

Lazarmus - Projektergebnisse aus technischer Sicht

20.07.2021

Kennzahlen

- Schwabensammlung:
 - 122.100 Dateien, 218 GB
- Igoras
 - 1590 Dateien, 267 MB
- imdas pro
 - 97.000 Dateien, 24 GB
- Cumulus
 - 290.000 Dateien, 4 TeraByte

Fachanwendungen

- Exporte der Rohdaten (Metadaten + Mediendateien) wurden von den Datenbankbetreibern geliefert
- Igoras: (MySQL Datenbank)
 - Export als CSV
- imdas pro: (Oracle Datenbank)
 - Export als MySQL, Umwandlung in CSV ; Teilmenge in XML
- Cumulus:
 - Exportroutine in XML bereits vorhanden

Übernahme

- Schaffung der nötigen Infrastruktur (Serverlaufwerk, externe Festplatten)

Herausforderungen der Übernahme

- Gesamtmenge (4 TB)
- Sehr große Einzeldateien in XML (> 2 GB)
- Übernahme von > 250.000 Dateien in einer einzelnen Verzeichnisebene

Datenanalyse, -aufbereitung

- Neustrukturierung der flach übernommenen Cumulus-Bilder in Verzeichnisstruktur
- Analyse der Schwaben-Dateisammlung, Dokumentation der originalen Verzeichnisstruktur als XML-Hierarchie
- Überprüfung aller Übernahmen auf inhaltliche Vollständigkeit
- Analyse von XML-Dateien und CSV-Dateien auf Wohlgeformtheit und syntaktische Validität
- Migration von Dateiformaten (z.B. XML 1.1. in XML 1.0 ; SQL in CSV)
- Überführung von CSV-Tabellen in Datenbankschema und Nachbildung der Relationen zwischen Tabellen

```
<rootdir path="T:\LAZARMUS\Schwaben">
  <directory name="2016DieSchwaben">
    <directory name="S 00 Allgemein">
      <directory name="Adressen">
        <file name="140610 Alle e-mail-Adressen Beirat.docx"/>
        <file name="151106 Adressen Die Schwaben.xlsx"/>
        <directory name="Adresslisten alt">
          <file name="131210 Adressen1.xls"/>
          <file name="140115 Adressen Die Schwaben.xlsx"/>
          <file name="140825 Adressen Die Schwaben, aktuell.xlsx"/>
          <file name="Adressen (3) Muster.xlsx"/>
        </directory>
        <file name="Einladung Nachzügler.docx"/>
        <file name="Thumbs.db"/>
      </directory>
      <directory name="Andere Ausstellungen">
        <directory name="Baden 900, Karlsruhe">
          <file name="Baden 900.pdf"/>
        </directory>
        <directory name="Beuren, Typisch Schwäbisch">
          <file name="2016-2-28-Schreiben zur Ausstellung-Entwurf Ewigleben.docx"/>
          <directory name="Beuren">
            <file name="Beuren-FLM_SchwabenAusstellung_2016-04-14_160617.jpg"/>
            <file name="Beuren-FLM_SchwabenAusstellung_2016-04-14_160952.jpg"/>
          </directory>
        </directory>
      </directory>
    </directory>
  </directory>
</rootdir>
```

person	[table]
person_id	
nachname	
vorname	
notiz	
geschlecht	
gkdid	
modified	
anzeigenane	
swdid	
pndid	
langnotiz	
18.433 rows	8 >

thesaurustyp_id	
name	
modified	
95 rows	1 >

datumstyp	[table]
datumstyp_id	
name	
modified	
33 rows	3 >

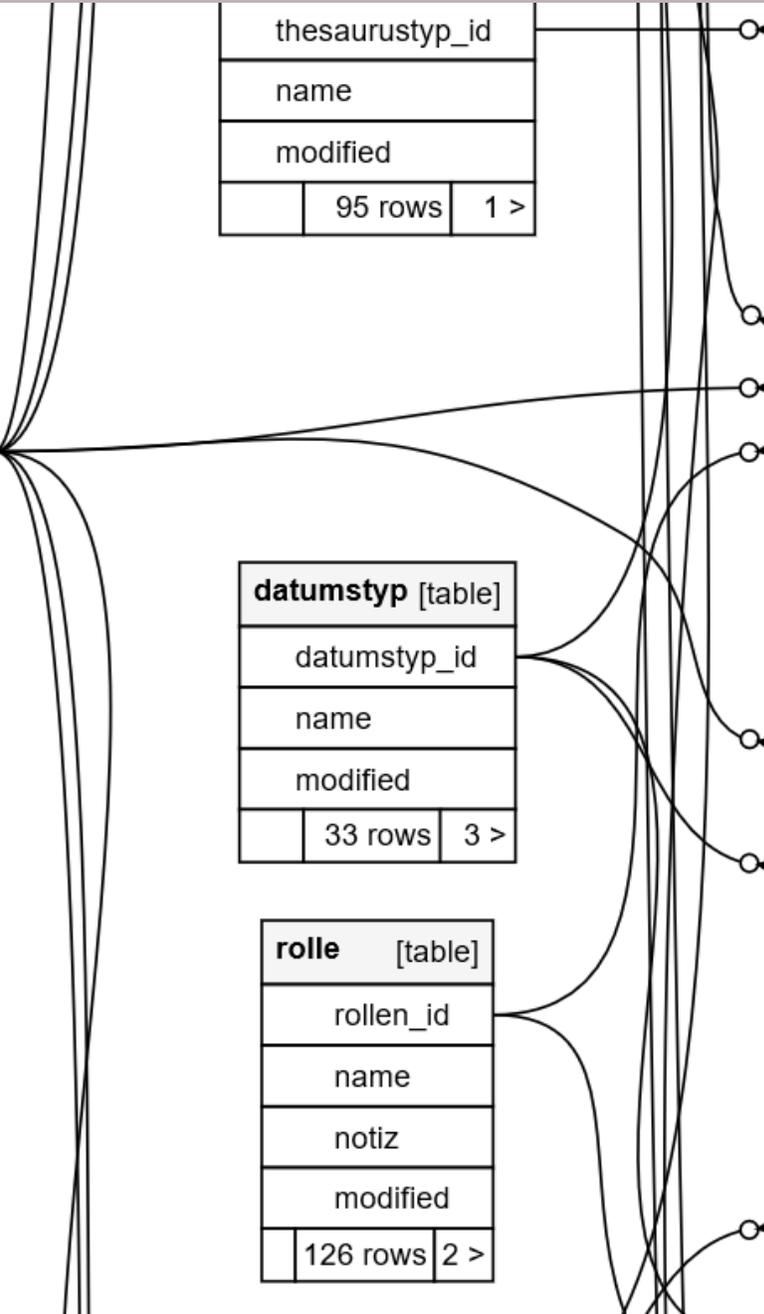
rolle	[table]
rollen_id	
name	
notiz	
modified	
126 rows	2 >

thesaurustyp_id	
modified	
< 4	737.475 rows

objekt_person	[table]
objekt_id	
personen_id	
rollen_id	
modified	
< 3	188.247 rows

person_datum	[table]
person_id	
datum	
datumstyp_id	
notiz	
modified	
< 2	8.473 rows

ereignisdatum	[table]
ereignis_id	
datum	



Archivalienbildung und Archivierung in DIMAG

- Ungewöhnlich große Archivalieneinheiten
- Programmierung eines automatisierten Ingestprozesses in das Langzeitarchivierungssystem DIMAG
- Dateiformaterkennung
- Erfassen von dateispezifischen Metadaten
- Technische Dokumentation je Archivalieneinheit

-	EL 230/9 Landesmuseum Württemberg: Digitale Objekte	
+	Mediendatenbank Cumulus	
-	Große Landesausstellung Die Schwaben (2016-2017)	
+	Orientierungshilfsmittel zur Dateisammlung	
-	Dateisammlung	
.....	EL 230/9_DO 4 Ausstellungsauf- und abbau	2016.10.19-2017.04.24
.....	EL 230/9_DO 5 Ausstellungstexte	2015.11.20-2016.11.25
.....	EL 230/9_DO 6 Auswertung und Abschlussbericht	2012.05.09-2017.08.09
.....	EL 230/9_DO 7 Drittmittel und Gremien	2014.02.13-2017.08.21
.....	EL 230/9_DO 8 Konzeptions- und Planungsphasen	2012.10.16-2017.06.16
.....	EL 230/9_DO 9 Museumspädagogik	2012.10.04-2017.05.31
.....	EL 230/9_DO 10 Objektdisposition	2013.05.13-2017.06.13
.....	EL 230/9_DO 11 Projektmanagement	2012.08.02-2018.04.13
.....	EL 230/9_DO 12 Volontariatsprojekt	2013.12.10-2017.02.22
.....	EL 230/9_DO 13 Öffentlichkeitsarbeit	2000.02.28-2018.04.13
.....	EL 230/9_DO 14 Grafik und Medien	2014.12.01-2017.08.01
.....	EL 230/9_DO 15 Ausstellungsdokumentation	2016.12.12-2017.06.08
+	Internetgestützte Objekterfassung und Recherche für das Ausstellungs- und Sammlungswesen (IGORas)	
+	Inventarisierungssoftware IMDAS Pro	

Dateisammlung

Signatur/AID



Hier einfügen: (Gekoppelt mit AFIS)

EL 230/9_DO 4			Ausstellungsauf- und abbau
EL 230/9_DO 15			Ausstellungsdokumentation
EL 230/9_DO 5			Ausstellungstexte
EL 230/9_DO 6			Auswertung und Abschlussbericht
EL 230/9_DO 16			Dokumentation der Dateisammlung
EL 230/9_DO 7			Drittmittel und Gremien
EL 230/9_DO 14			Grafik und Medien 
EL 230/9_DO 8			Konzeptions- und Planungsphasen
EL 230/9_DO 9			Museumspädagogik
EL 230/9_DO 10			Objektdisposition
EL 230/9_DO 13			Öffentlichkeitsarbeit
EL 230/9_DO 11			Projektmanagement
219-2449329			Protokoll zu 2-5922197 Dateisammlung
EL 230/9_DO 12			Volontariatsprojekt

📁 Dateisammlung
📁 Grafik und Medien
📁 R 1: Zugang

▲▼ Signatur/AID	▲▼ Titel
🔗	Hier einfügen: 📄 Primärdatenobjekt 📄 Dokumentationsobjekt 📄 Feldbeschreibung 📄 In
219-2463804	📄 ✂️ Setzkasten-LOOP.bpf
219-2463803	📄 ✂️ 969_I8N.m4a
219-2463802	📄 ✂️ 968_I8N.m4a
219-2463801	📄 ✂️ 967_I8N.m4a
219-2463800	📄 ✂️ 966_I8N.m4a
219-2463799	📄 ✂️ 965_I8N.m4a
219-2463798	📄 ✂️ 05Erfinderin_Schnitt002 FINAL HOCH TC KORREKTUR.mp4 
219-2463797	📄 ✂️ 05Erfinderin_Schnitt001_FINAL NUR HANDLUNG.mp4
219-2463796	📄 ✂️ 04Bankfrau_SCHNITT003 FINAL HOCH TC KORREKTUR.mp4
219-2463795	📄 ✂️ 03Hausfrau_SCHNITT_003 FINAL HOCH TC KORREKTUR.mp4
219-2463794	📄 ✂️ 02Hauslebauer_Schnitt_002 FINAL HOCH TC KORREKTUR.mp4
219-2463793	📄 ✂️ 01Pfarrer_Schnitt_001 FINAL HOCH TC KORREKTUR.mp4
219-2463792	📄 ✂️ Rechnung-Zeppelin-59387 Landesmuseum Württemberg Gebel.pdf

Metadaten

Archival ID (AID)	219-2463798
Parent ID (PID)	219-2463254
Objekt-Versionsnummer	1
Titel	05Erfinderin_Schnitt002 FINAL HOCH TC KORREKTUR.mp4
Ursprünglicher Dateiname	05Erfinderin_Schnitt002 FINAL HOCH TC KORREKTUR.mp4
Relativer Dateipfad	Grafik und Medien/Festplatte/R5 Setzkasten 5 Filme
Pronom Persistent ID	fmt/199
Objekttyp	 Primärdatenobjekt
Statuscode	 10 (In Bearbeitung)
Angelegt von	David Radlinger
Angelegt am	2020-12-11 07:29:07
Geändert von	David Radlinger
Geändert am	2020-12-11 07:29:07
Alternate ID	2aaaf346-3b7a-11eb-aedc-02bf86500189
Dateigröße in Bytes	1079732752
MD5-Wert der Datei	3c7aa25dbc6b2eaf115a6fe2e5348fd7
Letzte Dateiänderung (abg. Stelle)	2016-10-13 09:45:37
Format (Pronom)	MPEG-4 Media File
Pronom-Klassifikation	Positive (Specific Format)
MIME-Type (Pronom)	application/mp4, video/mp4
Pronom-Signature	DROID_SignatureFile_V97.xml
AV-Objekt Abspieldauer (MediaInfo)	00:09:21.05
AV-Objekt Containerformat (MediaInfo)	MPEG-4
AV-Objekt Gesamtbitrate (MediaInfo)	15.4 Mbps
Audio Bitrate (MediaInfo)	317 Kbps
Audio Kanäle (MediaInfo)	2 channels
Audio Abspieldauer (MediaInfo)	00:09:21.05
Audio Codec (MediaInfo)	AAC
Audio Abtastrate (MediaInfo)	48.0 KHz
Video Bitrate (MediaInfo)	15.1 Mbps
Video Farbmodell (MediaInfo)	YUV
Video Seitenverhältnis (MediaInfo)	16:9
Video Codec (MediaInfo)	AVC
Video Bildfrequenz (MediaInfo)	25.000 fps
Video Bildhöhe (MediaInfo)	1 080 pixels
Video Bildbreite (MediaInfo)	1 920 pixels

Fazit

- Technische Datenanalyse und -aufbereitung unerlässlich
- Automatisierte Prozesse und Prozesssteuerung notwendig
 - Prüfroutinen, technische Validierung
- Hierfür eingesetzte Mittel:
 - Programmiersprachen: Python, PHP, R
 - Datenbank Anwendungen: MySQL, SQLite
 - XML: X-Technologien