

Der Weg ins Portal

Erfahrungen der Fachstelle Archiv im Handling von Datenlieferungen

Workshop mit Herstellern archivischer Software
10. März 2016, Sächsisches Staatsarchiv
Dresden

Agenda

- 1. Ausgangssituation**
- 2. Anforderungen an Software-Exporte**
- 3. Herausforderungen und Problemstellung**
- 4. Wunschszenario: Optimaler Datenworkflow**
- 5. Fazit**
- 6. Ausblick**
- 7. Diskussion**

1. Ausgangssituation

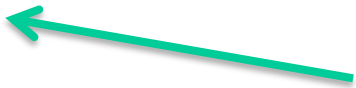
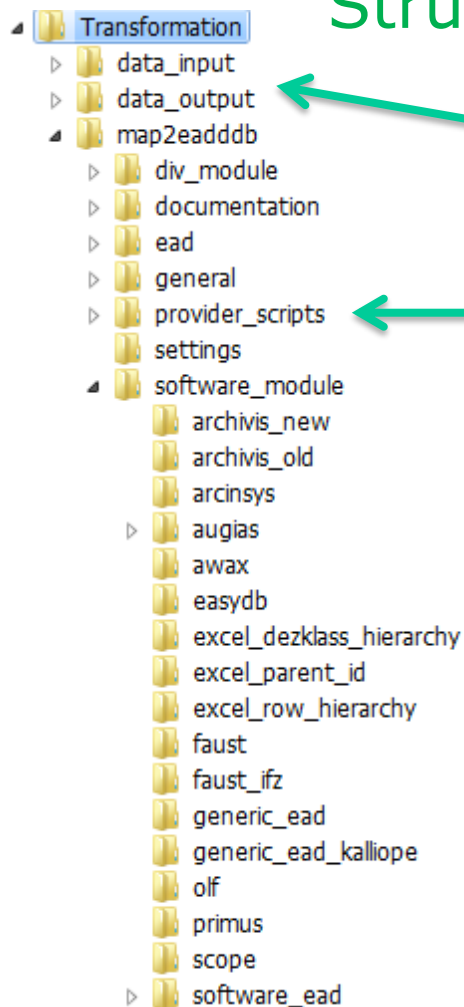
Grundbedingungen und Motivation

1. Ausgangssituation

- Vielzahl an Ausgangsformaten
- EAD-Export enthält nicht zwingend alle benötigten Informationen
- → bisheriges Vorgehen: Rückgriff auf proprietäre Exporte, Entwurf eigener Transformationsszenarien für einzelne Archivsoftware-Exporte
 - Ergänzt durch providerspezifische Skripte, um z.B. individuelle definierte Masken verarbeiten zu können
- Praktisch jede Lieferung wird noch einmal nachbearbeitet

1. Ausgangssituation

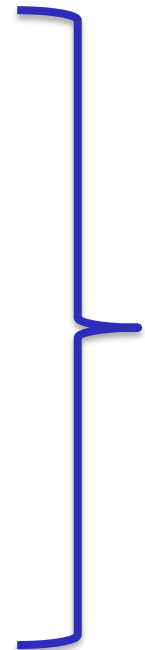
Strukturierung des Pre-Processings



Eingabe der Quell- und Ausgabe der Zieldaten



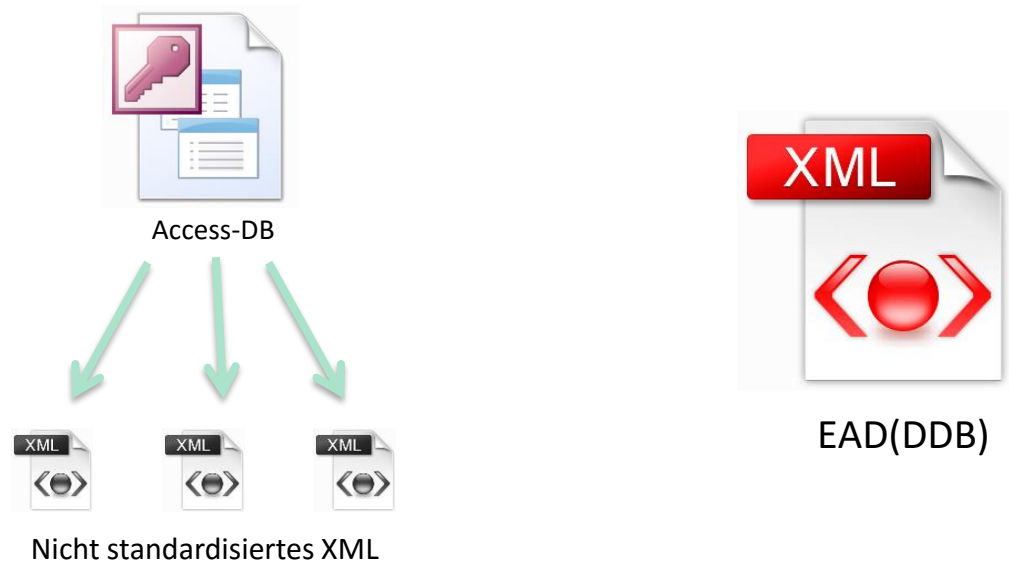
Providerspezifische Anpassungen



Softwarespezifische Module

Vom proprietären Datensatz zum Austauschformat EAD(DDB)

Exportformat	Verarbeitung
Proprietäre Binärformate	--
Proprietäre XML-Formate	-
Generische EAD-Formate	+
EAD(DDB) *1)	++



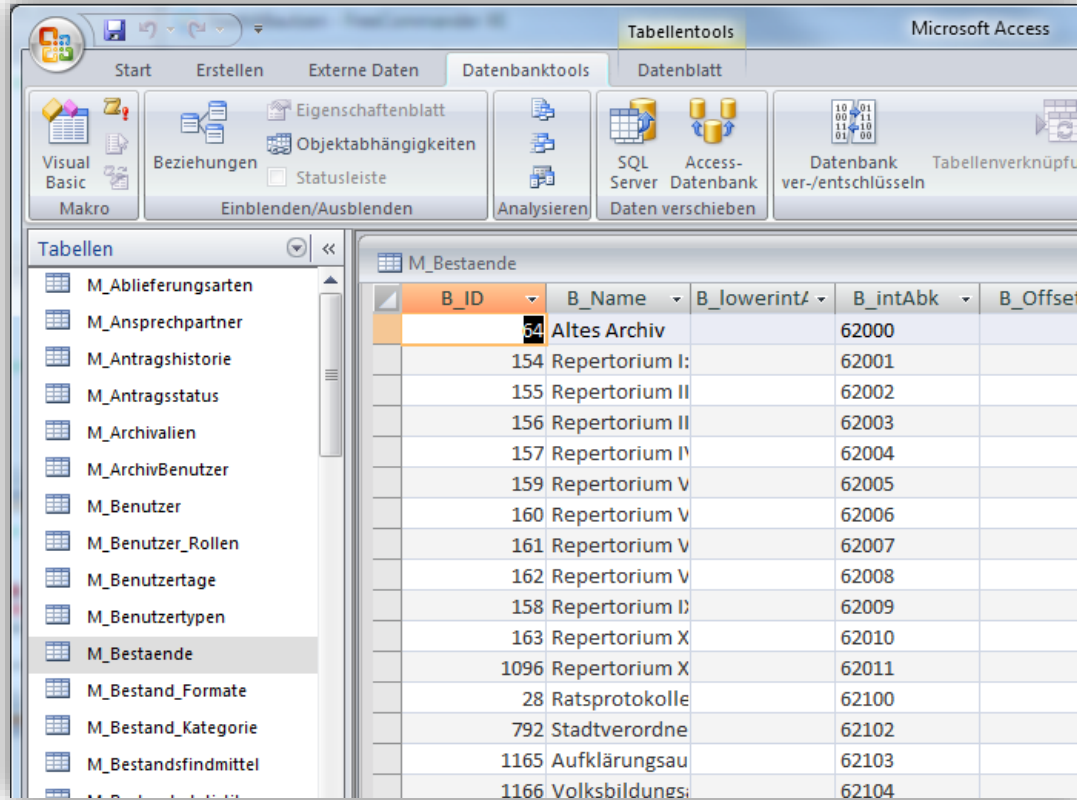
*1) EAD mit DDB-Anwendungsprofil

Vielzahl an Ausgangsformaten

Optionale Zwischenformate

Eingangsformat

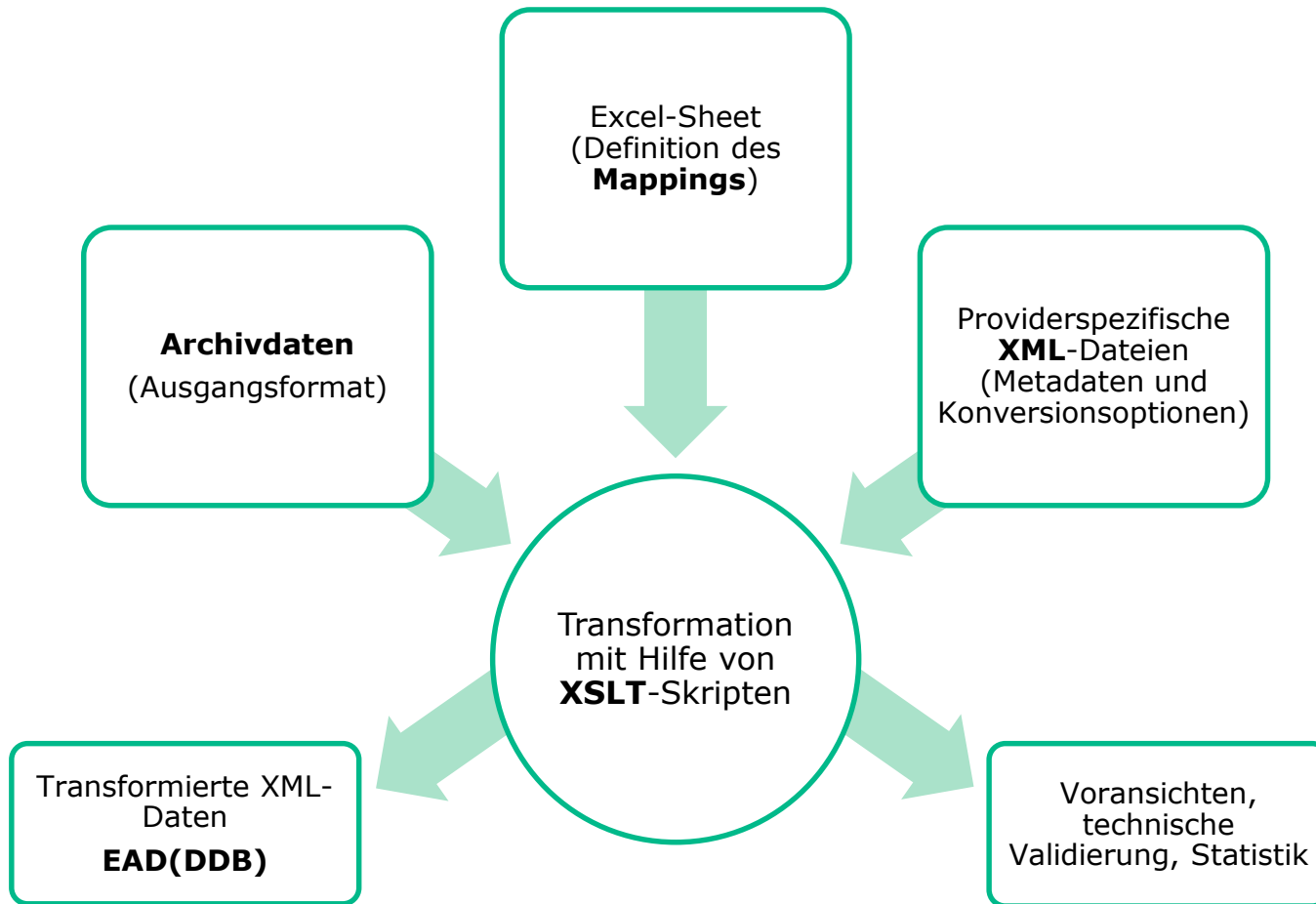
Vom proprietären Datensatz zum Austauschformat EAD(DDB)



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 IV		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 als Zwischenformat
- keine sprechenden Feldnamen
- Strukturierung abhängig von Formular-konfiguration
- Access ermöglicht XML-Export
 - nicht standardisiert, aber automatisiert weiterverarbeitbar
 - z.T. reichhaltiger als EAD-Export

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



2. Anforderungen an Software- Exporte

2. Anforderungen an Software-Exporte

- Stabile Identifier
- Konsistente Metadatendichte der Exportformate
- Ausgabe von wohlgeformtem XML
- Anbindung von Binärdaten
- Umgang mit Updates
 - Möglichkeit des Abgleichs/Versionierung
- Normdaten
- Kontextualisierung von Freitextdaten
- „Out-of-the-box“-Funktionalität: Vorkonfigurierter EAD(DDB) Basis-Export

3. Herausforderungen und Problemstellung

3. Herausforderungen und Problemstellung

- Fehlende/geänderte stabile Identifizier
 - Insbesondere bei Updates: Führt zu Dubletten im Portal
- Generischer EAD-Export oft weniger reichhaltig als proprietäre Formate bzw. XML-Anwendungsprofile
- z.T. keine Möglichkeit, Tektonik zu exportieren
- Umgang mit gesperrten Datensätzen
- Hohe Flexibilität führt zu Vielzahl an Variablen

3. Herausforderungen und Problemstellung

Beispiele

```
<c level="class" id="A90x76790040731430020150929142522055A90x80780237913131720150929142746090">
  <did>
    <unitid>
      <unittitle>I. a) Land-, Kreisausschuss- und St&#xE4;dtetagversammlungsakten und Nachrichten</unittitle>
    </unitid>
  </did>
  <c level="file" id="A90x39846742153167720150929142525830">
    <did>
      <unitid id="116">I a, Nr. 1</unitid>
      <unittitle>Landtagsrevers d. d. Grimma 1458</unittitle>
      <unitdate normal="1458"/>
    </did>
    <index>
      <indexentry>
        <geogname>Grimma</geogname>
      </indexentry>
    </index>
    <index>
      <indexentry>
        <other>Landtagsrevers</other>
      </indexentry>
    </index>
  </c>
```

- Auch wohlgeformte XML-Dateien enthalten syntaktische Fehler

3. Herausforderungen und Problemstellung

Beispiele

Schuldenzahlungsdeputation (Bestand)

Bestandssignatur:

Landesarchiv Baden-Württemberg Abt. Hauptstaatsarchiv Stuttgart, A 242

Bestandsbeschreibung:

Überlieferungsgeschichte

Die Deputation wurde während des Verfassungskampfes mit der Landschaft im Jahre 1768 errichtet und blieb nach dem Erbvergleich von 1770 als gemeinsame herrschaftliche und landschaftliche Deputation mit gemeinsamer Kasse zum Zwecke der Tilgung der Kammerschulden bestehen. Sie wurde 1803 aufgelöst, die Schuldenzahlungskasse bestand weiter und ging 1807 in der Generalstaatsschuldenkasse auf. Der Bestand bildete früher die Abteilung XIX des Finanzarchivs. Inhalt und Bewertung Protokolle der Deputation 1768-1793, Konzeptbücher der Kasse 1770-1806.

Online-Findbuch im Angebot des Archivs:

<http://www.landesarchiv-bw.de/plink/?f=1-3986>

Rechteinformation:

Es gelten die Nutzungsbedingungen des Landesarchivs Baden-Württemberg.

 Objekt beim Datenlieferanten anzeigen

Schuldenzahlungsdeputation (Bestand)

Bestandssignatur:

Landesarchiv Baden-Württemberg, Abt. Hauptstaatsarchiv Stuttgart, A 242

Bestandsbeschreibung:

Überlieferungsgeschichte

Die Deputation wurde während des Verfassungskampfes mit der Landschaft im Jahre 1768 errichtet und blieb nach dem Erbvergleich von 1770 als gemeinsame herrschaftliche und landschaftliche Deputation mit gemeinsamer Kasse zum Zwecke der Tilgung der Kammerschulden bestehen. Sie wurde 1803 aufgelöst, die Schuldenzahlungskasse bestand weiter und ging 1807 in der Generalstaatsschuldenkasse auf. Der Bestand bildete früher die Abteilung XIX des Finanzarchivs.

Inhalt und Bewertung

Protokolle der Deputation 1768-1793, Konzeptbücher der Kasse 1770-1806.

Online-Findbuch im Angebot des Archivs:

<http://www.landesarchiv-bw.de/plink/?f=1-3986>

Rechteinformation:

Es gelten die Nutzungsbedingungen des Landesarchivs Baden-Württemberg.

 Objekt beim Datenlieferanten anzeigen

Absätze werden im Portal nur angezeigt, wenn in `</lb>` umgewandelt

3. Herausforderungen und Problemstellung

Beispiele

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

Fehlende Kontextualisierung

3. Herausforderungen und Problemstellung

Beispiele

```
<?xml version="1.0"?>
```

```
<ead xmlns="urn:isbn:1-931666-22-9" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="urn:i  
<eadheader countryencoding="iso3166-1" dateencoding="iso8601" langencoding="iso639-2b" repositoryencoding="iso1  
<eadid url="www.zwickau.de">DE-Zwi18</eadid>  
<filedesc>  
  <titlestmt>  
    <titleproper>R1 - Rat der Stadt Altbestand Archivportal-D</titleproper>  
  </titlestmt>  
</filedesc>  
<profiledesc>  
  <creation>  
    <date normal="2015-10-02">2015-10-02+01:00</date>  
  </creation>  
</profiledesc>  
</eadheader>
```

Header: fehlende / falsch zugeordnete Informationen

3. Herausforderungen und Problemstellung

Beispiele

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

```
<ead xmlns="urn:isbn:1-931666-22-9" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="urn:is
```

```
<eadheader countryencoding="iso3166-1" dateencoding="iso8601" langencoding="iso639-2b" repositoryencoding="iso15
```

```
<eadid mainagencycode="DE-Zwi18" url="http://www.zwickau.de">A90x76790040731430020150929142522055</eadid>
```

```
<filedesc>
```

```
<titlestmt>
```

```
<titleproper>R1 - Rat der Stadt Altbestand Archivportal-D</titleproper>
```

```
</titlestmt>
```

```
</filedesc>
```

```
<profiledesc>
```

```
<creation>
```

```
<date normal="2015-10-02">2015-10-02+01:00</date>
```

```
</creation>
```

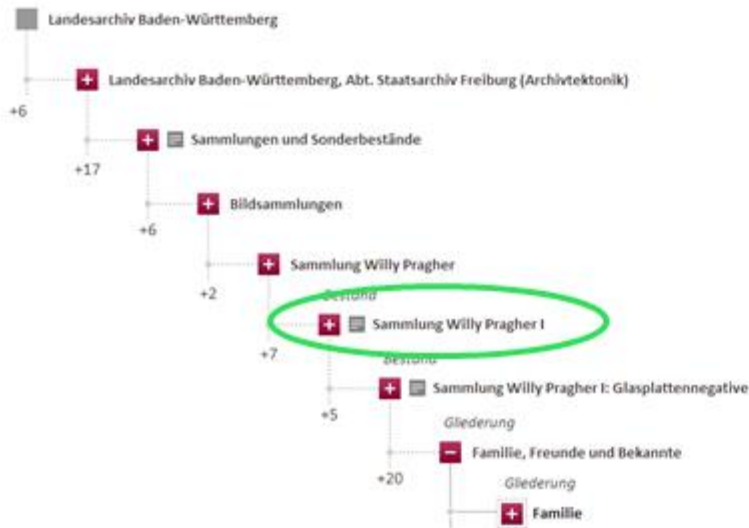
```
</profiledesc>
```

```
</eadheader>
```

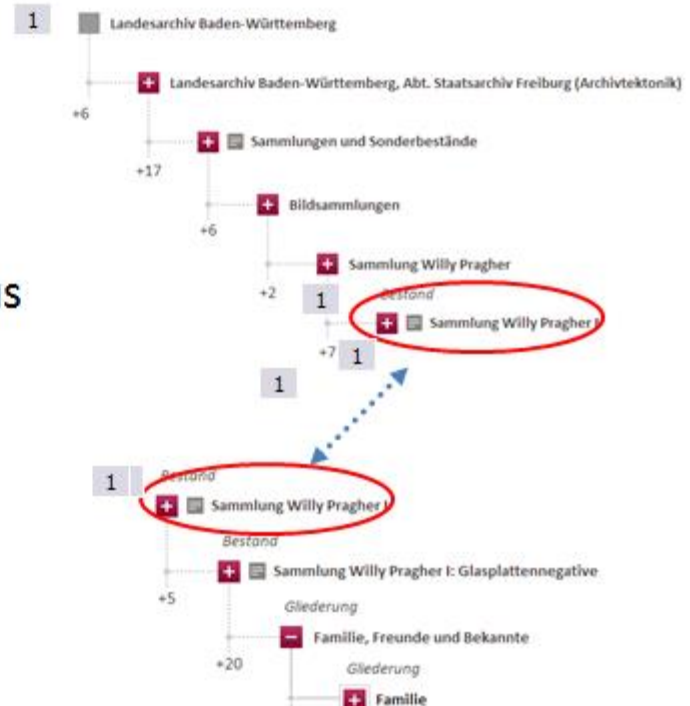
3. Herausforderungen und Problemstellung

Beispiele

Vollständige Hierarchien



1 Keine Verknüpfung zwischen Tektonik und Hierarchie



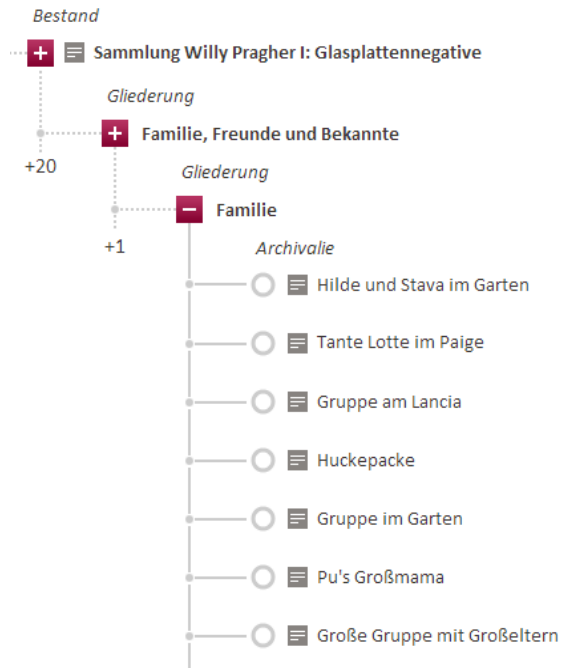
versus

fehlende Verknüpfung zwischen Findbuch und Tektonik

3. Herausforderungen und Problemstellung

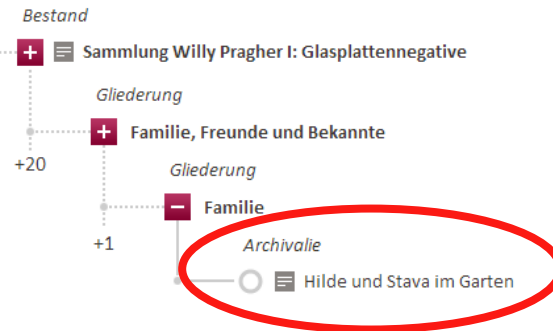
Beispiele

Stabile Identifier



Gleiche Identifier:

Daten werden überschrieben



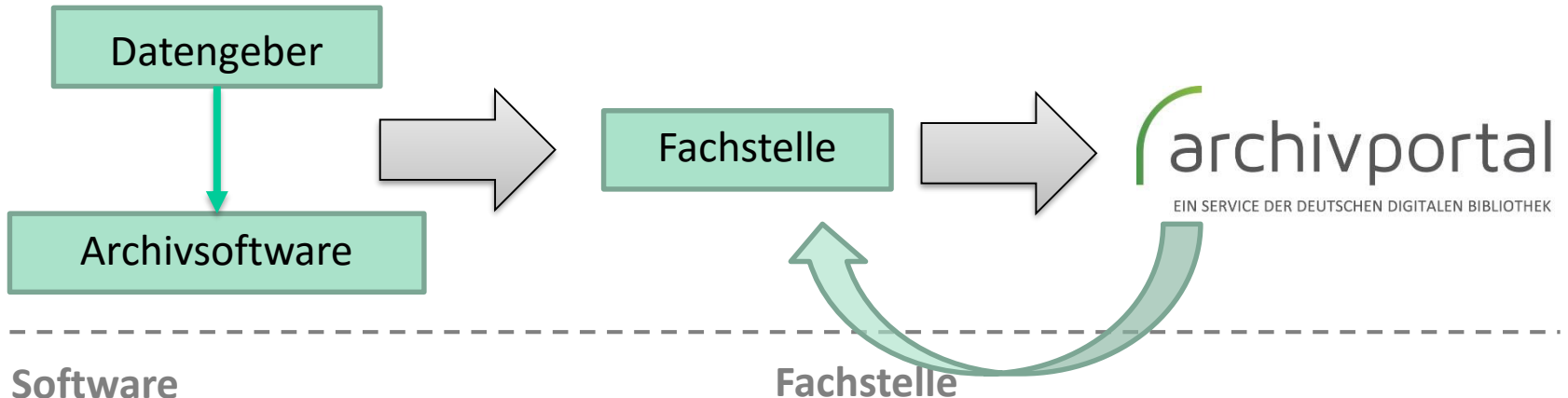
Geänderte Identifier:

Daten werden doppelt geladen



4. Skizze: Optimaler Datenworkflow

4. Skizze: Optimaler Datenworkflow



Software

- Schemavalidierung, Wohlgeformtheit
- Rollen bei Indexbegriffen
- EAD(DDB)-Export vorkonfiguriert
- Verknüpfung und korrekte Benennung sichergestellt
- Stabile Identifier als zentraler Bestandteil

Fachstelle

- Datenübertragung
- Auf Grundlage des Basis-EAD(DDB)-Exports weitere Anpassungen (bilateral)
- Weniger Datenverluste durch stabile Identifier
- Verarbeitung der Feldinhalte weiterhin durch Fachstelle (providerspezifische Skripte)
- Weniger Iterationen beim Laden ins Portal

5. Fazit

5. Fazit

- **Stabile Identifier** als zentrales Qualitätskriterium
 - Tiefere Integration auf Softwareebene wünschenswert
 - Umgang mit **Updates** bei vorangegangenen Lieferungen
- Proprietäre Exportformate z.T. reichhaltiger als Belegung des (generischen) EAD-Exports
 - Aufwendige und teils fragile Konversionsworkflows der Fachstelle als Übergangslösung
 - Langfristig: **valide EAD(DDB)-Exporte** benötigt
- Scheinbar geringe Anpassungen führen zu hohem Qualitätszuwachs

6. Ausblick

6. Ausblick

- Data Preparation Tool (DPT) als Lückenfüller bei „offiziellen“ EAD(DDB)-Export-Workflows
 - Behält seine Relevanz als Validierungs- und Auswertungssystem
- DDB plant Einführung von Selbstbedienungskomponenten, Validierung (Schematron) mit feinerer Granularität
 - Erfordert gewissen Mindeststandard an Datenqualität
- Exportschnittstellen der Softwareanbieter kommt hohe Bedeutung zu

Vielen Dank für Ihre Aufmerksamkeit!

... Fragen?

Oliver Götze, MSc.

Landesarchiv Baden-Württemberg

oliver.goetze@la-bw.de

Diskussion

In welchem Maße ist eine Anpassung Ihrer Software an die Anforderungen von EAD(DDB) möglich?