

# Formatauswahl

*denn sie wissen was sie tun*

nestor-Praktikertag 2021,  
virtuell,  
09.06.2021

Claire Röthlisberger-Jourdan



# Disposition

- Dateiformate
- Sinn und Zweck der Formatauswahl
- Welche Formatkategorien?
- Informationsquellen
- Katalog archiverischer Dateiformate (KaD)

# Die KOST

Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen

- Gemeinschaftsunternehmen diverser Archive der Schweiz sowie des Landesarchivs des Fürstentums Liechtenstein.
- Sie hat den Auftrag, ihre Träger bei der Archivierung elektronischer Unterlagen zu unterstützen
  - Standards und Richtlinien
  - Tools
  - Dienstleistungen zur Lösung konkreter Probleme
  - Studien und Veranstaltungen

# Die KOST

## Geschäftsstelle

- Georg Büchler, Claire Röthlisberger-Jourdan, nn

## PPEG (Preservation Planning Expert Group)

- Lambert Kansy, Staatsarchiv Basel Stadt
- Martin Lüthi, Staatsarchiv St. Gallen
- Kaya Demiroglu, Staatsarchiv Uri
- Romano Padeste, Staatsarchiv Zürich
- Pascal Schneiter, Staatsarchiv Neuenburg
- Daniel Jolliet, Bundesarchiv
- Geschäftsstelle der KOST

# Dateiformate

An welche Dateiendungen denken Sie beim Stichwort Dateiformate für die LZA?

➤ *Scheibt diese in den Chat*

A word cloud of file extensions. The most prominent words are 'pdf' (blue), 'jpg' (green), 'mp4' (red), 'mp3' (yellow), 'mov' (red), 'html' (purple), 'avi' (red), 'tiff' (green), 'xml' (purple), 'doc' (blue), 'wav' (yellow), 'mkv' (red), 'warc' (purple), 'jpx' (green), 'exe' (black), 'flac' (yellow), 'odt' (blue), 'xsl' (grey), 'sql' (grey), 'siard' (grey), 'arc' (purple), 'zip' (grey), 'txt' (blue), and 'csv' (grey).

# Sinn und Zweck der Formatauswahl

## Pro Format

- Wissen aneignen
  - Vor- und Nachteile
  - Risiken kennen
- Toolset erarbeiten
  - Erstellen
  - Betrachten
  - Erkennen
  - Exportieren
  - Validieren
  - *Beheben*

## ➤ Aufwand

# Sinn und Zweck der Formatauswahl

- Formatauswahl ist auch eine Strategiefraage
    - Langfristige Entscheide und Auswirkungen
  - Reduktion der Formate auf eine handhabbare Anzahl
    - Mind. ein Format pro vorhandene Formatkategorie
    - Entscheiden welche Formate im Archiv sein sollen
    - Abwägung der Risiken und Aufwand
    - Aufbauen des dazugehörenden Toolsets
- Trefft bewusste Entscheide und setzt diese um

# Sinn und Zweck der Formatauswahl

- *Haben Sie noch Fragen oder Ergänzungen zu «Sinn und Zweck der Formatauswahl»?*
- *Wenn Ja schreibt diese in den Chat*



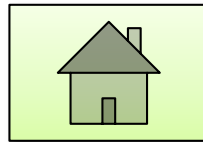
# Welche Formatkategorien?

- Unbestrittene Formatkategorien:

- Text



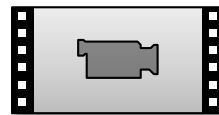
- Bild



- Audio



- Video



# Welche Formatkategorien?

- Weitere Formatkategorien je nach Archiv:
  - Tabellenkalkulation
  - Datenbanken
  - Hypertext
  - GIS
  - CAD/CAM
  - Medizinische Bildformate
  - eBook
  - Statistikdaten
  - ...

# Welche Formatkategorien?

- *Haben Sie noch Fragen oder Ergänzungen zu «Formatkategorien»?*
- *Wenn Ja schreibt diese in den Chat*

# Informationsquellen

Formatkataloge:

- Library of Congress, Sustainability of Digital Formats  
<https://www.loc.gov/preservation/digital/formats/>
- Katalog archivischer Dateiformate (KaD)  
[https://kost-ceco.ch/cms/kad\\_main\\_de.html](https://kost-ceco.ch/cms/kad_main_de.html)

# Informationsquellen

Informationen zu einzelnen Formaten / Formatgruppen:

- Just Solve the File Format Problem  
<http://fileformats.archiveteam.org/>
- Wikipedia  
<https://en.wikipedia.org/>

# Informationsquellen

- *Haben Sie noch Fragen oder Ergänzungen zu «Informationsquellen»?*
- *Wenn Ja schreibt diese in den Chat*

# Katalog archivischer Dateiformate (KaD)

KOST Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen

DE | FR

Suchen

Aktuell

Standards

Dateiformate  
(KaD)

Projekte

Publikationen

Tools & Services

Einleitung

Empfehlung

Text

Bild

Audio

Video

Tabellenkalkulation

Datenbanken

Hypertext

GIS

CAD/CAM

Weitere Kategorien

Bewertung

Hintergrund

Dateiformate (KaD)

## Katalog archivischer Dateiformate

### Einleitung

Der Katalog archivischer Dateiformate KaD beschreibt 52 wichtige und verbreitete Dateiformate und analysiert ihre Eignung für die digitale Archivierung.

### Empfehlung

Aufgrund der im KaD detailliert erläuterten Analysen, Bewertungen und Schlussfolgerungen empfiehlt die KOST für die verschiedenen Formatkategorien eine Auswahl von Dateiformaten als gegenwärtig archivtauglichste Lösung.

### Text

Unter "Text" verstehen wir Daten zur Repräsentation von wenig strukturiertem Text, wie sie u.a. von Textverarbeitungssoftware oder Texteditoren erzeugt werden.

### Bild

Die detaillierten Formatanalysen und -vergleiche des KaD beschränken sich im Bereich der Bilder auf Rastergrafiken.

### Audio

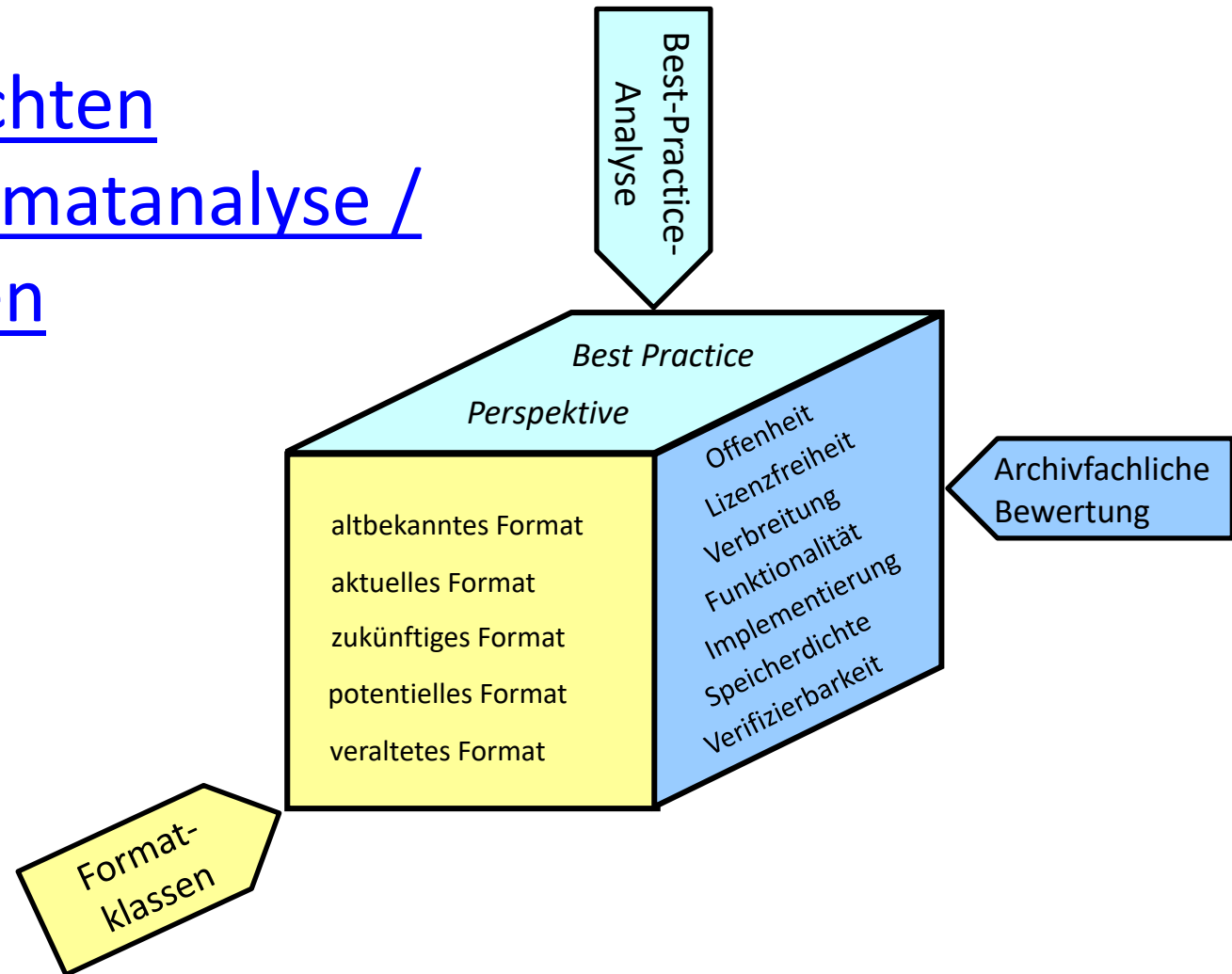
Audiodaten enthalten Informationen, die durch das menschliche Ohr wahrgenommen werden sollen.

### Video

Bei Videos ist je nach Entstehungszusammenhang, Archivauftrag und Bewertungsentscheid jeweils ein bestimmtes Format das in diesem Fall am besten geeignete.

# Katalog archivischer Dateiformate (KaD)

- Drei Sichten  
der Formatanalyse /  
Kriterien





# Katalog archivischer Dateiformate (KaD)

- Kriterienkatalog

Zu jedem Kriterium wird festgehalten:

- Name
- Beschreibung
- Gewichtung
- Bedingung für vollständige Erfüllung

Kriterien	Erläuterung	Gewicht
Offenheit	Die Spezifikation des Formats ist vollständig und öffentlich zugänglich	1
Lizenzfreiheit	Der Gebrauch des Formats ist durch keine Lizenzen eingeschränkt oder einschränkbar	1
Verbreitung	Das Format ist weit verbreitet; deshalb können lange Migrationszyklen erwarten werden	1
Funktionalität	Das Format kann die signifikanten Eigenschaften seiner Kategorie repräsentieren	1
Implementierung	Für das Format existieren mehrere Implementierungen	0.5
Speicherdichte	Das Format ermöglicht eine hohe Speicherdichte	0.5
Verifizierbarkeit	Es existieren Methoden und Werkzeuge zur Erkennung und Validierung	1
Best Practice	Das Format wird in Archiven bereits verwendet	0.5
Perspektive	Das Format ist im archivischen Sinn zukunftsfähig und hat noch weiteres Potenzial	0.5
Formatklasse	Einteilung in eine der sechs Formatklassen: A: altbekanntes Format    D: potentielles Format B: aktuelles Format        X: veraltetes Format C: zukünftiges Format      Ø: keine Aussage	

# Katalog archivischer Dateiformate (KaD)

- [Kriterienkatalog](#)

Dateiformate (KaD) > Hintergrund > Kriterienkatalog > Verifizierbarkeit

## Verifizierbarkeit

---

### [Verifizierbarkeit:](#)

- *Es existieren Methoden und Werkzeuge, die verifizieren können, dass eine Datei den Formatspezifikationen entspricht.*
  - Gewichtung: 1
- 

Hier geht es um die Validierung der gesamten Datei. Die Formatvalidierung muss bestätigen, dass eine Datei den Spezifikationen des Formats entspricht. Dabei müssen auch als problematisch erachtete eingebettete Objekte erkannt werden können.

### Bewertung

1 keiner: Es existiert kein Validator für das Format.

2 partiell: Teilaspekte können verifiziert werden, z.B. die Formaterkennung mittels Bytesequenz.

3 einer: Es existiert ein Validator für das Format.

4 mehrere: Es existieren mehrere unabhängige Validatoren für das Format.

Katalog archivischer Dateiformate

Version 6.1, August 2020

# Katalog archivischer Dateiformate (KaD)

- Bewertungs-  
matrix

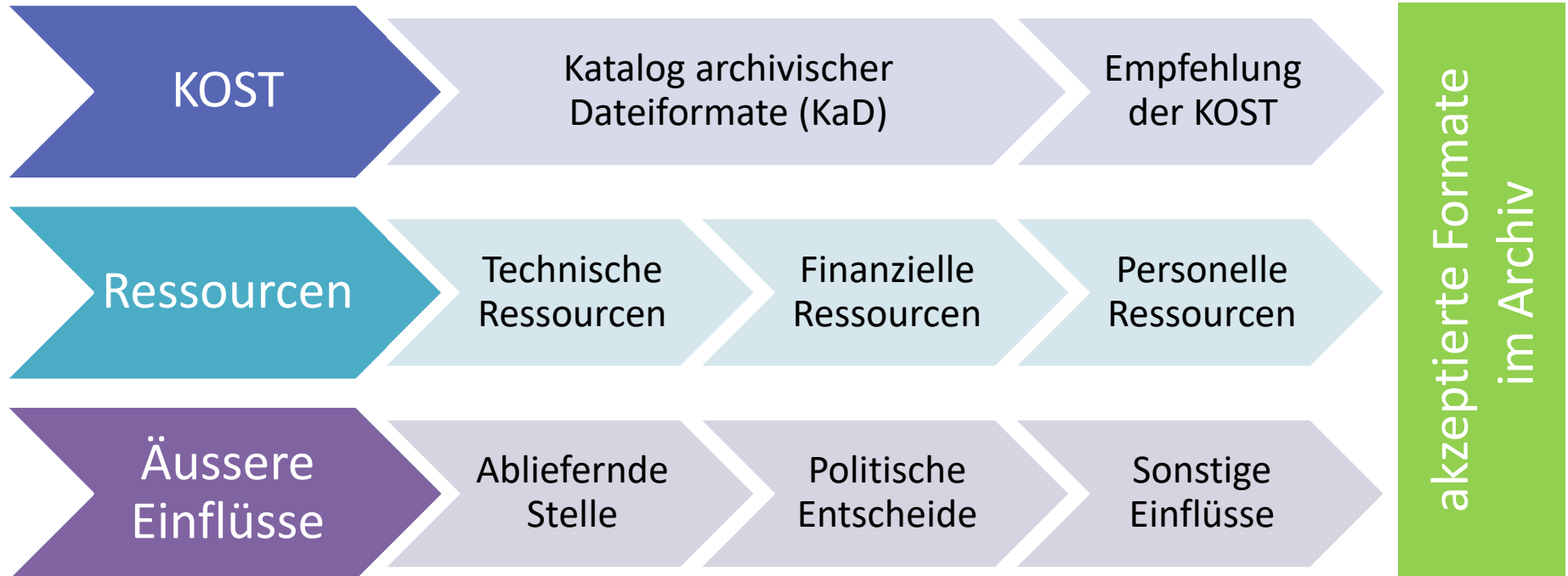
Kriterien	Erläuterung	Gewicht	Formate						
			Text						
			TXT	PDF	PDF/A-1	PDF/A-2	PDF/A-3	ODF	OOXML
Offenheit	Die Spezifikation des Formats ist vollständig und öffentlich zugänglich	1	4	4	4	4	4	4	4
Lizenzfreiheit	Der Gebrauch des Formats ist durch keine Lizenzen eingeschränkt oder einschränkbar	1	3	3	3	3	1	3	3
Verbreitung	Das Format ist weit verbreitet, deshalb können lange Migrationszyklen erwarten werden	1	4	4	4	4	1	2	4
Funktionalität	Das Format kann die signifikanten Eigenschaften seiner Kategorie repräsentieren	1	1	3	2	3	3	4	4
Implementierung	Für das Format existieren mehrere Implementierungen	0.5	4	4	4	4	3	4	3
Speicherdichte	Das Format ermöglicht eine hohe Speicherdichte	0.5	4	3	2	3	2	3	3
Verifizierbarkeit	Es existieren Methoden und Werkzeuge zur Erkennung und Validierung	1	2	4	4	4	2	3	2
Best Practice	Das Format wird in Archiven bereits verwendet	0.5	4	2	4	4	1	1	1
Perspektive	Das Format ist im archivischen Sinn zukunftsfähig und hat noch weiteres Potenzial	0.5	1	1	3	4	1	2	2
Formatklasse	Einteilung in eine der sechs Formatklassen: A: altbekanntes Format    D: potentielles Format B: aktuelles Format        X: veraltetes Format C: zukünftiges Format      Ø: keine Aussage		A	B	B	B	Ø	C	B
			TXT	PDF	PDF/A-1	PDF/A-2	PDF/A-3	ODF	OOXML
Gewichtete Summe / 4			0.73	0.82	0.84	0.91	0.52	0.75	0.77
Logarithmisch gewichtete Summe			0.86	1.36	1.55	2.87	0.41	0.94	1.04

# Katalog archivischer Dateiformate (KaD)

- Details zu den einzelnen Formaten:
  - Allgemeine Informationen
  - Beschreibung
  - Bewertung
  - Fazit
  - Referenz
  - Literatur

# Katalog archivischer Dateiformate (KaD)

- Einflüsse auf die Formatauswahl



# Katalog archivischer Dateiformate (KaD)

- Empfehlung



- Text

Text						
TXT	PDF	PDF/A-1	PDF/A-2	PDF/A-3	ODF	OOXML
0.73	0.82	0.84	0.91	0.52	0.75	0.77
0.86	1.36	1.55	2.87	0.41	0.94	1.04

- Text soll möglichst bereits im Amt in ein PDF-Format (PDF/A-1, PDF/A-2 oder PDF) umgewandelt und so archiviert werden.
- Eine Konvertierung von PDF-Dateien in ein PDF/A-Format ist zu einem späteren Zeitpunkt meist problemlos möglich.
- PDF/A-2u ist meine erste Wahl, da es die signifikanten Eigenschaften mit dem geringsten Aufwand am besten bewahrt.

# Katalog archivischer Dateiformate (KaD)

- Empfehlung

- Bild

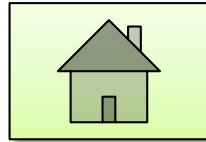


Bild					
TIFF	JPEG	JPEG2000	PNG	DNG	PDF/A-2
0.84	0.84	0.89	0.75	0.71	0.84
1.20	1.20	1.79	0.73	0.61	1.20

- Es gibt nicht ein einziges ideales Archivformat für Bilddaten. Es ist im Gegenteil unerlässlich, den konkreten Anwendungsfall und die archivische Bewertung in den Formatentscheid mit einzubeziehen.
- PDF/A-2 mit JPEG2000 ist meiner Meinung nach die beste Kombination bei neu zu erstellenden Digitalisaten.

# Katalog archivischer Dateiformate (KaD)

- [Empfehlung](#)

- Audio



Audio				
WAV	FLAC	ALAC	MP3	Ogg Vorbis
0.84	0.86	0.66	0.82	0.71
1.44	1.66	0.60	1.26	0.73

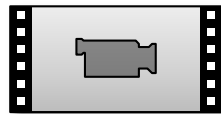
- Für die Archive empfiehlt es sich, entweder die anfallenden Audiodateien unkomprimiert im WAV-Format (genauer: WAVE PCM) oder als FLAC verlustfrei komprimiert zu archivieren.
- [FLAC](#) würde ich persönlich dem älteren WAV vorziehen.



# Katalog archivischer Dateiformate (KaD)

- Empfehlung

- Video



Video							
Uncompressed Video	Digital Video	MPEG-2	MPEG-4	MJPEG2000 (lossless)	MJPEG2000 (DCP)	ProRes	FFV1
0.64	0.61	0.63	0.67	0.66	0.69	0.52	0.69
1.00	0.91	0.97	1.12	1.08	1.21	0.69	1.21

- Bei Video ist die Migrationsvermeidung höher zu gewichten als die Reduktion der Formatanzahl im Archiv, und es können mehr Formate als (mindestens provisorisch) archivtauglich akzeptiert werden als für andere Formatkategorien.
- Existierende Videos zu konvertieren ist meistens nicht sinnvoll. Müssen Videos von vergänglichen Datenträgern (Kassetten, Minidisks etc.) übernommen werden, würde ich wenn möglich [FFV1](#), [MJPEG2000](#) oder [MPEG-4](#) wählen.

# Katalog archivischer Dateiformate (KaD)

- *Haben Sie noch Fragen oder Ergänzungen zum «Katalog archivischer Dateiformate»?*
- *Wenn Ja schreibt diese in den Chat*

# Katalog archivischer Dateiformate (KaD)

## Formatauswahl

*denn sie wissen was sie tun*

**KOST** Koordinationsstelle für die dauerhafte Archivierung  
elektronischer Unterlagen

---

Ein Gemeinschaftsunternehmen von Schweizer Archiven

## Claire Röthlisberger-Jourdan

Koordinationsstelle für die Archivierung elektronischer Unterlagen

KOST

c/o Schweizerisches Bundesarchiv

Archivstrasse 24

CH-3003 Bern

T +41 79 464 07 41

E [claire.roethlisberger@kost.admin.ch](mailto:claire.roethlisberger@kost.admin.ch)

W [kost-ceco.ch](http://kost-ceco.ch)