



„Was du heute kannst besorgen,
verschiebe nicht auf morgen.“

Preservation Planning als Bestandteil des
Pre-Ingest

Antje Lengnik



Gliederung

- Was ist digitale Bestandserhaltung?
- Warum ist Validierung so wichtig?
- Workflow im NLA
- Vorteile frühzeitiger Bestandserhaltung
- Tipps für den Einstieg



Was ist digitale Bestandserhaltung?

- Ein Prozess,
 - der schon mit der Übernahme der Daten beginnt.
 - der einen großen Eingriff in die Daten bedeutet.
 - der niemals abgeschlossen sein wird.
 - der die Benutzbarkeit prägt.
- Im NLA: Migrationsansatz (Konvertierung und Validierung)



Warum ist Validierung so wichtig?

- Konvertierungsfehler passieren häufiger, als man denkt, und können gravierend sein
 - Validierung ist die Qualitätskontrolle
- Archiv muss Anspruch der Authentizität und Integrität erfüllen
- Von Beginn an defekte Daten fallen oft erst bei Konvertierung/Validierung auf
- Auch ohne Konvertierung sollte validiert werden



Warum ist Validierung so wichtig?

Postkarte n: Hist. Karten Kirche 3 / Störtebeker 1
Zeichnungen Kirche: 6 x "Dom" (J.G. S.) 1 x Innenansicht (Foto von Zeichnung), dazu Negativ + Din-A-4

K
o
p
i
e
n

d
e
r

Z
e
i
c
h
n
u
n
g

(
6

b
u
n
t
,

l

s
w
)
)

L
ä
n
s
s
c
h
n
i
e

Konvertierung doc zu PDF/A
-> Text falsch formatiert und
abgeschnitten



Warum ist Validierung so wichtig?

Ochs...“ Das erinnert mich stark an Euch. Ich höre ihn, den Mozart. Man kann dabei so schön mit offenen Augen träumen und vergißt dann Zeit und Raum. – Ich bedauere jetzt immer, daß ich nicht singen kann und nichts über Musik in Worte fassen kann, sondern nur fühlen kann.

. Bei den klaren Sonnennachmittagen kann man schon in der Sonne sitzen. Es ist hier jetzt fast ganz ruhig, so daß wir uns so richtig von den Strapazen erholen können. Lange wird es aber nicht dauern, dann ziehen wir dort hin, wo Otto bisher war. Er selbst liegt jetzt mit seiner Einheit südl. des Ladogasees. Ich hatte gehofft, ihn besuchen zu könne; leider wird nun da nichts draus. – Ich

Original (doc)

Mehrere Zeilen
Text in eine Zeile
übertragen

ganz ruhig, so daß wir uns so richtig von den Strapazen erholen können. Lange wird es aber nicht dauern, dann ziehen wir dort hin, wo Otto bisher war. Er selbst liegt jetzt mit seiner Einheit südl. des Ladogasees. Ich hatte gehofft, ihn besuchen zu könne; leider wird nun da nichts draus. – Ich

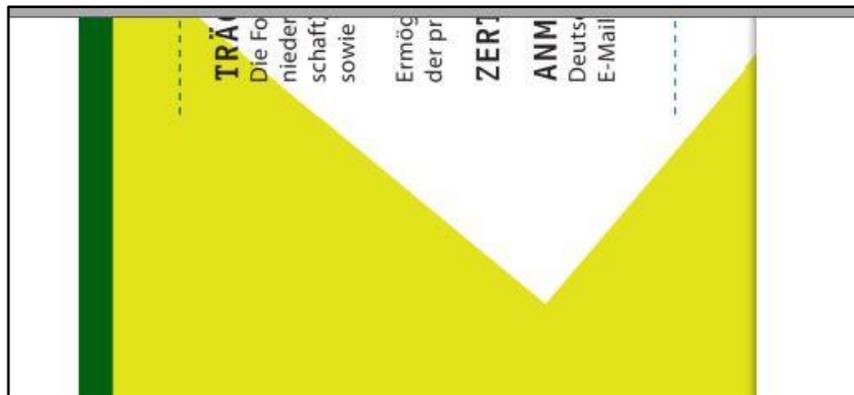
R 2 (pdf/A)



Warum ist Validierung so wichtig?



Original



R 2

Querformate:
Text wird
abgeschnitten



Warum ist Validierung so wichtig?

Validierungsergebnis KostVal:

K) Schrift-Validierung

Von den 46 im Dokument enthaltenen Zeichen sind 45 (97.83%) unbekannt und 0 (0%) undefiniert.

Producer: GPL Ghostscript 9.07

Font: Calibri Full name: XYHSOL+Calibri (object no 27) Type: Type1 Font file: CFF

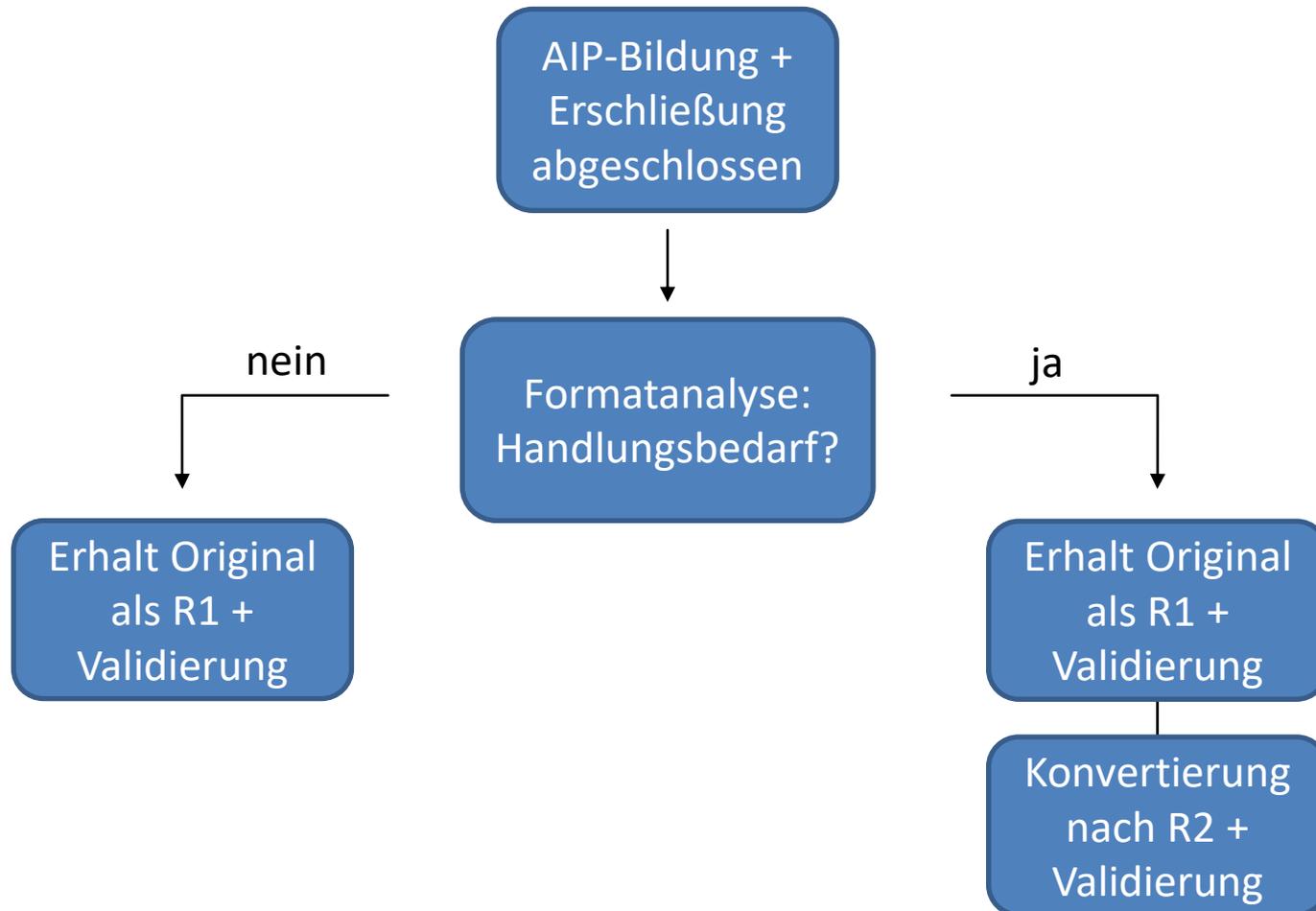
Invalides PDF/A:

Schriften werden nicht erkannt

-> kann zu Problemen bei späteren Konvertierungen führen!



Workflow im NLA





Workflow im NLA

Hilfsmittel

Software (Auswahl)	Funktion
TreeSize	Analyse
DIMAG IngestList	Analyse, Validierung
XnView	Analyse, Validierung, Konvertierung
Callas pdfaPilot	Konvertierung
Microsoft Excel	Konvertierung
KostVal	Validierung
VeraPDF	Validierung



Workflow im NLA

Hilfsmittel

- Dokumentation von Erfahrungen/Vorgehensweisen

Vermerke etc.

<input type="checkbox"/>	Konvertierung exe mit Bildern
<input type="checkbox"/>	Konvertierung html Website nach warc
<input type="checkbox"/>	Konvertierung msg, eml
<input type="checkbox"/>	Konvertierung nach CSV
<input type="checkbox"/>	Konvertierung nach jpg und tif
<input type="checkbox"/>	Konvertierung nach mp4 bzw. mkv
<input type="checkbox"/>	Konvertierung NIST nach wav
<input type="checkbox"/>	Konvertierung odt
<input type="checkbox"/>	Konvertierung Pagemaker pmd pm6 p65
<input type="checkbox"/>	Konvertierung pdf
<input type="checkbox"/>	Konvertierung ppt, odp
<input type="checkbox"/>	Konvertierung ps
<input type="checkbox"/>	Konvertierung psd
<input type="checkbox"/>	Konvertierung von AOL ART Dateien
<input type="checkbox"/>	Konvertierung von Excel-Dateien
<input type="checkbox"/>	Konvertierung von psd Dateien mit Ebenen
<input type="checkbox"/>	Konvertierung von txt nach txt mit UTF-8

„Karteikarten“ für Dateiformate zum Nachschlagen

XLSX, XLS	
Kategorie 2	
Dateiendung	.xlsx, .xls (bei älteren Excel-Versionen)
Formatname	Microsoft Excel for Windows
<u>PronomID</u>	<u>fmt/214</u>
Zielformat	siehe Kategorie 1 für strukturierte Informationen oder gestalteter Text
Konvertierung	Microsoft Office Excel, LibreOffice Calc
Visualisierung	Microsoft Office Excel, LibreOffice Calc
Sonstiges	
Ob eine Excel-Datei besser als strukturierte Informationen oder als gestalteter Text konvertiert wird, hängt stark von den darin enthaltenen Informationen ab. Eine Entscheidungshilfe bieten diese Unterlagen .	



Vorteile frühzeitiger Bestandserhaltung

- Größtmögliche Chance, Informationsverlust zu vermeiden
 - Je älter ein Format/eine Datei, desto höher das Risiko
- Händische Prozesse sind Voraussetzung für spätere Automatisierung
- Benutzbarkeit kann leichter sichergestellt werden
- Fragen von abgebenden Stellen zu Formaten können besser beantwortet werden



Tipps für den Einstieg

- Hashwertprüfung, Virenskan, Backup
- Wissensmanagement
 - Erfahrungen/Vorgehensweisen dokumentieren
 - Informationsangebote nutzen (z.B. PRONOM Registry, nestor Toolbox, COPTR, dpc Handreichungen)
- Grundstock an archivfähigen Formaten definieren
- Je exotischer, desto gefährdeter
- Learning by Doing



Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

Antje Lengnik

Niedersächsisches Landesarchiv – Abt. Zentrale Dienste

nla.dimag@nla.niedersachsen.de