

Datenaufbereitung im Archivportal-D

Wie aus Findbuch und Tektonik
digitalisiertes Archivgut entsteht

Workshop „Vom Dateningest zur Präsentation –
Wege ins Archivportal-D durch Aggregatoren“
18. Februar 2016, Duisburg

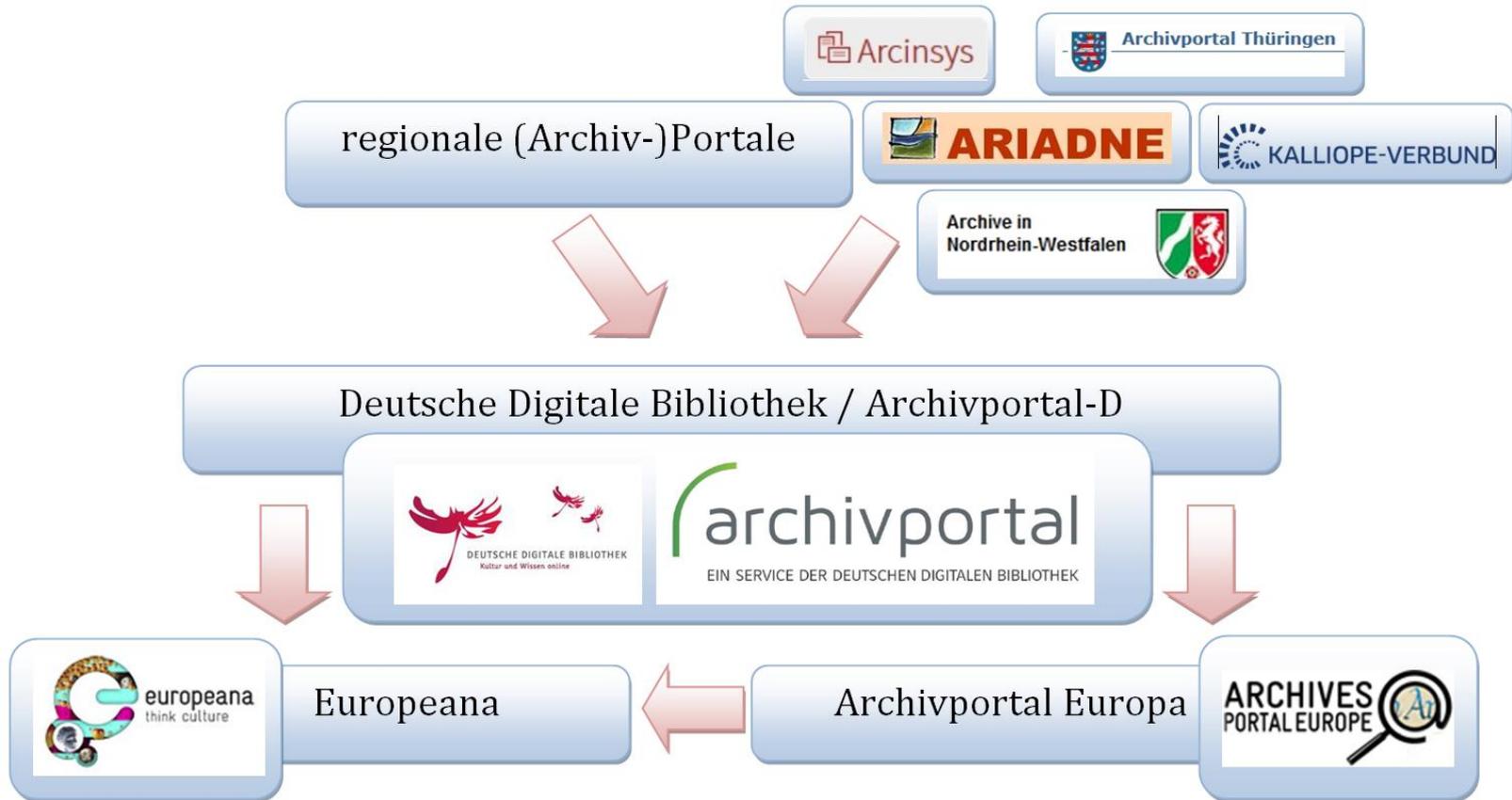
Agenda

- 1. Verortung des Archivportals-D in der Datenhaltung der DDB**
- 2. Vom proprietären Datensatz zum Austauschformat EAD(DDB)**
- 3. Herausforderungen und Lösungsansätze**
- 4. Demo: Datenintegration mit Hilfe des DPT**
- 5. Fazit**
- 6. Ausblick**
- 7. Diskussion**

1. Verortung des Archivportals-D in der Datenhaltung der DDB

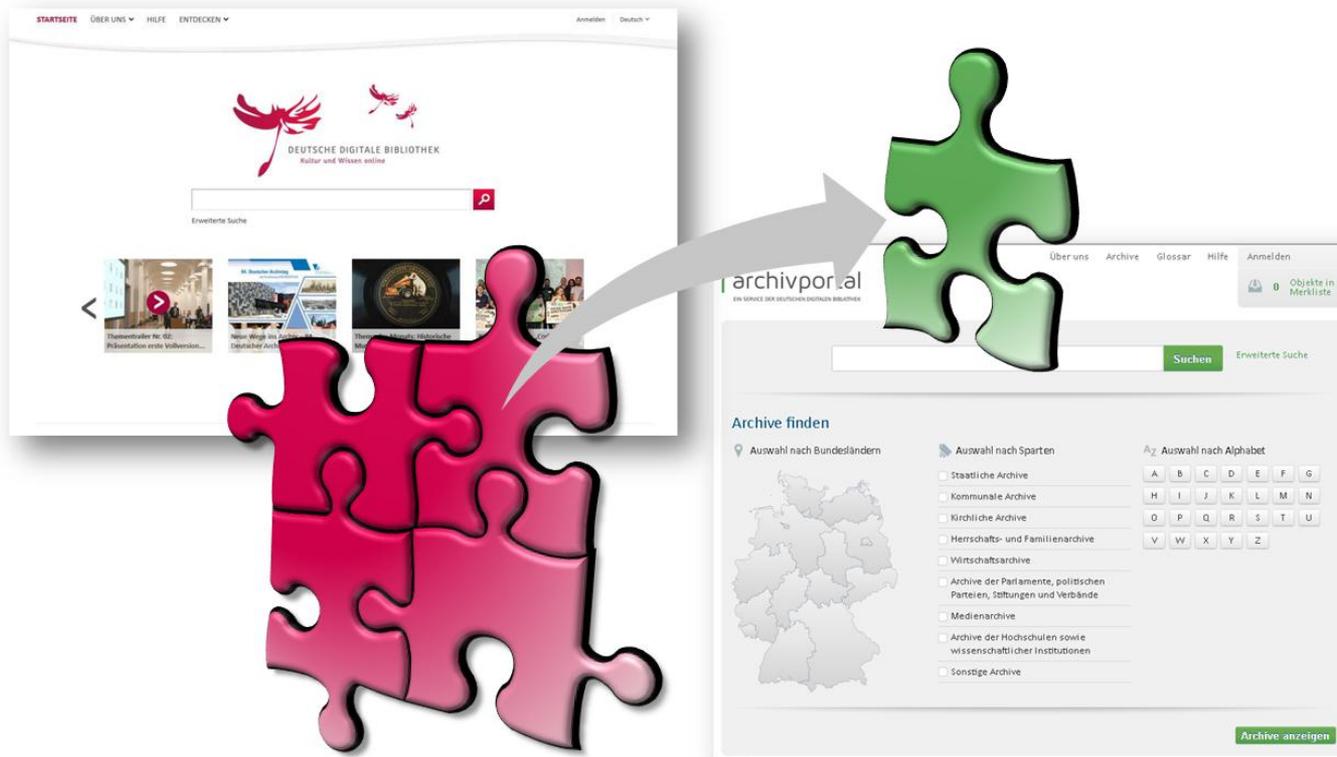
1 Backend, 2 Präsentationsschichten

(Archiv-)Portallandschaft



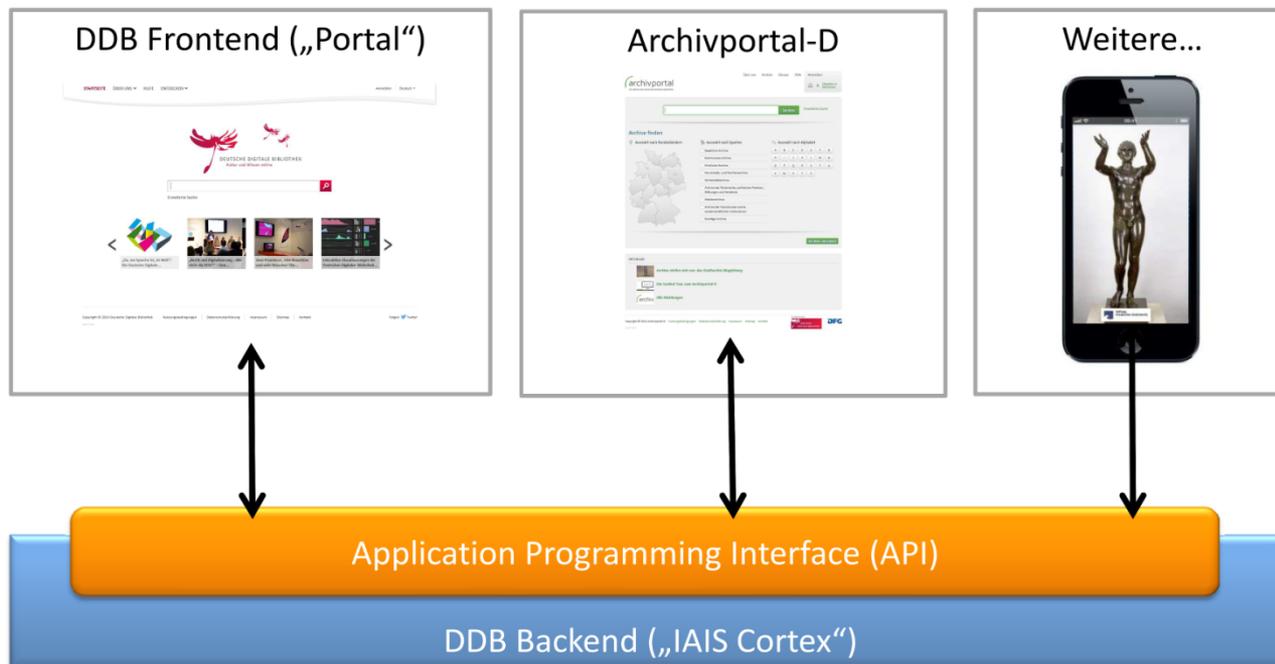
Verortung des Archivportals-D in der Datenhaltung der DDB

DDB: ein Backend – mehrere Nutzersichten



Verortung des Archivportals-D in der Datenhaltung der DDB

DDB: ein Backend – mehrere Nutzersichten



Quelle: FIZ Karlsruhe

2. Vom proprietären Datensatz zum Austauschformat EAD(DDB)

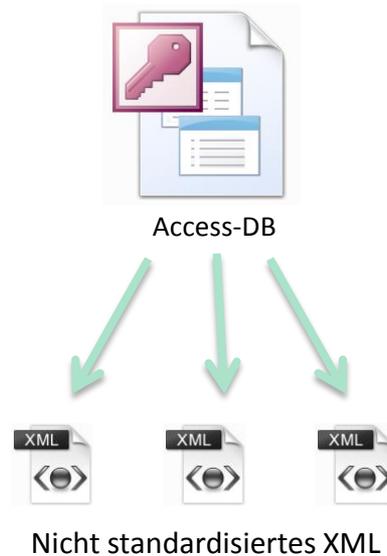
Transformation durch modulare
Ingestwerkzeuge

Das Austauschformat EAD(DDB)

- **Archivisches Standardformat** für die DDB
- Definiert von der EAD-Arbeitsgruppe
- Ziel: ein **einheitliches**, möglichst **überschaubares**, aber trotzdem **flexibles** EAD-Profil für Deutschland
 - Statt 150 rund 50 Elemente
- Möglichkeit, **Tektonik**, **Findbuch** und **Digitalisate** abzubilden

Vom proprietären Datensatz zum Austauschformat EAD(DDB)

Software-Plattform	Exportformat
AUGIAS-Archiv	Access/mdb, EAD, EAD(DDB)
Faust	EAD, EAD(DDB) Faust-XML
Scope	EAD(DDB), Scope-XML
MidosaxML	EAD
easyDB	EAD(DDB)
Excel	Aus xls generiertes XML

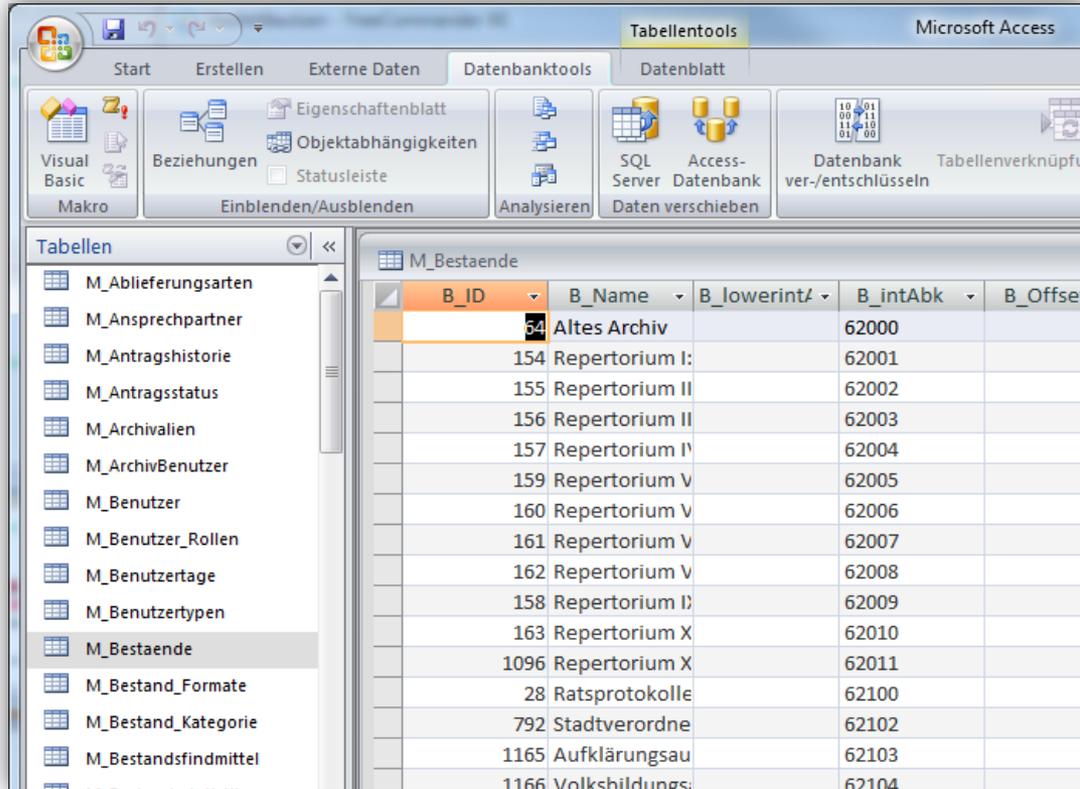


Vielzahl an Ausgangsformaten

Optionale Zwischenformate

Eingangsformat

Vom proprietären Datensatz zum Austauschformat EAD(DDB)



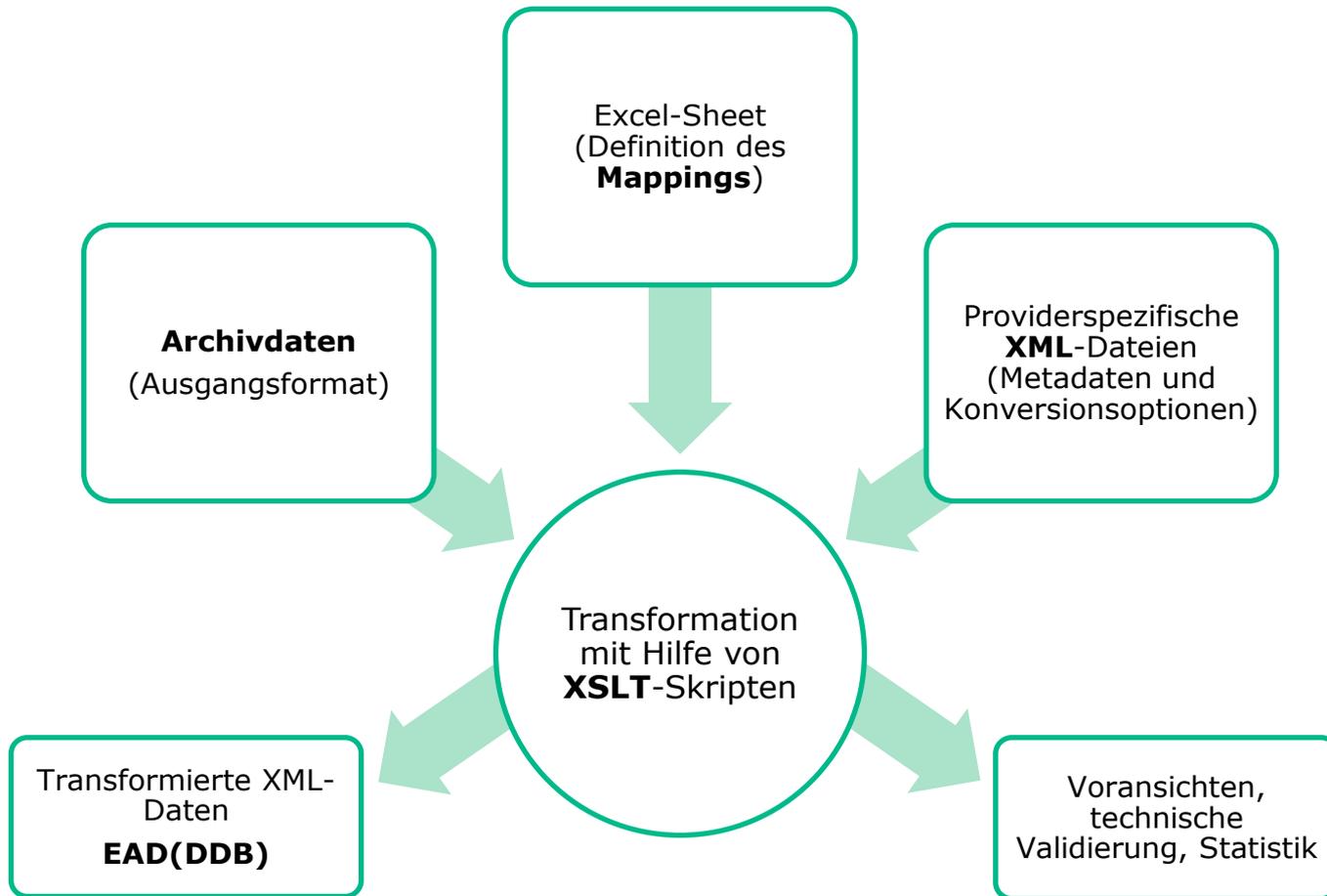
B_ID	B_Name	B_lowerint/	B_intAbk	B_Offset
34	Altes Archiv		62000	
154	Repertorium I:		62001	
155	Repertorium II		62002	
156	Repertorium II		62003	
157	Repertorium I'		62004	
159	Repertorium V		62005	
160	Repertorium V		62006	
161	Repertorium V		62007	
162	Repertorium V		62008	
158	Repertorium I)		62009	
163	Repertorium X		62010	
1096	Repertorium X		62011	
28	Ratsprotokolle		62100	
792	Stadtverordne		62102	
1165	Aufklärungsaus		62103	
1166	Volksbildungs		62104	

- Spezialfall: Access-Export (z.B. aus Augias-Archiv) als Zwischenformat
- keine sprechenden Feldnamen
- Strukturierung abhängig von Augias-Formularkonfiguration
- Access ermöglicht XML-Export
 - nicht standardisiert, aber automatisiert weiterverarbeitbar

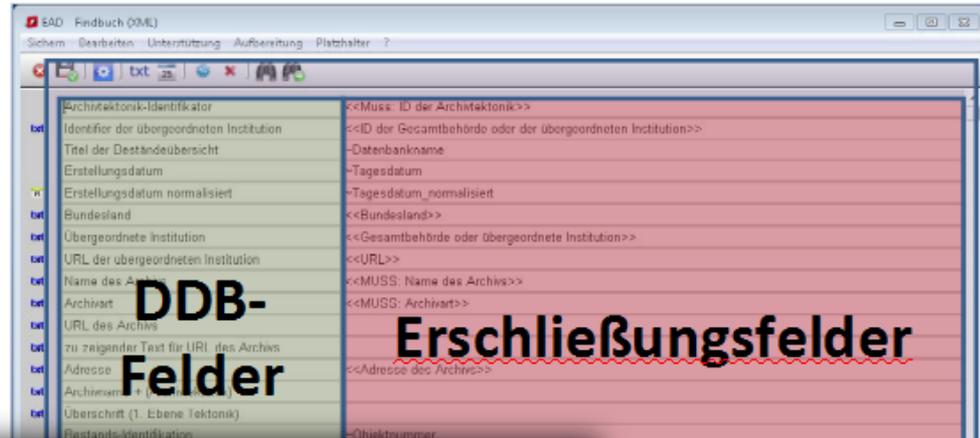
Vom proprietären Datensatz zum Austauschformat EAD(DDB) (2)

- Transformationsmechanismen für verschiedenste Ausgangsformate benötigt
- Erster Schritt: Datenmapping
 - konzeptionell
 - technisch
- Mittelfristiges Ziel: Notwendigkeit von Zwischenformaten und Sonderlösungen reduzieren
- Modularer Aufbau/Ablauf Grundvoraussetzung für Flexibilität

Vom proprietären Datensatz zum Austauschformat EAD(DDB) (3)



Konzeptionelles Mapping am Beispiel des Faust EAD-Exports und der internen Excel-Mappings



VERZEICHNUNGSSTUFE	FELDNAME (XML/DATENBANK)	BEISPIEL	FELDUORDNUNG DDB	SPEZIFIZIERUNG (UNTERUEBERSCHRIFT)	SORTIERUNG	SUCHFACETTENZUORDNUNG
Archivale/Vorgang	Provenienzstelle		Provenienz			
Archivale/Vorgang	Rücklink	Häufigkeit: 2501 mal Be	Rücklink			
Archivale/Vorgang	Signatur		Signatur			
Archivale/Vorgang	Ablaufgeschwindigkeit		Sonstige Erschließungsinfor	Ablaufgeschwindigkeit		
Archivale/Vorgang	Anwendungszeitraum		Sonstige Erschließungsinfor	Anwendungszeitraum		
Archivale/Vorgang	Auflage		Sonstige Erschließungsinfor	Auflage		
Archivale/Vorgang	Auftraggeber		Sonstige Erschließungsinfor	Auftraggeber		Person
Archivale/Vorgang	Ausführungsart		Sonstige Erschließungsinfor	Ausführungsart		
Archivale/Vorgang	Autor		Sonstige Erschließungsinfor	Autor		
Archivale/Vorgang	Befestigung		Sonstige Erschließungsinfor	Befestigung		
Archivale/Vorgang	Beglaubigungen		Sonstige Erschließungsinfor	Beglaubigungen		
Archivale/Vorgang	Beschreibstoff		Sonstige Erschließungsinfor	Beschreibstoff		
Archivale/Vorgang	Bildbeschreibung		Sonstige Erschließungsinfor	Bildbeschreibung		
Archivale/Vorgang	Datumszeit		Sonstige Erschließungsinfor	Datumszeit		
Archivale/Vorgang	Edition/Regest		Sonstige Erschließungsinfor	Edition/Regest		
Archivale/Vorgang	Entstehungszeitraum		Sonstige Erschließungsinfor	Entstehungszeitraum		
Archivale/Vorgang	Form		Sonstige Erschließungsinfor	Form		
Archivale/Vorgang	Format		Sonstige Erschließungsinfor	Format		
Archivale/Vorgang	Fotograf		Sonstige Erschließungsinfor	Fotograf		Person
Archivale/Vorgang	Geburtsdatum		Sonstige Erschließungsinfor	Geburtsdatum		
Archivale/Vorgang	Geburtsort		Sonstige Erschließungsinfor	Geburtsort		
Archivale/Vorgang	Gericht Staatsanwaltschaft		Sonstige Erschließungsinfor	Gericht/Staatsanwaltschaft		Person
Archivale/Vorgang	Haftzeit		Sonstige Erschließungsinfor	Haftzeit		

Zuordnung

Erschließungsfelder

DDB-Felder

Vom proprietären Datensatz zum Austauschformat EAD(DDB) (4)

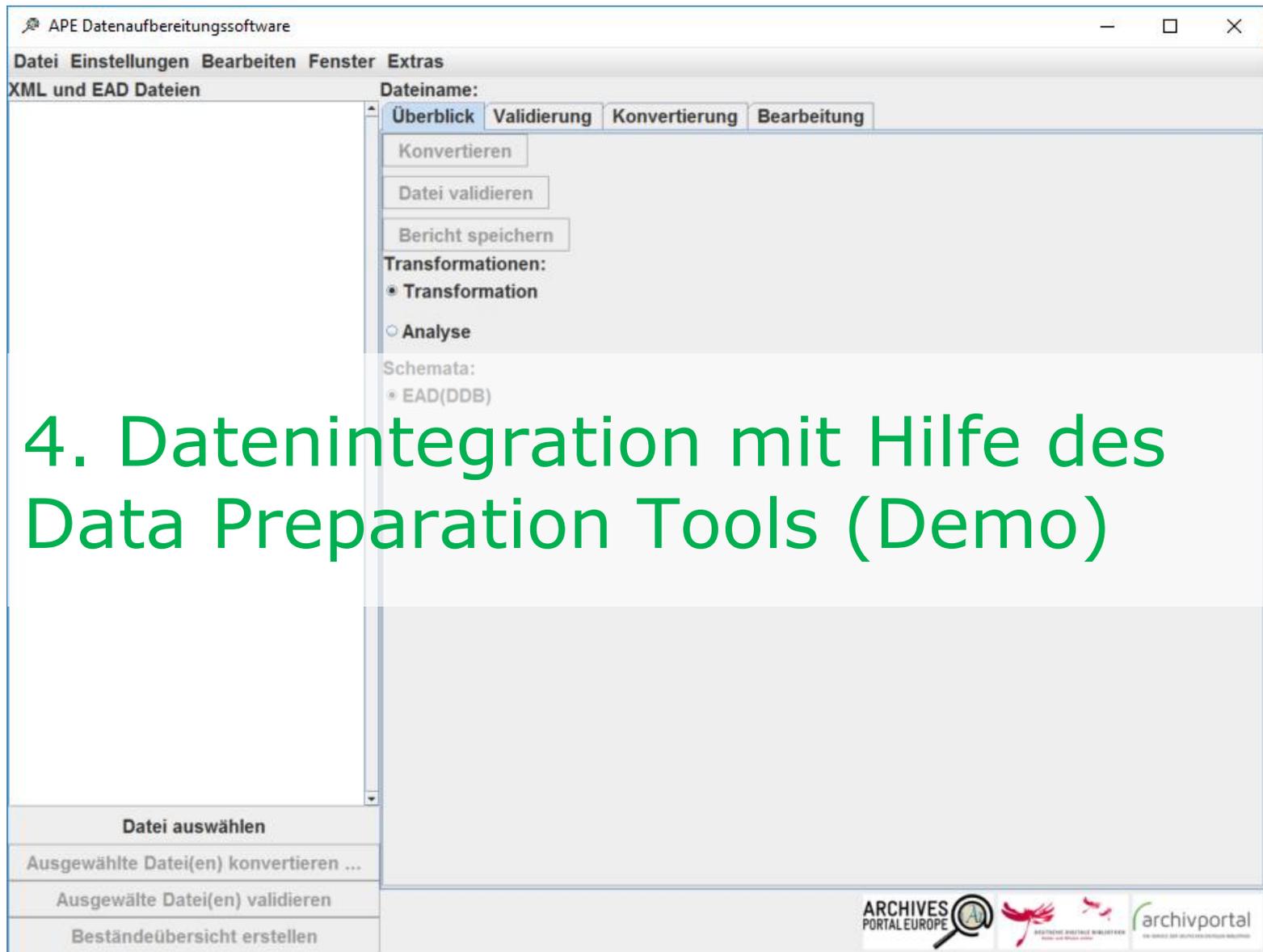
- Kontrolle bzw. Verbesserung der Datenqualität durch:
 - Schema-Validierung der XML-Struktur von Findbuch und Tektonik
 - Generierung von lokalen Voransichten (grundsätzliche Kontrolle, ohne Daten ins Testsystem laden zu müssen)
 - Analyse der statistischen Auswertung, technische Validierung
 - Untersuchung der Darstellung im Testsystem
- **Primär jedoch:** Kontrolle durch datenliefernde Archive selbst

3. Herausforderungen und Lösungsansätze

Fragen der Datenqualität, Konsistenz und Einbeziehung von Softwareherstellern

Herausforderungen und Lösungsansätze

- Stabile Identifier
- Heterogene Datenformate der Archivsoftware
- Weiterlieferung der Daten an Europeana
- Datenqualität und Konsistenz
 - Vielzahl an Faktoren: Softwarehersteller, Verarbeitung beim Datengeber, der Fachstelle und in der DDB
- Teilweise aufwendige Verarbeitung archivischer Daten

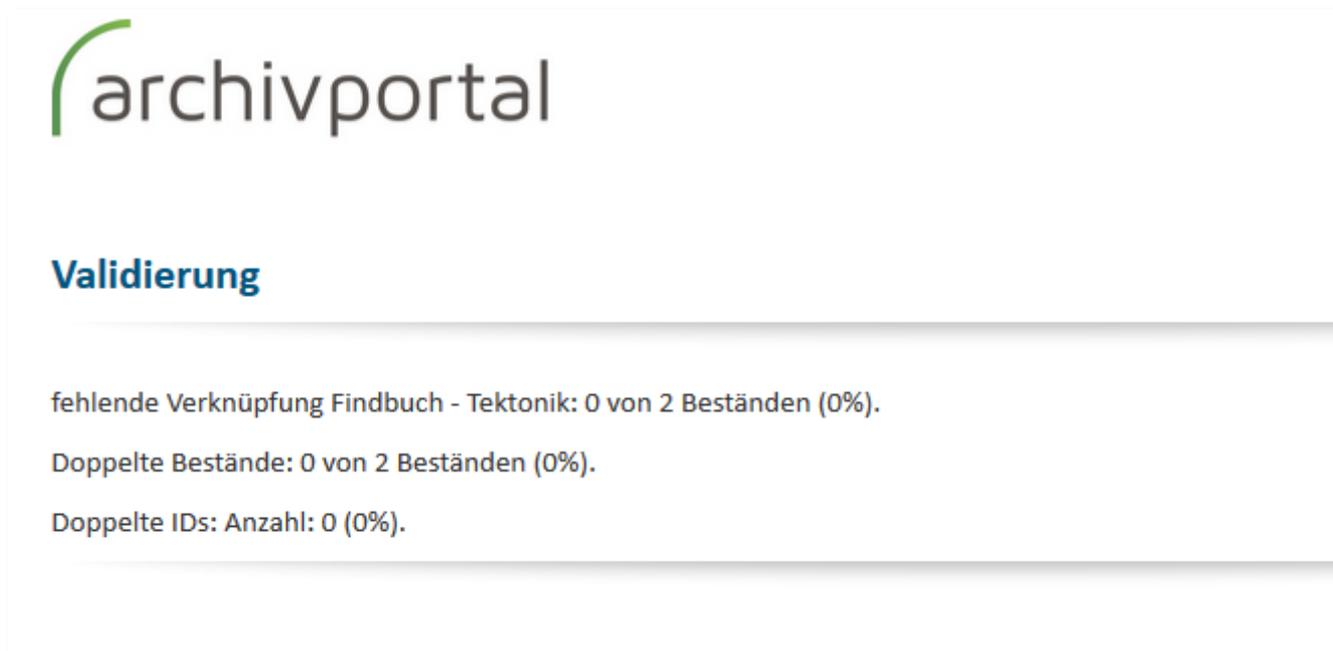


4. Datenintegration mit Hilfe des Data Preparation Tools (Demo)

4. Datenintegration mit Hilfe des Data Preparation Tools

- Ursprünglich entwickelt im Projekt *Archivportal Europa*, weiterentwickelt im APD-Projekt
- Modularität: zentrale Funktionalitäten der Mapping-Skripte über grafische Oberfläche zugänglich
 - vereint Transformations- sowie Validierungsfunktion
- Kein separater (ggf. kostenpflichtiger) XSLT-Prozessor benötigt
- Bietet Voransichten für transformierte Daten, ohne diese ins Testsystem laden zu müssen

4. Datenintegration mit Hilfe des Data Preparation Tools



The screenshot shows the 'Validierung' (Validation) section of the archivportal interface. It lists three validation items, all with 0% error rates:

- fehlende Verknüpfung Findbuch - Tektonik: 0 von 2 Beständen (0%).
- Doppelte Bestände: 0 von 2 Beständen (0%).
- Doppelte IDs: Anzahl: 0 (0%).

Technische Validierung

4. Datenintegration mit Hilfe des Data Preparation Tools



The screenshot shows the Archivportal website interface. At the top right, there are links for "Über uns" and "Glossar". The main heading is "Metadatenauswertung". Below it, the section "Insgesamt" is displayed, followed by "Hierarchieebenen". A list of results is shown, including counts for "(Teil-)Bestände" and "Anzahl Objekte mit eigenem DDB-View".

Über uns Glossar

archivportal

Metadatenauswertung

Insgesamt

Hierarchieebenen

- (Teil-)Bestände - Insgesamt: 3 mal
 - Max: 1 mal (ID: Heuchelheim)
 - Beispiele:
 - : Insgesamt: 1 mal (ID:)
 - : Insgesamt: 1 mal (ID:)
 - : Insgesamt: 1 mal (ID:)
- Anzahl Objekte mit eigenem DDB-View (VZ/Vorgäng + Serien/Gliederungsstufen mit Beschreibung) - Insgesamt: 2841 mal
 - Max: 2484 mal (ID: Heuchelheim)
 - Beispiele:
 - : Insgesamt: 2484 mal (ID:)

Auswertung der Metadatenstrukturierung



Erweiterte Suche



0

Objekte in Merkliste

[◀ Zurück zur Ergebnisliste](#)

[Anfang](#) [Zurück](#) [Ergebnis 3 von 5](#) [Weiter](#) [Ende](#)

HIERARCHIE

Kreisarchiv Gießen (Archivtektonik) (automatisch generierter Testbestand)

▼ [Gemeindearchiv Heuchelheim, Heuchelheim](#)

3 Verhältnis zum Deutschen Reich und zu anderen Staaten

3.1 Verhältnis zum Deutschen Reich

▼ [Plakat mit dem Aufruf des Reichspräsidenten von Hindenburg an das Deutsche Volk vom 13.3.1930](#)





Plakat mit dem Aufruf des Reichspräsidenten von Hindenburg an das Deutsche Volk vom 13.3.1930

Verzeichnungsstufe:

Archivale

Archivaliensignatur

Kreisarchiv Gießen, Heuchelheim / 101

Lokale Voransicht transformierter Bestände

4. Datenintegration mit Hilfe des Data Preparation Tools

Grundlegende Datei- und Ordnerstruktur

Dateneingabe:

/data_input/{Archivname}/Bestand1.xml

/data_input/{Archivname}/Tektonik.xml

Datenausgabe:

/data_output/{Archivname}

Einstellungen:

/map2eadddb/provider_scripts/ead-{ISIL}/settings_provider.xml

/map2eadddb/settings/general_settings_p1.xml

/map2eadddb/settings/general_settings_p2.xml

- ▼ Transformation
 - > data_input
 - > data_output
 - ▼ map2eadddb
 - > div_module
 - > documentation
 - > ead
 - > general
 - ▼ provider_scripts
 - ead-ArchiveNRW
 - ead-arcinsys
 - ead-DE-105
 - ead-DE-476
 - ead-DE-1951
 - ead-DE-1985
 - ead-DE-2052
 - ead-DE-2088
 - ead-DE-2089

5. Fazit

Fazit

- Besondere Anforderungen archivischer Daten an die Repräsentation
- Einige Softwarehersteller investieren bereits in die Entwicklung einer EAD(DDB)-Schnittstelle
 - Allerdings noch nicht beim Endnutzer angekommen
- Data Preparation Tool schließt diese Lücke
- Eine spartenspezifische Lösung bietet einen deutlichen Mehrwert im Vergleich zum „kleinsten gemeinsamen Nenner“

6. Ausblick

Ausblick

- Sukzessive Etablierung einer Selbstbedienungsfunktion
 - **Data Preparation Tool** als Möglichkeit für Archive, ihre Daten lokal aufzubereiten und testen zu können
 - **Tutorials** für den Datenexport
 - → Archiven **mehr Kontrolle** in die Hand geben
- Nachhaltigkeit: Betreuung durch die Fachstelle Archiv ist auch über die Projektlaufzeit hinaus gewährleistet



„Stop hugging your data“

Sir Tim Berners-Lee, 2009

Vielen Dank für Ihre Aufmerksamkeit!

... Fragen?

Oliver Götze, MSc.

Landesarchiv Baden-Württemberg

archiv@deutsche-digitale-bibliothek.de

Diskussion (Vorschlag)

Inkonsistente
Daten, dafür
mehr
Information
enthalten

oder

Konsistente,
strikte
Datenstruktur,
auf Kosten
archivspezifischer
Anpassungen

Bildnachweise

Folie 27: „Stop hugging your data“. Paul Miller: *Why Linked Data?* <http://de.slideshare.net/cloudofdata/why-linked-data-2457996>