preservation formats, but they are not obsolete either.

Keywords - Obsolescence; Audio-Visual Material; Preservation Planning; Technology Watch

Conference Topics - The Cutting Edge: Technical Infrastructure and Implementation

I. INTRODUCTION

The German National Library of Science and Technology (TIB) has a collection of audio-visual (AV) material. TIB acquires new content from different producers. This multitude of producers leads to a variety of different file formats which are preserved in TIB's digital archive. The poster will describe how the risk of obsolescence of born-digital AV files is determined for the three file formats in which the majority of born-digital AV material in TIB's holdings are available.

A. Background

AV material usually consists of a container (e.g. mp4), which wraps one or more content streams. The content streams are video/ audio stream(s). They can be encoded by different audio and video codecs [1, p. 137] (e.g. Pulse Code Modulation, FFv1). I will refer to file format as the combination of container, video codec and audio codec and differentiate on

the level of format version as indicated by MediaInfo [2]. The examined file formats are *MPEG-4/AVC/AAC, Version 4*, as well as *WebM/VP8/Vorbis* and *MPEG-PS/MPEG Video, Version 2/MPEG Audio, Version 1*.

B. Research question

Preservation Planning as described in the OAIS covers questions regarding the obsolescence of file formats [3, 4.15]. File formats can have three different states: they are either *preferred as preservation format*, or *not preferred, but not obsolete*, or they can be *obsolete*. Which state are the examined file formats in? The poster describes the verification of the following hypotheses:

1. The examined file formats are not preferred as preservation formats.
2. The examined file formats are not obsolete.

II. ATTRIBUTES OF PREFERRED FORMATS

A literature study reveals different approaches in order to assess preferences regarding a file format. Todd combines the findings of different studies and concludes the most common criteria for file format selection are adoption, platform independence, disclosure or documentation, transparency, and metadata support [4, p. 10].

A. Criteria for Suitability as Preservation Format

Each of the five criteria is broken down into preferred, acceptable and critical factors. For each of the factors measurements were developed. These measurements consider the requirements of TIB's

designated community. E.g. metadata support is rated acceptable if technical metadata is embedded in the format. This was tested through the extraction of technical metadata with MediaInfo [2].

B. Classification of TIB's holdings

Each container, video and audio codec is weighted according to the developed measurements. An overall weighting reveals if a container or codec was considered preferred, accepted or critical. The three examined file formats are rated critical for preservation purposes, which verifies the first hypothesis.

III. OBSOLESCENCE

A file format is obsolete if it is at risk to become inaccessible [5, p. 93] to our designated community. According to Ryan there is one factor to measure obsolescence: if no rendering software is available a file format is obsolete [6, p. 14].

A. Availability of Rendering Software

The National Library of the Netherlands uses view paths as a formal approach of evaluating the availability of rendering software. A View path consists of the information on the hardware platform, the operating system and the viewer application (incl. version) which enables the designated community to access the content of the file [7, p. 48].

B. Classification of TIB's holdings

For each of the examined formats two view paths were documented. These view paths work independent from each other, meaning that no component is used twice. With two view paths for each file format the second hypothesis - the examined file formats are not obsolete - is verified.

IV. RESULTS & CONCLUSION

The majority of TIB's born-digital AV holdings come in a file format which is not preferred for preservation, but not obsolete. Both hypotheses are verified.

Although this is true for the file format, it is not necessarily true for each file. Not all files are implemented according to the file format specification and therefore valid. There is a critical lack of (open source) validation software for AV files. AV playback software is tolerant to implementation errors so that testing the render ability cannot replace validation [8, p. 28].

V. FUTURE WORK

Regarding the view paths TIB must evaluate if testing and documenting view paths for all (AV) file formats add a crucial value in order to determine the right point in time for migration. A regular check if the view paths still apply to the designated community should be scheduled. This presumes a deep insight of the equipment and requirements of the designated community.

It could be evaluated, if the lack of validation software can be counterbalanced by tentatively migrating into a file format which is preferred for preservation purposes. Further research should bring into focus the automated evaluation of (digital to digital) migration of AV content.

REFERENCES

- [1] B. E. Koenig and D. S. Lacey, "Forensic Authentication of Digital Audio and Video Files," in Handbook of digital forensics of multimedia data and devices, A. T. S. Ho, Ed., Hoboken: Wiley, 2015, 133-181.
- [2] MediaArea, MediaInfo. [Online] Available: <https://mediainfo.net/de/MediaInfo>. Accessed on: Nov. 01 2018.
- [3] Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-M-2, 2012.
- [4] M. Todd, "File formats for preservation: Technology Watch Report," https://www.dpconline.org/component/docman/?task=doc_download&gid=375, 2009.
- [5] D. Pearson and C. Webb, "Defining File Format Obsolescence: A Risky Journey," IJDC, vol. 3, no. 1, pp. 89-106, <https://doi.org/10.2218/ijdc.v3i1.44>, 2008.
- [6] H. M. Ryan, "Who's afraid of File Format Obsolescence?: Evaluating File Format Endangerment Levels and Factors for the Creation of a File Format Endangerment Index," School of Information and Library Science, University of North Carolina, Chapel Hill, 2014.
- [7] J. F. Steenbakkens, "Digital Archiving in the Twenty-First Century: Practice at the National Library of the Netherlands," Library Trends, vol. 54, no. 1, pp. 33-56, <http://muse.jhu.edu/article/193231>, 2005.
- [8] J. Houpert, P. Melas, W. Bailer, and P. Walland, "Recommendations and techniques for content in a 'born robust' form," 2015. Accessed on: Jun. 25 2018.