

Schnittstellen, Metadatenmodell und Austauschformate

Jakob Voß

Verbundzentrale des GBV (VZG)

LaZAR Workshop, Jena, 2016-04-19

Schnittstellen

Arten von Schnittstellen

- ▶ Nutzerschnittstellen / Benutzeroberflächen für Menschen
- ▶ Interne Programmierschnittstellen für Programmierer
- ▶ Web-APIs / Web-Services für Programme

Arten von Schnittstellen

- ▶ ~~Nutzerschnittstellen~~ / ~~Benutzeroberflächen~~ für Menschen
- ▶ ~~Interne Programmierschnittstellen~~ für Programmierer
- ▶ **Web-APIs** / **Web-Services** für Programme

Technische Grundlage

- ▶ Meist einfache URL-Anfragen über HTTP[S]:
`http://example.org/gib/mir/?datensatz=123`
`https://example.org/etwas/damit?machen=los`
- ▶ Allgemeinen Datenformate wie CSV, JSON, XML, RDF...

Buzzwords: *REST*, *Mashup*, ...

Wofür werden Schnittstellen benutzt?

- ▶ **Daten rein** (Import, Anreicherung...)
- ▶ **Daten raus** (Export, Abfrage...)
- ▶ **Aktion anstoßen** (Freischalten, Benachrichtigen, Sperren...)

... und zwar automatisierbar!

Beispiel: Abfrage von Daten

Abfrage von Briefen bei `correspSearch`:

```
http://correspsearch.bbaw.de/api/v1/csv.xml  
?sender=http://d-nb.info/gnd/118554700
```

⇒ Liste digitalisierter Briefe von Alexander von Humboldt

Beispiel: URIs und Kombination von Schnittstellen

Suche nach GND-Normdaten bei <http://lobid.org/>:

<http://lobid.org/person>

[?name=Alexander%20von%20Humboldt&format=ids](http://lobid.org/person?name=Alexander%20von%20Humboldt&format=ids)

⇒ Liste der GND-URIs verschiedener Alexander von Humboldts

⇒ <http://d-nb.info/gnd/118554700>

Beispiel: URIs und Kombination von Schnittstellen

Suche nach Geonames-URIs bei <http://geonames.org/>:

[http://api.geonames.org/search
?q=Jena&maxRows=10](http://api.geonames.org/search?q=Jena&maxRows=10)

⇒ Liste der Geonames-URIs von Jena

⇒ <http://www.geonames.org/2895044>

Beispiel: URIs und Kombination von Schnittstellen

```
http://correspSearch.bbaw.de/api/v1/tei-xml.xql  
?sender=http://d-nb.info/gnd/118554700  
&placeSender=http://www.geonames.org/2895044
```

⇒ Ein in Jena von Alexander von Humboldt abgeschickter Brief

Die Bücher, wie Götter, Erd- u. Himmel
in der Pflanzung des, wie die von Buch
2. Buche gegeben. Alle die sind in
in der Pflanzung die ich geistig dargestellt
die Welt in 2 in 2 dargestellt ist.
die ist die. Die Welt ist in der Pflanzung
Menschheit 2 Götter. Man hat die
man in der Pflanzung die in der Pflanzung. In
ist die Pflanzung die in der Pflanzung von dem. Die
und man muss wissen über die Pflanzung
die von der Pflanzung. Man hat die
gegeben. Man hat die alle die, die die ist
gegeben können. ^{Man} ^{Alte} ^{Wamboldt}
Jahre 18 Jhr. 95.

AP 2-3: Entwicklung von Exportschnittstellen

Ziel: Entwicklung von Exportschnittstellen zur Datenbereitstellung

Beteiligung: VZG

Zeit: Monat 15-24 (im ursprünglichen Antrag)

Grundlage: Repository; Datenmodell (AP 1-3)

Arbeitsschritte:

- ▶ Analyse bestehender Exportschnittstellen
- ▶ Entwicklung Exportschnittstellen (OAI-PMH und LOD)
- ▶ Test und Evaluation der Exportschnittstellen
- ▶ Bereitstellung der Daten

Umsetzung und Möglichkeiten

- ▶ easyDB greift selbst auf APIs der VZG zu (GND & Geonames)
- ▶ easyDB bietet eigene API für praktisch alle Funktionen (allerdings eher als Programmierschnittstelle)
- ▶ VZG setzt darauf aufbauend eine Auswahl von frei zugänglichen Standard-Schnittstellen um
 - ▶ OAI-PMH
 - ▶ Linked Open Data (LOD)

→ *Lesende Schnittstellen auch indirekt über Datenexport möglich!*

OAI Protocol for Metadata Harvesting (OAI-PMH)

- ▶ Etablierter Standard (2002) zum Austausch von Metadaten aus Repositories
- ▶ Für Forschungsdaten-Repositories notwendig
- ▶ Im Wesentlichen nur eine Art von Anfrage möglich:
 - ▶ Was ist seit ... hinzugekommen oder hat sich geändert?
 - ▶ ggf. Einschränkung auf vordefinierte Gruppen von Inhalten

Linked Open DATA (LOD) aka “Semantic Web”

- ▶ Vereinheitlichung aller Daten mittels URIs
 - ▶ DOI, GND, Geonames...
- ▶ Nur für frei verfügbare Daten
- ▶ Schwerpunkt auf Verknüpfungen
 - ▶ z.B. welche Urheber (GND-URI) haben zu welchen Themen (GND-URI) an welchen Orten (Geonames-URI) erhoben?
- ▶ Genaue Art des Export noch offen

Metadatenmodell und Austauschformate

Was ist ein Metadatenmodell?

- ▶ Mögliche Arten von
 - ▶ Inhalten (Objekttypen)
 - ▶ Beziehungen (Verknüpfungen, Relationen)
 - ▶ Werten (Datentypen und Auswahllisten)
- ▶ Grundlage von konkreten Datenformate
- ▶ Beschränkung auf *eine mögliche Art* der Beschreibung

Beispiel: Urheber

- ▶ Jedes Objekt (Datei, Konvolut, Ausschnitt) hat *mindestens einen* **Urheber**
- ▶ Ein **Urheber** muss eine GND-ID oder ORCID-ID haben.
- ▶ Ein **Urheber** kann eine Affiliation haben (Freitext)

LaZAR-Metadatenmodell

- ▶ Allgemeine Objekttypen
 - ▶ Dateien (Video, Audio, Bild, Text, Sonstiges)
 - ▶ Konvolute (Sammlung von Dateien und/oder Unterkonvoluten)
- ▶ Datenfelder mit verschiedenen Beschränkungen
- ▶ Einige Auswahllisten

⇒ Genauere Anpassung erfordert tatsächliche Forschungsdaten

- ▶ Export für Langzeitarchivierung
- ▶ Export für Suche in publizierten Forschungsdaten (DataCite)
 - ▶ Für DOI-Vergabe
 - ▶ DataCite-Format (notwendig)
<https://schema.datacite.org/>
- ▶ Ggf. weitere Exporte (ZIP-Archive o.Ä.)

⇒ Abbildung des internen Datenformat aus easyDB nach DataCite

Zusammenfassung

- ▶ Schnittstellen zur Integration mit anderen Programmen
 - ▶ OAI-PMH
 - ▶ LOD
- ▶ Zwingend notwendig
 - ▶ Verknüpfung mit GND/ORCID-URLs
 - ▶ Export per OAI-PMH Schnittstelle
 - ▶ DataCite als ein Exportformat