

Datenlieferung in EAD(DDB)

Martin Reisacher (Landesarchiv
Baden-Württemberg)

Workshop: Beteiligung an der
Deutschen Digitalen Bibliothek
und am Archivportal-D

Frankfurt, 29.04.2014



archivportal



DEUTSCHE DIGITALE BIBLIOTHEK
Das Netzwerk für
Kultur und Wissenschaft

- Allgemeiner Teil
 - Allgemeine Ziele bzgl. Datenlieferungen
 - Probleme beim EAD(DDB)-Export
 - Langfristige Ziele
 - Übergangslösungen!
 - Wo sind die Grenzen?
 - Was müssen Sie tun?
- Konkrete Datenlieferung
 - Datentransfer
 - Datentransformation mit DPT (Data Transformation Tool von APEx)
 - Validierung mit DPT
 - Datenkontrolle

Ziele!

- Einfache Lieferprozesse für Sie (und uns)!
- Gute und effiziente Kontrollprozesse hinsichtlich der Darstellung Ihrer Daten!
- → Ihre Erschließungsinformationen für die Nutzer in der bestmöglichen Qualität abzubilden und durchsuchbar zu machen!
- Lieferung in EAD(DDB) um diese Ansprüche zu erreichen
- → aber: ...

Probleme beim EAD(DDB)-Export

- XML-Exporte die nicht weiterverarbeitet werden können

```
<?xml version="1.0" encoding="UTF-8"?>
```

Fehlende Informationen!

Ergebnisse filtern ? Ergebnisse pro Seite 20 Sortieren nach Relevanz

Zeit > 1 - 20 von 77.245 Ergebnissen Seite 1 von 3.863 Weiter > Ende >

Ort

Nach Häufigkeit geordnet Seite 1 Weiter >

Freiburg im Breisgau (77.228)	Landwasser : Freiburg im Breisgau FR (240)
Freiburg im Breisgau FR (1.208)	Italien (229)
Berlin (941)	Insel Mainau : Litzelstetten, Konstanz KN (188)
Stuttgart S (695)	Spanien (164)
Landschaften (492)	Kehl OG (146)

1...

Person/Organisation >

Stichwort >


Sprache >

Medientyp >

Sparte >

Datenlieferant >

hiv
ve
hlung Willy
ekannte >>
berg, Abt.
?f=5-91010-



Suchfilter

versus

Ergebnisse filtern ? Ergebnisse pro Seite 20 Sortieren nach Relevanz

Zeit > 1 - 20 von 77.245 Ergebnissen Seite 1 von 3.863 Weiter > Ende >

Ort

Keine Werte verfügbar Seite 1

Mama Prager

19. Dezember 1928, Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Freiburg, W 134 (Glaspl.) Sammlung Willy Prager I: Glasplattennegative

...Sammlung Willy Prager I: Glasplattennegative >> Familie, Freunde und Bekannte >> Familie...

...Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Freiburg, W 134 (Glaspl.) Sammlung Willy Prager...

Person/Organisation >


Stichwort >

Sprache >

Medientyp >

Sparte >

Datenlieferant >



Keine
Suchfilter

Fehlende Kontextualisierung

Sonstige

Erschließungsangaben:

Ausstellungsort: Ravensburg
Aussteller: Albrecht I., König
Empfänger: Kloster Weißenau
Überlieferungsart: Ausfertigung
Beschreibstoff: Pergament
Siegelbeschreibung: Siegel liegt bei
Druck: WUB Bd. XI, Nr. 5240, Seite 217

versus

Sonstige

Erschließungsangaben:

Ravensburg
Albrecht I., König
Kloster Weißenau
Ausfertigung
Pergament
Siegel liegt bei
WUB Bd. XI, Nr. 5240, Seite 217

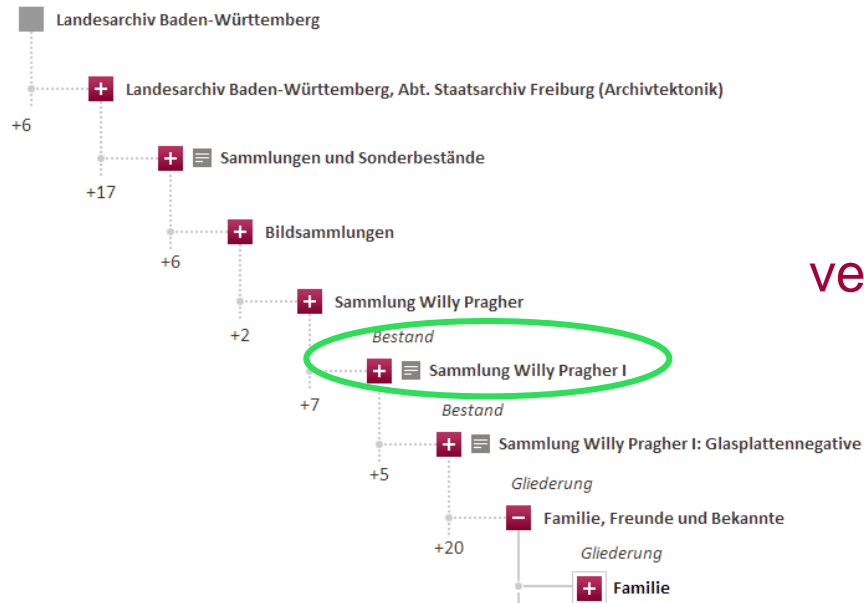
Falsche Zuordnung der Inhalte

Jagstbrücke Jagsthausen

Archivaliensignatur:	Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Ludwigsburg, EL 20/4 III a Nr 172
Kontext:	Regierungspräsidium Stuttgart: Glasplatten zu Straßen- und Wasserbau (bis 9x12 cm) >> Brücken >> Brücken über Flüsse >> Jagst >> Jagstbrücke Jagsthausen
Enthältvermerke:	1 Glasplatte; Format: 6,5x9 cm

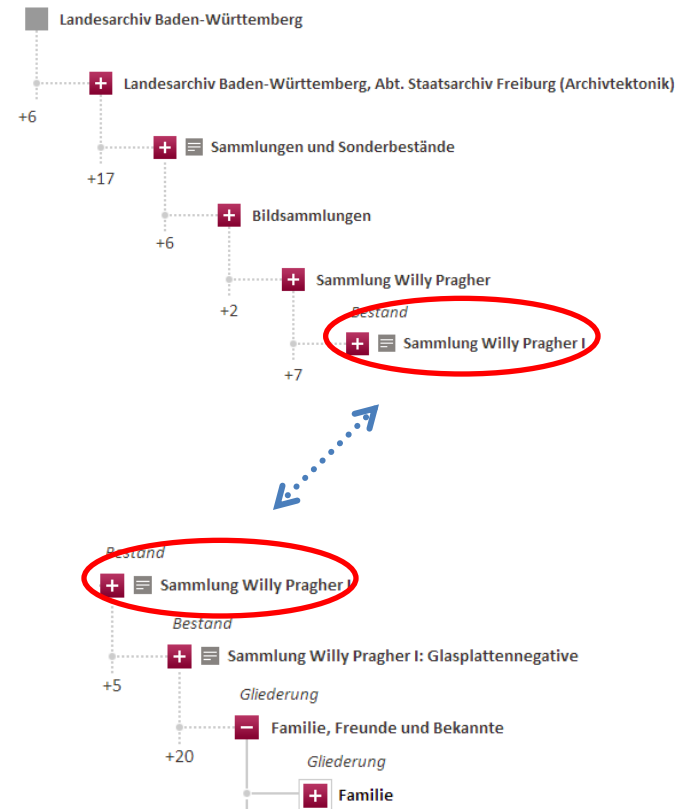
Fehlende Verknüpfungen

Vollständige Hierarchien



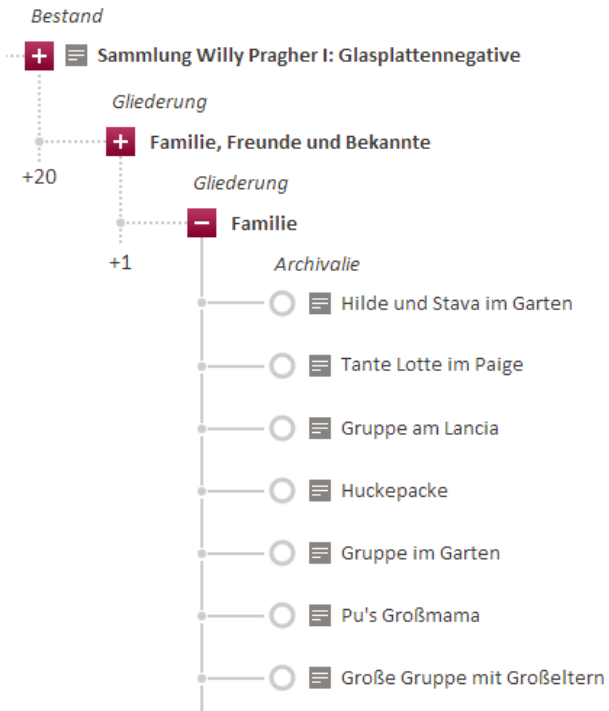
versus

Keine Verknüpfung zwischen Tektonik und Hierarchie

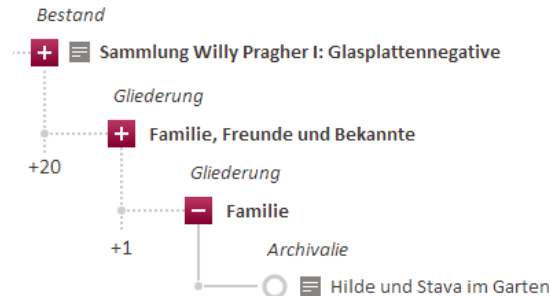


Keine stabilen Identifier

Stabile Identifier (Signatur)



Gleiche Identifier: Daten werden überschrieben



Geänderte Identifier: Daten werden doppelt geladen



- Workshops mit Softwareherstellern
 - Herbst 2012 und Herbst 2013
- Fortschritte, aber noch immer zentrale Probleme
- Herausforderung: individuelle Konfigurierbarkeit der Masken
- → oftmals Fehlerursache nicht ersichtlich (Schnittstelle oder Nutzer?)
- Status quo: Müssen praktisch jede EAD-Lieferung nachbearbeiten

Langfristige Ziele!



- Valide und korrekte Exporte in EAD(DDB)!
- Vollständige Exporte in EAD(DDB)!
- Stabile Identifier! (auch bei einem Software-Wechsel)
- Stabile Links zu Digitalisaten!
- Langfristig müssen die Schnittstellen valides EAD(DDB) nach diesen Anforderungen bedienen

- Notwendig da:
 - Stabile Exporte wohl erst in nächsten Softwareversionen/Updates
 - für Archive, die sich (noch) keine neue Software leisten können
 - für Softwarelösungen ohne EAD-Export
- → Transformation wird von der Fachstelle übernommen
- Daten in Ihrem Exportformat
 - EAD, XML, CSV, Excel, Access
 - **Aber: in archvisch strukturierter Form**
- Muss nach XML umwandelbar sein (kein doc/pdf)

Data Preparation Tool



- Wird entwickelt im Rahmen des APEx-Projekts
- Dient zur Transformation und Validierung von XML-Dateien beim Datenlieferant
- Es können eigene Skripte und Schema eingebunden werden

Zwischenlösung 2

- Findbücher oder ganze Bestandsgruppen
- Empfehlung in EAD und Software eigenem Format zu liefern, da es die Kommunikation vereinfacht
- Individuelle Anpassung des DP-Tools (APEX), damit Sie für Updates selbst EAD(DDB) erzeugen und liefern können
- Nicht unbedingt notwendig, dass Sie EAD(DDB) verstehen, sondern, dass Sie wissen, unter welchem Label Ihre Daten in der DDB angezeigt werden sollen
- Informationsmaterial und Definition von EAD(DDB) auf der Seite des LABW: <http://www.landesarchiv-bw.de/ead>

Archivaliensignatur:

Alt-/Vorsignatur:

Kontext:

Laufzeit:

Enthältvermerke:

Provenienz:

Vorprovenienz:

Umfang:

Maße:

Material:

Urheber:

Archivalientyp:

Sprache der Unterlagen:

Sonstige

Erschließungsangaben:

Bemerkungen:

Indexbegriff Person:

Indexbegriff Ort:

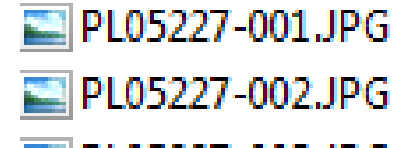
Indexbegriff Sache:

Digitalisat im Angebot des
Archivs:

Weitere Hilfestellungen

- Verknüpfung von Images und VZEs, falls diese eindeutig über den Namen zuzuordnen sind

```
<unitid>PL05227-001</unitid>  
<objvdesec>Ausführungsort: Tübingen
```



- Zusätzliches Befüllen der Suchfilter

```
<odd>  
  <head>Künstler Bild</head>  
  <p>Feld Künstler Bild</p>  
</odd>
```



Ort	>
Person/Organisation	>
Stichwort	>

- Auswahl von Testdatensätze zur effizienten Kontrolle

- Rund 15 Archive im Testsystem
 - Transformation nach EAD(DDB) klappt gut
 - EAD, XML, Excel, Datenbanken, teilweise sogar txt
 - ActaPro/Augias/Faust EAD Exporte zu EAD(DDB)
 - Scope zu EAD
 - Augias zu EAD
 - Faust zu EAD
 - Excel zu EAD
 - Mittlerweile Anpassungen relativ zügig, da generische Skripte
 - Spezielle Anpassungen können nur im Rahmen der vorhandenen Ressourcen berücksichtigt werden

Was können wir nicht anbieten?

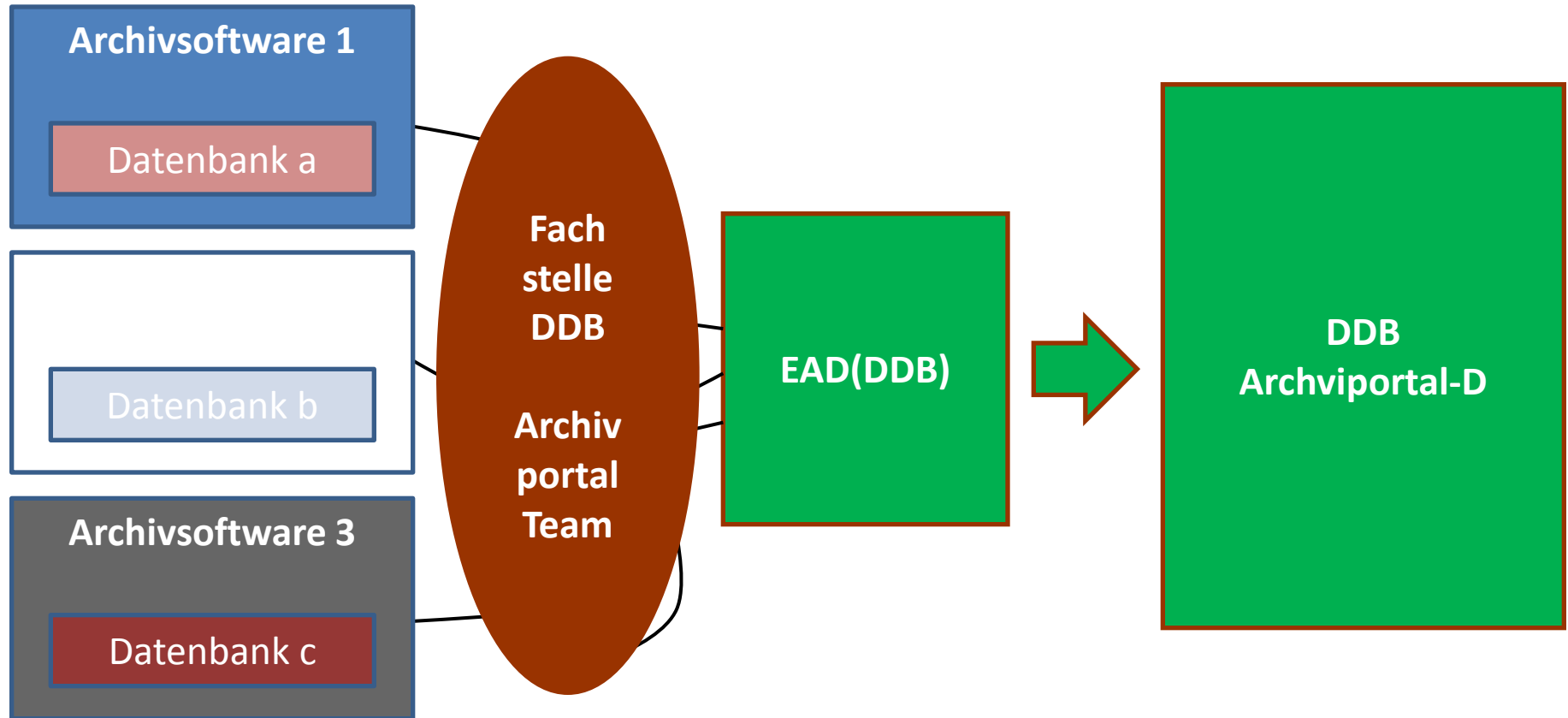


- Keine Dauerlösung!!! – langfristig stabile EAD(DDB) Exporte notwendig
- Zwischenlösung durch Ressourcen aus Archivportal-D Projekt möglich
- Keine inhaltlichen Korrekturen der Daten
- Gewisse Transformationszeit beim Export
- Begrenzte Ressourcen

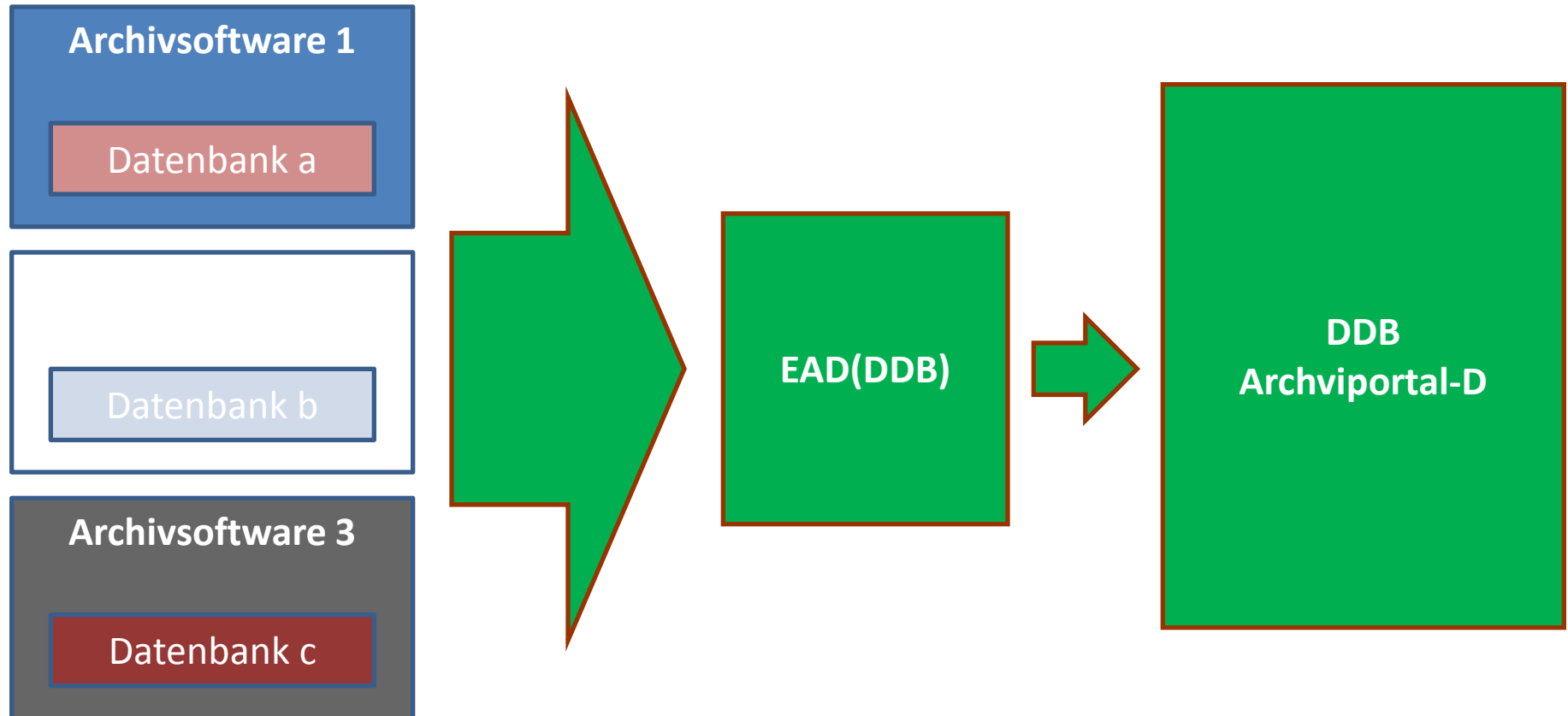
Was müssen Sie tun?

- Unterstützung beim Mapping: Hilfe bei der Zuordnung
- Lieferung von EAD und internem Format, da es die Kommunikation mit Ihnen und den Softwareherstellern vereinfacht
- Definition welche Bestände nicht geliefert werden sollen (Sperrungen)
- Kontrolle der Daten im Testsystem
- **!!!Softwareherstellern auf Bedeutung guter EAD(DDB)-Exporte hinweisen!!!**

Datenfluss momentan



Datenfluss zukünftig



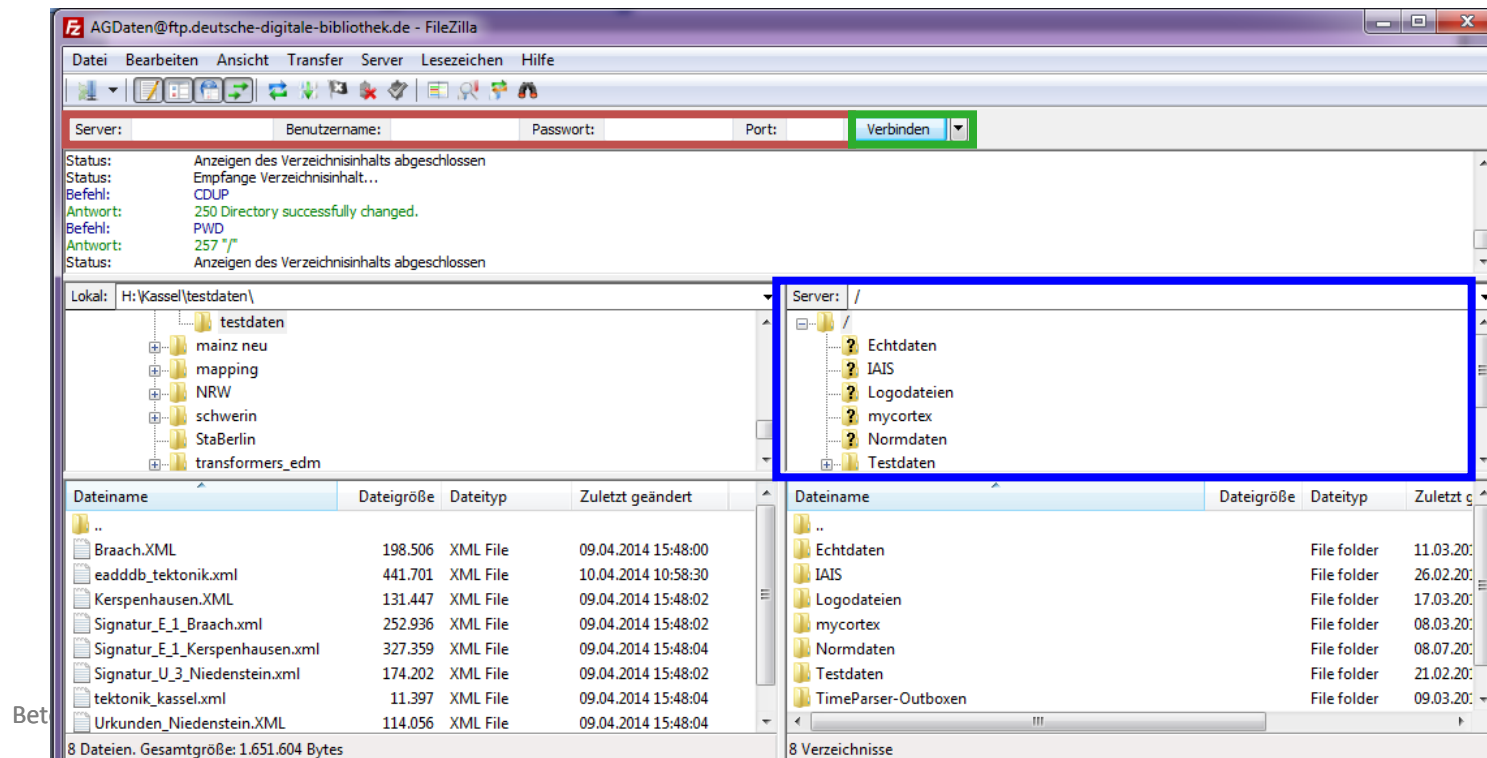
Datenlieferung aus Ihrer Sicht



- Export der Daten
- Lieferung der Daten
- Transformation der Daten
- Transformation der Daten bei Ihnen
- Datenkontrolle

Datentransfer

- Per FTP: FileZilla (freies Programm, mit dem Sie Daten auf den FTP-Server der DDB verschieben können)
- Per Mail (bei kleinen Datenmengen)
- Per Datenträger (eher vermeiden)



- In einem ersten Schritt analysieren wir die Struktur der Daten
- Nachnutzen von bereits vorhandenen Regeln
- Erstellung erster Transformationsregeln
 - Feldzuordnungen
 - Verzeichnungsgebe: Bestand, Gliederung, VZE
 - Archivalientyp: Urkunde, Plan, Akte
- Abgleich der Regeln mit Ihnen im Testsystem

Hierarchieebene	ElementName in Scope	Label in der DDB / Archivportal-D	Ueberschrift	Indexbegriff
collection	Archivierungsgeschichte	Bestandsbeschreibung	Archivierungsgeschichte	
collection	Hinweise zu Findhilfsmitteln	Bestandsbeschreibung	Hinweise zu Findhilfsmitteln	
collection	Kommentierte Beständeübersicht	Bestandsbeschreibung	Bestandsbeschreibung	
collection	Verwaltungsgeschichte	Bestandsbeschreibung	Verwaltungsgeschichte	
collection	Bestandsaufnahme			
collection	Signatur	Bestandssignatur		
collection	xxxxxxxxxx	Provenienz		
collection	xxxxxxxxxx	Sprache der Unterlagen		
collection	Titel	Titel		
collection	Laufmeter	Umfang	Laufmeter	
collection	xxxxxxxxxx	Urheber	Autor/Fotograph	
collection	Veröffentlichungen	Verwandte Bestände und Literatur	Literatur	
collection	Verwandte Verzeichnungseinheiten	Verwandte Bestände und Literatur	Verwandte Verzeichnungseinheiten	
collection	xxxxxxxxxx	Vorprovenienz		
class	Angeklagte / Beklagte	Beschreibung	Angeklagte / Beklagte	
class	Darin	Beschreibung		
class	Hinweise Strafakten	Beschreibung	Hinweise Strafakten	
class	Titel	Titel		
file	Bild/Ton	Allgemeine Formalbeschreibung	Technik	
file	Technik	Allgemeine Formalbeschreibung	Technik	
file	Signatur	Archivaliensignatur		
file	Bemerkungen	Bemerkungen		
file	Angeklagte / Beklagte	Enthältvermerke	Angeklagte / Beklagte	Person
file	Darin	Enthältvermerke	Darin	
file	Dienstbezeichnung	Enthältvermerke	Dienstbezeichnung	
file	Dienststelle	Enthältvermerke	Dienststelle	
file	Dienstzeit	Enthältvermerke	Dienstzeit	
file	Geburtsname	Enthältvermerke	Rufname	
file	Inhalt	Enthältvermerke	Regest	
file	Ortsteil	Enthältvermerke	Ortsteil	
file	Regest	Enthältvermerke	Regest	
file	Rufname	Enthältvermerke	Rufname	
file	Thema	Enthältvermerke	Thema	

DDB-Label | Überschrift

Bestandsbeschreibung | Archivierungsgeschichte

Bestandsbeschreibung | Entstehungsgeschichte

Bestandsbeschreibung | Bestandsbeschreibung

Bestandsbeschreibung | Inhaltliche Beschreibung

Bestandslaufzeit

Bestandssignatur

Provenienz

Sprache der Unterlagen

Titel

Umfang | Laufmeter

Urheber | Autor/Fotograph

Verwandte Bestände und Literatur | Literatur

Verwandte Bestände und Literatur | Verwandte Verzeichnungseinheiten

Vorprovenienz

Beschreibung | Angeklagte / Beklagte

Beschreibung

Beschreibung | Hinweise Straftaten

Titel

Allgemeine Formalbeschreibung | Technik

Allgemeine Formalbeschreibung | Technik

Archivaliensignatur

Bemerkungen

Enthältvermerke | Angeklagte / Beklagte

Enthältvermerke | Darin

Enthältvermerke | Dienstbezeichnung

Enthältvermerke | Dienststelle

Enthältvermerke | Dienstzeit

Enthältvermerke | Rufname

Enthältvermerke | Regest

Enthältvermerke | Ortsteil

Enthältvermerke | Regest

Enthältvermerke | Rufname

Enthältvermerke | Thema

Enthältvermerke | Vermerke

Laufzeit

Maße | Breite (cm)

Transformationsregeln durch Anlegen von Testdatensätzen

archivportal



AUGIAS-Archiv 8.2 - [Ebene: 1: Vorhandene Verzeichnungseinheiten in Ebene 2: 0; Bilder: 0; PDF: 0; Multimedia-Files: 0; Indexbegriffe: 0]

Bestand	Aktuelle Sign.	Alte Archiv-Sign.	I. Num.	v. Num.	Datierung von	Datierung bis	Dat. => Findbuch
01-1 - Akten							

Klassifikationsgruppe 1	Klassifikationsgruppe 2

Titel	Enthält

Alter Aktentitel	Provenienz

Umfang	Band	Edition

Erhaltung

Bemerkung	Standort	gesperrt bis	gesperrt für

Entnommen	Verzeichnungsprotokoll

Microfilm/-fiche	Registratursignatur	Reservefeld

[CC-BY-SA 3.0](#) Autor: [Maybe 89](#)

DPT-Tool Vorstellung



- → Wird von uns je Datenlieferant angepasst (Konfigurationsdatei) und zur Verfügung gestellt werden
- Ausgabe von Fehlermeldungen bei Problemen
- Momentan erfolgen noch Anpassungen am DPT für das Archivportal-D
- individuelle Version je Datenlieferant sobald Transformationsregeln abgeschlossen sind und Anpassungen erfolgt sind
- APEX Version verfügbar unter:
http://www.bundesarchiv.de/archivgut_online/standards_werkzeuge/03641/index.html

- Sind nur Element enthalten, welche Teil von EAD(DDB) sind?

```
</langmaterial>
<origination>Provenienz</origination>
<origination label="pre">Vorprovenienz</origination>
<origination label="Fotograph">
  <name source="Normvokabularname" authfilenumber="Normvokabular_Nummer">Urheber</name>
</origination>
</did>
<scopecontent encodinganalog="Einleitung">
  <head>Bestandseinleitung_Überschrift</head>
  <p>Bestandseinleitung_Text</p>
</scopecontent>
<relatedmaterial>
  <head>Verweis auf verwandte Bestände/Literatur Überschrift</head>
  <p>Verweis auf verwandte Bestände/Literatur Überschrift Text</p>
</relatedmaterial>
<biographist>
  <p>Biographische Angaben</p>
</biographist>
<accessrestrict>
  <head>Zugangsbeschränkung Überschrift</head>
  <p>Zugangsbeschränkung</p>
</accessrestrict>
```

Technische Datenkontrolle 2



- Passt die Verknüpfung zwischen Findbuch und Tektonik?
- Sind die Identifier/Signaturen eindeutig?
- Vergleich Software-EAD – erzeugtes EAD: sind alle Daten enthalten?
- Inhaltliche Kontrolle: passen die Einträge zum Label?

Kontrolle im Testsystem: manueller Testdatensatz

Serientitel: Titel der Archivalie

Archivaliensignatur:	Name des Archivs, Archivaliensignatur
Alt-/Vorsignatur:	Altsignatur
Kontext:	Bestandstitel >> Gliederungsüberschrift >> Serientitel
Laufzeit:	Laufzeit
Enthältvermerke:	enthält u.a.: Enthältvermerk
Provenienz:	Provenienz
Vorprovenienz:	Vorprovenienz
Umfang:	Umfang
Maße:	Maße
Material:	Material
Urheber:	Urheber
Formalbeschreibung:	Allgemeine Formalbeschreibung
Archivalientyp:	Archivalientyp
Sprache der Unterlagen:	Sprache der Unterlagen1
Sonstige Erschließungsangaben:	<i>Sonstiges Erschließungsangaben Überschrift:</i> Sonstiges Erschließungsangaben Text <i>Sonstiges Erschließungsangaben in Datumform Überschrift (z.B.: Geburtsdatum):</i> Datum
Bemerkungen:	Unspezifische_Bemerkungen1
Indexbegriff Person:	Indexbegriff Person
Indexbegriff Ort:	Indexbegriff Ort
Indexbegriff Sache:	Indexbegriff Sache
Digitalisat im Angebot des Archivs:	Digitalisat beim Datenlieferanten vorhanden
Bestand:	Name des Archivs, Bestandssignatur Bestandstitel

Datenqualität

- Sichtbare Datenqualität
 - „reichste“ Erschließungsinformationen
 - Kontextualisierte Information
 - Indexbegriffe
 - Korrekt aufgebaute Hierarchien
 - Verwendung von Normdaten!
- Unsichtbare Datenqualität
 - Stabile eindeutige Identifier
 - Verknüpfung Findbuch - Tektonik
 - Stabile Links

hierarchischer Aktenplan

1 Gliederung 1	
10 Gliederung 10	
100 Gliederung 100	
101 Gliederung 101	
2 Gliederung 2	
20 Gliederung 20	
200 Gliederung 200	
201 Gliederung 201	

flacher Aktenplan

1 Gliederung 1
10 Gliederung 10
100 Gliederung 100
101 Gliederung 101
2 Gliederung 2
20 Gliederung 20
200 Gliederung 200
201 Gliederung 201

- Anreicherung und Normierung
 - Zeitwerte

Zeit	
1931 bis 1940	×
1941 bis 1950	×
<input type="text"/>	
Ort	>
Person/Organisation	>

Nach Häufigkeit geordnet	
20. Jahrhundert	(885.307)
19. Jahrhundert	(406.520)
18. Jahrhundert	(285.641)
1951 bis 1960	(271.510)
21. Jahrhundert	(164.371)



Zeit

01.01.1939 - 31.12.1945 x

Von

1. Januar 1939

Bis

31. Dezember 1945

Eingrenzung ?

☒ Unschärf ☐ Genau

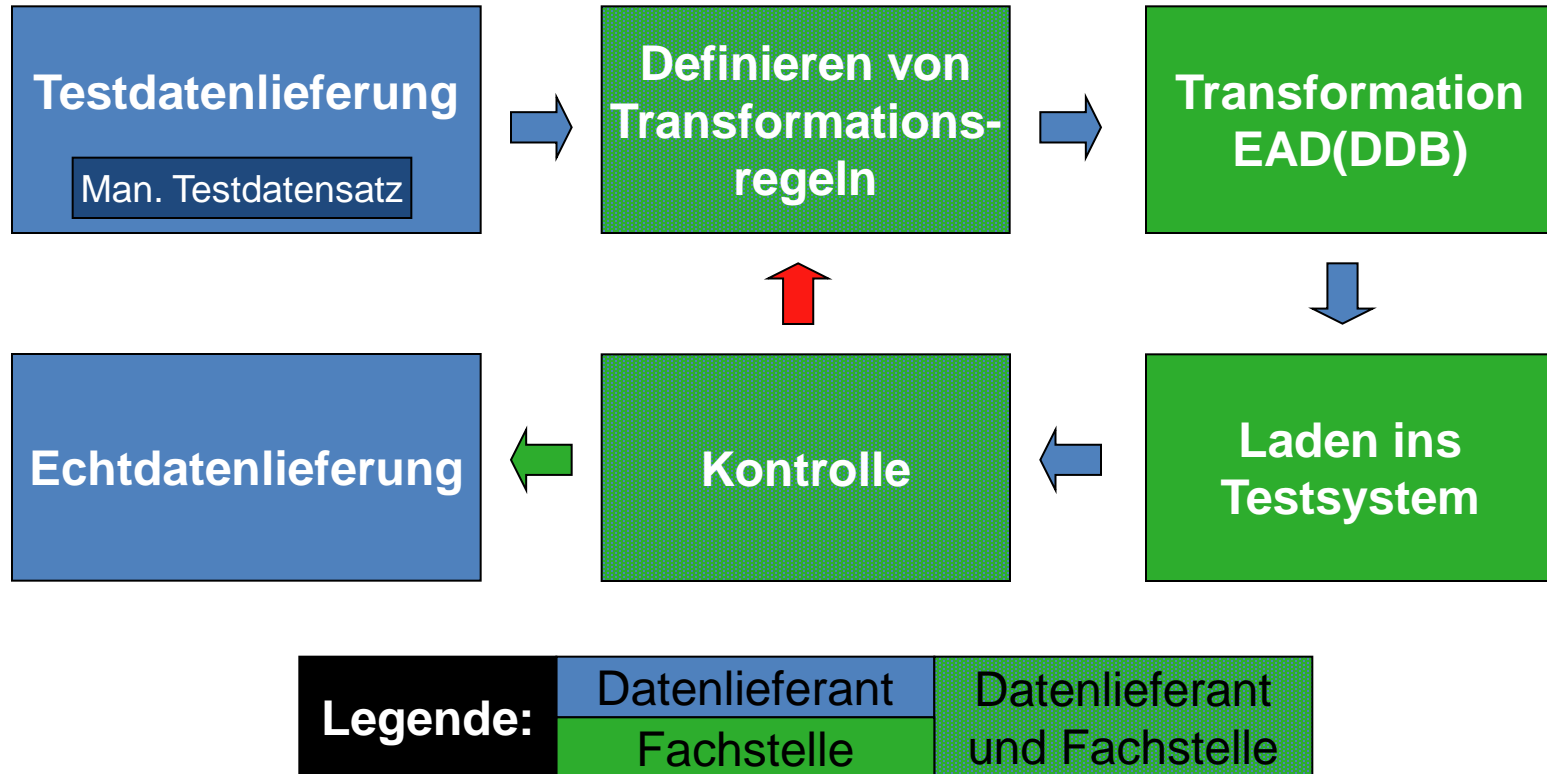
Anwenden

- Normalisierte Archivalientypen/Medientypen
- Anreicherung mit Normdaten

Workflow

- Analyse der Struktur: Ziffern werden zu #
1ste Hälfte 19 Jh. → #ste Hälfte ## Jh.
- Wie oft kommen die Zeitwerte vor?
0,5 %
- Wurde bereits eine Regeln definiert?
Nein
- Daten werden zum Internen Standard normiert
Yste Hälfte ZZ Jh. → Z. Hälfte YY. Jahrhundert
- Und in Zeitwerte umgewandelt
1. Hälfte 19. Jahrhunder → 1801/1850

Ablauf-Diagramm



Testdaten und Contentfragebogen möglichst frühzeitig liefern, erleichtert uns stark die Planung

Vielen Dank für Ihre Aufmerksamkeit! Fragen? Anmerkungen?

Martin Reisacher
Fachstelle Archiv DDB
Landesarchiv Baden-Württemberg
martin.reisacher@la-bw.de