Technische Hochschule Wildau
Technical University of Applied Sciences

Masterarbeit

zur Erlangung des akademischen Grades
Master

Technische Hochschule Wildau

Fachbereich Wirtschaft, Informatik, Recht

Studiengang Bibliotheksinformatik M. Sc.

Thema (deutsch): Preservation Planning für die digitalen audio-visuellen Materialen der Technischen Informationsbibliothek Hannover - Entwicklung eines Kriterienkataloges und eines Migration Plugins

Thema (englisch): Preservation Planning for Digital Audio-Visual Material at the German National Library of Science and Technology (TIB) – Developing a Catalogue of Criteria and a Migration Plugin

Autor/in: Merle Friedrich

Matrikel-Nr.: 50041765

Seminargruppe: BIM/16

Betreuer/in: M. Eng. Marcel-Dominique Paul Block

Zweitgutachter/in: Prof. Dr. rer. nat. Janett Mohnke

Eingereicht am: 03.03.2019

Bibliographic Description

Preservation Planning for Digital Audio-Visual Material at the German National Library of Science and Technology (TIB) – Developing a Catalogue of Criteria and a Migration Plugin
Masterarbeit 2019, 81 Seiten, 12 Abbildungen; 15 Tabellen; 4 Anlagen
Technische Hochschule Wildau, Fachbereich WIR

Goal: Evaluation of the risk of obsolescence for the most common audio-visual file formats in TIB's holdings and development of a migration plugin

Inhalt: Die Technische Informationsbibliothek Hannover (TIB) betreibt ein Langzeitarchivierungssystem um die digitalen Bestände, unter anderem audiovisuelle Materialien, zu erhalten. Um obsolete audiovisuelle Formate zu identifizieren wird ein Kriterienkatalog entwickelt. Diejenigen Formate, welche einen Großteil der audiovisuellen Bestände der TIB ausmachen, werden unter Berücksichtigung der Anforderungen der Nutzergruppen (Designated Community) der TIB untersucht. Die Analyse ergibt, dass die untersuchten Formate zwar nicht für die Langzeitarchivierung empfohlen werden, aber auch nicht von Obsoleszenz bedroht sind. Ergänzend wurde ein Migrations-Plugin für das Langzeitarchivierungssystem entwickelt und erfolgreich eingesetzt. Das Migrations-Plugin kann während des Preservation Planning eingesetzt werden um die untersuchten Formate in das von der TIB für die Langzeitarchivierung bevorzugte Format zu migrieren.

Abstract: The German Nation Library of Science and Technology (TIB) runs a digital archive in order to preserve its digital holdings which consists amongst others of audio-visual material. In order to identify obsolete audio-visual file formats a catalogue of criteria is developed and file formats which represent the majority of TIB's audio-visual holdings are examined. The assessment reveals that the examined file formats are not a preferred file format for digital preservation, but on the other hand they are not at risk to become obsolete to TIB's designated community. Furthermore, a migration plugin for the digital archive's software environment is developed and successfully deployed. The migration plugin can be used during preservation planning in order to migrate the examined formats into TIB's chosen archival file format if a migration becomes necessary.

Keywords: Langzeitarchivierung, Digital Preservation, audio-visual material, audio-visuelle Medien, Preservation Planning, Migration

Eidesstattliche Erklärung:

Ich erkläre hiermit, dass ich die vorliegende Abschlussarbeit eigenständig angefertigt und nur die angegebenen Quellen und Hilfsmittel verwendet habe.
Hannover, 20. Februar 2019

_____

Merle Mareike Friedrich

Anmerkung zur Veröffentlichung:

Für diese Veröffentlichung wurden Kapitel 1, 2 und 6 überarbeitet. Die Überarbeitung fand nach der Bewertung statt.

Note on publication:
Chapters 1, 2 and 6 have been revised for this publication. The revision took place after the evaluation.

List of Contents

Abbreviations

AAC	Advanced Audio Codec
AIP	Archival Information Package
AONS	Automatic Obsolescence Notification System
AV	Audio-Visual
AVC	Advanced Video Codec
COPTR	Community Owned digital Preservation Tool Registry
CPU	Computer Processing Unit
CRC	Cyclic Redundancy Check
DIP	Dissemination Information Package
DIPRec	Digital Preservation Recommender System
DROID	Digital Record Object Identification
FFmpeg	Fast Forward MPEG
ffv1	FFmpeg Videocodec 1
GOP	Group of Picture
GUI	Graphical User Interface
IE	Intellectual Entity
IETF	Internet Engineering Task Force
INFORM	Index for Risk Management
ISO / IEC	International Organization for Standardization/ International Electrotechnical Commission
JAR	Java Archive
KB	National Library of the Netherlands (Koninklijke Bibliotheek)
KOST	Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen
LoC	Library of Congress
LTO	Linear Tape Open
MD5	Message-Digest Algorithm 5
mkv	Matroska
MPC-HC	Media Player Classic – Home Cinema
MPEG	Moving Picture Experts Group
nestor	Network of Expertise in long-term STOrage and availability of digital Resources in Germany
OAIS	Open Archival Information System

MacOS	Macintosh Operating System
PAR	Preservation Action Registry
PCM	Pulse Core Modulation
PLATO	Preservation Planning Tool
PS	Program Stream
PUID	Pronom Unique Identifier
RI	Representation Information
SHA	Secure Hash Algorithm
SIP	Submission Information Package
SSD	Solid State Drive
TIB	German National Library of Science and Technology
VLC	VideoLAN Client

List of Figures

List of Tables

Attachments

Attachment 1 CD: Migration Plugin

Content CD:

- Migration Plugin
  - bin
    - ConvertToFfv1MkvPCM.sh
  - META-INF
    - MANIFEST.MF
  - PLUGIN-INF
    - Metadata_ffv1Converter.xml
  - Ffv1MigrationTools.jar
  - LICENSE
  - README.md
- Tests
  - InputFiles
    - A_10310.mp4
    - A_10310.mp4.mediainfo.xml
    - B_15854.webm
    - B_15854.webm.mediainfo.xml
    - C_10368.VOB
    - C_10368.VOB.mediainfo.xml
  - OutputFilesTest1
    - A_10310.mkv
    - A_10310.mkv.mediainfo.xml
    - B_15854.mkv
    - B_15854.mkv.mediainfo.xml
    - C_10368short.mkv
    - C_10368short.mkv.mediainfo.xml
  - OutputFilesTest2
    - A_10310.mkv
    - A_10310.mkv.mediainfo.xml
    - B_15854.mkv
    - B_15854.mkv.mediainfo.xml
    - C_10368short.mkv
    - C_10368short.mkv.mediainfo.xml
  - logs.log
- Masterthesis_Friedrich.pdf

1 Introduction

Digital preservation has the goal of keeping information from today accessible to an audience in the future. The information or content consists of a representation in the form of a digital object (file) which is encoded in a file format. In order to access and use the information even after a change in technology or a change in the audience[1], a back-up strategy which aims at the preservation of the bit-stream will fail, because accessing the content depends on recent hard- and software.

The team digital preservation has the task to preserve the digital holdings of the German National Library of Science and Technology (TIB) for the long term[2]. TIB therefore operates a digital archive. TIB provides information on the subjects of science and technology to academics and business.[3] This information comes in textual materials as well as non-textual materials such as audio-visual (AV) material.[4]

AV material enters the archive in a variety of file formats. When analyzing a file format from a preservation point of view, it is relevant to analyze on one hand, if a file format is favorable, accepted or critical for preservation purposes. On the other hand it is useful to evaluate if a file format is widespread or obsolete in the designated community. With these two factors an archive can determine if a preservation action is necessary in order to preserve the information and keep it accessible. If a format at risk has been identified, applicable handling methods like migration are required and shall be supplied by TIBs archival system.

How can memory institutions like TIB identify suitable formats for preservation?[5] And how can the information be passed over into this format? The following shall answer these questions for TIB's digital AV material.

[1] Recommended Practice CCSDS 650.0-M-2, 1.1.
[2] Technische Informationsbibliothek 2017a, p. 22.
[3] Bähr and Schwab 2018.
[4] Bähr and Schwab 2018.
[5] Stanescu 2005, p. 62.

2 Research Questions and Methodology

Is it possible to migrate digital audio-visual material from the holdings of TIB which is obsolete?

This overall question can be answered in two separate parts:

1. Are there file formats in TIB's audio-visual holdings which are obsolete?

Not all file formats are suitable for preservation; nevertheless they can enter an archive for reasons of authenticity or rights. When a file format is obsolete migration is necessary. Based on a literature study a catalogue of criteria for suitability and risk of obsolescence is developed. The most common file formats in TIB's archive are examined in order to verify the following hypothesis:

1. a) The majority of file formats within the TIB AV-holdings are not suited as preferred preservation formats.

1. b) The majority of file formats within the TIB AV-holdings are not obsolete.

Referring to the first part of the overall question the second research question is derived:

2. How can the migration to a preservation-suitable format be achieved within the infrastructure of digital preservation at TIB?

The given infrastructure of TIB's archive can be enhanced by custom plugins. I want to develop a plugin in order to migrate the majority of TIB's AV holdings to a format suitable for archival needs for audio-visual material. After successfully testing the plugin, I can verify the hypothesis:

2. A plugin can be developed and integrated in TIB's archive software environment in order to migrate to a suitable format.

To answer the research questions and confirm the hypothesis the following approaches are chosen:

1. How is it possible to determine if a format has a (high) risk to become obsolete?

- Literature Review: Development of a catalogue of criteria to determine the suitability for preservation
- Literature Review: Development of a catalogue of criteria to determine file format obsolescence
- Evaluation of the codecs and containers in the TIB AV-holdings according to the developed catalogues

2. How can the migration to a preservation-suitable format be achieved within the infrastructure of the digital archive at TIB?

- Testing preservation planning within the given software environment
- Requirement analysis for the plugin and its interactions within the system
- Feasibility Study: Developing a migration plugin (including deploying and testing)

In order to develop an awareness of the discussed problem the features of audio-visual material are displayed from a preservation point of view in chapter 3. This chapter is followed by an overview of the theoretical model of digital preservation in chapter 4. As the research question is part of the preservation planning function of an archive, this function is described in more detail. After the introduction to theory the following chapter focuses on practical tasks, introducing digital preservation at TIB with a spotlight on preservation planning.

The verification of the hypotheses is subject to the next chapters. In chapter 6 the developed catalogue of criteria is introduced and provides a basis for the classification of TIB's holdings regarding the suitability as archival format. But even if a format is evaluated as critical, this does not mean that action must be taken immediately. In order to judge if a format is obsolete, which would make a preservation action necessary, chapter 7 examines if rendering software is available. In order to enable migration as preservation action a migration plugin was developed, implemented and tested in TIB's archival software environment. These steps as well as the evaluation of the requirements are summarized in chapter 8.

Although the development of the migration plugin is tailored to TIB's archival software environment and holdings, the applied criteria catalogues can be reused for preservation planning by memory institutions with born-digital, audio-visual holdings.

3 Digital Audio-Visual Material

The German National Library of Science and Technology collects audio-visual material (AV material) from science and technology in the form of simulations, documentation of lectures, experiments etc. These videos are accessible via the AV-Portal (https://av.tib.eu) and preserved in TIB's digital archive. In order to evaluate file formats for digital AV material one has to consider the different elements of a file format, which are introduced in chapter 3.1. TIB's AV holdings are thereafter introduced in chapter 3.2 with a focus on these elements. The analysis of the holdings forms the basis of the classification of file formats for both suitability (chapter 6.2) and obsolescence (chapter 7.2). But an understanding of the elements of AV file formats is also crucial in order to develop a migration plugin (chapter 8.3).

3.1 Elements of Audio-Visual Material

Digital audio-visual (AV) material has two features: the wrapper (container layer) and the bit stream or content streams (codec layer).[6, 7] When I speak of file format I refer to the combination of wrapper and content stream(s). When I address a feature I speak of the wrapper or container, video stream or video codec, and audio stream or audio codec.

The content stream holds the information for the video or audio. Raw or uncompressed data streams are uncommon as moving images have a high amount of information.[8] In most cases the information is stored encoded, compressed by a coder. During playback the information is decoded[9] with the help of a decoder.[10] The coding program (coder and decoder) is referred to as a codec. [11] Data streams are either encoded lossy, which means that information from the raw data cannot be restored bit per bit, or lossless, which means that all information can be restored.[12]

An object can consist of multiple streams e.g. one video stream, an English audio stream and a German audio stream. Additionally it is possible to store subtitles, closed captioning or other information. The container wraps up the different streams and is therefore also called a

[6] Houpert et al. 2015, p. 12.
[7] Ho 2015, p. 137.
[8] Newmarch 2017, p. 11.
[9] Newmarch 2017, p. 11.
[10] Watkinson 2001, p. 200.
[11] Watkinson 2001, p. 200
[12] Watkinson 2001, p. 201.

wrapper. It can also include technical metadata as time code, and descriptive metadata like title, chapters etc. [13,14]

The container can be identified by file ending (e.g. .mp4 for an MPEG 4 Container). But as the file ending can be manipulated all technical metadata should be examined by a tool (e.g. MediaInfo.[15]) that can not only identify the container, but also the video and audio encoding/format and other technical aspects like frame rate or display aspect ratio.

It is important to note that a video or audio stream in one format can differentiate depending on the format version and in some cases versions are split in different format profiles.[16] For an archive it is therefore necessary to export the technical metadata on a very granular level. For this thesis I will differentiate on the level of format version.

3.2 Audio-Visual Holdings at TIB

The German National Library of Science and Technology collects information in the field of architecture, chemistry, computer science, mathematics and physics[17] . Customers are academics in general and business on a national level as well as the University of Hanover [18] . The TIB also holds a growing collection of digital AV material in the form of lectures, documentation of experiments, simulations etc[19] covering the different subjects. The digital videos are administered in a Media Asset Management System, and delivered to the user via the AV portal (https://av.tib.eu/). Videos are preserved in TIB's digital archive.

The collection consists of ca 13000 videos at the time of the writing. The collection is enhanced by videos of the same subjects which are not part of our holdings (via linking) and is steadily growing. TIB acquires the videos from different producers. As there are no restrictions on the file formats, there is a wide range of AV file formats in the digital archive. The extraction of technical metadata with MediaInfo[20] on the format version level reveals that there are 45 different combinations of container, video codecs, and in most cases audio codecs (some are without audio). A detailed overview can be found in Appendix A. I will

[13] Wright 2012, p. 12
[14] Newmarch 2017, p. 11.
[15] MediaArea.
[16] Arms et al. 2018.
[17] Technische Informationsbibliothek 2017a, p. 8.
[18] Technische Informationsbibliothek 2017a, p. 8.
[19] Technische Informationsbibliothek.
[20] MediaArea.

differentiate container, video and audio codec on the level of format version. Considering only the format and format version is a generalization and not adequate for preservation actions like migration[21] (see chapter 4). Nevertheless one can get a sufficiently detailed impression of the collection in order to analyze the collection for suitability and to develop a migration plugin.

The three most numerous file formats cover approximately 73 % of TIB's AV holdings and are therefore examined, see Table 1. More than 55 % of the holdings consist of the Advanced Video Codec (AVC) with Advanced Audio Codec (AAC) in version 4, wrapped in a MPEG-4 container. The next most common formats are found more than 1000 times in TIB's holdings: the WebM-Container with the video codec VP8 and Vorbis audio codec, and the MPEG-PS Container with MPEG Video in version 2 and MPEG Audio in Version 1.

Furthermore the following target format for a migration is investigated: FFmpeg Videocodec 1 (ffv1) in version 3, with Pulse Core Modulation (PCM) audio codec in a matroska (mkv) container. The qualities of the target format for preservation purposes are illustrated in chapter 6.3.

Table 1 Examined File Formats

	Container	Video	Audio	Occurances in Archival Holdings
Input	MPEG-4	AVC	AAC, Version 4	7408
	WebM	VP8	Vorbis	1062
	MPEG-PS	MPEG Video, Version 2	MPEG Audio, Version 1	1006
Output	Matroska (mkv)	ffv1, version 3	PCM	-

4 Digital Preservation – Theoretical Approach

TIB, as with many other memory institutions, has committed itself to Digital Preservation as one pillar in its strategy. Digital Preservation has the goal to manage information "to be ac-

[21] Arms and Fleischhauer 2005, p. 1.

cessible, usable, and understandable in the long term".[22] Besides the advantages of storing information digital there are risks of losing information due to a lack of restoring, rendering or interpreting possibilities. [23] Risks occur on the bitstream level (e.g. bit flips on hard drives, scratches on CD), on the logical level (e.g. no software available to render the file format correctly) and on the semantic level (e.g. missing descriptive metadata hinders the interpretation of a simulation).[24]

Long term does not refer to a given time period, but to overcoming the challenges of changing technologies, new data formats or a change in the user community. [25] The user community or designated community plays an important role in the archival decisions regarding preservation, as the information is preserved for their use and must therefore be understandable by them.[26]

In 2003, the Consultative Committee for Space Data Systems developed a reference model for an Open Archival Information System (OAIS) which describes "an organization […] of people and systems that has accepted the responsibility to preserve information and make it available for a **Designated Community".**[27] Chapter 4.1 gives an introduction to the reference model. The reference model also describes tasks of preservation planning which are related to the first research question of obsolete formats and is subsequently described in chapter 4.2 Preservation Planning – Functional Overview and 4.3 Preservation Planning - Workflow, complemented by further research regarding preservation planning. As the second research question wants to answer the question of whether migration is possible within TIB's archive, chapter 4.4 illuminated the reasons for migration and risks of migration. Another approach to preserve the accessibility of files in obsolete file formats is the emulation of the rendering software and possibly software environment which comes with different strategies, advantages and drawbacks than migration. As the research questions focus on the migration of the files, emulation is out of the scope of this work.

[22] Pearson and Webb 2008, p. 90.
[23] Recommended Practice CCSDS 650.0-M-2, 1.3.
[24] Becker and Rauber 2011, p. 1009.
[25] Recommended Practice CCSDS 650.0-M-2, 1.1.
[26] Recommended Practice CCSDS 650.0-M-2, 1.11.
[27] Recommended Practice CCSDS 650.0-M-2, 1.1.

Whereas chapter 4 focuses on the theoretical approach, chapter 5 introduces digital preservation at TIB and the realization of preservation planning. This forms the basis of the development of the migration plugin.

4.1 Reference Model for an Open Archival Information System

The reference model for an Open Archival Information System (OAIS) is adopted broadly in the digital preservation community and has been under review since 2017.[28] It offers an in-depth description and requirements of the functions and information flows within an OAIS as well as within its environment. The model describes an OAIS-compliant archive as "an organization […] of people and systems that has accepted the responsibility to preserve information and make it available for a **Designated Community".**[29] The term designated community is used to describe a group of consumers who shall be capable of comprehending the information over the long-term.[30] TIB has defined its designated community in its preservation policy (see chapter 5.1).

Objects that shall be preserved are stored in information packages. High-level external interactions describe the workflow from the submission of an information package, archiving and dissemination of the information object. A producer delivers an information package to the archive. The Submission Information Package (SIP) consists of the object and may include other descriptive information.[31] The information package is transferred into an Archival Information Package (AIP) which contains in addition preservation description information and is preserved within the archive. [32] If a consumer requests information, a Dissemination Information Package (DIP) is generated from AIPs and made available to the consumer.[33]

It is in the responsibilities of an archive that the information is understandable to the designated community and to ensure its authenticity by granting evidence through the preservation description information.[34] In order to achieve this, the archive has to provide different functions, which are displayed in the functional model in Figure 1 OAIS Functional Model.

---

[28] Kilbride et al. 2015.
[29] Recommended Practice CCSDS 650.0-M-2, 1.1.
[30] Recommended Practice CCSDS 650.0-M-2, 3.1.
[31] Recommended Practice CCSDS 650.0-M-2, 1.15.
[32] Recommended Practice CCSDS 650.0-M-2, 1.9.
[33] Recommended Practice CCSDS 650.0-M-2, 1.11.
[34] Recommended Practice CCSDS 650.0-M-2, 3.1.

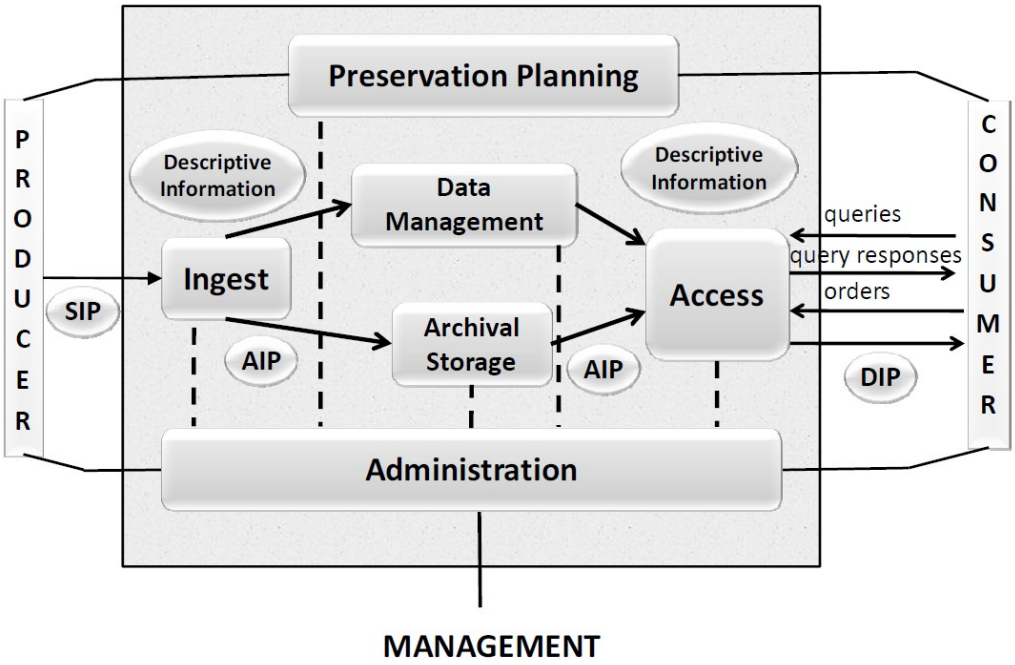


Figure 1 OAIS Functional Model[35]

The producer hands a SIP to the OAIS. The ingest functional entity accepts SIPs from Producers, performs quality control, extracts representation information (RI, like technical metadata), and generates an AIP.[36] The archival functional entity manages the permanent storage of the AIPs, including safety measurements like error checking, renewing storage technology, and retrieval of AIPs[37] , while the data management functional entity contains all the functions and processes to maintain a central information storage like a data base, which holds the descriptive information regarding the holdings of the archive, and the administrative data to manage the archive.[38] The access functional entity allows the designated community to find and receive DIPs while implementing legal restrictions (e.g. access restrictions)[39] Administrative functions cover the contact with the Producer concerning the negotiation of submission agreements and quality control of the SIPs. It is responsible to maintain archives standards and policies. It must deliver functions to inventory, report and migrate the holdings

[35] Arms et al. 2017g.
[36] Recommended Practice CCSDS 650.0-M-2, 4.1-4.2.
[37] Recommended Practice CCSDS 650.0-M-2, 4.2.
[38] Recommended Practice CCSDS 650.0-M-2, 4.2.
[39] Recommended Practice CCSDS 650.0-M-2, 4.2-4.3.

of the archive.[40] The preservation planning functional entity holds services to provide preservation plans and provides risk analysis reports. It is described in detail in the next chapter. [41]

4.2 Preservation Planning – Functional Overview

Preservation Planning is described in more detail as all research questions are tasks of preservation planning. As there are technical, financial and organizational aspects to consider, preservation planning can be considered as "often the most difficult part in digital preservation endeavours" [42]. As a loss of information e.g. of embedded metadata will always be part of a preservation process, it has to be taken into account. In the OAIS model the Preservation Planning functional entity therefore covers different services and functions. [43] The OAIS model was reviewed in 2018. During the review it is discussed to enhance the functions by the Planets (Preservation and Long-term Access through NETworked Services) functional model.[44] . Therefore the Planets functional model is introduced later in this chapter. While the Reference Model offers a great overview of the functions, the order of these steps, the implementation as well as the technologies is not prescribed. And not all functions can be executed or supported by software. The process of preservation planning with Rosetta is therefore described separately in chapter 5.2.

Preservation planning seeks to evaluate actions to access digital content in the course of changing technological environments, consumer expectations or organizational capabilities.[45] For AV material "this will involve retaining a knowledge of the encoding conditions of the file and the ongoing changes in standards, formats […], and playout technology".[46]

In the OAIS the main functions in the preservation planning functional entity are "Develop Preservation Strategies and Standards", "Develop Packaging Designs & Migration Plans", "Monitor Technology" and "Monitor Designated Community" which are displayed in Figure 2. The chart also shows the interactions with the administration functional entity as well as with consumers and producers. An extensive description of all function can be found in the

[40] Recommended Practice CCSDS 650.0-M-2, 4.2.
[41] Recommended Practice CCSDS 650.0-M-2, 4.2.
[42] Strodl et al. 2007, p. 32.
[43] Recommended Practice CCSDS 650.0-M-2, 1.14.
[44] Sierman 2016.
[45] Becker and Rauber 2011, p. 1009.
[46] Houpert et al. 2015, p. 10.

OAIS[47]. In the following I want to map the research questions to the preservation planning functions.

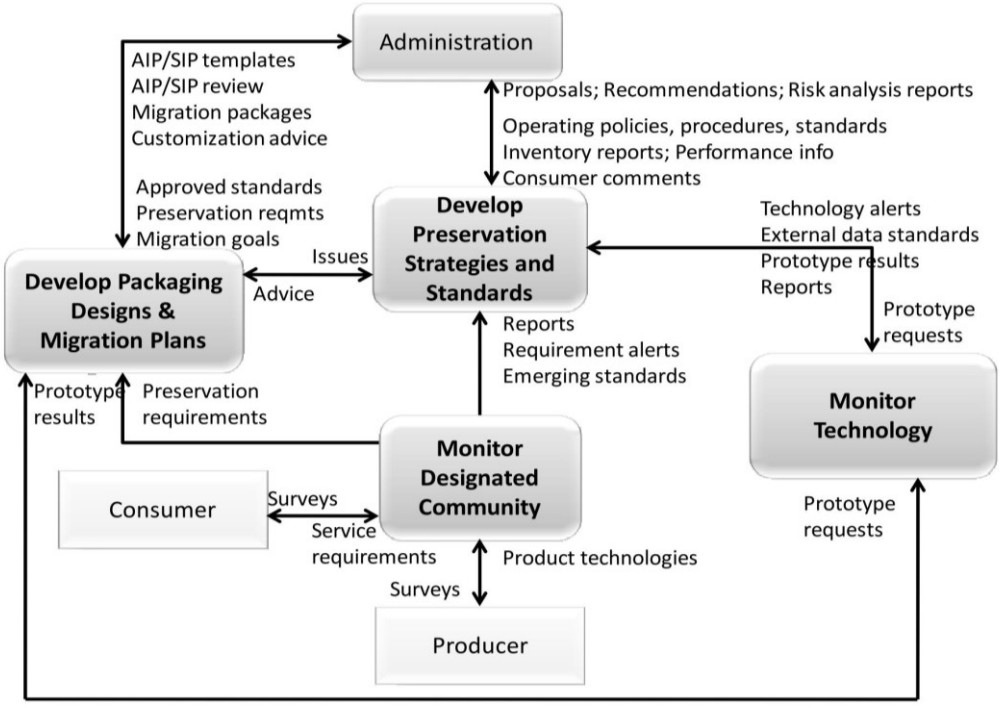


Figure 2: The Functions of the Preservation Planning Functional Entity[48]

The first research question *"Are there formats in TIB's audio-visual holdings which are obsolete?"* is located within Develop Preservation Strategies and Standards. This function collects representation information (e.g. technical metadata concerning the format) on the inventory from the Administration, as well as technology alerts from the Monitor Technology function. In combination with the information about emerging standards from the Monitor Designated Community function it is possible to identify if formats are obsolete. If this is the case, the Develop Preservation Strategies and Standards sends recommendations to Administration.[49]

The second research question *"How can the migration to a preservation-suitable format be achieved within the infrastructure of digital preservation at TIB?"* is part of the Develop Packaging Designs and Migration Plans function. This function develops detailed migration plans on the basis of the preservation requirements and migration goals from the Administra-

[47] Recommended Practice CCSDS 650.0-M-2, 4.14.
[48] Recommended Practice CCSDS 650.0-M-2, 4.14.
[49] Recommended Practice CCSDS 650.0-M-2, 4.15.

tion. [50] Therefore it can rely on prototype software from Monitor Technology function, and on reports from Monitor Designated Community function.[51] The migration plan is sent to Administration. After the approval the Administration schedules and performs the migration.[52] Preservation planning as defined in the OAIS is a constant process in an archive.[53]

Another model has been developed during the EU funded Planets Project from 2006 until 2010.[54] This model describes three functions: Preservation Watch (which extends "Monitor Designated Community" and "Monitor Technology" [55]), Preservation Planning and Preservation Action.[56] The relations between these functions are described in Figure 3.

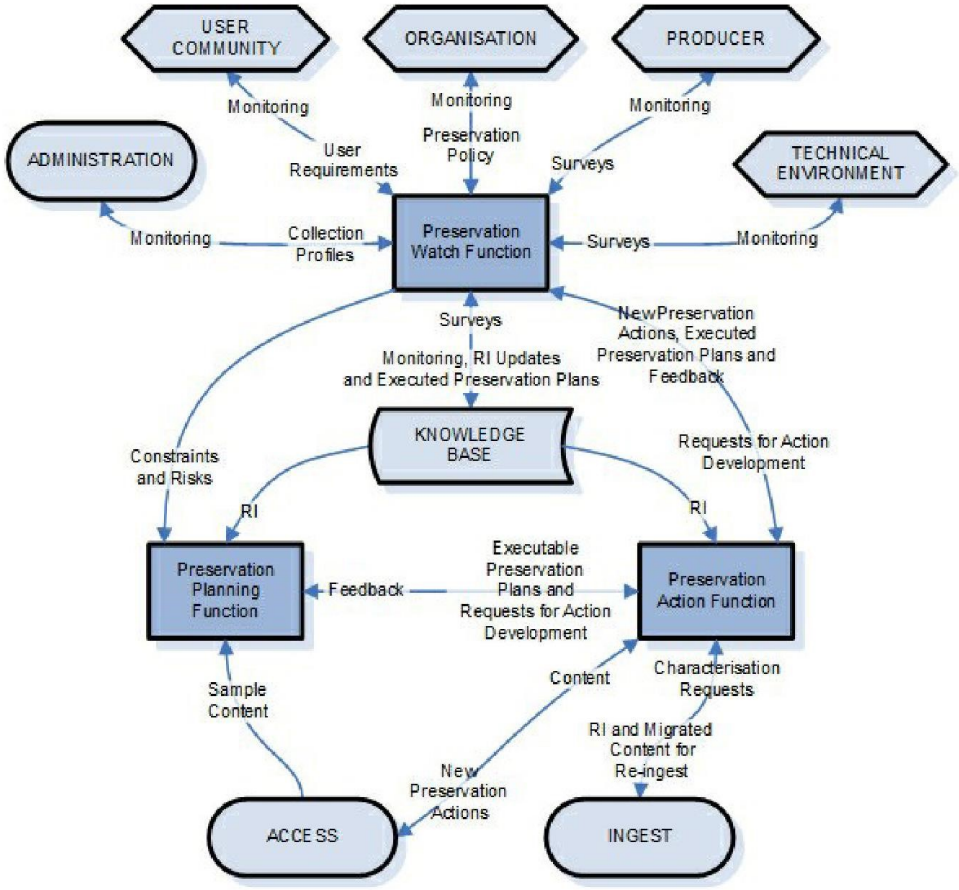


Figure 3 Planets Functional Model [57]

[50] Recommended Practice CCSDS 650.0-M-2, 4.15.
[51] Recommended Practice CCSDS 650.0-M-2, 4.15-4.16.
[52] Recommended Practice CCSDS 650.0-M-2, 4.15-4.16
[53] Becker and Rauber 2011, p. 1027.
[54] Farquar 2007.
[55] Sierman 2016.
[56] Sierman 2016.
[57] Wheatley and Sierman 2010, p. 8.

The first research question concerning file format obsolescence is located within the Preservation Watch function which gathers information from Administration, from the user community (or designated community), the organization, the producers and the technical environment. Representation information (RI) which describes the object and its technical dependencies is stored in a Knowledge Base. [58] Preservation Watch function also provides a testbed and performs risk analysis.[59] Risks that are critical or imminent are reported to the Preservation Planning function.[60] If a format within TIB's AV holdings is obsolete the Preservation Watch function would report this to Preservation Planning.

The Preservation Planning function analyses appropriate preservation solutions and is therefore the allocation of the second research question concerning the migration.[61] Based on the risks reported by Preservation Watch and the RI of the knowledge base alternatives are tested and evaluated. On this basis the Preservation Planning function sends executable preservation plans to the Preservation Action function. The Planets function model locates the migration in the Preservation Action function, and therefore separated from preservation planning, just as like the OAIS.[62]

Both models offer an overview of the functions and tasks an archive performs. The steps which are performed while planning a preservation action and their order according to Planets is described in the next chapter.

4.3 Preservation Planning - Workflow

Preservation Planning consists of four consecutive steps, each can be broken into more detailed workflows. Figure 4: Preservation Planning according to Planets as Implemented by the Tool PLATO (stands for Preservation PLAnning TOol) gives a detailed insight of the process.

[58] Todd 2009, p. 22.
[59] Wheatley and Sierman 2010, p. 9.
[60] Wheatley and Sierman 2010, p. 9
[61] Wheatley and Sierman 2010, p. 11.
[62] Wheatley and Sierman 2010, p. 28.

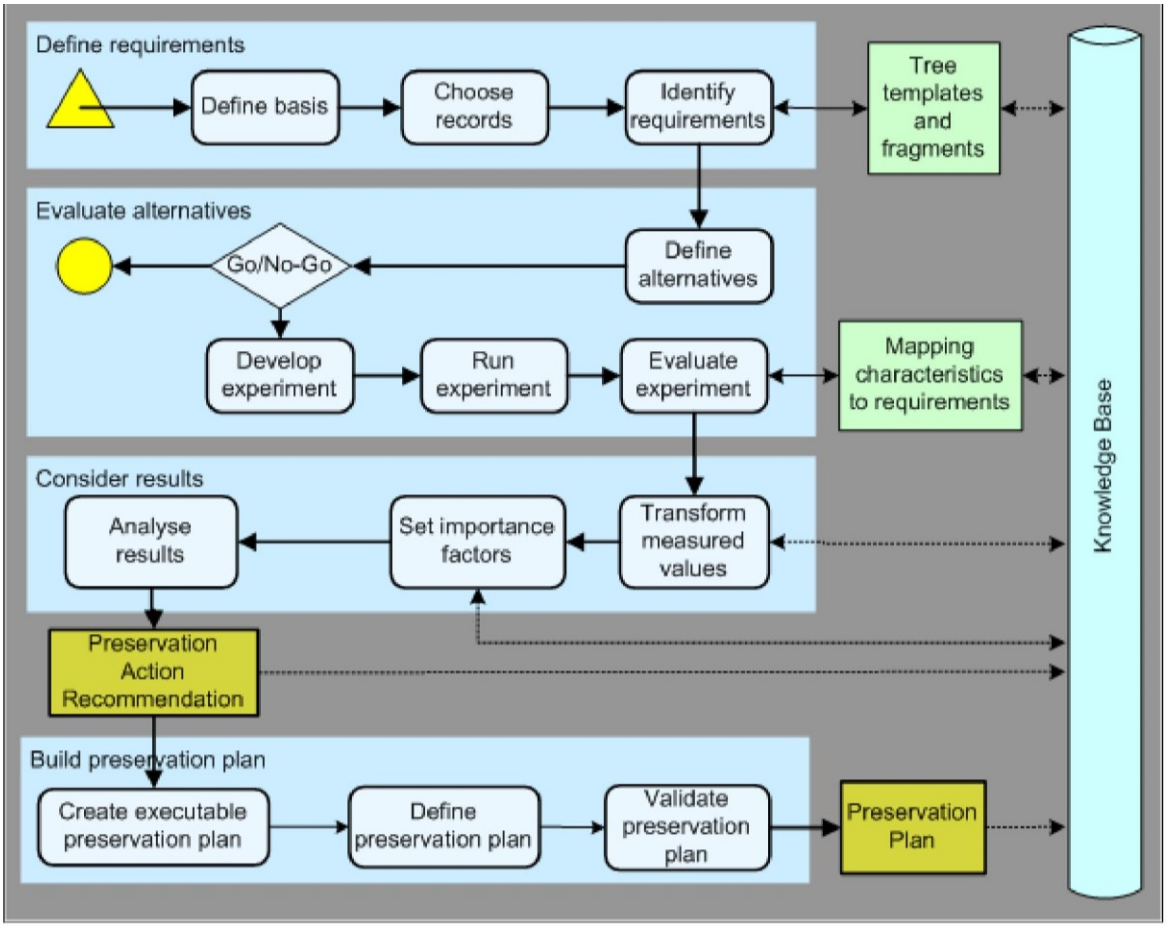


Figure 4: Preservation Planning according to Planets as Implemented by the Tool PLATO[63]

The first step is to **define requirements** through the formulation of measureable criteria.[64] The definition of the basis consists of a description of the collection, rights and the archives policies. [65] Choose records refers to the constitution of a representative test set of objects. [66] The last task during the define requirements step is to identify the demands. [67] This includes interactions with the knowledge base that holds information from the designated community, the institutions policy and the producer. [68] One possibility is to store the detailed criteria in a tree structure. [69]

[63] Planets 2010, p. 12.
[64] Becker and Rauber 2011, pp. 1011–1012.
[65] Strodl et al. 2007, p. 33.
[66] Strodl et al. 2007, p. 33.
[67] Strodl et al. 2007, p. 33.
[68] Becker et al. 2008, p. 368.
[69] Becker et al. 2008, p. 368

In a second step **alternatives are evaluated** through controlled experiments on a test set of the archive's objects. Each of the criterions from the first step is evaluated for each alternative solution.[70] The archive defines different alternative preservation solutions, e.g. tools for migration or emulators. [71] The necessary resources (amount of work, time, money) in order to run the experiments are estimated. [72] A decision is made on which alternative shall be evaluated, cancelled or postponed.[73] For the persisting alternatives a detailed plan is developed consisting of the workflow, the technical environment and measurement of the outcome. [74] The experiment is conducted on the defined test set.[75] At last the experiment is evaluated against the identified requirements from the first step. [76]

Thirdly the **results are analyzed**. Each criterion is weighted according to the archive's preservation policy or according to the organizations capabilities, as in the case of weighting e.g. processing time or cost of software. An overall rating factor can be calculated and leads to a recommendation of an alternative.[77] Therefore each of the results must be adjusted to a uniform scale in order to be comparable.[78] The archive assigns an importance factor to the criterions. [79] Through this a comparable rating factor can be calculated. [80]

The last step is the **building of a preservation plan**. [81] The preservation plan includes the decisions as well as a complete description of the first three steps. [82] The preservation plan should be simulated before actually executing it. While the execution all parameters must be equal to the simulation.[83]

All steps are accompanied by preservation watch: monitoring the designated community, technology and institutional requirements. [84]

---

[70] Becker and Rauber 2011, pp. 1011–1012
[71] Strodl et al. 2007, p. 33.
[72] Strodl et al. 2007, p. 33
[73] Strodl et al. 2007, pp. 33–34.
[74] Strodl et al. 2007, p. 34.
[75] Strodl et al. 2007, p. 34.
[76] Strodl et al. 2007, p. 34.
[77] Becker and Rauber 2011, pp. 1011–1012
[78] Strodl et al. 2007, p. 34.
[79] Strodl et al. 2007, p. 34
[80] Strodl et al. 2007, p. 34
[81] Becker and Rauber 2011, pp. 1011–1012
[82] Becker et al. 2008, pp. 369–370.
[83] Bailer et al. 2014, p. 358.
[84] Becker and Rauber 2011, pp. 1011–1012

Software can support only certain aspects of preservation planning: it can help to examine the objects of the test set before and after migration, as well as conducting the migration. For both software tools which offer an extensive report are recommended[85] Rosetta offers the integration of tools and a dedicated workflow for preservation planning, which is introduced in chapter 5.2.

4.4 Migration

The OAIS defines Digital Migration as the transfer of digital information with the purpose of preserving the full information content[86]. The intent is to replace the old representation. The archive is in control and responsible for the transfer. [87] Becker et al. (2008) add that migration is used to transform objects at risk into better suitable information objects.[88]

The archive decides the point in time of the migration. Migration can take place at the creation of the digital object[89] (e.g. digitization of a 16mm-film). If migration takes place before[90] or directly after the ingest[91] into the archive it is also called normalization[92]. Another point in time for migration is on-demand or at obsolescence.[93] In its preservation policy an archive defines when migration takes places. The decision for a file format depends on the resources, the relationship with the producer and designated community at this point in time.[94]

Lawrence et al. (2000) list five reasons for migration: Format obsolescence, obsolescence of underlying hardware, proprietary formats with no published specifications, normalization to gain oversight, and simplified metadata management through embedded metadata.[95] The migration to a born robust format which distinguished through integrity measurements and redundancy is another reason for migration.[96] Wright (2012) states that migration for compressed AV material will be driven by the obsolescence of the compression[97].

---

[85] Strodl et al. 2007, p. 30.
[86] Recommended Practice CCSDS 650.0-M-2, 1.11.
[87] Recommended Practice CCSDS 650.0-M-2, 1.11
[88] Becker et al. 2008, p. 367.
[89] Todd 2009, p. 20.
[90] Wright 2012, p. 10.
[91] Stanescu 2005, p. 72.
[92] Wright 2012, p. 10
[93] Todd 2009, p. 20
[94] Todd 2009, p. 20
[95] Lawrence et al. 2000, p. 22.
[96] Houpert et al. 2015, p. 17.
[97] Wright 2012, p. 11.

Migration from one format into another comes with risks[98]. A loss in quality can occur due to (repeated) compression of the AV material.[99] It is therefore recommended not to choose a compressed format as target format. [100] Errors in the migration process can be caused by the conversion software, both through technological error in the conversion software, but also through human error by e.g. using the wrong parameters within the conversion process. [101] The choice for conversion software should be made carefully and the software should be tested with a set of samples. [102, 103] As not all formats are interchangeable a loss of information (e.g. in the embedded metadata) imposes a risk to the object[104]. Therefore a migration plan defines "significant properties" that must be preserved.[105]

Significant properties (also called characteristics) consist of a property (e.g. display aspect ratio) and a value (e.g. 16:9), see Figure 5. After a migration the significant property must have the same value. They are determined by the producer, the archive and the user who has access to the digital object (the designated community).[106]

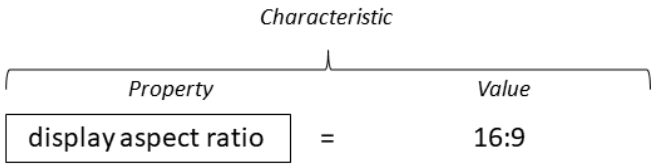


Figure 5: Properties and Characteristics, adapted from Dappert and Farquhar (2009) [107]

In order to reduce the risk, the migration of the content leads to a new AIP version. The first version of the AIP is retained for authenticity purposes. [108] An extensive quality control reduces the risk further.[109]

---

[98] Recommended Practice CCSDS 650.0-M-2, 5.5.
[99] Wright 2012, p. 16.
[100] Wright 2012, p. 16
[101] Wright 2012, p. 16
[102] Lawrence et al. 2000, p. 21.
[103] Houpert et al. 2015, p. 14.
[104] Lawrence et al. 2000, p. 21.
[105] Becker and Rauber 2011, p. 1013.
[106] Dappert and Farquhar 2009, p. 302.
[107] Dappert and Farquhar 2009, p. 299.
[108] Recommended Practice CCSDS 650.0-M-2, 5.6.
[109] Becker and Rauber 2011, p. 1025.

5 Digital Preservation at TIB

Digital Preservation is part of TIB's strategy.[110] TIB is aware that this strategy needs to be adapted to given and future circumstances in order to preserve the information.[111] The preservation policy (chapter 5.1) gives further information on TIB's principles for digital preservation and its designated community.

With their partners "ZB MED Information Centre for Life Sciences" and "ZBW Leibniz Information Centre for Economics" TIB operates a running digital preservation system (or digital archive) on the base of Ex Libris Rosetta (ProQuest).[112] The digital preservation system consists of a productive system, which must always be identical to the test system. For the development of new workflows or tools a development system is used.[113] The development and testing of the plugin took place in the development environment. As the plugin is part of preservation planning, the preservation planning workflow with Rosetta is described in chapter 5.2 "Preservation Planning with Rosetta".

5.1 Preservation Policy

TIB has its own preservation policy, which is reviewed yearly and is the basis on which decisions concerning digital preservation are made.[114] The policy therefore is relevant for the covered topics: preservation planning and migration. TIB describes the missions and principles of its digital archive. The designated community consists of three groups: patrons, employees and data producers.[115] TIB's patrons are members of the university of Hanover, research institutions and the industry.[116] Preservation Watch is one of twelve principles that TIB defines for digital preservation.[117] Migration as preservation strategy is set as one possibility in TIB's preservation strategy and leads to a new version of the AIP.[118] The authenticity of the objects is achieved by keeping the original, document any changes that are made and versioning of the AIP.[119] Through collaboration in community networks, conferences and

[110] Technische Informationsbibliothek 2017b.
[111] Bähr and Schwab 2018.
[112] Bähr and Schwab 2018.
[113] Bähr and Schwab 2017b.
[114] Technische Informationsbibliothek 2017b.
[115] Technische Informationsbibliothek 2017b.
[116] Technische Informationsbibliothek 2017b.
[117] Technische Informationsbibliothek 2017b.
[118] Technische Informationsbibliothek 2017b.
[119] Technische Informationsbibliothek 2017b.

exchange with partner organizations as well as scientific publications TIB conducts community, and technology watch.[120]

5.2 Preservation Planning with Rosetta

Preservation planning in general was described in chapters 4.2 and 4.3. The preservation planning process with Rosetta is introduced as it is part of testing the migration plugin. The composition of an elaborated preservation plan for the examined materials is out of scope.

The preservation module of Rosetta consists of four sub-modules: the format library, risk analysis, preservation planning and preservation execution.[121] The workflow is described in Figure 6 Workflow Preservation Planning with Rosetta. As the preservation execution or preservation action is assigned to the administration functional entity according to the OAIS[122], the illustrated workflow ends with a signed-off preservation plan and not with the execution of a plan. The full workflow can be seen in Appendix B.

Preservation planning is carried out by the digital preservation staff and supported by the digital preservation system (Rosetta). The process starts when a risk is identified. In the format library each format is identified with PUID (Pronom Unique Identifier) or internal ExLibris-Identifier. [123] Staff can add a risk identifier to the format, and specify one or more risk parameters. A risk parameter consists of a characteristic (e.g. File Format, Creating Application Version, Validity,), an operator (e.g. equals, contains, …) and a parameter (string, integer, boolean depending on the characteristic). The characteristic can consist of any metadata field that is populated and indexed in Rosetta.[124]

After the risk has been activated, the staff can schedule a risk report. The risk report lists all AIPs as well as the number of files that are affected by the risk.[125] At this point the staff can decide if it is necessary to proceed. It might not be necessary, if only derivative copies are affected, and the preservation master is not associated with a risk. If action is necessary the preservation staff creates a preservation set which contains all affected files. The launch of

[120] Bähr and Schwab 2017a.
[121] Ex Libris Documentation Department 2018, pp. 9–10.
[122] Recommended Practice CCSDS 650.0-M-2, p. 58.
[123] Ex Libris Documentation Department 2018, p. 23.
[124] Ex Libris Documentation Department 2018, p. 41.
[125] Ex Libris Documentation Department 2018, pp. 94–95.

the preservation plan includes acquiring input from stakeholders, the preservation policy and other policies that might exist.[126]

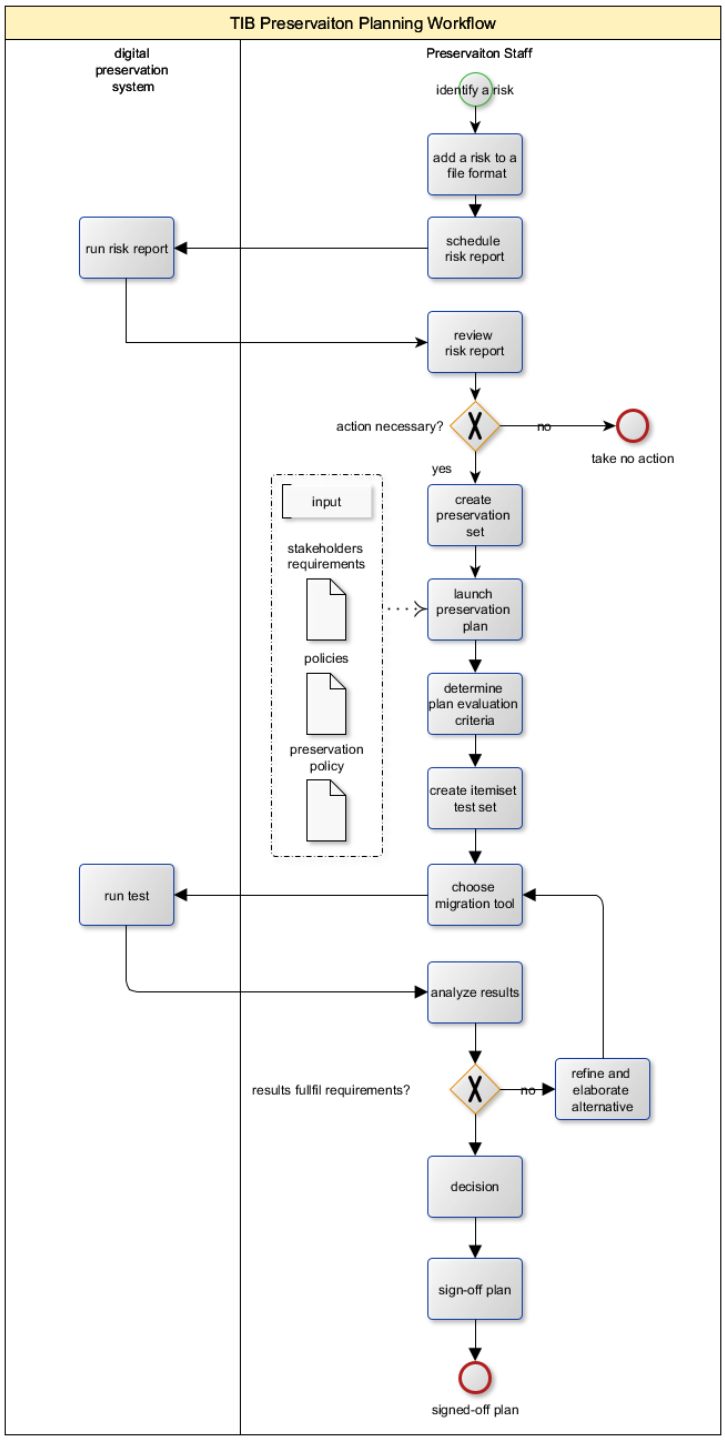


Figure 6 Workflow Preservation Planning with Rosetta adpted from Bähr and Schwab (2017a)

[126] Bähr and Schwab 2017a.

In order to be able to compare different alternatives, preservation staff can determine plan evaluation criteria. Criteria can compromise criteria for the format (e.g. traceability of change), organizational criteria (e.g. software / hardware costs, local staff), and of the migration (comparing the significant properties). Preservation staff can refine the test set when creating the itemized test set.[127]

A migration tool can be chosen: either an internal plugin or external software. Tools can be implemented into the software as an internal plugin which then conducts the migration. When choosing external migration, Rosetta exports the test set and provides a folder for the import of the migrated files.[128] TIB decides which tool shall be used considering goals, formats and the designated community as well as software availability and capability.[129] In case an error occurs, the file appears in the "technical issues" section. The possibilities to handle the files can be to rerun the process, revalidate the file or abort of the process depending on the error.[130]

After a successful migration, the staff can start evaluation. As far as possible, criteria are evaluated by the system. This comprises "source vs target" evaluation, "target characteristics" and "file comparison plugin". Manual evaluation is possible as the migrated files can be downloaded (if they were migrated by the system). A detailed report of the preservation plan can be downloaded.[131] The files of the test set are only saved temporarily and are no longer available after completing evaluation of a plan.

At this stage it is possible to refine the preservation plan and test it again, with other settings or another migration tool. When the preservation staff decides that a migration plan shall be executable it can be signed off. Only a signed-off plan can migrate the files permanently and thus create a new version of an AIP.[132]

Rosetta's Preservation Planning follows along the same four steps as described in chapter 4.3: Defining the requirements and the preservation set/ records, evaluating the alternatives by running an experiment, consider the results by evaluating the results and signing off an exe-

[127] Ex Libris Documentation Department 2018, p. 105.
[128] Ex Libris Documentation Department 2018, p. 119.
[129] Bähr and Schwab 2017a.
[130] Ex Libris Documentation Department 2018, p. 177.
[131] Ex Libris Documentation Department 2018, pp. 124–126.
[132] Bähr and Schwab 2017a.

cutable preservation plan. The process of preservation planning has been conducted multiple times for testing purposes during the development of the migration plugin described in chapter 8. After this introduction into the theory and practice of preservation planning the following chapter focuses on the first research question.

6 Attributes of Suitable Formats

The first research questions aims at examining TIB's holdings of AV-material in regards to the suitability for long-term preservation. On the one hand it is obvious that not all formats are preferred for preservation, on the other hand there cannot be a universally applicable "right" format for digital preservation.[133] Many attributes of formats as well as exterior factors (e.g. storage space, available tools) play a role in weighting formats for long-term preservation within an archive. [134] In the following chapter the criteria for the evaluation of a file format are introduced. The most common formats in TIB's AV holdings are assessed by means of these criteria. In 6.3" Classification of mkv/ffv1/PCM as Preferred Archival Format" not only the file format criteria will be illuminated but also factors that played a role in the institutions decision for mkv/ffv1/PCM as the preferred format.

6.1 Criteria for Suitability as Archival Format

Todd (2009) compares different approaches for judging if a file format is suitable for preservation in the Digital Preservation Coalitions Technology Watch Series Report in 2009. He combines amongst others publications from Arms & Fleischhauer (2005), Rog and van Wijk (2008) and McLellan (2007).

The summarizing report indicates five main selection criteria: adoption, platform independence, disclosure or documentation, transparency, and metadata support.[135] I will therefore focus on these five criteria, as they give enough insights in order to evaluate the suitability. If one of the investigated file formats meets all the five criteria, other criteria including reusability, robustness, stability and rights management would be analyzed.[136] The weighting of the criteria depend on the archive, and is therefore described in chapter 6.2 "Classification of TIB's holdings".

---

[133] Addis et al. 2010, p. 5.
[134] Addis et al. 2010, p. 5
[135] Todd 2009, p. 13.
[136] Todd 2009, p. 10.

Adoption
A file format has a high degree of adoption, if it is frequently used in the designated community.[137, 138] A widespread format is less likely to become obsolete, tools for validation and migration are probably developed for a wider market and not only for archival use.[139]
Favorable
• accepted as preferred format by other archiving institutions [140] • different tools for the playback of the file format are available[141] • tools for migration are available and free of charge
Acceptable
• different tools for the playback of the file format are available[142] • tools for migration are available and free of charge
Critical
• no tool for manipulation or migration available • available tools for manipulation or migration are subject to charge

[137] Arms et al.
[138] Arms and Fleischhauer 2005, p. 3.
[139] Arms and Fleischhauer 2005, p. 3.
[140] Arms et al.
[141] Arms et al.
[142] Arms et al.

Platform independence
A format can be dependent on a platform or other external resources in different ways, regarding particular hardware, operating system, or software libraries. [143] A file format with more external dependencies is more at risk, due to obsolete hardware or external resources that are no longer available. [144]
Favorable
• may be rendered and saved on different kinds of recent hardware (standard office hardware e.g. flash drive, hard drive, Linear Tape Open (LTO)) • playable on different recent operating systems (Windows, Macintosh Operating System (MacOS)) • native browser-support of the format
Acceptable
• may be rendered and saved on different kinds of recent hardware (e.g. flash drive, hard drive, LTO) • playable on different recent operating systems (Windows, MacOS)
Critical
• depending on hardware which is no longer produced • depending on operating system which is no longer maintained

[143] Arms and Fleischhauer 2005, p. 4.
[144] Rog and van Wijk 2008, p. 4.

Transparency
A file format is called transparent when it is possible to analyze the content with basic tools. [145] Transparency describes a basic, ideally uncompressed or lossless compressed, datastream structure as opposed to a lossy / storage-optimized stream.[146] One must consider that born-digital files might compress the information upon creation, so that uncompressed information never existed.[147] Nevertheless they should be migrated into a lossless compressed or uncompressed format in order to mitigate the risk of information loss.[148] . Due to the amount of information in audio-visual material lossless compression is accepted in the preservation community. Internal validity and integrity checks add to the inspectability of a format.[149] Encryption and other digital rights management methods impose additional dependencies which hinder transparency. [150] Encryption must be evaluated per file and cannot be evaluated on file format level, because it is optional and its use depends on the producer of the file.
Favorable
• uncompressed files/codec OR • lossless compressed files/codec with internal integrity checks
Acceptable
• lossless compressed files/codec
Critical
• lossy compressed files/codec

[145] Arms and Fleischhauer 2005, p. 3.
[146] Arms and Fleischhauer 2005, p. 3.
[147] Arms et al.
[148] Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen et al. 2017d.
[149] Todd 2009, p. 18.
[150] Todd 2009, p. 18.

Standard / Disclosure / Documentation
A file format is disclosed if its complete specification is available.[151] Open documentation will foster the development of validation tools.[152] Furthermore a documented structure of the file format helps reduce the necessary investment while preservation planning.[153] Non-proprietary formats with published specifications are recommended by most institutions. Proprietary formats with published specifications are accepted.[154] Standardization is not a must but acts as an indication that the development of the file format is managed through an official body and is stable.[155] Besides the documentation of the file format, there might be patents on (parts of) file formats.[156] This might slow down the development of open source decoders or lead to license fees for migration software. [157]
Favorable
• non-proprietary format[158] • validation tool available[159] • open documentation OR standard available without charge
Acceptable
• standard / documentation with charge OR proprietary format with published documentations[160]
Critical
• no official documentation available • patents on (part of) a format

[151] Arms and Fleischhauer 2005, p. 3.
[152] Arms and Fleischhauer 2005, p. 3.
[153] Rog and van Wijk 2008, p. 3.
[154] Evelyn Peters McLellan 2006.
[155] Todd 2009, p. 14.
[156] Arms et al.
[157] Arms and Fleischhauer 2005, p. 4.
[158] Evelyn Peters McLellan 2006
[159] Arms and Fleischhauer 2005, p. 3
[160] Evelyn Peters McLellan 2006

Metadata Support
Metadata support describes the possibility to store metadata within the file format rather than in a sidecar file. Metadata might include descriptive, but also technical and administrative metadata.[161] Metadata support simplifies management and monitoring of the files, as well as their usage.[162] A file format suitable for preservation must contain representation information, which allows the correct rendering of the file. The possibility to embed other metadata (at the time of the creation of the file) is beneficial for preservation purposes.[163] It is important to note that embedded metadata cannot replace metadata management within the archive.[164]
Favorable
• possibility to embed descriptive metadata[165] (Container) • fixity metadata embedded (for integrity checks)[166] • technical metadata embedded[167]
Acceptable
• technical metadata embedded[168]
Critical
• sidecar file for representation information

[161] Arms and Fleischhauer 2005, p. 4
[162] Arms et al.
[163] Arms et al.
[164] Todd 2009, p. 19.
[165] Arms et al.
[166] Arms et al.
[167] Arms et al.
[168] Arms et al.

6.2 Classification of TIB's holdings

For each of the criteria information must be collected, rated and an overall decision must be made. The balancing of the importance of the different criteria depends on the local preservation strategy or policy of the archive[169] and on the designated community[170]. Both should be under review regularly. [171] In this case the evaluation must not result in a ranking, but shall illustrate if a format can be recommended for digital preservation. Therefore the weighting is simple. If one of the five main criteria is to be judged "critical", the format is not recommended – no matter how the other criteria are rated. On all five formats, the lowest rating determines the overall weighting of the format. Should all criteria be acceptable or favorable for one format, other factors like reusability, robustness, complexity etc. should be examined. [172] A format is classified as "critical" if one of elaborated critical criterion from chapter 6.1 applies. In order to be classified as "acceptable" or "favorable" all criteria must apply (exceptions are marked.)

A good starting point for information on a file format is the file format registry PRONOM , where each format is given a Pronom Unique Identifier (PUID)[173]. Format identification tools like Digital Record Object Identification (DROID) use the information from Pronom and report the file format in terms of a PUID[174]. PRONOM offers information on container formats for AV material, but not necessarily information on video codecs, rendering software, documentation etc.[175] Therefore the DROID analysis of the examined files only reports a PUID for the container. MediaInfo was used in order to extract information on video and audio codec. Both tools are recommended for AV file identification.[176]

The Library of Congress' (LoC) information page on file formats cover most of the criteria, only the metadata support is not explicitly specified but can be found in the general description of the format if applicable[177].

---

[169] Todd 2009, p. 33.
[170] Arms and Fleischhauer 2005, p. 6.
[171] Todd 2009, p. 33
[172] Todd 2009, p. 10
[173] The National Archives 2014c.
[174] The National Archives 2014a.
[175] Brown 2007, p. 10.
[176] Houpert et al. 2015, p. 23.
[177] Arms et al.

The following Tables 2 - 6 give an overview on the measurements for each of the criteria. The measurements are based on TIB's preservation policy and designated community.

Table 2 Measurement Adoption

Adoption	Measurement
Favorable	
• accepted as preferred format by other archiving institutions [178] • different tools for the playback of the file format are available[179] • tools for migration are available and free of charge	• LoC[180], British Film Institute[181], network of expertise in long-term storage and availability of digital resources in germany (nestor) AG Media[182], KOST[183] • Examples Windows Media Player, VideoLAN Client (VLC), MPC-HC (see chapter 7.2) • The tool Fast Forward MPEG (FFmpeg) supports decoding[184]
Acceptable	
• different tools for the creation and manipulation of the file format are available[185] • tools for migration are available and free of charge	• Examples Windows Media Player, VLC, Media Player Classic – Home Cinema (MPC-HC, see chapter 7.2) • Ffmpeg supports decoding[186]
Critical	
• no tool for manipulation or migration available • available tools for manipulation or migration are subject to charge	• examined only if acceptable criteria do not apply

[178] Arms et al.
[179] Arms et al.
[180] Library of Congress 2017.
[181] British Film Institute 2011.
[182] Barteleit et al. 2016.
[183] Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen et al. 2017d.
[184] FFmpeg 2018a.
[185] Arms et al.
[186] FFmpeg 2018a

Table 3 Measurement Platform Independence

Platform independence	Measurement
Favorable	
• may be rendered and saved on different kinds of recent hardware • playable on different recent operating systems (Windows, MACOS) • native browser-support of the format	• tested on local hard drive and solid state drive (SSD) • availability of player software for different operating system, like VLC Media Player[187] • supported by Chrome and Firefox[188]
Acceptable	
• may be rendered and saved on different kinds of recent hardware • playable on different recent operating systems (Windows, MACOS)	• tested on local hard drive and SSD • availability of player software for different operating system, like VLC Media Player[189]
Critical	
• depending on hardware which is no longer produced • depending on operating system which is no longer maintained	• examined only if acceptable criteria do not apply

[187] videoLAN 2018.
[188] Mozilla 2018.
[189] videoLAN 2018

Table 4 Measurement Transparency

Transparency	Measurement
Favorable	
• uncompressed files/codec OR • lossless compressed files/codec with internal integrity checks	• information from the LoC[190] or Wikipedia[191],[192] • information from the format's documentation
Acceptable	
• lossless compressed files/codec	
Critical	
• lossy compressed files/codec	

Table 5 Measurement Standard / Disclosure / Documentation

Standard / Disclosure / Documentation	Measurement
Favorable	
• validation tool available[193] • non-proprietary format[194] • open documentation OR standard available without charge	• Tool available according to Community Owned digital Preservation Tools Registry (COPTR)[195] • information from the LoC[196] • other documentation
Acceptable	
• standard / documentation with charge OR proprietary format with published documentations[197]	
Critical	
• no official documentation available • patents on (part of) a format	

[190] Library of Congress 2017
[191] Wikipedia 2018b.
[192] Wikipedia 2018a.
[193] Arms and Fleischhauer 2005, p. 3
[194] Evelyn Peters McLellan 2006
[195] COPTR contributors 2018.
[196] Library of Congress 2017
[197] Evelyn Peters McLellan 2006

Table 6 Measurement Metadata Support

Metadata Support	
Favorable	
• possibility to embed descriptive metadata[198] (Container) • fixity metadata embedded (for integrity checks)[199] • technical metadata embedded[200]	• information from the LoC[201] and other documentation (search terms: "metadata", "title", "year") • information from the LoC, section self-documentation[202] or other documentation (search terms: "fixity", "checksum","CRC"for Cyclic Redundancy Check, "MD5" for Message-Digest Algorithm 5, "SHA" for Secure Hash Algorithm) • extraction of technical metadata with MediaInfo[203] is possible
Acceptable	
• technical metadata embedded[204]	• extraction of technical metadata with MediaInfo[205] is possible
Critical	
• sidecar file for representation information	• examined only if acceptable criteria do not apply

After defining measurements for each weighting, each container, video codec and audio codec was analyzed individually in Table 7 "Rating of Formats - Container", Table 8 "Rating of Formats - Video Codecs" and Table 9 "Rating of Formats - Audio Codecs". As mentioned in chapter 3.2 Audio-Visual Holdings at TIB the investigated three formats constitute more than

[198] Arms et al.
[199] Arms et al.
[200] Arms et al.
[201] Library of Congress 2017
[202] Arms et al.
[203] MediaArea.
[204] Arms et al.
[205] MediaArea

70 % of TIB's AV holdings. A quick overview of the results can be found in Table 10 "Classification on TIB's AV holdings".

Table 7 Rating of Formats - Container

Container	
	Container: MPEG-4 (PUID: fmt/199[206])
Adoption	Favorable: different tools for playback are available and tools for migration are free of charge, preferred format by KOST[207]
Platform Independence	Favorable: Playback from different hardware and operating systems works, native browser support[208]
Transparency	Not rated – no compression on container level
Standard / Disclosure / Documentation	Acceptable: Standard from the International Organization for Standardization / International Electrotechnical Commission (ISO/IEC) 14496:14 - MP4 File Format is available for charge[209]
Metadata Support	Acceptable: Technical metadata can be extracted Although embedding of descriptive metadata is supported it cannot be rated favorable because no information on embedded fixity information referring LoC[210].
	Container: WebM (PUID: fmt/573[211])
Adoption	Acceptable: different tools for playback are available (see chapter 7.2) and tools for migration are free of charge cannot be marked favorable because it is not a preferred format for other archiving institutions

[206] The National Archives 2010.
[207] Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen et al. 2017c.
[208] Mozilla 2018.
[209] International Organization for Standardization 2003.
[210] Arms et al. 2017f.
[211] The National Archives 2013.

Platform Independence	Favorable: Playback from different hardware and operating systems works, native browser support[212]
Transparency	Not rated – no compression on container level
Standard / Disclosure / Documentation	Favorable: Non-proprietary format with open documentation[213], a validation tool is available[214]
Metadata Support	Acceptable: Technical metadata can be extracted Although embedding of descriptive metadata is supported it cannot be rated favorable because no information on embedded fixity information[215]
	Container: MPEG-PS (PUID: x-fmt/385[216])
Adoption	Favorable: different tools for playback are available and tools for migration are free of charge, conditionally preferred format by KOST[217]
Platform Independence	Acceptable: Playback from different hardware and operating systems works cannot be marked favorable because no native browser support[218]
Transparency	Not rated – no compression on container level
Standard / Disclosure / Documentation	Acceptable Standard ISO/IEC 11172-1 - System is available for charge[219]
Metadata Support	Acceptable: Technical metadata can be extracted Descriptive metadata and fixity metadata cannot be embedded[220]

[212] Mozilla 2018.
[213] The WebM Project 2017.
[214] Bunkus 2018.
[215] webm wiki 2012.
[216] The National Archives 2014b.
[217] Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen et al. 2017b.
[218] Mozilla 2018.
[219] International Organization for Standardization 1993a.
[220] Arms et al. 2017e.

The rating of video codecs can be found in the following table. An overall rating can be found at the end of this chapter.

Table 8 Rating of Formats - Video Codecs

Video Codecs	
	Video: AVC
Adoption	Acceptable: different tools for playback are available and tools for migration are free of charge cannot be marked favorable because it is not a preferred format for other archiving institutions
Platform Independence	Favorable: Playback from different hardware and operating systems works, native browser support[221]
Transparency	Critical: Although a lossless version exists[222], it is most likely that most of TIB's holdings are lossy compressed. A detailed analysis is necessary.
Standard / Disclosure / Documentation	Acceptable Although the standard ISO/IEC 14496:10 – Advanced Video Coding is free of charge, it is a proprietary format, as licenses cover the sale of software.[223]
Metadata Support	Acceptable: Technical metadata can be extracted cannot be rated favorable because no information on embedded fixity information referring LoC[224].
	Video: VP8
Adoption	Acceptable: different tools for playback are available and tools for migration are free of charge cannot be marked favorable because it is not a preferred format for other archiving institutions

[221] Mozilla 2018.
[222] Bubestinger-Steindl et al. 2013.
[223] Arms et al. 2017f.
[224] Arms et al. 2017f

Platform Independence	Favorable: Playback from different hardware and operating systems works, native browser support[225]
Transparency	Critical: Lossy compression[226]
Standard / Disclosure / Documentation	Acceptable: Non-proprietary format with open documentation[227], but no validation tool known to the author
Metadata Support	Acceptable: Technical metadata can be extracted no information on fixity metadata in the documentation[228]
	Video: MPEG Video, Format version : Version 2
Adoption	Favorable: different tools for playback are available and tools for migration are free of charge, preferred format by KOST[229] and conditionally preferred by LOC[230]
Platform Independence	Acceptable: Playback from different hardware and operating systems works cannot be marked favorable because no native browser support[231]
Transparency	Critical: Lossy compression[232]
Standard / Disclosure / Documentation	Acceptable Standard ISO/IEC 11172-2 - Video is available for charge[233]
Metadata Support	Acceptable: Technical metadata can be extracted fixity metadata cannot be embedded[234]

[225] Mozilla 2018.
[226] Wikipedia 2018b.
[227] Bankoski et al. 2011.
[228] Bankoski et al. 2011.
[229] Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen et al. 2017b.
[230] Arms et al. 2017e.
[231] Mozilla 2018.
[232] Arms et al. 2018.
[233] International Organization for Standardization 1993b.

After the rating of audio codecs in the following table the overall rating is evaluated.

Table 9 Rating of Formats - Audio Codecs

Audio Codecs	
	Audio: AAC, Version 4
Adoption	Favorable: supported by Windows Media Player, accepted format for born-digital files by LoC[235], different tools for playback are available and tools for migration are free of charge
Platform Independence	Favorable: Playback from different hardware and operating systems works, native browser support[236]
Transparency	Critical: Lossy compression[237]
Standard / Disclosure / Documentation	Acceptable Standard ISO/IEC 14496:3 - MP4 File Format is available for charge[238]
Metadata Support	Acceptable: Technical metadata can be extracted cannot be rated favorable because no information on embedded fixity information referring LoC[239].
	Audio: Vorbis
Adoption	Acceptable: different tools for playback are available and tools for migration are free of charge cannot be marked favorable because it is not a preferred format for other archiving institutions

[234] Arms et al. 2018.
[235] Arms et al. 2017a.
[236] Mozilla 2018.
[237] Wikipedia 2018a.
[238] International Organization for Standardization 2009.
[239] Arms et al. 2017f

Platform Independence	Favorable: Playback from different hardware and operating systems works, native browser support[240]
Transparency	Critical: Lossy compression[241]
Standard / Disclosure / Documentation	Acceptable: Non-proprietary format with open documentation[242], but no validation tool known to the author
Metadata Support	Acceptable: Technical metadata can be extracted cannot be rated favorable because no information on embedded fixity information according to the documentation[243]
	Audio: MPEG Audio, Format version : Version 1
Adoption	Favorable: accepted format for born-digital files by LOC[244], different tools for playback are available and tools for migration are free of charge
Platform Independence	Acceptable: Playback from different hardware and operating systems works cannot be marked favorable because no native browser support[245]
Transparency	Critical: Lossy compression[246]
Standard / Disclosure / Documentation	Acceptable: Standard ISO/IEC 11172-3 - Audio is available for charge[247]
Metadata Support	Acceptable: Technical metadata can be extracted fixity metadata cannot be embedded referring LoC[248]

[240] Mozilla 2018.
[241] Wikipedia 2018a.
[242] Xiph.Org Foundation 2015.
[243] Xiph.Org Foundation 2015.
[244] Arms et al. 2017e.
[245] Mozilla 2018.
[246] Arms et al. 2017d.
[247] International Organization for Standardization 1993c.

The general overview in Table 10 suggests that all container formats would be acceptable, whereas video- and audio-codecs are considered critical. But as the file format consists of container, video- and audio-codec, none of the file formats can be rated acceptable. The thesis 1 a) "*The majority of formats within the TIB AV-holdings are not widely recommended as preferred preservation formats."* is verified.

Table 10 Classification on TIB's AV holdings

Format	Adoption	Platform Independence	Transparency	Disclosure	Metadata Support	Overall
Container						
MPEG-4	++	++	o	+	+	+
WebM	+	++	o	++	+	+
MPEG-PS	++	+	o	+	+	+
Video-Codec						
AVC	+	++	-	+	+	-
VP8	+	++	-	+	+	-
MPEG Video, Version 2	++	+	-	+	+	-
Audio-Codec						
AAC, Version 4	++	++	-	+	+	-
Vorbis	+	++	-	+	+	-
MPEG Audio, Version 1, Layer 2	++	+	-	+	+	-
Legend: favorable: ++ acceptable: + critical - not rated: o						

[248] Arms et al. 2017d.

6.3 Classification of mkv/ffv1/PCM as Preferred Archival Format

While Pulse Core Modulation (PCM) is an established archival format for audio in the digital preservation community[249] , there are two commonly used archival formats for audio-visual material: JPEG2000 in MXF container or ffv1 in a Matroska container[250] . Ffv1 was adopted early as archival format by the Österreichische Mediathek[251]. During the EU funded Preforma project the tool MediaConch[252] was developed in order to validate ffv1 in a Matroska container[253]. Furthermore the standardization by the Internet Engineering Task Force (IETF) was initiated through this project. Today the format in adopted as archival format by memory institutions worldwide[254]. TIB collected experiences with the format during migration tests and decided to adopt ffv1 version 3 in a matroska container as archival format for the digitization of analogue films. All files produced within or for TIB are without DRM measures.

In order to compare the formats the five criteria which were elaborated in chapter 6.1 "Criteria for Suitability as Archival Format" are evaluated first in Table 11 Rating of Preferred Archival Format. Furthermore other criteria are evaluated, but they are not included into the weighting.

Table 11 Rating of Preferred Archival Format

Preferred Archival Format	
	Container: Matroska
Adoption	Favorable: different tools for playback are available and tools for migration are free of charge, preferred format by nestor AG Media[255]
Platform Independence	Acceptable: Playback from different hardware and operating systems works cannot be marked favorable because no native browser support[256]

[249] Justrell et al. 2017, p. 99.
[250] Lorrain 2014.
[251] Jaks 2018.
[252] MediaArea.
[253] Justrell et al. 2017, p. 97.
[254] Wikipedia 2018c.
[255] Barteleit et al. 2016, p. 20.
[256] Mozilla 2018.

Transparency	Not rated – no compression on container level
Standard / Disclosure / Documentation	Favorable: MediaConch as validation tool available[257], non-proprietary format[258], ongoing standardization by IETF[259]
Metadata Support	Favorable: Descriptive metadata, technical and fixity metadata is embedded[260]
	Video: ffv1, Version 3
Adoption	Favorable: different tools for playback are available and tools for migration are free of charge, preferred format by nestor AG Media[261] and KOST[262]
Platform Independence	Acceptable: Playback from different hardware and operating systems works cannot be marked favorable because no native browser support[263]
Transparency	Acceptable: Lossless compressed file, internal integrity checks,[264] Cannot be marked favorable because it is not uncompressed
Standard / Disclosure / Documentation	Favorable: MediaConch as validation tool available[265], non-proprietary format[266], ongoing standardization by IETF[267]
Metadata Support	Favorable: technical and fixity metadata (SliceCRC) is embedded[268]

[257] MediaArea.
[258] Arms et al. 2017c.
[259] Bunkus et al. 2018.
[260] Bunkus et al. 2018.
[261] Barteleit et al. 2016, p. 20
[262] Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen et al. 2017a.
[263] Mozilla 2018.
[264] Arms et al. 2017b.
[265] MediaArea
[266] Arms et al. 2017b.
[267] Rice et al. 2018.
[268] Rice et al. 2018.

	Audio: PCM
Adoption	Favorable: different tools for playback are available and tools for migration are free of charge, preferred format by nestor AG Media[269]
Platform Independence	Favorable: Playback from different hardware and operating systems works, native browser support[270]
Transparency	Favorable: uncompressed codec[271]
Standard / Disclosure / Documentation	Favorable: MediaConch as validation tool available[272], non-proprietary format with open documentation[273]
Metadata Support	Acceptable: Technical metadata can be extracted fixity metadata cannot be embedded referring LoC[274]

The weighting is summarized in Table 12. The overall weighting is acceptable for the container, video codec and audio codec. Therefore the format is judged acceptable. This means that a migration from the existing formats, which are judged critical (see previous chapter 6.2), to the acceptable format can be considered as a preservation action (see chapter 4.4).

[269] Barteleit et al. 2016, p. 67.
[270] Mozilla 2018.
[271] Arms et al. 2010.
[272] MediaArea
[273] Arms et al. 2010.
[274] Arms et al. 2010.

Table 12 Classification of Preferred Archival Format

Format	Adoption	Platform independence	Transparency	Disclosure	Metadata support	overall
Container						
Matroska	++	+	o	++	++	+
Video-Codec						
Ffv1, Version 3	++	+	+	++	++	+
Audio-Codec						
PCM	++	++	++	++	+	+
Legend: favorable: ++ acceptable: + critical - not rated: o						

In order to have a more complete picture of the format, other criteria are considered. They include reusability, robustness, stability and rights management.[275]

Reusability is a given: one the one hand the format has the advantage that playback is possible with open source tools like ffmpeg[276]. And even though it might not be supported by commonly used cutting tools in the moving image industry[277] the drawback is not important to TIB: the designated community does not include the moving image industry (see chapter 5.1 “Preservation Policy”). If there should be a change in the designated community, ffv1 can be converted to uncompressed (as there was no (additional) information loss during the migration to ffv1) and from uncompressed to any other suitable format.

Ffv1 Version 3 as well as matroska come with fixity metadata which foster the robustness of the format. While the container matroska has one included checksum which covers the whole content[278], ffv1 in version 3 works with checksums on a more granular level. During the migration to ffv1 one can determine an amount of slices. Each frame is divided into the defined amount of slices e.g. 16. For each of these slices a CRC checksum is created and embedded in the file. This enables not only fixity checks on a very granular level, but also a possibility

[275] Todd 2009, p. 10.
[276] FFmpeg 2018a.
[277] Jaks 2018.
[278] Bunkus et al. 2018.

to repair a slice if a bit flip (a change from 1 to 0 or vice versa on bit stream level) occurred[279].

Stability is describes as backward compatibility as well as a managed release cycle for new format versions[280] . Both are most important for proprietary formats.[281] . During the standardization of both ffv1 and matroska backwards compatibility is ensured[282] . The release cycle of new version can be neglected as they are not proprietary formats.

Rights management covers the ability to include rights metadata in the format[283] which is a given in matroska[284]. On the other hand protection like digital rights management (DRM) hinders preservation[285] when files are acquired by a third party. It is possible to embedd DRM into matroska[286]. As long as TIB does not receive matroska files from the provider but generates them as a new version of the preservation master DRM is not used and thus does not impose drawbacks.

Todd (2009) also defines three "absent" criteria: costs, extent and ability to represent the full content". The costs for migration are not listed in detail but are comparable to the migration into other formats, maybe even less. This is because open source tools for migration and validation already exist[287], which reduces the costs for development or licensing, and the format's documentation is available free of charge, which reduces the costs for buying standards. The extent of the files is another criteria which contributes to the decision if a migration is necessary. The migration from lossy compressed codecs to the lossless compressed ffv1 version 3 leads to files which are multiple times as big as the original preservation master as seen in tests with TIB's holdings. The migration therefore leads to a permanent higher costs for storage space. As seen in chapter 4.4 "Migration" this must be taken into account, and thus not only the judgment of the current formats are important, but also the risk of obsolescence as described in chapter 7 "Attributes of Obsolete Formats". The ability to represent the full content is a combination of the embedded metadata and the complexity or transparency

[279] MediaArea.
[280] Todd 2009, p. 15.
[281] Todd 2009, p. 15.
[282] Richardson et al. 2018.
[283] Todd 2009, p. 15.
[284] Bunkus et al. 2018.
[285] Todd 2009, p. 15.
[286] Bunkus et al. 2018.
[287] Justrell et al. 2017, p. 97.

of the format amended by a discussion about representation information in general.[288]. Ffv1 Version 3 in a matroska container includes sufficient technical metadata to allow playback with current software or the retrieval of current playback software.

The classification as seen in Table 12 Classification of Preferred Archival Format as well the evaluation of the other criterions reveal that mkv/ffv1/PCM is a suitable archival format. For the second research question concerning the migration from TIB's AV holdings it is set as target format (see chapter 8 "Migration Plugin").

7 Attributes of Obsolete Formats

In the previous chapter criteria for archival suitability were examined. But if a format is not preferred for preservation, this does not mean that preservation action is needed immediately. A file format can have the following three conditions:

- preferred for preservation
- not preferred for preservation, not obsolete
- not preferred for preservation, obsolete

Only in the last case preservation action must be performed for loss of the content to be prevented.[289] Regardless of the conditions of a file format testing, the possibilities for migration, like the development of a migration plugin (chapter 8), is part of preservation planning as described in chapter 4.

In the literature two terms are used to describe a file format, which is at risk of becoming inaccessible[290]: obsolescence and endangerment. Obsolescence can be defined as "the phenomenon that occurs when information stored in a particular file format is no longer accessible using current technology."[291] But there are broader definitions: "Obsolescence describes a state of becoming obsolete, rather than a state of already being obsolete."[292] This definition is very close to the definition of endangerment: "the possibility that information stored in a par-

[288] Todd 2009, p. 16.
[289] Bähr and Schwab 2017a.
[290] Pearson and Webb 2008, p. 94.
[291] Ryan 2014, p. 2.
[292] Pearson and Webb 2008, p. 93.

ticular file format will not be interpretable or renderable using standard methods within a certain timeframe."[293]

Hereinafter I will use the term obsolete to describe a *file format that is at risk to become inaccessible by our designated community*.

Evaluating if a file format is obsolete therefore means to examine how difficult it is to access the content.[294, 295]

Defining if a file format is obsolete can differ based on expertise and organizational conditions and circumstances. [296,297] Even if a preservation action is recommended, one person might schedule the preservation action immediately, while another might schedule it at a later moment.[298] Research has developed different frameworks and tools in order to measure file format obsolescence or the risk that comes with a file format.

A workbook has been developed by Lawrence et al. (2000). It covers not only the risk of an obsolete file format but also an organizational risk concerning the archive, and the risk during migration.[299] Another approach is the Index for Risk Management (INFORM) methodology by Stanescu (2005)[300], where the risks are more granular subdivided into file format risk, software risk and hardware risk, and associated organizations of software and hardware. Organizational risks and migration risks are considered as well.[301] Using this methodology an organization shall develop a questionnaire which has to be answered by a group of experts.[302] Reviewing their answers enables an informed preservation action.[303] Risk analysis with both methods would involve a team of specialists[304], but due to limited resources conducting expert interviews is not feasible within the scope of this work.

---

[293] Ryan 2014, p. 3.
[294] Ryan 2014, p. 211.
[295] Pearson and Webb 2008, p. 94.
[296] Stanescu 2005, p. 74.
[297] Stanescu 2005, p. 74.
[298] Stanescu 2005, p. 72.
[299] Lawrence et al. 2000, p. 6.
[300] Stanescu 2005, p. 62.
[301] Stanescu 2005, p. 64.
[302] Stanescu 2005, p. 68.
[303] Stanescu 2005, p. 63.
[304] Lawrence et al. 2000, p. 6.

In order to collect information from experts, and make them generally available, attempts have been made to automate risk assessment for file format obsolescence. The project Automatic Obsolescence Notification System (AONS) II from the National Library of Australia and the Australian Partnership for Sustainable Repositories was carried out in 2007. It "aimed to refine and develop a software tool that would automatically find and report indicators of obsolescence risks".[305] Unfortunately the tool is no longer available.[306] In a European context the Digital Preservation Recommender System (DiPRec) was developed as part of the Assets Project (2010-2012). DiPRec collected information on file formats through linked open data in order to facilitate risk analysis.[307] This service is no longer available as well.[308] Both tools relied on the information of one or more file format registries for digital preservation purposes. Until today the most common file format registry in the digital preservation community is PRONOM, by the national archives.[309] While there is the possibility to provide information on format risks, technical environment, etc. [310], this is missing for most registered formats. An automated risk analysis thus is not possible.

Ryan (2014) took another approach. She comes to the conclusion that file formats can be compared to species. She therefore applied methods used to analyze species extinction in order to measure file format obsolescence. [311] She examines 21 factors in several expert interviews.[312] Her findings suggest that a lack of rendering software is the only factor that puts a file format at risk. [313] She suggests that the factors "specifications available" and "community / 3rd party support" (adoption) should be taken into account when no rendering software is available.[314] In the following chapters I will therefore focus on the factor if rendering software is available.

It is not only criteria on a file format level that must be taken into considerations, but also the validity of each file. To validate a file means to examine if it follows the format specifica-

[305] Pearson and Webb 2008, p. 89.
[306] See: http://apsr.anu.edu.au/aons2/index.htm
[307] Gordea et al., p. 51.
[308] http://www.assets4europeana.eu/
[309] The National Archives 2004.
[310] The National Archives 2004.
[311] Ryan 2014, p. 54.
[312] Ryan 2014, pp. 74–75.
[313] Ryan 2014, p. 210.
[314] Ryan 2014, p. 210

tions. [315] A file can be well-formed (syntactically correct) and valid (semantically correct)[316], it can be well-formed but not valid, or it can be not well-formed. Ryan does not take the validity of a single file into account. She examines if the ease of validation is an important factor for file format obsolescence, and comes to the conclusion that a file format can be obsolete regardless of whether it is easy to validate a file in this format or not.[317]

Nevertheless, an invalid file holds the risk that migration might lead to errors. Unfortunately, few validation tools for AV material are available, and none for the majority of file formats in TIB's archive.[318] Additionally the ability to render of a file cannot act as an indicator for validity, as the playback software for AV material is very tolerant regarding invalid files.[319] The validation of files in TIB's holdings is due to missing tools out of scope.

7.1 Rendering Software Available

In order to exclude obsolescence it is not enough that rendering software exists. Software must be available[320] and reproduce the digital object authentically, which means "able to be used in a way that retains the object's significant characteristics".[321] A film with a ratio of 4:3 must not be stretched to 16:9 during playback (if the user does not want it). Furthermore the software should be used by or available to the designated community[322].

Evaluating if rendering software is available can be achieved by determining the view-paths for a file format. A view-path consists of a technical description of how a file format can be rendered. It holds information on the hardware platform, operating system, and the viewer application. [323] A view-path must be available in the institution as well as for the designated community.[324]

[315] Brown 2007, p. 7.
[316] Brown 2007, p. 7
[317] Ryan 2014, p. 146.
[318] Houpert et al. 2015, p. 28.
[319] Houpert et al. 2015, p. 28.
[320] Ryan 2014, p. 14.
[321] Vermaaten et al. 2012.
[322] Graf and Gordea 2014, p. 3.
[323] Steenbakkers 2005.
[324] Pearson and Webb 2008, p. 100.

The National Library of the Netherlands (KB) guarantees at least two different view-paths for their holdings.[325] Following their example I want to provide at least two different view-paths for the majority of the AV file formats in TIB's archive.

7.2 Classification of TIB's Holdings

In order to test the renderability of the file formats two files of each examined file formats were picked randomly from the holdings. Following the example from the KB "Intel Pentium […] NT […] Acrobat Reader 3.0"[326] I noted the Computer Processing Unit (CPU) as hardware platform.

For each view path – file combination several aspects were tested:

- Is the video displayed correctly?
- Is the audio played correctly?
- Is skipping to a later frame possible?
- Is pausing possible?

If one of these aspects was not fulfilled in the combination of hardware, operation system and software, then there is no valid view path. E.g. the Windows Media Player did not allow skipping to a later frame when displaying WebM / VP8 / Vorbis. The combination was not examined further. The software MPC-HC fulfilled all of the aspects and represents thus a valid view path for this container-codec-combination. A more detailed overview of the tested view paths can be found in Appendix C on page F.

Table 13 displays the examined, valid view paths for the file formats. They are written in the order 1. hardware platform (CPU) – 2. operating system – 3. viewer application. In order to have two independent view paths for each format, none of the three elements of a view path should be doubled, e.g. when one view path has Windows 10 as operating system, the other one must consist of another operating system. All of the hardware platforms, operating system and viewer applications must be available within TIB and our designated community.

[325] Steenbakkers 2005
[326] Steenbakkers 2005.

Table 13 View Paths for TIB's AV holdings

View Paths	
File Format	Container: MPEG-4 Video: AVC Audio: AAC, Version 4
View Path	Intel® Core™ i7-8700 – Windows 10, Version 1803 – Windows Media Player Version 12
View Path	Intel® Core™ i5-3470 – Windows 8.1 Pro, Version 6.3 – VLC Media Player Version 2.2.2
File Format	Container: WebM Video: VP8 Audio: Vorbis
View Path	Intel® Core™ i7-8700 – Windows 10, Version 1803 – MPC-HC Version 1.7.13
View Path	Intel® Core™ i5-3470 – Windows 8.1 Pro, Version 6.3 – VLC Media Player Version 2.2.2
File Format	Container: MPEG-PS Video: MPEG Video, Format version : Version 2 Audio: MPEG Audio, Format version : Version 1
View Path	Intel® Core™ i7-8700 – Windows 10, Version 1803 – Windows Media Player Version 12
View Path	Intel® Core™ i5-3470 – Windows 8.1 Pro, Version 6.3 – VLC Media Player Version 2.2.2

Rosetta offers the possibility to describe software on the level of software versions, including their dependence on operating systems. Within the so called application library it is possible to

link the formats to render software. [327] This can be used as possibility to save the view paths within the archive software. As it is possible to add local fields to the application library[328] ,

[327] Ex Libris Documentation Department 2018, p. 47.
[328] Ex Libris Documentation Department 2018, p. 47.

one can easily store the value for hardware platforms (CPU) and operating system as well (see Figure 7 Application Library with Local Fields for View Path).

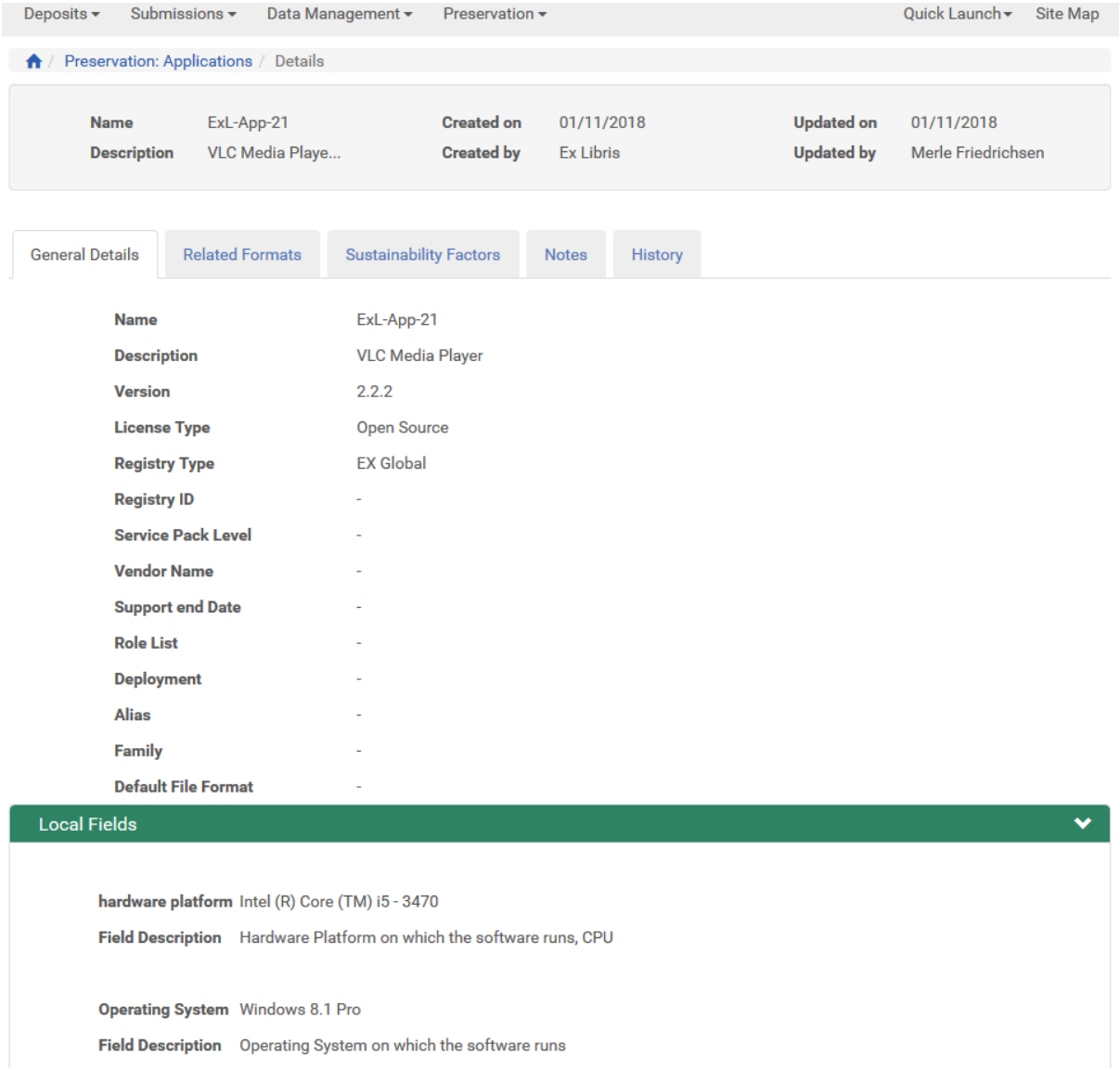


Figure 7 Application Library with Local Fields for View Path

For each format two valid view paths have been successfully examined, and thus the formats are not obsolete. The thesis *1. b) The majority of formats within the TIB AV-holdings are not obsolete.* is verified.

8 Migration Plugin

The obsolescence risk for TIB's digital AV material was evaluated in chapter 7.2. Although the majority of formats in TIB's holdings are not obsolete, it is worth testing the possibilities

to migrate the content into a suitable format. The format matroska with ffv1 and PCM turned out to be a suitable archival format as derived in chapter 6.3. Rosetta offers a preservation planning module, as described in chapter 5.2. A plugin which can convert the majority of TIB's AV holdings into the preferred archival format (mkv/ffv1/PCM) is not available. Migration can be performed externally, as Rosetta offers the possibility to export a given set of files, and import the migrated equivalents.[329] Therefore the user which executes the migration needs explicit knowledge on the migration software.

The purpose of the development of the migration plugin is to interact with the migration software in a determined way. The plugin can be used by authorized users of TIB's digital preservation team, without requiring knowledge of the migration software. E.g. the user does not need to know the exact migration software command to migrate a video, because this is coded in the plugin (see Figure 8 Screenshot: Choose Internal Plugin)

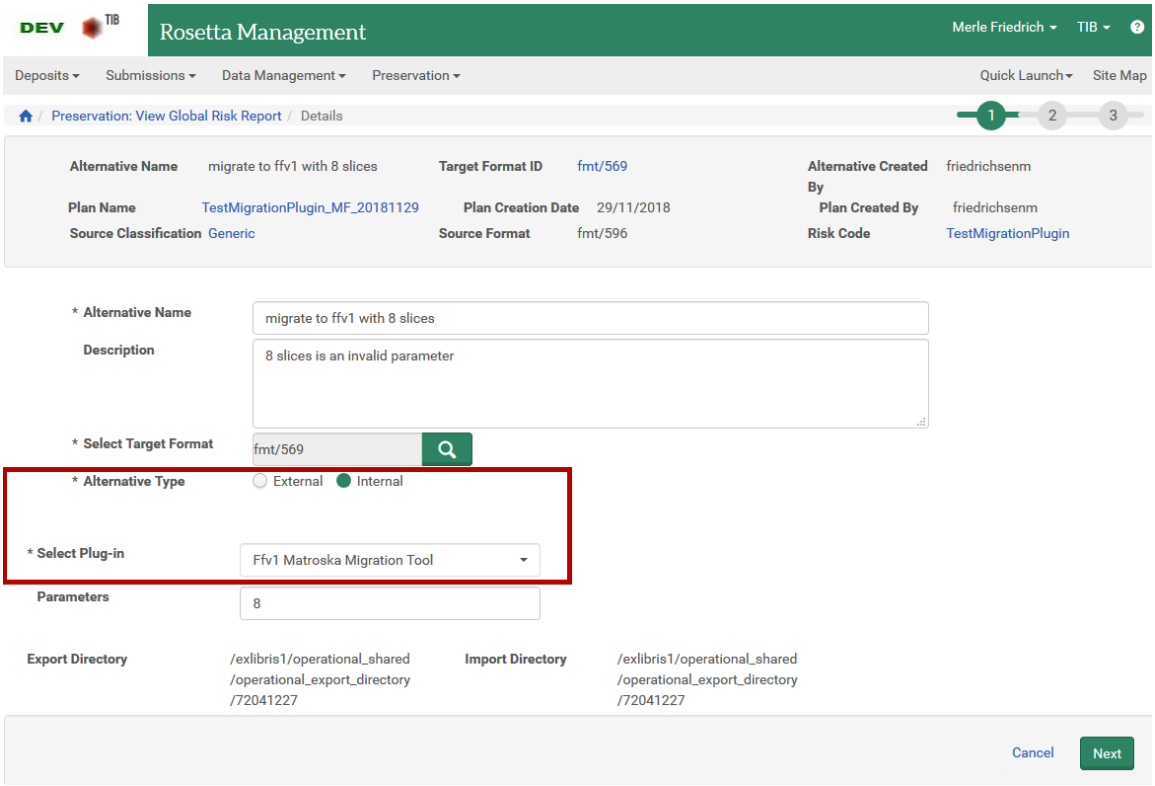


Figure 8 Screenshot: Choose Internal Plugin

[329] Ex Libris Documentation Department 2018.

The plugin must convert from the input formats, which represent the majority of TIB's digital AV holdings, to the given output format as seen in Table 14. The exact parameters for the migration can be found in the requirement analysis, chapter 8.2.

Table 14 Input and Output Format for Migration Plugin

	Container	Video Codec	Audio Codec
Input	MPEG-4 (mp4)	AVC	AAC, Version 4
	WebM (webm)	VP8	Vorbis
	MPEG-PS	MPEG Video, Version 2	MPEG Audio, Version 1
Output	Matroska (mkv)	ffv1, version 3	PCM

The development of the migration plugin also includes testing as described in chapter 8.4 "Evaluation of the Requirements ". The expendability of the plugin as well as the dissemination is discussed in chapter 9.2 "Review of the Development of the Plugin".

8.1 Prerequisites

In order to analyze the requirements for the migration plugin, two fields have to be assessed: the integration from plugins within Rosetta and tools which are able to perform a migration to the preferred archival format and can be embedded into the plugin environment.

Rosetta offers two types of migration plugins: a java plugin or a script plugin. [330] In addition to the Software Development Kit[331,332] different plugins[333, 334, 335] were reviewed regarding their structure and handling of custom parameter. As there are more examples for script plugins it was decided to develop a script plugin.

Regarding the migration tool, tests had already been conducted. For the transcoding of a file to mkv/ffv1/PCM TIB uses FFmpeg. FFmpeg is a command line tool which is commonly used for working with audio-visual material. When migrating to ffv1 Version 3 with FFmpeg

[330] ExLibris - Rosetta 2018a.
[331] ExLibris - Rosetta 2018b.
[332] ExLibris - Rosetta 2018a.
[333] National Library of New Zealand.
[334] National Library of New Zealand.
[335] Ott.

the format offers several options. These options will be introduced, and used in chapter 8.2 ”Analysis of Requirements for Migration Plugin”. The ffv1 version 3 is chosen with the command “-level 3”[336]. The number of frames which are compressed in a group of pictures (GOP) is defined with the command “-g 1”[337]. For archival purposes a GOP of 1 is recommended, this means that each frame is compressed individually, and does not rely on information of a prior or following frame[338]. There are different algorithms for compression which can be chosen with the option “-coder”[339]. Some recommendation for archival purposes rely on the Range coder with the command “-coder 1”[340]. The number of slices is set with “-slices”, and the option with CRC per slice is confirmed with “-slicecrc 1”[341]. The audio codec is chosen with “-c:a pcm_s24le” which means Pulse Code Modulation with signed bit, 24 bit bit depth and little endian[342]. Previous tests confirmed that these options to lead to an adequate archival format, which is also the selected file format as preservation master for digitized films at TIB.[343]

As FFmpeg is distributed under a GNU Lesser General Public License[344] using this tool does not require the payment of license fees. Moreover FFmpeg runs on a multitude of system environments which will facilitate the future use of the plugin by other institutions.

After assessing the prerequisites it was possible to formulate the requirements.

8.2 Analysis of Requirements for Migration Plugin

The formulation of the requirements is based on “Institute of Electrical and Electronics Engineers (1998)”. A requirement should be required and must not contradict external documentation or software specifications,[345] or internal requirements.[346]In order to be complete, a requirement must name all possible input parameters and external requirements.[347] It is im-

[336] FFmpeg Wiki 2014.
[337] FFmpeg Wiki 2014.
[338] FFmpeg Wiki 2014.
[339] FFmpeg Wiki 2014.
[340] Kromer 2019.
[341] FFmpeg Wiki 2014.
[342] FFmpeg Wiki 2012.
[343] Reiche 2018.
[344] FFmpeg 2018b.
[345] Institute of Electrical and Electronics Engineers 1998, p. 4.
[346] Institute of Electrical and Electronics Engineers 1998, p. 6.
[347] Institute of Electrical and Electronics Engineers 1998, pp. 5–6.

portant to formulate a requirement clear and unambiguous.[348] Each requirement is weighted from essential (E) to conditional (C) to optional (O).[349] A requirement must be verifiable, or measurable. During the evaluation a column can be added in order to check if the requirement is met (yes or no).[350] In order to enhance the traceability, each requirement gets an identifier that is referred to throughout the document.[351] The identifier can be used to track all changes concerning one requirement, if they do occur during the project. This is relevant for the documentation, and enhances the possible modification of requirements.[352]

Requirements that emerge from software and hardware environment can be found in the section of external interfaces. As the migration plugin must fulfill the requirements of Rosetta, these requirements cannot be formulated solution independent. Only requirements concerning script plugins, not java plugins, are recorded in “external interfaces” section.

[348] Institute of Electrical and Electronics Engineers 1998, p. 4.
[349] Institute of Electrical and Electronics Engineers 1998, pp. 6–7.
[350] Institute of Electrical and Electronics Engineers 1998, p. 7.
[351] Institute of Electrical and Electronics Engineers 1998, p. 8.
[352] Institute of Electrical and Electronics Engineers 1998, p. 8.

ID	Requirement	External documentation	Weight
F	Functional Requirements "What is the software supposed to do?" [353]		
F-1	The plugin must migrate all video and audio streams in the input file to the following format: Container: Matroska (mkv) Videocodec: ffv1, Version 3 Audiocodec: PCM		E
F-2	The video codec must use the value 1 for the Group Of Picture (GOP)		E
F-3	The video codec must use the value 1 (Range coder) for the coder		E
F-4	The video codec must use the value 1 (large context) for context		E
F-5	The video codec must use CRC per slice		E
F-6	Default value for slices is 16		C
F-7	A valid custom parameter must influence the number of slices per frame		C
F-8	An invalid custom parameter must be reported to the log		C
F-9	An invalid custom parameter must not migrate the file		C
F-10	An invalid custom parameter must lead to an return value other than 0		C
F-11	The audio codec must be signed		E
F-12	The audio codec must be little endian		E
F-13	The audio codec must have a bit depth of 24 bits		E

[353] Institute of Electrical and Electronics Engineers 1998, p. 3.

F-14	The plugin must migrate the input when choosen as an alternative, internal plugin in a preservation plan		E
E-	External Interfaces "How does the software interact with people, the system's hardware, other hardware, and other software? "[354]		
E-1	The plugin must accept the parameter transferred by Rosetta: $n = output directory path[355]	https://developers.exlibrisgroup.com/rosetta/sdk/plugins/MigrationTool	E
E-2	The plugin must accept the parameter transferred by Rosetta: $(n-1) = input file name/path[356]	https://developers.exlibrisgroup.com/rosetta/sdk/plugins/MigrationTool	E
E-3	The plugin must accept the parameter transferred by Rosetta: $1 - $(n-2) = custom parameters (from Transformation Profile UI)[357]	https://developers.exlibrisgroup.com/rosetta/sdk/plugins/MigrationTool	E
E-4	The user should be able to influence the number of slices.		C
E-5	Script plugins as well as Java plugins are installed in Rosetta from a JAR file. "Each plugin [Java Archive] JAR file must contain a metadata XML file in the /PLUGIN-INF/ directory. The filename should contain 'metadata' and have an xml extension (e.g. metadata_myFirstPlugin.xml)".[358]	https://developers.exlibrisgroup.com/rosetta/sdk/plugins	E

[354] Institute of Electrical and Electronics Engineers 1998, p. 3
[355] ExLibris - Rosetta 2018a.
[356] ExLibris - Rosetta 2018a
[357] ExLibris - Rosetta 2018a
[358] ExLibris - Rosetta 2018b.

E-6	"In case of a script error (or an initiated exit with a return value other than 0) - the script's echo messages will be printed to the log as an ERROR message in the following format: Execution of {script_full_path} failed: {echo_message}"[359]	https://developers.exlibrisgroup.com/rosetta/sdk/plugins	E
E-7	"Scripting language should be specified using the Shebang line."[360]	https://developers.exlibrisgroup.com/rosetta/sdk/plugins	E
E-8	A message on the start of the plugin should appear in the Rosetta server logs		E
E-9	If errors do occur a message should appear in the Rosetta server logs		E
E-10	A message on the success of a migration should appear in the Rosetta server logs		C
P-	Performance "What is the speed, availability, response time, recovery time of various software functions, etc.?" [361]		
P-1	The plugin must be available anytime Rosetta is available.		E
P-2	The plugin should log the start of the process within the first 20 Minutes after "Run Test" was clicked in the GUI (Graphical User Interface).		E
S-	Software System Attributes		
	"What are the portability, correctness, maintainability, security, etc. considerations? "[362]		
S-1	External documentation of the custom parameters must be available.		E

[359] ExLibris - Rosetta 2018b.
[360] ExLibris - Rosetta 2018b.
[361] Institute of Electrical and Electronics Engineers 1998, p. 3
[362] Institute of Electrical and Electronics Engineers 1998, p. 3

S-2	Comments in the script shall illustrate the concept of the plugin.		C
D-	Design Constraints		
	"Are there any required standards in effect, implementation language, policies for database integrity, resource limits, operating environment(s) etc.?"[363]		
D-1	The plugin must run on a Solaris server.		E

8.3 Preparations and Development

Preparations included getting access with reading, writing and executing rights to the Solaris server on which the development system of Rosetta runs. In the course of the development further decisions had to be made. These decisions, benefits and other options are described exemplary. The complete source code can be found on GitHub at https://github.com/TIBHannover/ffv1_Migration_Tool.

It was decided to install the migration software FFmpeg on the solaris server where Rosetta runs. Another option would be an FFmpeg build inside the migration plugin. Installing FFmpeg on the server has benefits. FFmpeg comes with a number of different builds of different environments e.g. a solaris build. Installing FFmpeg on the server means that the migration plugin can easily be reused by other institutions which may rely on another FFmpeg build. Also wrapping the build into the migration plugin would mean to make a new version of the plugin with every new version of FFmpeg (in order to keep the plugin up-to-date).

The plugin is written in Bash as it is supported by the underlying operating system as well as by other operating systems of other Rosetta customers, which adds to the reusability of the plugin.

[363] Institute of Electrical and Electronics Engineers 1998, p. 3

Studying existing plugins[364,365,366] revealed the underlying structure of a migration plugin which was adopted:

1. Read in Parameters
2. Check directories
3. Migrate
4. Report migration / Report errors
5. Exit with appropriate exit code

The interaction with the plugin is controlled by the custom parameters. As described in the requirements the number of slices is influenced by giving no parameter (F-6) or giving a valid parameter (F-7). There are two possible ways custom parameters can be used: either the customer enters a command for migration or one can work with a profile. Advantages and disadvantages of both methods can be found in Table 15 Custom Parameter Handling.

Table 15 Custom Parameter Handling

	command	**Profile**
Advantage	More flexible as the command line can be enhanced without restrictions	Validation of custom parameters is limited to the number of profiles
		User needs no command line knowledge
Disadvantage	Validation of the custom parameters is more extensive because there are more options and combinations possible	Adjustments to the command line have to be made in the plugin and require a new deployment

The advantages of working with the custom parameter as a profile prevail. As only one parameter shall be influenced it was possible to have all possible values as profiles and name them according to their values. It would also be possible to give speaking names like “maxSlices” or “minSlices”.

[364] National Library of New Zealand.
[365] National Library of New Zealand.
[366] Ott.

The custom parameters are saved into profiles in line 10-17. If there are three input parameters, the first parameter is the custom parameter, see requirement E-3. If the user gives no custom parameter, the profile is set to default as seen in Figure 9 Code Block: Create Profile.

Figure 9 Code Block: Create Profile

```
9   #create profile according to user input or set default
10  if [ "$#" -eq 3 ] ; then

11          profile="$1"
12  elif [ "$#" -gt 3 ] ; then
13          retval="3"
14          echo "more than one custom parameter, only one allowed"
15  else
16          profile="default"
17  fi
```

The FFmpeg command is modified according to the profile. In line 32-35 the number of slices is changed according to the chosen profile, see Figure 10 Code Block: Use of Custom Parameter. As the name of the profile and the value of the slices are the same, it is not necessary to save the value for the number of slices in a separate variable. If speaking names like "maxSlices" were applied, it would have been necessary to work with another variable which stores the value according to the profile.

Figure 10 Code Block: Use of Custom Parameter

```
32  #use custom slice parameter if valid user input
33  elif [ $profile = "4" ] || [ $profile = "6" ] || [$profile = "9" ] || [$profile = "12" ] || [$profile = "16" ] || \
34  [ $profile = "24" ] || [ $profile = "30" ]; then
35      /exlibris/dps/bin/ffmpeg -nostdin -i "$in_filename" -map 0 -c:v ffv1 -level 3 -g 1 -coder 1 -context 1 \
36      -slices $profile -slicecrc 1 -c:a pcm_s24le "$out_filepath"
37      retval="$?"
38      echo "Custom parameter is set to $profile slices"
```

The script plugin receives at least two or more parameters, as describes in requirement E-3. The two final parameter are the input / output paths to the file. If no custom parameter is given, these are the only two parameter. If one or more custom parameter is given (separated by a space character), they are handed over first. Therefore the paths are not always in the same place and cannot be adressed with a fixed number like "$1". Instead they are adressed with the last parameter "${*: -1:1}" and the second to last parameter "${*: -2:1}" (see line 3-4). As the script plugin works with profiles, it is not necessary for the user to hand over more than one parameter, therefore more than three parameter (equates to more than one custom

paramter) are reported to the logs (see requirement F-8, line 12-14), does not lead to a migration (see requirement F-9) and exits with error code 3 (see requirement F-10).

The development of the plugin was accompanied by conversations with colleagues about best practice for development of a plugin[367],[368], and testing with Rosetta[369]. This led to a deeper insight on these fields as well as a general understanding of the interaction between Rosetta and plugins in general.

8.4 Evaluation of the Requirements

During the development of the plugin as well as for the evaluation of the requirements, testing was necessary. There were three testing routines. General testing was performed directly in the shell on the server, with the advantage that no login to Rosetta was required and errors were found easily.

The basic testing routine was always necessary when a minor change to the plugin had been made. The following steps were executed:

1. creating a test-set with files
2. generate a basic preservation plan
3. testing the preservation plan
4. evaluating the outcome
5. checking the server logs

When a major change in the software was made, an enhanced testing routine was conducted. This included all steps from the basic testing routine, and additionally:

1. changing the version in PLUGIN-INF/metadata_ffv1Converter.xml
2. creating a new jar-file (required for the installation routine, see requirement E-5)
3. upgrading plugin and restarting Rosetta (with the help of Administration)
4. ingesting selected files beforehand (once) in order to test the behavior of the plugin
5. making an elaborated preservation plan with testing of several technical parameter
6. downloading files after migration to perform quality control

[367] Wilson 2018.
[368] Ott 2018.
[369] Ott 2018.

During testing some technical obstacles were observed: a not reproducible error led to a locked AIP and locked files during the running of a preservation plan. These IEs could not be unlocked by the user but had to be reported to the ExLibris support. Also at one point the indexing of the technical metadata of newly ingested IE was not possible, therefore the risk analysis did not catch them, and the IE was not found by a preservation plan. This error was fixed bx ExLibris and resolved with the installation of a new Rosetta version.

The evaluation of the requirements leads to the conclusion, that all requirements are met. A testing protocol can be found in Appendix D “Test Protocol Migration Plugin”. The tests were conducted with three different randomly chosen input files, each of the most common file formats in TIB’s AV holdings was represented. As the tests examine the requirements of the plugin it is sufficient to test its functions with one example file for each file format. The test parameters varied: either no custom parameter was given, a valid custom parameter, multiple parameter separated by a space character, or an invalid custom parameter (see Figure 11 Migration Plugin: Invalid Custom Parameter). Testing included the analysis of the output files, the logs, the source code and documentation as well as monitoring the GUI.

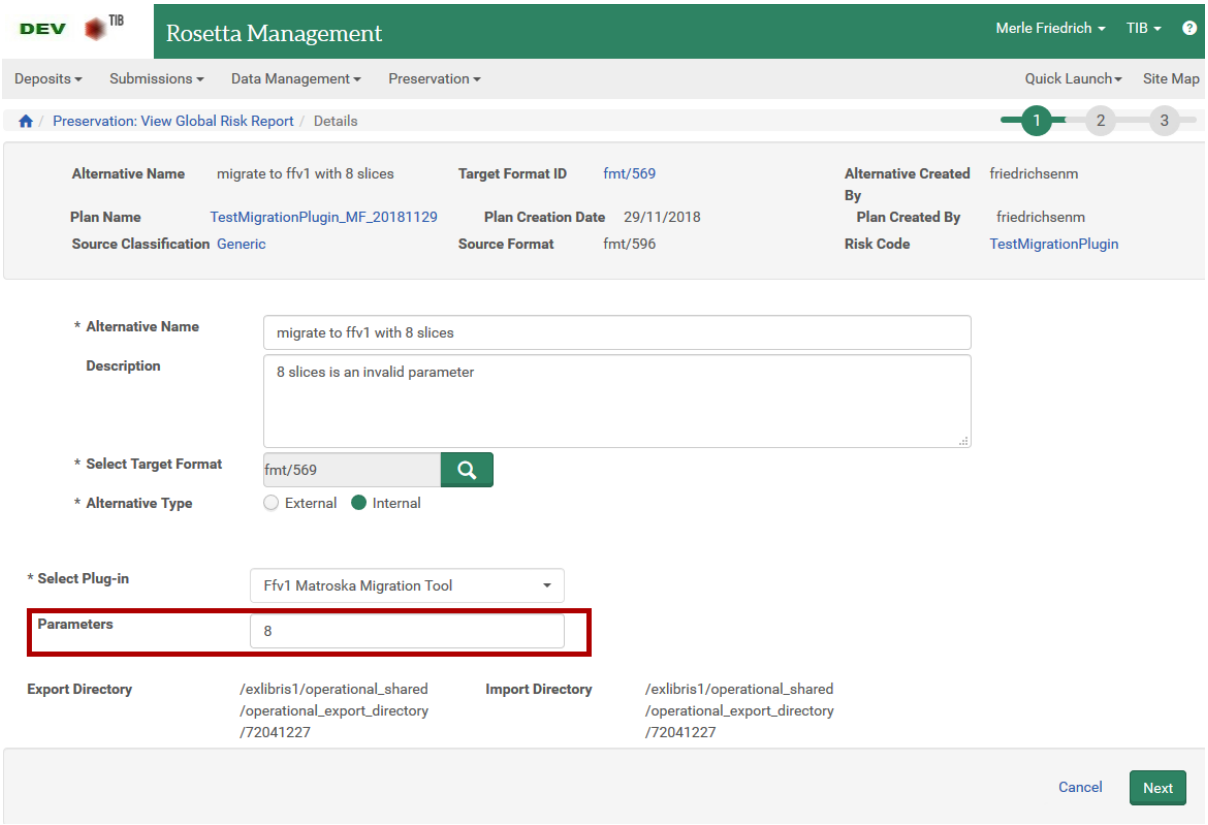


Figure 11 Migration Plugin: Invalid Custom Parameter

Except for one requirement, all requirements were fulfilled. This includes functional requirements, external interfaces, performance, software system attributes as well as design constraints. The exception is requirement P-2: *"The plugin should log the start of the process within the first 20 minutes after "Run Test" was clicked in the GUI"*. It turned out, that all messages from the plugin are reported to the log after the plugin returned the exit code. So if a migration takes longer than 20 minutes, the plugin will not report to the log in the first 20 minutes.

On the other hand, there are two indicators that migration started as seen in Figure 12 Section of the Logs: Rosetta reports to the log that conversion started, and the FFmpeg messages appear as an error in the logs. Although the requirement is not met, the user still can retrace that migration has started, which was the intention of the requirement.

Figure 12 Section of the Logs

```
2018-11-29 18:09:08,840 INFO[com.exlibris.preservation.task.ConvertFormatPreservationTask] (GENER-
    IC_PRSRV_Q Queue Job Receiver 74) []| PRSV-rosetta02.develop.lza.tib.eu | Converting file
    /exlibris1/operational_shared/operational_delivery_shared/convert_temp/REP1223926/A_10310.mp4 to:
    /exlibris1/operational_shared/operational_export_directory/72041227/72041233/import/REP1223926
2018-11-29 18:09:09,027 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: ffmpeg version 4.0.2 Copyright (c) 2000-2018 the FFmpeg developers
2018-11-29 18:09:09,027 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: built with gcc 7.3.0 (GCC)
2018-11-29 18:09:09,027 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: configuration: --prefix=/tib/user/usern/src/ffmeg/dist --enable-nonfree
2018-11-29 18:09:09,030 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: libavutil56. 14.100 / 56. 14.100
2018-11-29 18:09:09,030 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: libavcodec 58. 18.100 / 58. 18.100
2018-11-29 18:09:09,030 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: libavformat58. 12.100 / 58. 12.100
2018-11-29 18:09:09,030 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: libavdevice58.3.100 / 58.3.100
2018-11-29 18:09:09,030 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: libavfilter 7. 16.100 /7. 16.100
2018-11-29 18:09:09,030 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: libswscale5.1.100 /5.1.100
2018-11-29 18:09:09,030 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: libswresample 3.1.100 /3.1.100
2018-11-29 18:09:09,069 INFO[com.exlibris.core.infra.svc.api.scriptRunner.ExecWrapper] (Thread-35) []
    ERROR: Input #0, mov,mp4,m4a,3gp,3g2,mj2, from
    '/exlibris1/operational_shared/operational_delivery_shared/convert_temp/REP1223926/A_10310.mp4':
```

Although this requirement was not fulfilled, there seems to be no possibility to change the plugin to report earlier to the logs. As Rosetta logs the start of the conversion, the user can trace back the beginning of the migration in the logs in another way than the intended. As the requirement was ranked conditionally, there is no need to find another solution. Nevertheless one should investigate if the ffmpeg command can be modified to give less output, in order to avoid unnecessary error messages and long logs.

With the evaluation of the requirements it was proven that the developed migration plugin meets the requirements. The second thesis *"A plugin can be developed and integrated in TIB's archive software environment in order to migrate to a suitable format."* is verified.

In order to achieve reusability it was decided to publish the plugin under a MIT-License: [https://github.com/TIBHannover/ffv1_Migration_Tool/blob/master/LICENSE]. The publication on Github is an easy way of sharing the code and making it reusable in an easy manner for other Rosetta customers. The MIT-License grants the possibility to change the code according to other needs.

9 Conclusion

The intention is to answer if it is possible to migrate obsolete AV material from TIB's holdings. As described in chapter 2 the "Research Questions and Methodology" this question was split into two research questions to which the background was illustrated in chapters 3 to 5: "Digital Audio-Visual Material", "Digital Preservation – Theoretical Approach" and "Digital Preservation at TIB". The questions were answered in chapters 6 to 8: "Attributes of Suitable Formats", "Attributes of Obsolete Formats" and "Migration Plugin". The conclusion summarizes the approaches to the questions and reviews the answers to the research questions. The choice of methodologies led to the expected results. All research theses were verified in the course of the chapters. The review shows up limitations of the research as well as implications for future work.

9.1 Review of the Catalogue of Criteria

The first research question "*Are there file formats in TIB's audio-visual holdings which are obsolete?*" was answered in two chapters, subdivided into the questions if the majority of file format are suitable for digital preservation and if the file formats are obsolete.

The literature review in chapter 6 "Attributes of Suitable Formats" revealed five main selection criteria for suitable archival formats according to Todd (2009). Each of the criteria - *adoption*, *platform independence*, *disclosure or documentation*, *transparency*, and *metadata support*[370] - were broken down to measurable indicators in order to examine the majority of TIB's file formats which make up more than 70% of TIB's digital AV holdings. The evaluation has shown, that the majority of file formats are not suited as preferred archival format and thus the thesis 1.a) was verified.

[370] Todd 2009, p. 13

In chapter 7 “Attributes of Obsolete Formats“ the thesis 1. b) *“The majority of file formats within the TIB AV-holdings are not obsolete”* was verified. The term obsolete was defined to describe a *file format that is at risk to become inaccessible by our designated community*. Literature review had shown that missing rendering software is the only factor which leads to obsolete file formats.[371] In order to assess obsolescence *view paths* (information on the hardware platform, operating system, and the viewer application) [372] were tested and documented. For each of the examined format two independent view-paths exist, and therefore the file formats are not obsolete.

The evaluation of the file formats gives a first overview on the condition of TIB’s AV collection. The results must be seen in the context of TIB’s preservation policy and archives environment and cannot be generalized. Nevertheless other institutions can adopt the measurements according to their preservation policy and designated community.

The results of chapters 6.2 "Classification of TIB’s holdings”, 6.3 and 7.2 can be applied to file formats. But reliable statements can only be made on the basis of each individual file. Not only the file format and version, but also the profile as well as creation software can make a difference at the file level. Also whether a file is valid has a huge impact on the decision for a preservation action. Unfortunately the digital preservation community faces a problematic lack of validation tools for the different AV file formats. "If transcoding the original file leads to errors, this can serve as a first indicator of problematic files due to unclean implementation of either codec and/or container data."[373]

Although the evaluation must be made at the file level, the catalogue of criteria gives a good overview on the question if there are obsolete formats in the archive. Adding view paths into Rosetta will help checking obsolescence regularly and should be enhanced for all file formats in TIB’s archive. As every archiving institution has its own designated community the view paths are not transferable. But they can serve as a starting point and be modified according to other archives’ hardware and software environment.

[371] Ryan 2014, p. 210
[372] Steenbakkers 2005
[373] Bubestinger-Steindl 2018.

9.2 Review of the Development of the Plugin

The development of a plugin for TIB's archive software environment was described in chapter 8 "Migration Plugin". The requirements for the plugin in chapter 8.2 were divided into functional requirements, requirements which concern the external interfaces, the performance, software system attributes as well as design constraints. The requirements consider the integration with Rosetta as well as requests emerging from the chosen archival format which is described in chapter 6.3 "Classification of mkv/ffv1/PCM as Preferred Archival Format".

The testing of the plugin presupposes knowledge concerning preservation planning with Rosetta as described in chapter 5.2 "Preservation Planning with Rosetta". Except for one conditional requirement all requirements are met, and the second thesis *"A plugin can be developed and integrated in TIB's archive software environment in order to migrate to a suitable format."* was verified.

Due to the custom parameters handling as profiles, the plugin can be enhanced to future requirement e.g. concerning the audio codec. Another possible enhancement concerns the testing of a successful migration. With the command "framemd5"[374] one can create a checksums for each (decoded) frame of the input file. As the plugin transcodes to the lossless compressing codec ffv1, the framemd5 checksums from the output file must match the framemd5 checksums of the input file. In this way one can verify that no information (concerning the visual output, not the metadata) was lost during migration. This would be different when transcoding to another lossy video codec. In this case a migration could lead to a loss of information and thus to different framemd5 checksums.

Even though testing the plugin revealed that it fulfills most requirements, it will not be in use in the running operations in the near future. This is because the majority of AV file formats is not obsolete as described in chapter 6.2. A migration therefore is not necessary and would lead to an unnecessary amount of storage space as the migration from a lossy compressed format to a lossless compressed format leads to larger file size. If the archive wants to ensure that a migration is possible at a later point in time a copy of the FFmpeg build could be saved.

---

[374] FFmpeg Wiki 2013.

Nevertheless the development of the plugin has led to deeper insights on how Rosetta interacts with plugins and how testing is performed best. This knowledge is helpful for future integration of other tools.

The plugin was published on Github to facilitate distribution of the plugin. The developed plugin therefore can be reused and altered by other Rosetta customers who are interested in a migration to mkv/ffv1/PCM. It might also be helpful for other archiving institutions to find the FFmpeg command line which is used. Regarding the reusability of tasks and tools it is worth watching the efforts of the "Preservation Action Registry"[375]. This projects aims to unify the description of tasks and tools among different archiving software and contributes in this way to share knowledge and reuse good practice preservation actions[376].

[375] Artefactual et al. 2018.
[376] Artefactual et al. 2018.

10 Source Code

```
#!/bin/bash
#read input parameter, set last and next to last parameter
in_filename="${*: -2:1}"
output_dir="${*: -1:1}"
out_filename="$(basename "$in_filename")"
out_filename_no_ext="${out_filename%.*}"
out_filepath="${output_dir}/${out_filename_no_ext}.mkv"

#create profile according to user input or set default
if [ "$#" -eq 3 ] ; then
        profile="$1"
elif [ "$#" -gt 3 ] ; then
        retval="3"
        echo "more than one custom parameter, only one allowed"
else
        profile="default"
fi

#make output directory
if [ ! -d "$output_dir" ]; then
        mkdir -p "$output_dir";
fi

echo "Migrating $in_filename to $out_filepath"

#select ffmpeg command line configuration according to profile
if [ $profile = "default" ]; then
        /exlibris/dps/bin/ffmpeg -nostdin -i "$in_filename" -map 0 -c:v ffv1 -level 3 -g 1 -coder 1 -context 1 \
         -slices 16 -slicecrc 1 -c:a pcm_s24le "$out_filepath"
        retval="$?"
        echo "Default parameter is set to 16 slices"
#use custom slice parameter if valid user input
elif [ $profile = "4" ] || [ $profile = "6" ] || [ $profile = "9" ] || [ $profile = "12" ] || [ $profile = "16" ] || \
[ $profile = "24" ] || [ $profile = "30" ]; then
        /exlibris/dps/bin/ffmpeg -nostdin -i "$in_filename" -map 0 -c:v ffv1 -level 3 -g 1 -coder 1 -context 1 \
        -slices $profile -slicecrc 1 -c:a pcm_s24le "$out_filepath"
        retval="$?"
        echo "Custom parameter is set to $profile slices"
#no valid user parameter parsed
else
        retval="2"
        echo "no valid custom parameter"
fi

#report the status of the migration to the log
if [ $retval == "0" ]; then
        echo "STATUS=0"
        echo "CONTENT=migration of $in_filename to $out_filepath succeeded"
else
        echo "STATUS=$retval"
        echo "CONTENT=execution failed with return code $retval"
fi

exit $retval;
```

11 Publication bibliography

Addis, Matthew; Allasia, Walter; Bailer, Werner; Boch, Laurent; Gallo, Francesco; Wright, Richard (2010): 100 million hours of audiovisual content. In *Proceedings of the 1st International Digital Preservation Interoperability Framework Symposium*, pp. 1–8. DOI: 10.1145/2039263.2039266.

Arms, Caroline; Fleischhauer, Carl (2005): Digital formats. Factors for sustainability, functionality, and quality. In *Society for Imaging Science and Technology* 2005, 2005 (1).

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate: Sustainability of Digital Formats: Planning for Library of Congress Collections. Sustainability Factors. Edited by Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml, checked on 1/4/2019.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2010): PCM, Pulse Code Modulated Audio. Edited by Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000016.shtml, updated on 6/1/2010, checked on 11/8/2018.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2017a): Advanced Audio Coding (MPEG-4). Edited by Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000114.shtml, updated on 6/22/2010, checked on 11/8/2018.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2017b): FF Video Codec 1. Edited by Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000341.shtml, updated on 12/21/2015, checked on 11/8/2018.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2017c): Matroska Multimedia Container. Edited by Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000342.shtml, updated on 12/21/2015, checked on 11/8/2018.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2017d): MPEG-1 and MPEG-2 Layer II Audio Encoding. Edited by Library of Congress. Available

online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000338.shtml, updated on 2/27/2017, checked on 11/8/2018.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2017e): MPEG-2 Encoding Family. Edited by Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000335.shtml, checked on 1/4/2019.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2017f): MPEG-4 File Format, Version 2. Edited by Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000155.shtml, updated on 3/2/2012, checked on 11/8/2018.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2017g): MPEG-4, Advanced Video Coding (Part 10) (H.264). Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000081.shtml, updated on 12/5/2011, checked on 11/8/2018.

Arms, Caroline R.; Fleischhauer, Carl; Murray, Kate (2018): MPEG-2 Video Encoding (H.262). Edited by Library of Congress. Available online at https://www.loc.gov/preservation/digital/formats/fdd/fdd000028.shtml, updated on 6/7/2018, checked on 11/8/2018.

Artefactual; Arkivum; Preservica; Jisc; Open Preservation Foundation (Eds.) (2018): RDSS PAR | RDSS-PAR. Available online at http://parcore.org/, updated on 9/25/2018, checked on 1/21/2019.

Bähr, Thomas; Schwab, Franziska (2017a): Preservation Management - Digitale Langzeitarchivierung an der TIB - TIB Wiki. Edited by Technische Informationsbibliothek (TIB). Available online at https://wiki.tib.eu/confluence/display/lza/Preservation+Management, checked on 1/31/2019.

Bähr, Thomas; Schwab, Franziska (2017b): Technische Infrastruktur - Digitale Langzeitarchivierung an der TIB - TIB Wiki. Edited by Technische Informationsbibliothek (TIB). Available online at https://wiki.tib.eu/confluence/display/lza/Technische+Infrastruktur, checked on 1/31/2019.

Bähr, Thomas; Schwab, Franziska (2018): Digitale Langzeitarchivierung an der TIB. Edited by Technische Informationsbibliothek (TIB). Available online at https://wiki.tib.eu/confluence/display/lza/Digitale+Langzeitarchivierung+an+der+TIB, updated on 2/20/2018, checked on 1/31/2019.

Bailer, Werner; Hall-May, Martin; Veres, Galina V. (2014): Metadata Representation and Risk Managment Framework for Preservation Processes in AV Archives. In Serena Coates, Ross King, Steve Knight, Christopher Lee, Peter McKinney, Erin O'Meara, David Pearson (Eds.): iPRES 2014 - Proceedings of the 11th International Conference on Preservation of Digital Archives. iPRES. Melbourne, 6-10 October, pp. 357–359. Available online at https://hdl.handle.net/11353/10.378066, checked on 1/31/2019.

Bankoski; James; Wilkins; Paul; Xu; Yaowu (2011): VP8 Data Format and Decoding Guide (2070-1721). Available online at https://tools.ietf.org/html/rfc6386, updated on 11/4/2018, checked on 11/14/2018.

Barteleit, Sebastian; Bähr, Thomas; Ernst, Volkmar; Hasler, Tim; Keiper, Jürgen; Lippert, Wolfram et al. (2016): Leitfaden für die digitale Langzeitarchivierung audiovisueller Medien. verfasst von der nestor-Arbeitsgruppe Media. nestor -Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland. Berlin, checked on 11/13/2018.

Becker, Christoph; Kulovits, Hannes; Rauber, Andreas; Hofman, Hans (2008): Plato. A Service Oriented Decision Support System for Preservation Planning. New York, NY: ACM, checked on 1/31/2019.

Becker, Christoph; Rauber, Andreas (2011): Decision criteria in digital preservation. What to measure and how. In *Journal of the American Society for Information Science and Technology* 62 (6), pp. 1009–1028. DOI: 10.1002/asi.21527.

British Film Institute (Ed.) (2011): BFI Collection Policy. November 2011. London. Available online at https://www.bfi.org.uk/about-bfi/policy-strategy/bfi-national-archive-collections-policy, checked on 11/13/2018.

Brown, Adrian (2007): Developing Practical Approaches to Active Preservation. In *IJDC* 2 (1), pp. 3–11. DOI: 10.2218/ijdc.v2i1.10.

Bubestinger-Steindl, Peter (2018): On the Validation of AV Material. Conversation to Merle Friedrich. London, 10/25/2018.

Bubestinger-Steindl, Peter; Lewetz, Hermann; Jaks, Marion (2013): Comparing video codecs and containers for archives. Available online at http://download.das-werk-statt.com/pb/mthk/info/video/comparison_video_codecs_containers.html#codec_h264_lossless, updated on 4/19/2017, checked on 11/13/2018.

Bunkus; Moritz; Lhomme; Steve; Rice; Dave (2018): Matroska Specifications. Available online at https://tools.ietf.org/html/draft-ietf-cellar-matroska-01, updated on 12/2/2018, checked on 12/13/2018.

Bunkus, Moritz (2018): mkvalidator tool | Matroska. Available online at https://www.matroska.org/downloads/mkvalidator.html, updated on 10/30/2018, checked on 11/14/2018.

COPTR contributors (2018): Community Owned digital Preservation Tool Registry (COPTR). Edited by COPTR. Available online at http://coptr.digipres.org/Main_Page, updated on 11/1/2018, checked on 11/14/2018.

Dappert, Angela; Farquhar, Adam (2009): Significance Is in the Eye of the Stakeholder. In Maristella Agosti, José Borbinha, Sarantos Kapidakis, Christos Papatheodorou, Giannis Tsakonas (Eds.): Research and advanced technology for digital libraries. 13th European conference, ECDL 2009, Corfu, Greece, September 27 - October 2, 2009 ; proceedings, vol. 5714. Berlin: Springer (Lecture Notes in Computer Science, 5714), pp. 297–308, checked on 1/31/2019.

Evelyn Peters McLellan (2006): Final Report. Selecting Digital File Formats for Long-Term Preservation, checked on 1/31/2019.

Ex Libris Documentation Department (2018): Preservation Guide, checked on 1/31/2019.

ExLibris - Rosetta (2018a): Migration Tool. Available online at https://developers.exlibrisgroup.com/rosetta/sdk/plugins/MigrationTool, checked on 1/31/2019.

ExLibris - Rosetta (2018b): Plugins. Available online at https://developers.exlibrisgroup.com/rosetta/sdk/plugins, updated on 2018, checked on 1/31/2019.

Farquar, Adam (2007): PLANETS - Preservation and Long-Term Access through Networked Services. Available online at https://www.planets-project.eu/, checked on 1/31/2019.

FFmpeg (Ed.) (2018a): FFmpeg Supported File Formats, Codecs or Features. Available online at https://www.ffmpeg.org/general.html#toc-Supported-File-Formats_002c-Codecs-or-Features, checked on 1/31/2019.

FFmpeg (Ed.) (2018b): FFmpeg License and Legal Considerations. Available online at https://www.ffmpeg.org/legal.html, updated on 5/29/2018, checked on 12/14/2018.

FFmpeg Wiki (Ed.) (2012): audio types – FFmpeg. Available online at https://trac.ffmpeg.org/wiki/audio%20types, updated on 2017, checked on 1/3/2019.

FFmpeg Wiki (2013): framemd5 Intro and HowTo – FFmpeg. Available online at https://trac.ffmpeg.org/wiki/framemd5%20Intro%20and%20HowTo, checked on 1/21/2019.

FFmpeg Wiki (Ed.) (2014): Encode/FFV1 – FFmpeg. Available online at https://trac.ffmpeg.org/wiki/Encode/FFV1, updated on 2018, checked on 1/3/2019.

Gordea, Sergiu; Lindley, Andrew; Graf, Roman: Computing Recommendations for Long Term Data Accessibility basing on Open Knowledge and Linked Data. In *Joint proceedings of the RecSys 2011 Workshop on Human Decision Making in Recommender Systems* 2011, checked on 1/31/2019.

Graf, Roman; Gordea, Sergiu (2014): A Model for Format Endangerment Analysis using Fuzzy Logic. In Serena Coates, Ross King, Steve Knight, Christopher Lee, Peter McKinney, Erin O'Meara, David Pearson (Eds.): iPRES 2014 - Proceedings of the 11th International Conference on Preservation of Digital Archives. iPRES. Melbourne, 6-10 October, checked on 1/31/2019.

Ho, Anthony T. S. (2015): Handbook of digital forensics of multimedia data and devices. Hoboken: Wiley. Available online at https://onlinelibrary.wiley.com/doi/book/10.1002/9781118705773.

Houpert, Jörg.; Melas, Panos; Bailer, Werner; Walland, Paul (2015): Recommendations and techniques for content in a 'born robust' form, checked on 1/31/2019.

Institute of Electrical and Electronics Engineers (1998): IEEE recommended practice for software requirements specifications. Approved 25 June 1998. New York, NY: IEEE (IEEE Std, 830-1998), checked on 1/31/2019.

International Organization for Standardization (1993a): ISO/IEC 11172-1:1993. Available online at https://www.iso.org/standard/19180.html, checked on 11/14/2018.

International Organization for Standardization (Ed.) (1993b): ISO/IEC 11172-2:1993. Available online at https://www.iso.org/standard/22411.html, checked on 11/14/2018.

International Organization for Standardization (Ed.) (1993c): ISO/IEC 11172-3:1993. Available online at https://www.iso.org/standard/22412.html, checked on 11/14/2018.

International Organization for Standardization (Ed.) (2003): ISO/IEC 14496-14:2003. Available online at https://www.iso.org/standard/38538.html, checked on 11/14/2018.

International Organization for Standardization (Ed.) (2009): ISO/IEC 14496-3:2009. Available online at https://www.iso.org/standard/53943.html, checked on 11/14/2018.

Jaks, Marion (2018): FFV1 an der Mediathek. Edited by Österreichischen Mediathek. Available online at https://www.mediathek.at/digitalisierung/ffv1-an-der-mediathek-dt/, checked on 1/3/2019.

Justrell, Börje; Fresa, Antonella; Prandoni, Claudio (2017): The preforma handbook. validating formats, a prerequisite for preserving digital objects, checked on 1/3/2019.

Kilbride, Wiliam; Wheatley, Paul; Sierman, Barbara; L'Hours, Hervé (2015): About the OAIS Community Forum - wiki.dpconline.org. Available online at

http://wiki.dpconline.org/index.php?title=About_the_OAIS_Community_Forum, updated on 2/16/2018, checked on 1/31/2019.

Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen; Centre de coordination pour l'archivage à long terme des documents électroniques; KOST; CECO (2017a): FFV1. Available online at https://kost-ceco.ch/cms/ffv1.html, checked on 12/13/2018.

Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen; Centre de coordination pour l'archivage à long terme des documents électroniques; KOST; CECO (2017b): MPEG-2. Available online at https://kost-ceco.ch/cms/kad_mpeg-2_de.html, checked on 11/13/2018.

Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen; Centre de coordination pour l'archivage à long terme des documents électroniques; KOST; CECO (2017c): MPEG-4. Available online at https://kost-ceco.ch/cms/mpeg-41.html, checked on 11/13/2018.

Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen; Centre de coordination pour l'archivage à long terme des documents électroniques; KOST; CECO (2017d): Videoformate: Einleitung. Available online at https://kost-ceco.ch/cms/videoformate-einleitung.html, checked on 11/13/2018.

Kromer, Reto (2019): Video into Matroska/FFV1. Available online at https://avpres.net/FFmpeg/im_MKV.html, updated on 12/15/2018, checked on 1/3/2019.

Lawrence, Gregory W.; Kehoe, William R.; Rieger, Oya Y.; Walters, William H.; Kenney, Anne R. (2000): Risk management of digital information. A file format investigation. Washington, DC: Council on library and information resources.

Library of Congress (2017): Format Descriptions for Moving Images. Available online at https://www.loc.gov/preservation/digital/formats/fdd/video_fdd.shtml, checked on 11/13/2018.

Lorrain, Emanuel (2014): A short guide to choosing a digital format for video A short guide to choosing a digital format for video archiving masters. Edited by PACKED. Available online at http://scart.be/?q=en/content/short-guide-choosing-digital-format-video-archiving-masters, checked on 1/3/2019.

MediaArea: Documentation: How To Use MediaConch. Available online at https://mediaarea.net/MediaConch/Documentation/HowToUse, checked on 1/3/2019.

MediaArea: MediaConch. Available online at https://mediaarea.net/MediaConch, checked on 12/13/2018.

MediaArea: MediaInfo. Available online at https://mediaarea.net/de/MediaInfo, checked on 11/1/2018.

Mozilla (Ed.) (2018): Media formats for HTML audio and video. Available online at https://developer.mozilla.org/en-US/docs/Web/HTML/Supported_media_formats, checked on 1/31/2019.

National Library of New Zealand: FLV-MP4MergerMigrationTool Version 0.1. Version 0.1: (unpublished). Available online at unpublished, checked on 12/21/2018.

National Library of New Zealand: MP4toWaveMigrationTool Version 3.1. Version 3.1: (unpublished), checked on 12/21/2018.

Newmarch, Jan (Ed.) (2017): Linux sound programming. New York, New York, New York, NY: Apress; Distributed by Springer Science+Business Media New York.

Ott, Konstantin: PDFAPilot Version 1.15. Migration Plugin. Version 1.15: ZBW (unpublished), checked on 1/31/2019.

Ott, Konstantin (2018): Migration Plugin für Rosetta. Email to Merle Friedrich, 8/10/2018.

Pearson, David; Webb, Colin (2008): Defining File Format Obsolescence: A Risky Journey. In *IJDC* 3 (1), pp. 89–106. DOI: 10.2218/ijdc.v3i1.44.

Planets (2010): Planets Preservation Planning Tool: Plato 3.0. User Manual, checked on 1/31/2019.

Recommended Practice CCSDS 650.0-M-2, June 2012: Reference Model for an Open Archival Information System (OAIS). Available online at https://public.ccsds.org/pubs/650x0m2.pdf, checked on 1/31/2019.

Reiche, Miriam (2018): Zielformate: Archivmaster und Derivative Copy. Edited by Technische Informationsbibliothek (TIB). TIB. Available online at https://projects.tib.eu/delft/digitalisierung/zielformate-archivmaster-und-derivative-copy/, checked on 1/31/2019.

Rice; Dave; Niedermayer; Michael; Martinez; Jerome (2018): FFV1 Video Coding Format Version 0, 1, and 3. Available online at https://tools.ietf.org/html/draft-ietf-cellar-ffv1-05, updated on 12/3/2018, checked on 12/13/2018.

Richardson, Michael; Terriberry, Tim; Campbell, Ben (2018): Codec Encoding for LossLess Archiving and Realtime transmission (cellar) -. Edited by Internet Engineering Task Force. Available online at https://datatracker.ietf.org/wg/cellar/about/, updated on 2019, checked on 1/3/2019.

Rog, Judith; van Wijk, Caroline (2008): Evaluating File Formats for Long-term Preservation. In : 4th International Conference on Preservation of Digital Objects, vol. 24, pp. 1–11. Available online at https://www.kb.nl/sites/default/files/docs/KB_file_format_evaluation_method_27022008.pdf, checked on 1/31/2019.

Ryan, Heather M. (2014): Who's afraid of File Format Obsolescence? Evaluating File Format Endangerment Levels and Factors for the Creation of a File Format Endangerment Index. University of North Carolina, Chapel Hill. School of Information and Library Science. Available online at https://cdr.lib.unc.edu/record/uuid:86e05605-97e3-40de-927a-ce5f5a6fa4d0, checked on 1/31/2019.

Sierman, Barbara (2016): Preservation Watch. Edited by DigitalPreservationCoalition.

Stanescu, Andreas (2005): Assessing the durability of formats in a digital preservation environment. In *OCLC Systems & Services* 21 (1), pp. 61–81. DOI: 10.1108/10650750510578163.

Steenbakkers, Johan F. (2005): Digital Archiving in the Twenty-First Century: Practice at the National Library of the Netherlands. In *Library Trends* 54 (1), pp. 33–56. DOI: 10.1353/lib.2006.0010.

Strodl, Stephan; Becker, Christoph; Neumayer, Robert; Rauber, Andreas (2007): How to choose a digital preservation strategy. Evaluating a Preservation Planning Procedure. New York, NY: ACM, checked on 1/31/2019.

Technische Informationsbibliothek (Ed.): TIB AV-Portal. Available online at https://av.tib.eu/, checked on 1/31/2019.

Technische Informationsbibliothek (Ed.) (2017a): Jahresbericht 2017, checked on 11/29/2018.

Technische Informationsbibliothek (Ed.) (2017b): Preservation Policy der Technischen Informationsbibliothek (TIB) - Technische Informationsbibliothek (TIB). Technische Informationsbibliothek (TIB). Available online at https://www.tib.eu/de/service/tib-preservation-policy/, checked on 1/31/2019.

The National Archives (2004): The National Archives | PRONOM | About. The National Archives. Available online at http://www.nationalarchives.gov.uk/aboutapps/PRONOM/default.htm, checked on 1/31/2019.

The National Archives (2010): mpeg 4. The National Archives, Kew, Surrey TW9 4DU. Available online at http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=924&strPageToDisplay=summary, checked on 11/8/2018.

The National Archives (2013): webM. The National Archives, Kew, Surrey TW9 4DU. Available online at http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=1361&strPageToDisplay=summary, checked on 11/8/2018.

The National Archives (2014a): DROID. The National Archives. Available online at http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/, checked on 1/31/2019.

The National Archives (2014b): MPEG PS. The National Archives, Kew, Surrey TW9 4DU. Available online at http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=659, checked on 11/8/2018.

The National Archives (2014c): PRONOM | WebM. The National Archives, Kew, Surrey TW9 4DU. Available online at http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=1361&strPageToDisplay=summary, checked on 1/31/2019.

The WebM Project (Ed.) (2017): WebM Container Guidelines. Available online at https://www.webmproject.org/docs/container/, updated on 10/10/2018, checked on 11/14/2018.

Todd, Malcolm (2009): File formats for preservation. Technology Watch Report. Available online at https://www.dpconline.org/component/docman/?task=doc_download&gid=375, checked on 1/31/2019.

Vermaaten, Sally; Lavoie, Brian; Caplan, Priscilla (2012): Identifying Threats to Successful Digital Preservation: the SPOT Model for Risk Assessment (9/10). Available online at http://dlib.org/dlib/september12/vermaaten/09vermaaten.print.html, updated on 5/9/2013, checked on 1/31/2019.

videoLAN (Ed.) (2018): VLC media player. Available online at https://www.videolan.org/vlc/, checked on 11/21/2018.

Watkinson, John (2001): An introduction to digital video. 2nd ed. Oxford England, Boston: Focal Press, checked on 1/31/2019.

webm wiki (Ed.) (2012): Global Metadata in WebM. Available online at http://wiki.webmproject.org/webm-metadata/global-metadata, updated on 11/1/2018, checked on 11/15/2018.

Wheatley, Paul Robert; Sierman, Barbara (2010): Evaluation of Preservation Planning within OAIS, based on the Planets Functional Model, checked on 11/1/2018.

Wikipedia (Ed.) (2018a): Comparison of audio coding formats - Wikipedia. Available online at https://en.wikipedia.org/w/index.php?oldid=866214860, updated on 11/8/2018, checked on 1/31/2019.

Wikipedia (Ed.) (2018b): Comparison of video codecs - Wikipedia. Available online at https://en.wikipedia.org/w/index.php?oldid=860105259, updated on 11/8/2018, checked on 11/13/2018.

Wikipedia (Ed.) (2018c): FFV1 - Wikipedia. Available online at https://en.wikipedia.org/w/index.php?oldid=875129757, updated on 12/27/2018, checked on 1/3/2019.

Wilson, Carl (2018): Development of a migration plugin. Telephone Call to Merle Friedrich, 8/9/2018.

Wright, Richard (2012): Preserving Moving Pictures and Sound: DPC Technology Watch Report 12-01: Digital Preservation Coalition, checked on 1/31/2019.

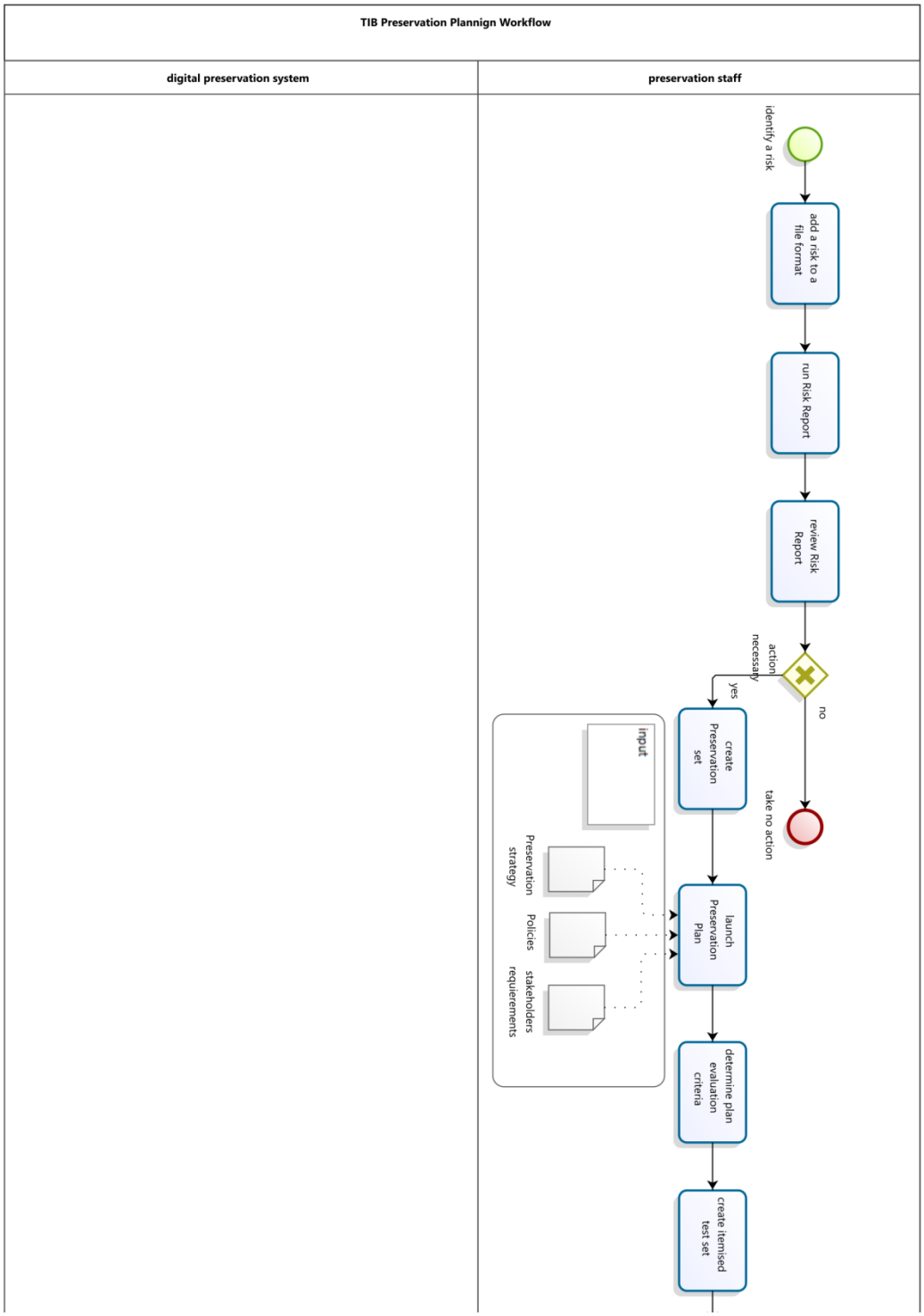
Xiph.Org Foundation (Ed.) (2015): Vorbis I specification. Available online at https://xiph.org/vorbis/doc/Vorbis_I_spec.html, updated on 2/27/2015, checked on 1/31/2019.

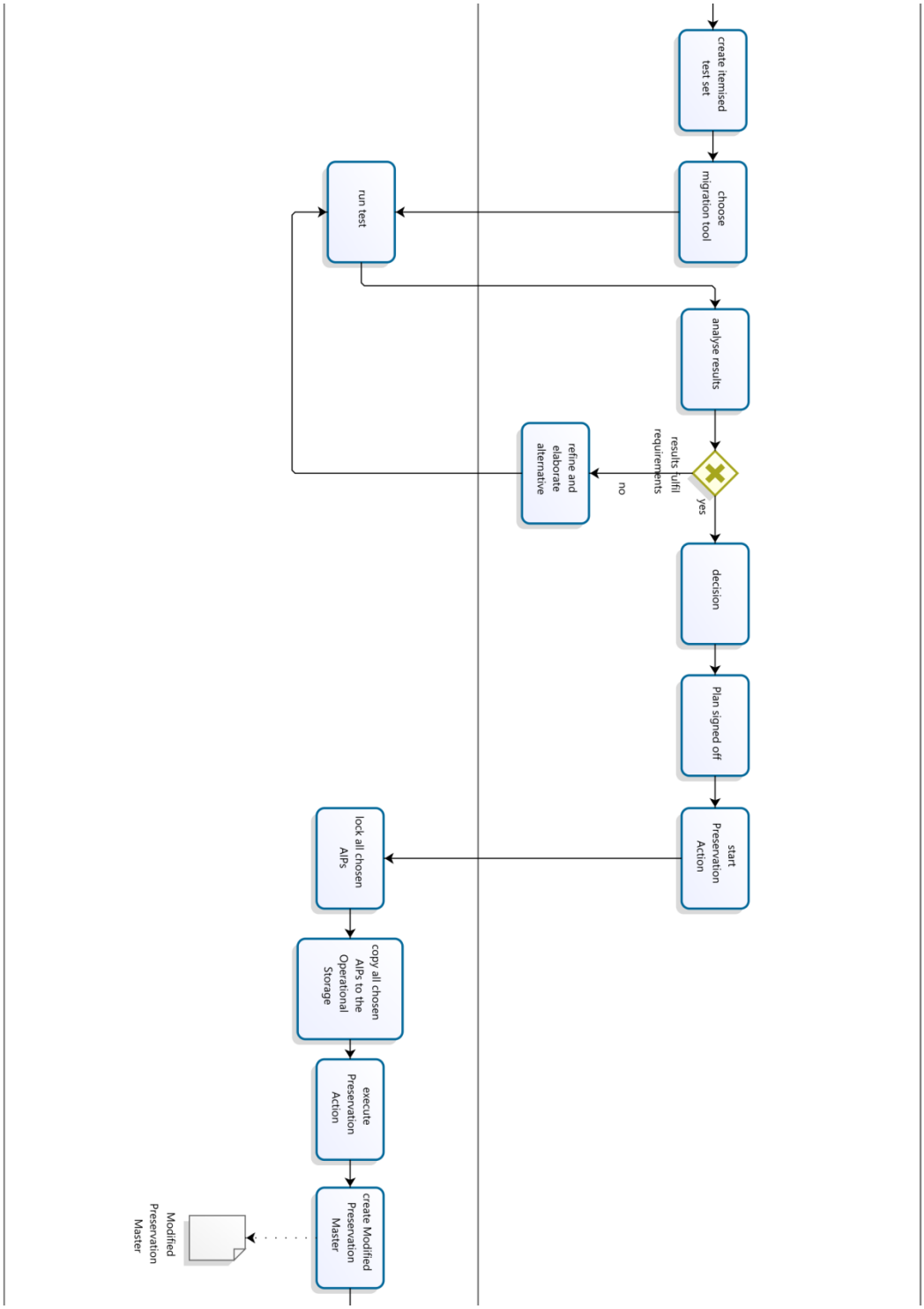
Appendix A Overview of File Formats in TIB's AV holdings

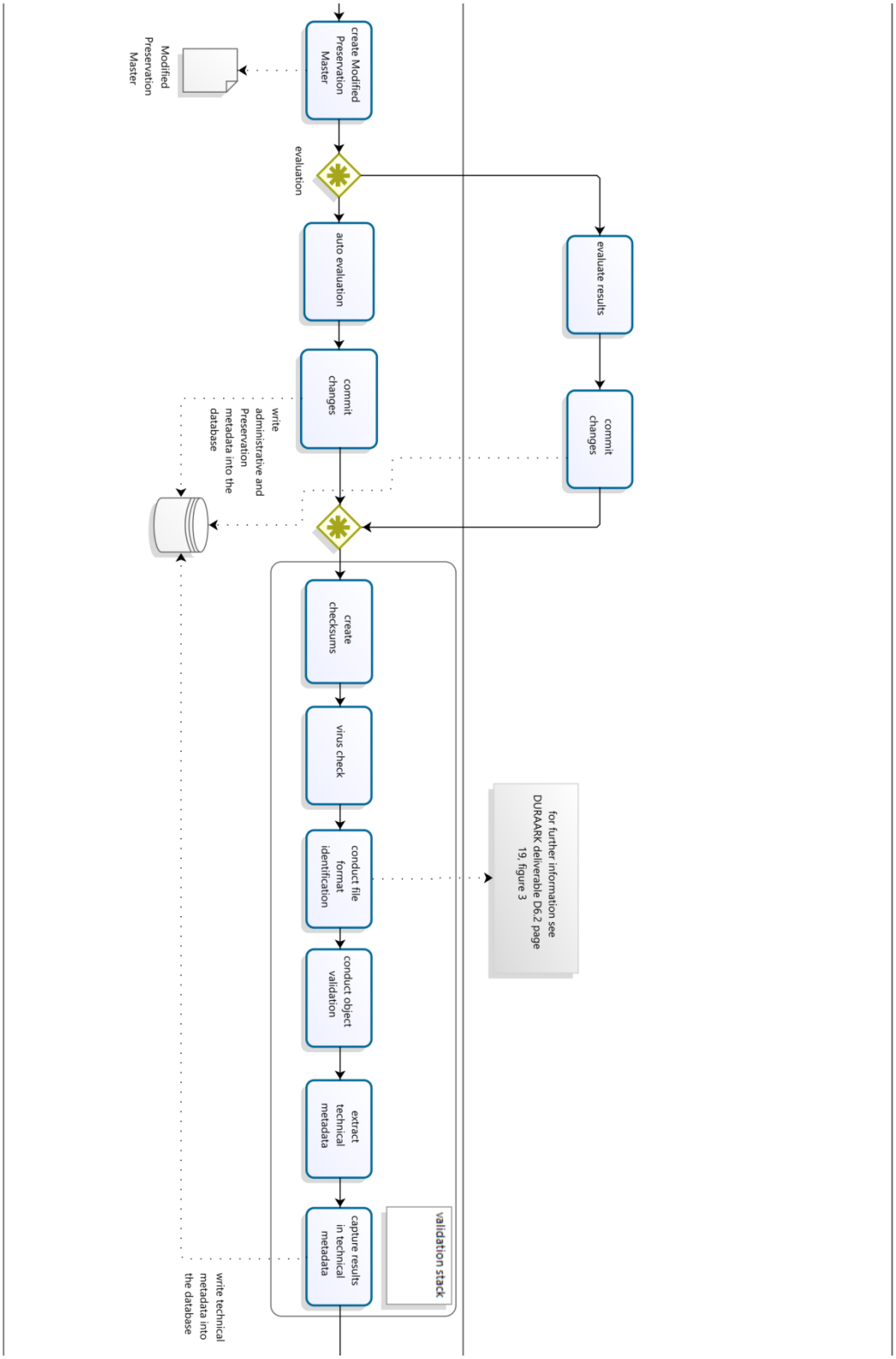
Count	General	Video	Audio
7408	MPEG-4	AVC	AAC,Version 4
1062	WebM	VP8	Vorbis
1006	MPEG-PS	MPEG Video, Version 2	MPEG Audio, Version 1
768	MPEG-4	AVC	AAC
706	MPEG-PS	MPEG Video, Version 1	MPEG Audio,Version 1
568	MPEG-4	MPEG-4 Visual	AAC, Version 4
425	MPEG-4	MPEG-4 Visual	
210	WebM, Version 2	VP8	Vorbis
175	MPEG-TS	AVC	AC-3
134	MPEG-PS	MPEG Video, Version 2	
93	AVI	AVC	PCM
67	MPEG-PS	MPEG Video, Version 2	MPEG Audio, Version 1
64	MPEG-PS	MPEG Video, Version 2	AC-3
44	Flash Video	VP6	MPEG Audio, Version 1
33	Flash Video	VP6	MPEG Audio, Version 2
31	MPEG-4	AVC	
31	MPEG-4	MPEG-4 Visual	AAC
17	AVI	MPEG-4 Visual	
12	Flash Video	AVC	
11	AVI	AVC	MPEG Audio,Version 1
11	MPEG-4	MPEG-4 Visual	MPEG Audio, Version 1
8	MPEG-4	AVC	MPEG Audio,Version 1
8	Windows Media	VC-1	WMA, Version 2
7	AVI	DV	PCM
5	MPEG-PS	MPEG Video, Version 2	PCM
5	MPEG-PS	MPEG Video, Version 2	AC-3
4	AVI	MPEG-4 Visual	PCM
4	BDAV	AVC	AC-3
4	Flash Video	Sorenson Spark	
4	MPEG-PS	MPEG Video, Version 1	
3	AVI	MPEG Video, Version 2	MPEG Audio,Version 1
3	AVI	MPEG-4 Visual	AAC,Version 4
3	MPEG-4		
3	QuickTime	AVC	AAC,Version 4
3	Windows Media	VC-1	WMA
2	AVI	M-JPEG	
2	MPEG Video, Version 2	MPEG Video, Version 2	
2	MPEG-4	MPEG Video, Version 1	
1	AVI	JPEG	
1	AVI	MPEG Video, Version 2	AC-3
1	AVI	MPEG Video, Version 2	
1	Matroska	AVC	
1	MPEG-4	AVC	AAC
1	MPEG-PS	MPEG Video, Version 1	MPEG Audio,Version 2
1	QuickTime	AVC	
12953	**total**		

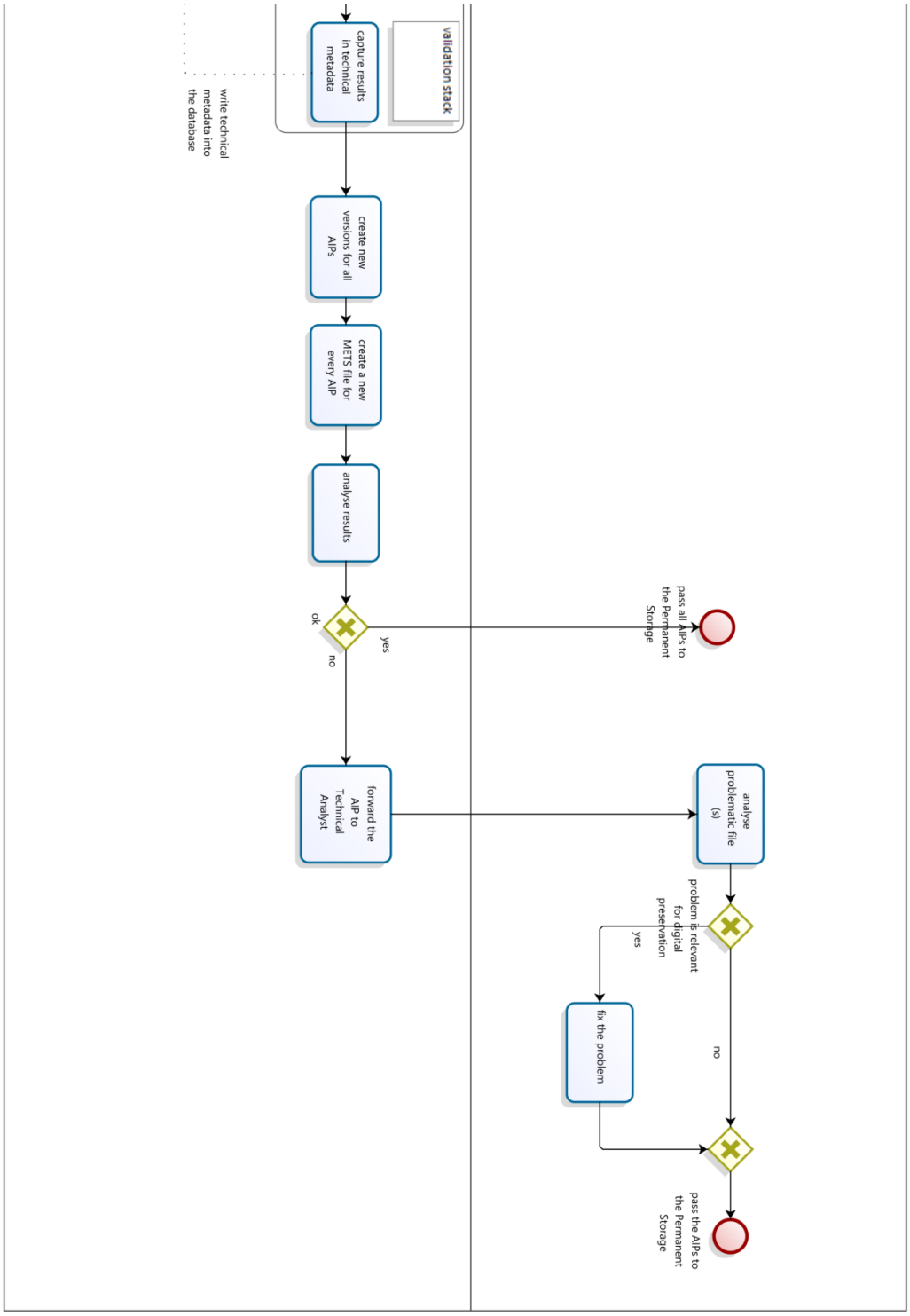
Appendix B Detailed Preservation Planning with Rosetta

Also available under: https://wiki.tib.eu/confluence/display/lza/Preservation+Management




create itemised test set
choose migration tool
run test
analyse results
results fulfil requirements
yes
no
refine and elaborate alternative
decision
Plan signed off
start Preservation Action
lock all chosen AIPs
copy all chosen AIPs to the Operational Storage
execute Preservation Action
create Modified Preservation Master
Modified Preservation Master


create Modified Preservation Master
Modified Preservation Master
evaluation
auto evaluation
commit changes
write administrative and Preservation metadata into the database
evaluate results
commit changes
create checksums
virus check
conduct file format identification
conduct object validation
extract technical metadata
capture results in technical metadata
validation stack
write technical metadata into the database
for further information see DURAARK deliverable D6.2 page 19, figure 3


validation stack
capture results in technical metadata
write technical metadata into the database
create new versions for all AIPs
create a new METS file for every AIP
analyse results
ok
yes
no
pass all AIPs to the Permanent Storage
forward the AIP to Technical Analyst
analyse problematic file (s)
problem is relevant for digital preservation
yes
no
fix the problem
pass the AIPs to the Permanent Storage

Appendix C View-Paths Tests

	View Path	Tested Video (internal MAM-ID)	Comment
Container: MPEG-4 Video: AVC Audio: AAC, Version 4	Intel® Core™ i7-8700– Windows 10, Version 1803 – Windows Media Player Version 12	15432, 15463	
	Intel® Core™ i5-3470 – Windows 8.1 Pro, Version 6.3 – VLC Media Player Version 2.2.2		
Container: WebM Video: VP8 Audio: Vorbis	Intel® Core™ i7-8700– Windows 10, Version 1803 – MPC-HC Version 1.7.13	15241 WebM-Derivative, 15398 WebM-Derivative	
	Intel® Core™ i5-3470 – Windows 8.1 Pro, Version 6.3 – VLC Media Player Version 2.2.2		
	Intel® Core™ i7-8700– Windows 10, Version 1803 – Windows Media Player Version 12		Skipping not possible
Container: MPEG-PS Video: MPEG Video, Format version : Version 2 Audio: MPEG Audio, Format version : Version	Intel® Core™ i7-8700– Windows 10, Version 1803 – Windows Media Player Version 12	14469, 14492	
	Intel® Core™ i5-3470 – Windows 8.1 Pro, Version 6.3 – VLC Media Player Version 2.2.2		

Appendix D Test Protocol Migration Plugin

The input files, output files, their MediaInfo report and logs can be found in Attachment 1: CD Migration Plugin. The source code is additionally published at https://github.com/TIBHannover/ffv1_Migration_Tool/

Input Files

	File Name	Container	Video Codec	Audio Codec
A	10310	MPEG-4	AVC	AAC, Version 4
B	15854	WebM	VP8	Vorbis
C	10368	MPEG-PS	MPEG Video, Version 2	MPEG Audio, Version 1

Test Number

	Parameter Type	Custom Parameter
1	No parameter	
2	Valid parameter	4
3	Multiple Parameter, first one valid*	16 8 19 (file B) 9 test (file C)
4	Invalid Parameter	8

*conducted with Files B and C

Test method description

	Test method
AO	Analyze output files with MediaInfo
AL	Analyze logs
MG	Monitor GUI (Number of Screenshot)
AS	Analyze sourcecode available at GibHub
AD	Analyze documentation available at GibHub

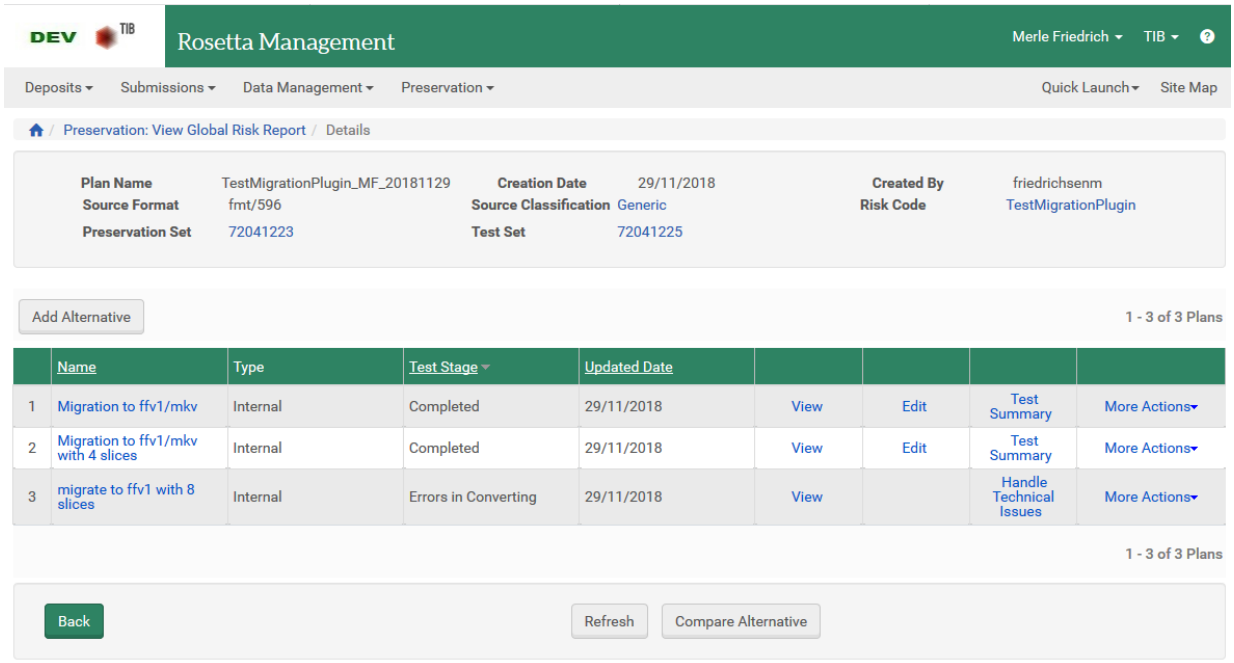
Evaluation of the Requirements

ID	Requirement	Test method	Test number	fulfilled
F	Functional Requirements			
F-1	The plugin must migrate all video and audio streams in the input file to the following format: Container: Matroska (mkv) Videocodec: ffv1, Version 3 Audiocodec: signed PCM, 24 bit, little endian	AO, AS	1,2	yes
F-2	The video codec must use the value 1 for the Group Of Picture (GOP)	AO, AS	1,2	yes
F-3	The video codec must use the value 1 (Range coder) for the coder	AO, AS	1,2	yes
F-4	The video codec must use the value 1 (large context) for context	AS	1,2	yes
F-5	The video codec must use CRC per slice	AO, AS	1,2	yes
F-6	Default value for slices is 16	AO, AS	1	yes
F-7	A valid custom parameter must influence the number of slices per frame	AO, AS	2	yes
F-8	An invalid custom parameter must be reported to the log	AL	3,4	yes
F-9	An invalid custom parameter must not migrate the file	AL, AO	3,4	yes
F-10	An invalid custom parameter must lead to an return value other than 0	AS, MG (I)	3,4	yes
F-11	The audio codec must be signed	AO, AS	1,2	yes
F-12	The audio codec must be little endian	AO, AS	1,2	yes
F-13	The audio codec must have a bit rate of 24 kbits/s	AO, AS	1,2	yes
F-14	The plugin must migrate the input when choosen as an alternative, internal plugin in a preservation plan	AL	1,2	yes
E-	External Interfaces			
E-1	The plugin must accept the parameter transferred by Rosetta: $n = output directory path	AL	1,2,3,4	yes
E-2	The plugin must accept the parameter transferred by Rosetta: $(n-1) = input file name/path	AL	1,2,3,4	yes
E-3	The plugin must accept the parameter transferred by Rosetta: $1 - $(n-2) = custom parameters (from Transformation Profile UI)	AL	2,3,4	yes

E-4	The user should be able to influence the number of slices.	AO, AS, MG (II)	2	Yes
E-5	“Each plugin JAR file must contain a metadata XML file in the /PLUGIN-INF/ directory. The filename should contain 'metadata' and have an xml extension (e.g. metadata_myFirstPlugin.xml)”.	AS	general	yes
E-6	“In case of a script error (or an initiated exit with a return value other than 0) - the script’s echo messages will be printed to the log as an ERROR message in the following format: Execution of{script_full_path} failed: {echo_message}”	AL	3,4	yes
E-7	“Scripting language should be specified using the Shebang line.”	AS	general	yes
E-8	A message on the start of the plugin should appear in the Rosetta server logs	AL	1,2,3,4	yes
E-9	If errors do occur a message should appear in the Rosetta server logs	AL	3,4	yes
E-10	A message on the success of a migration should appear in the Rosetta server logs	AL	1,2	yes
P-	Performance			
P-1	The plugin must be available anytime Rosetta is available.	AL	1,2,3,4	yes
P-2	The plugin should log the start of the process within the first 20 Minutes after “Run Test” was clicked in the GUI.	AL	1,2,3,4	no*
S-	Software System Attributes			
S-1	External documentation of the custom parameters must be available.	AD	general	yes
S-2	Comments in the script shall illustrate the concept of the plugin.	AS	general	yes
D-	Design Constraints			
D-1	The plugin must run on a Solaris server.	AL	1,2,3,4	yes

*the logs show that all messages from the script are printed into the logs after the script was completed, therefore the intended message does only show up after the script ends. Nevertheless the FFmpeg command leads to output which is shown in the log. Although this is not intended, this can be used as indicator that the migration started.

Screenshot I:



Screenshot II: