



LUDWIGSBURG

# **Musterworkflow zur Bearbeitung von Dateisammlungen für DIMAG-Anwender**

Stadtarchiv Ludwigsburg  
Carolin Leitzbach M.A.

Medienkonservatorin

22.03.2023

# Drei Phasenmodell



Kai Naumann: Welche Schritte erfordert die Aufbereitung von Dateisammlungen und welche Querschnitts- und Spezialwerkzeuge werden gebraucht?, in: Kai Naumann; Michael Puchta: Kreative digitale Ablagen und die Archive, Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23. November 2016 in der Generaldirektion der Staatlichen Archive Bayerns, München 2017



# 1. Analyse

## 1.1 Grobe Vorbewertung

- Was wird von wem angeboten?
- Wie wird es angeboten? (CDs, Dateiablage, Datenbank, etc.)
- Sind archivwürdige Informationen zu erwarten?

## 1.2 Virenscan

- Bei Übernahme externer Datenträger
- Mit hauseigenem Virens Scanner

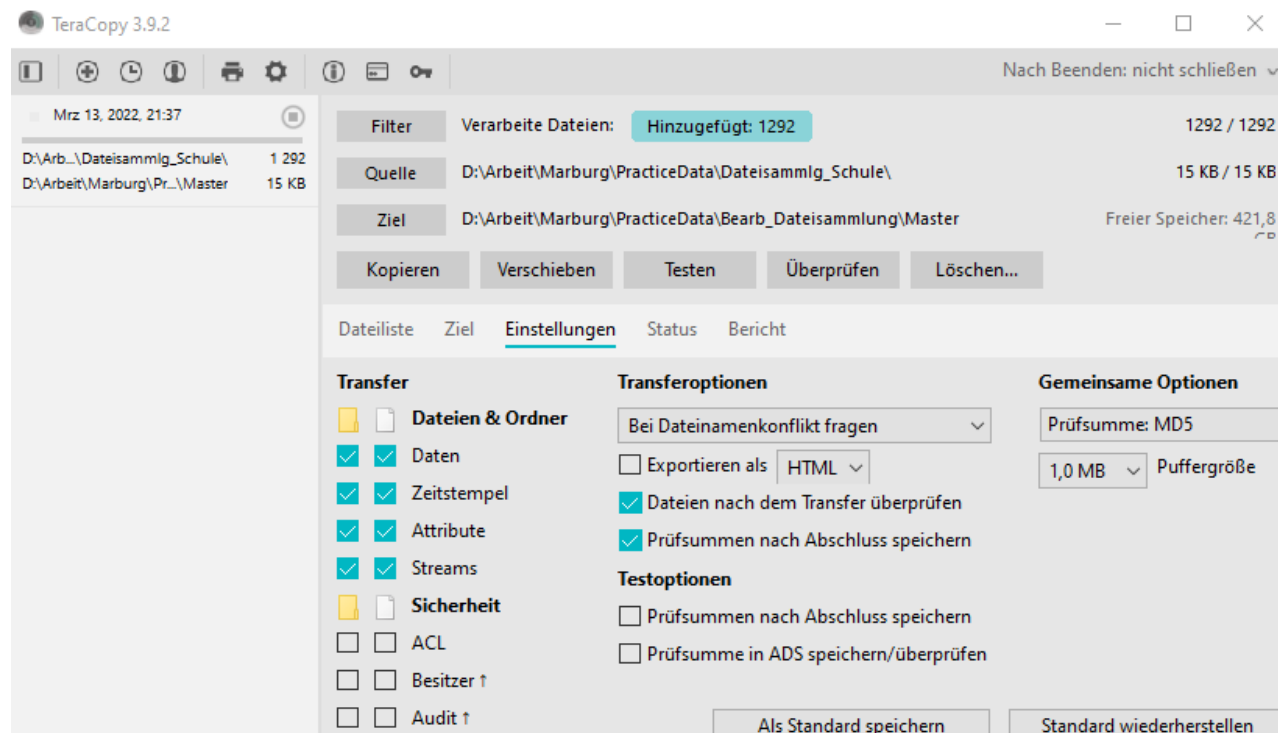
# 1. Analyse

## 1.3 Prüfsummen und Backup

- Erstellung einer Masterkopie und einer Arbeitskopie
- Die Kopiererstellung sollte prüfsummenkontrolliert durchgeführt werden, um Kopierfehler auszuschließen.

# 1. Analyse

## 1.3 Prüfsummen und Backup



Tool: TeraCopy

- Freeware-Tool
- Einsatzzweck: Kopiert ganze Ordnerstrukturen prüfsummenkontrolliert.

# 1. Analyse

## 1.3 Prüfsummen und Backup

Tool: FreeCommander

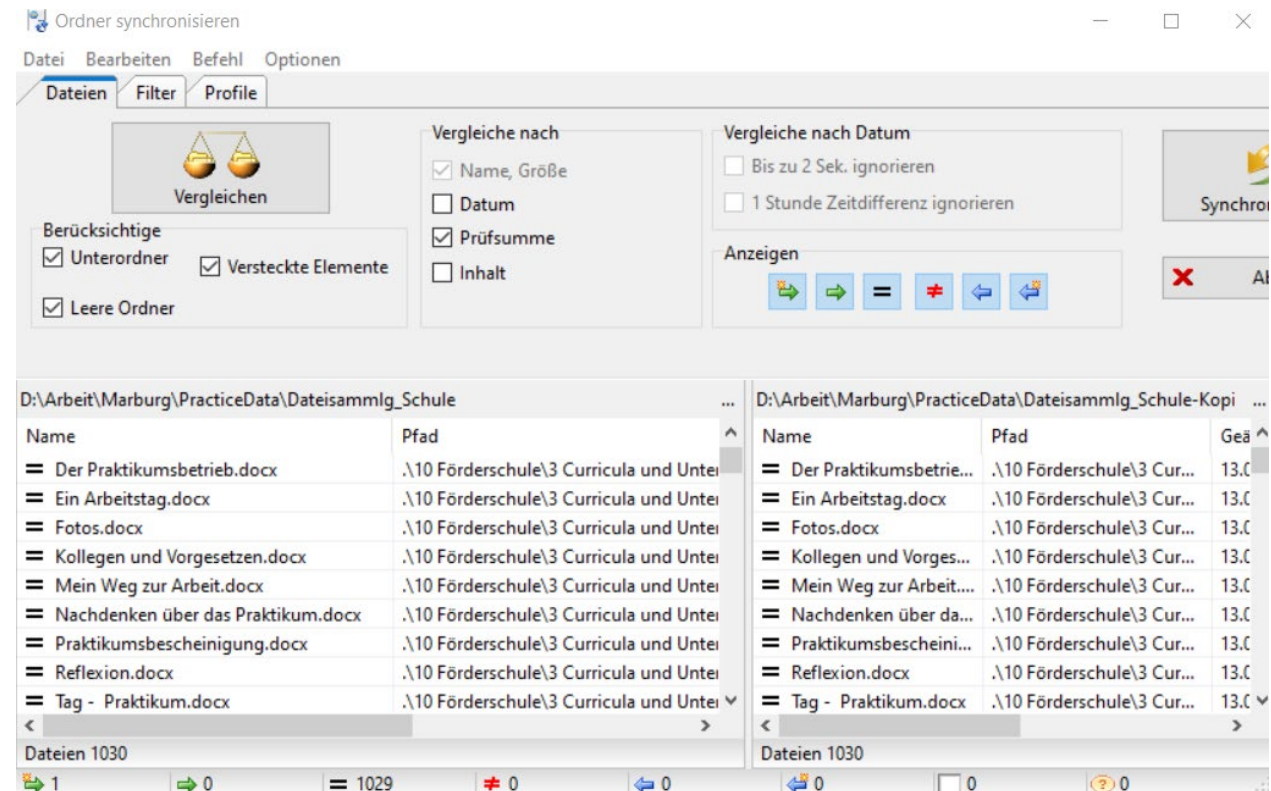
- Freeware-Tool

- Einsatzzweck

- Prüfsummen vergleichen

- Mehrfachumbenennen-Tool:  
Namen normalisieren,  
Sonderzeichen bearbeiten

- Ordnerstrukturen abbauen



# 1. Analyse

## 1.4 Analyse der Ordner und Dateien

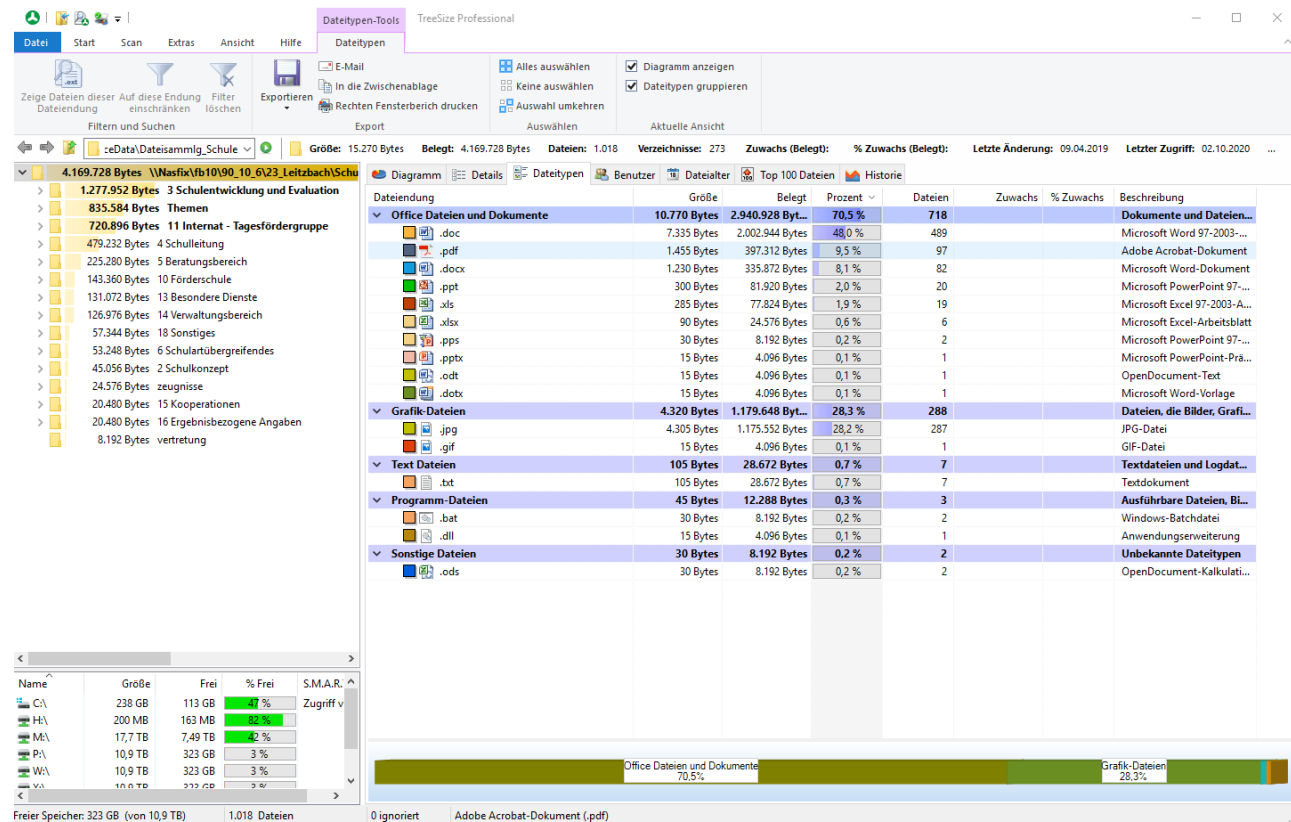
- Containerformate entpacken
- Überblick über Dateiformate erhalten
- Verteilung des Speichervolumens
- Erzeugen einer Dateiliste

# 1. Analyse

## 1.4 Analyse der Ordner und Dateien

Tool: TreeSize Professional

- Kostenpflichtiges Tool (ca. 50€)
- Einsatzzweck:
  - Überblick über Ordner/Dateien erhalten
  - Speicherverteilung auslesen
  - Duplikate/Sonderformate finden/löschen



The screenshot shows the TreeSize Professional interface. On the left, a file tree displays a directory structure with folders like 'Schulentwicklung und Evaluation' and 'Themen'. The main window shows a detailed analysis of file types, including a table with columns for file ending, size, growth, and percentage.

Dateiendung	Größe	Belegt	Prozent	Dateien	Zuwachs	% Zuwachs	Beschreibung
<b>Office Dateien und Dokumente</b>	<b>10.770 Bytes</b>	<b>2.940.928 Bytes</b>	<b>70,5 %</b>	<b>718</b>			<b>Dokumente und Dateien...</b>
.doc	7.335 Bytes	2.002.944 Bytes	48,0 %	489			Microsoft Word 97-2003...
.pdf	1.455 Bytes	397.312 Bytes	9,5 %	97			Adobe Acrobat-Dokument
.docx	1.230 Bytes	335.872 Bytes	8,1 %	82			Microsoft Word-Dokument
.ppt	300 Bytes	81.920 Bytes	2,0 %	20			Microsoft PowerPoint 97-...
.xls	285 Bytes	77.824 Bytes	1,9 %	19			Microsoft Excel 97-2003-A...
.xlsx	90 Bytes	24.576 Bytes	0,6 %	6			Microsoft Excel-Arbeitsblatt
.pps	30 Bytes	8.192 Bytes	0,2 %	2			Microsoft PowerPoint 97-...
.pptx	15 Bytes	4.096 Bytes	0,1 %	1			Microsoft PowerPoint-Prä...
.odt	15 Bytes	4.096 Bytes	0,1 %	1			OpenDocument-Text
.dotx	15 Bytes	4.096 Bytes	0,1 %	1			Microsoft Word-Vorlage
<b>Grafik-Dateien</b>	<b>4.320 Bytes</b>	<b>1.179.648 Bytes</b>	<b>28,3 %</b>	<b>288</b>			<b>Dateien, die Bilder, Grafi...</b>
.jpg	4.305 Bytes	1.175.552 Bytes	28,2 %	287			JPG-Datei
.gif	15 Bytes	4.096 Bytes	0,1 %	1			GIF-Datei
<b>Text Dateien</b>	<b>105 Bytes</b>	<b>28.672 Bytes</b>	<b>0,7 %</b>	<b>7</b>			<b>Textdateien und Logdat...</b>
.bat	105 Bytes	28.672 Bytes	0,7 %	7			Textdokument
<b>Programm-Dateien</b>	<b>45 Bytes</b>	<b>12.288 Bytes</b>	<b>0,3 %</b>	<b>3</b>			<b>Ausführbare Dateien, Bl...</b>
.bat	30 Bytes	8.192 Bytes	0,2 %	2			Windows-Batchdatei
.dll	15 Bytes	4.096 Bytes	0,1 %	1			Anwendungsweiterung
<b>Sonstige Dateien</b>	<b>30 Bytes</b>	<b>8.192 Bytes</b>	<b>0,2 %</b>	<b>2</b>			<b>Unbekannte Dateitypen</b>
.ods	30 Bytes	8.192 Bytes	0,2 %	2			OpenDocument-Kalkulati...



## 2. Nachbewertung und SIP-Formierung

- Auswahl der Dateien und Verzeichnisse für die Archivierung
- Formaterkennung und anschließende Migration und Validierung
- Formierung der Archivalieneinheiten und Anreicherung mit Metadaten

## 2. Nachbewertung und SIP-Formierung

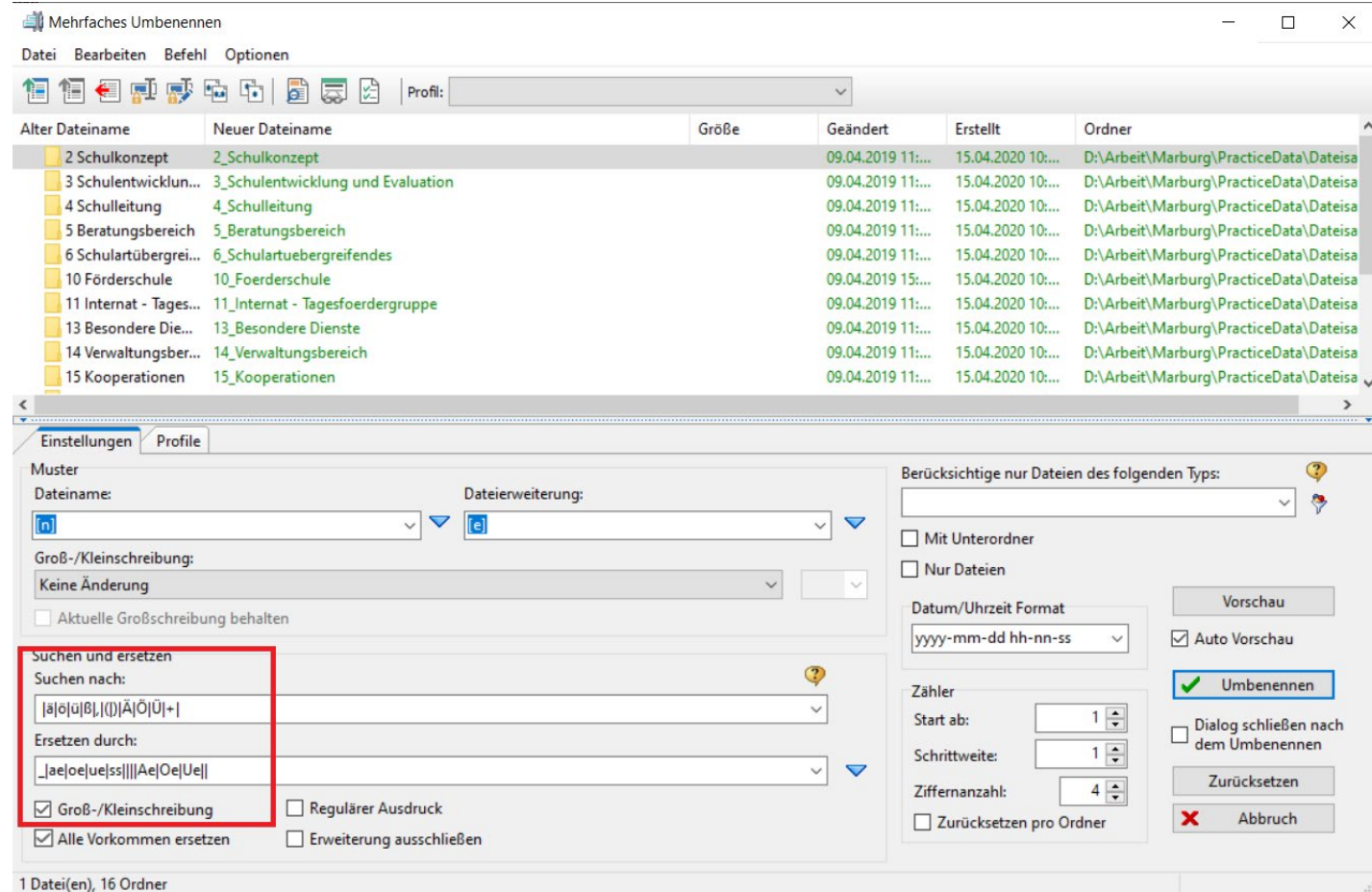
### 2.1 Namen normalisieren

- Normalisierung bedeutet das Ersetzen von Umlauten und anderen Sonderzeichen, die teilweise beim Arbeiten mit weiterer Software Probleme bereiten können.
- Erlaubt sind folgende Zeichen: A-Z, a-z, 0-9, \_ und -

# 2. Nachbewertung und SIP-Formierung

## 2.1 Namen normalisieren

Tool: FreeCommander



Mehrfaches Umbenennen

Datei Bearbeiten Befehl Optionen

Alter Dateiname	Neuer Dateiname	Größe	Geändert	Erstellt	Ordner
2 Schulkonzept	2_Schulkonzept		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
3 Schulentwicklun...	3_Schulentwicklung und Evaluation		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
4 Schulleitung	4_Schulleitung		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
5 Beratungsbereich	5_Beratungsbereich		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
6 Schulartübergrei...	6_Schulartuebergreifendes		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
10 Förderschule	10_Foerderschule		09.04.2019 15:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
11 Internat - Tages...	11_Internat - Tagesfoerdergruppe		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
13 Besondere Die...	13_Besondere Dienste		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
14 Verwaltungsber...	14_Verwaltungsbereich		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa
15 Kooperationen	15_Kooperationen		09.04.2019 11:...	15.04.2020 10:...	D:\Arbeit\Marburg\PracticeData\Dateisa

Einstellungen Profile

Muster

Dateiname: [n] Dateierweiterung: [e]

Groß-/Kleinschreibung: Keine Änderung

Aktuelle Großschreibung behalten

Suchen und ersetzen

Suchen nach: [a|ä|ö|ü|ß|(|)|Ä|Ö|Ü|+]

Ersetzen durch: [\_]a|e|o|e|s|s|\_|A|E|O|U|]

Groß-/Kleinschreibung  Regulärer Ausdruck

Alle Vorkommen ersetzen  Erweiterung ausschließen

Berücksichtige nur Dateien des folgenden Typs:

Mit Unterordner

Nur Dateien

Datum/Uhrzeit Format: yyyy-mm-dd hh-nn-ss

Auto Vorschau

Zähler

Start ab: 1

Schrittweite: 1

Zifferanzahl: 4

Zurücksetzen pro Ordner

Vorschau

Umbenennen

Dialog schließen nach dem Umbenennen

Zurücksetzen

Abbruch

1 Datei(en), 16 Ordner

## 2. Nachbewertung und SIP-Formierung

### 2.2 Nachbewertung

- Manuelle Nachbewertung
  - Händisches Verschieben der zu kassieren Dateien/Ordner in einen vorübergehenden „kassiert“-Ordner
  - Erzeugung einer Dateiliste des „kassiert“-Ordners zu Dokumentationszwecken
  - Tool: FreeCommander

## 2. Nachbewertung und SIP-Formierung

### 2.2 Nachbewertung

- Automatisierte Nachbewertung
  - Doppelte Verzeichnisse und Dateien finden und bei Bedarf löschen
  - Obsolete Dateien wie temporäre Dateien löschen
  - Benutzerdefinierte Datei- und Objekttypen löschen, falls notwendig
  - Tool: TreeSize Professional

## 2. Nachbewertung und SIP-Formierung

### 2.3 Formaterkennung

- Die Formaterkennung ist ein wichtiger Bestandteil der digitalen Erhaltung. Sie ist notwendig:
  - um zu ermitteln welches Format tatsächlich vorliegt.
  - um eine geeignete Software zur Sichtung bei unbekanntem Dateiformat zu finden.
  - um das Erhaltungsrisiko des Formats abzuschätzen.

## 2. Nachbewertung und SIP-Formierung

### 2.3 Formaterkennung

Tool: IngestList

Ändere:  IngestList:

DROID:

Jhove:

Multimedia:

File:

name	pronom-file-format	pronom-format-id	pronom-classification	pronom-mime
1791000_0_VN_1975019.pdf	Acrobat PDF 1.4 - Portable Doc...	fmt/18	Positive (Specific Format)	application/pdf
1791000_0_VN_1975020.pdf	Acrobat PDF 1.4 - Portable Doc...	fmt/18	Positive (Specific Format)	application/pdf
1791000_0_VN_1975021.pdf	Acrobat PDF 1.4 - Portable Doc...	fmt/18	Positive (Specific Format)	application/pdf
1791000_0_VN_1975022.pdf	Acrobat PDF 1.4 - Portable Doc...	fmt/18	Positive (Specific Format)	application/pdf
1791000_0_VN_1975023.pdf	Acrobat PDF 1.4 - Portable Doc...	fmt/18	Positive (Specific Format)	application/pdf
1791000_0_VN_1976000.pdf	Acrobat PDF 1.4 - Portable Doc...	fmt/18	Positive (Specific Format)	application/pdf

# 2. Nachbewertung und SIP-Formierung

## 2.3 Formaterkennung

Tool: Droid

DROID v6.2.1

File Edit Run Filter Report Tools Help

New Open Save Export Add Remove Start Pause Filter On Report

Untitled-2

Resource	Extension	Size	Last modified	Ids	Format	Version	Mime type	PUID	Method	Hash
D:\Works...			11.02.19 13:30							
Instagra...	gz	41,2 MB	03.09.18 14:33		GZIP Format		application/gzip	<a href="#">x-fmt/266</a>	Signature	
Beteiligu...	pdf	1,8 MB	10.03.15 09:17	(38)	"Acrobat PDF/E ... "2u", "3a", "3b",..."		application/pdf	"fmt/20", "fmt/1...	Extension	
Beteiligu...	pdf	1,8 MB	09.02.18 10:39		Acrobat PDF 1.4...	1.4	application/pdf	<a href="#">fmt/18</a>	Signature	
fake.pdf	pdf	17 bytes	08.02.19 14:19		Acrobat PDF 1.6...	1.6	application/pdf	<a href="#">fmt/20</a>	Signature	
Filelist_...	bxt	3,4 KB	08.02.19 14:19	(2)	"TRIM Context R...		text/plain	"fmt/1085", "x-f...	Extension	
JPEG_20...	jpg	1,8 MB	16.01.19 13:47		JPEG File Inter...	1.02	image/jpeg	<a href="#">fmt/44</a>	Signature	
JPEG_20...	tiff	1,8 MB	16.01.19 13:47		JPEG File Inter...	1.02	image/jpeg	<a href="#">fmt/44</a>	Signature	
md5s20...		1,6 KB	08.02.19 14:19							
Standpu...	pdf	1,4 MB	07.08.18 17:46		Acrobat PDF 1.4...	1.4	application/pdf	<a href="#">fmt/18</a>	Signature	
Videogr...	mkv	1,7 GB	27.11.18 15:42		Matroska	1-4		<a href="#">fmt/569</a>	Signature	
Videogr...	mts	173,7 MB	16.10.14 11:42		M2TS			<a href="#">fmt/1055</a>	Signature	

tiff

Unstimmigkeit zwischen Extension und Dateiformat

(37)

Kein eindeutiges Ergebnis



Eindeutiges Ergebnis



Erkennung nicht möglich



## 2. Nachbewertung und SIP-Formierung

### 2.4 Formate vereinheitlichen

- Veraltete Dateiformate, die nicht mehr von der gängigen Software unterstützt werden, stellen eines der größten Risiken bei der digitalen Archivierung dar.
- Je einfacher der Aufbau, je weniger Variante, Eigenschaften und Eigenheiten ein Format hat, umso besser eignet es sich für den Erhalt digitaler Informationen.

## 2. Nachbewertung und SIP-Formierung

### 2.4 Formate vereinheitlichen

Format	Empfehlung	Begründung
JPEG	Nicht konvertieren	Verlust schon vorhanden, JPEG ist ein bekanntes Format
TIFF	Nicht konvertieren	Archivformat
PDF	Nach PDF/A-2b konvertieren	Nicht archivfähig
PDF/A-1a/b	Nicht konvertieren	Es wäre nur ein Versionswechsel
PDF/A-2a/b/u	Nicht konvertieren	Archivformat
Word	Nach PDF/A-2b konvertieren	Nicht archivfähig
Excel	Nur wenn die hinterlegten Formeln nicht wichtig sind, konvertieren nach PDF/A-2b Möglich auch: CSV	
PowerPoint	Nur wenn die Funktionen nicht wichtig sind, konvertieren nach PDF/A-2b	

## 2. Nachbewertung und SIP-Formierung

### 2.4 Formate vereinheitlichen

Format	Empfehlung	Begründung
MP3	Nicht konvertieren	Verlust schon vorhanden, MP3 ist ein bekanntes Format
Wave	Nicht konvertieren	Archivformat
MP4	Nicht konvertieren	Verlust schon vorhanden, MP4 ist ein bekanntes Format
MKV FFv1.3	Nicht konvertieren	Archivformat

Für weitere Informationen und Empfehlungen für die Migration von Videoformaten siehe:  
[https://kost-ceco.ch/cms/kad\\_recommendation\\_de.html](https://kost-ceco.ch/cms/kad_recommendation_de.html) und  
<https://memoriav.ch/de/film/empfehlungen-film/>



## 2. Nachbewertung und SIP-Formierung

### 2.4 Formate vereinheitlichen

Archivformate	Empfehlung
JPEG, TIFF, PNG	Irfanview
PDF/A	Adobe Pro oder PDF24
MKV FFV1	FFMPEG

## 2. Nachbewertung und SIP-Formierung

### 2.5 Formate validieren

- Das Layout eines Dateiformats wird über die Formatspezifikation bestimmt. Diese legt fest, welche Informationen an welcher Stelle innerhalb der Datei hinterlegt sind.
- Die Validierung überprüft, ob auch der inhaltliche Aufbau einer Datei mit den vorgegebenen Spezifikationen übereinstimmt.
- Validatoren werden für einzelne Dateiformate entwickelt.

## 2. Nachbewertung und SIP-Formierung

### 2.5 Formate validieren

Format	Tool
PDF	JHOVE
PDF/A	VeraPDF, Kost-Val oder Preflight
JPEG-/TIFF-Dateien	Kost-Val, DPF Manager
WAVE-Format	shntool
MKV-Format	MediaConch

## 2. Nachbewertung und SIP-Formierung

### 2.6 SIP-Formierung

- Gruppierung von Informationsobjekten, beispielsweise auf Basis der Ordnerstruktur oder der Dateinamen
  - Tool: IngestTool
- Händische Formierung nur im Ausnahmefall, wenn die vorliegende Struktur ungenügend ist.
  - Tool: FreeCommander

## 2. Nachbewertung und SIP-Formierung

### 2.6 SIP-Formierung

- Anreicherung der Metadaten beispielsweise durch eine csv-Tabelle mit weiteren Informationen zu den abgelieferten Inhalten.

	A	B	C	D	E	F	G	H	I	J
1	Pfad	Ordnername	Signatur	Titel	Enthält:	Umfang	Datum von	Datum bis	Sperrfrist - Berechnung	
2	D:\Arbeit\Marburg\PracticeData\Dateisamml	10_Foerderschule	1	Förderschule	Geschäftsordnung, Berufs	16	2018	2018	2048	
3	D:\Arbeit\Marburg\PracticeData\Dateisamml	11_Internat - Tagesfoerdergruppe	2	Internat	Wohngruppen	159	2018	2018	2048	
4	D:\Arbeit\Marburg\PracticeData\Dateisamml	13_Besondere Dienste	3	Besondere Dienste	Hompage, IT, Medizinisch	32	2018	2018	2048	
5	D:\Arbeit\Marburg\PracticeData\Dateisamml	14_Verwaltungsbereich	4	Verwaltungsbereich		31	2018	2018	2048	
6	D:\Arbeit\Marburg\PracticeData\Dateisamml	15_Kooperationen	5	Kooperation		5	2018	2018	2048	
7	D:\Arbeit\Marburg\PracticeData\Dateisamml	16_Ergebnisbezogene Angaben	6	Ergebnisbezogene Angaben		5	2018	2018	2048	
8	D:\Arbeit\Marburg\PracticeData\Dateisamml	18_Sonstiges	7	Sonstiges		14	2018	2018	2048	
9	D:\Arbeit\Marburg\PracticeData\Dateisamml	2_Schulkonzept	8	Schulkonzept		11	2018	2018	2048	
10	D:\Arbeit\Marburg\PracticeData\Dateisamml	3_Schulentwicklung und Evaluation	9	Schulentwicklung und Evaluation		248	2018	2018	2048	
11	D:\Arbeit\Marburg\PracticeData\Dateisamml	4_Schulleitung	10	Schulleitung		73	2018	2018	2048	
12	D:\Arbeit\Marburg\PracticeData\Dateisamml	5_Beratungsbereich	11	Beratungsbereich		55	2018	2018	2048	
13	D:\Arbeit\Marburg\PracticeData\Dateisamml	6_Schulartuebergreifendes	12	Schulartenübergreifendes		13	2018	2018	2048	
14	D:\Arbeit\Marburg\PracticeData\Dateisamml	Themen	13	Themen		151	2018	2018	2048	
15	D:\Arbeit\Marburg\PracticeData\Dateisamml	vertretung	14	Vertretung		2	2018	2018	2048	
16										
17										
18										



## 3. SIP-Erzeugung für digitales Archiv und AFIS

- Ingest der Pirmärdaten und Metadaten in DIMAG, sowie Weitergabe der Metadaten an das AFIS.
- Eingesetztes Tool: IngestTool oder IngestList

# Weitere Informationen

## Nestor-Netzwerk

- <https://www.langzeitarchivierung.de>

## Linksammlung für den Einstieg:

- [https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor\\_Einstieg/faq\\_table.html;jsessionid=7A7662577192A078904E5C6E3C16566C.intranet672](https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Einstieg/faq_table.html;jsessionid=7A7662577192A078904E5C6E3C16566C.intranet672)

## Ansätze für die Übernahme von Dateisammlungen:

- [https://blha.brandenburg.de/wp-content/uploads/2020/03/Brandenburgische\\_Archive\\_36.pdf](https://blha.brandenburg.de/wp-content/uploads/2020/03/Brandenburgische_Archive_36.pdf)
- <https://www.kultur-machen.de/Dateisammlungen>



LUDWIGSBURG

**Vielen Dank für Ihre Aufmerksamkeit!**