

# Metadateninteroperabilität

Jakob Voß

Humboldt Universität, Berlin, 2017-06-27

## Kurzvorstellung

# Gemeinsamer Bibliotheksverbund (GBV)

- ▶ Größter Bibliotheksverbund in Deutschland
- ▶ Zentrale Katalogisierung (ab 2018 zusammen mit dem BSZ)

- ▶ Verbundzentrale des GBV mit Sitz in Göttingen
- ▶ Etwa 80 MitarbeiterInnen
- ▶ CBS & LBS-Hosting und -Betreuung
- ▶ Repositories, Portale, Discovery-Interfaces etc.

*Wir haben Praktika & Themen für Bachelorarbeiten!*

- ▶ Studium der Informatik & Bibliothekswissenschaft am IBI
- ▶ Stabstelle Forschung & Entwicklung an der VZG
- ▶ <http://jakobvoss.de>

# Interoperabilität

“Fähigkeit unterschiedlicher Systeme, möglichst nahtlos zusammenzuarbeiten” (Duden)

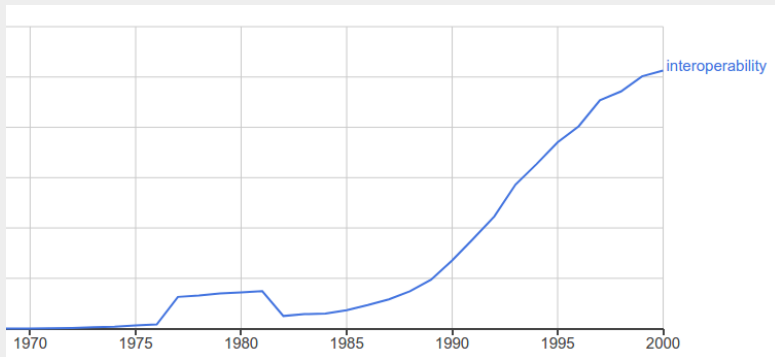
*Beispiele?*



CC BY-NC-SA fffff.at



# Interoperabilität



Quelle: <https://books.google.com/ngrams>

- ▶ Sozialgesetzbuch Gesetzliche Krankenversicherung

## **SGBV V § 291e Interoperabilitätsverzeichnis**

“Die Gesellschaft für Telematik hat bis zum 30. Juni 2017 ein elektronisches Interoperabilitätsverzeichnis für technische und semantische Standards, Profile und Leitfäden für informationstechnische Systeme im Gesundheitswesen aufzubauen und dieses Interoperabilitätsverzeichnis zu pflegen und zu betreiben. Das Interoperabilitätsverzeichnis dient der Förderung der Interoperabilität zwischen informationstechnischen Systemen.”

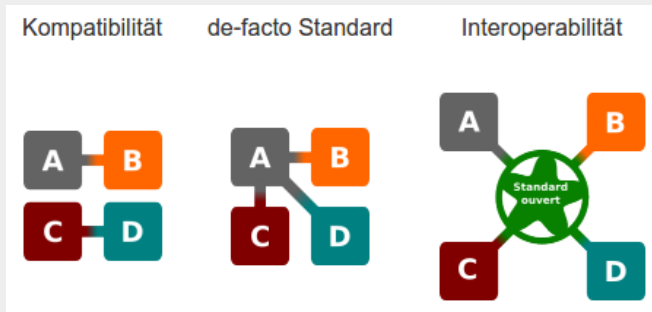
<http://www.vesta-gematik.de/>

- ▶ Deutsche Initiative für Netzwerkinformation (DINI),  
Kompetenzzentrum Interoperable Metadaten (KIM)

## **DINI AG KIM**

<https://dini.de/ag/standards/>

# Interoperabilität von IT-Systemen



<http://interoperability-definition.info/> (CC BY-SA)

⇒ Gemeinsame Standards/Schnittstellen unterschiedlicher Systeme

# Beispiele gemeinsamer Standards & Schnittstellen

- ▶ TCP/IP
- ▶ HTTP
- ▶ WWW
- ▶ ...

*Interoperabilität auf verschiedenen Ebenen!*

Interoperabilität basiert auf

- ▶ gemeinsamen Standards & Schnittstellen

# Beispiel aus der Praxis

## Document Availability Information API (DAIA)

*Aktuelle Verfügbarkeit von Dokumenten in Bibliotheken*

*Interoperabilität von Lokalsystemen und User Interfaces*

- ▶ Schnittstelle (HTTP-Anfrage mit Parametern)
- ▶ Datenmodell (Documents, Items, Services...)
- ▶ Datenformat (XML/JSON/RDF, inzwischen primär JSON)

<https://gbv.github.io/daia/>

## *Metadaten* interoperabilität

# Metadaten

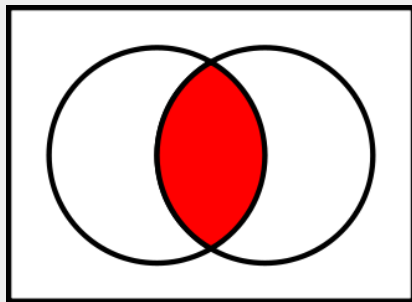
- ▶ Daten die andere Daten oder Objekte strukturiert beschreiben
- ▶ Alle [...] als Daten vorliegenden Formen der Erschließung
- ▶ Wesentlich [...] das Vorhandensein einer einheitlichen Struktur

Aus meinem Eintrag “Metadaten” im Lexikon der Bibliotheks- und Informationswissenschaft (LBI) von 2011



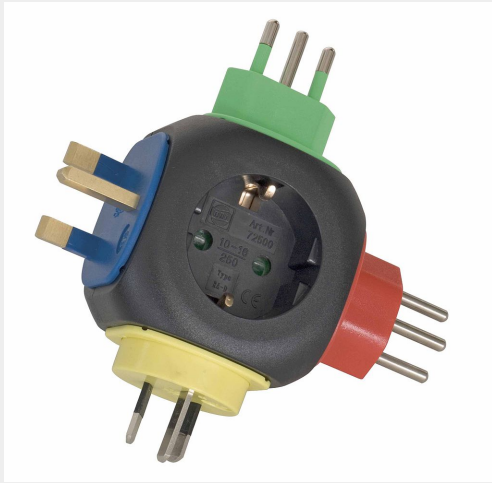
# Metadateninteroperabilität

- ▶ ein System hat (hoffentlich) eine einheitliche Struktur
- ▶ verschiedene Systeme haben verschiedene Strukturen



Schnittmenge

# Beispiele



CC BY-SA Cephira

## Beispiel: Adapter

- ▶ Einheitliche Struktur: Stecker, Spannung & Frequenz
- ▶ Gemeinsame Struktur: Netzspannung und -Frequenz

⇒ Grundlage: gemeinsame Schnittmenge auf einer Ebene

# Allgemeines Vorgehen

- ▶ Gemeinsame Schnittmengen von Metadaten-Strukturen finden
- ▶ Adapter/Mappings auf der passenden Ebene entwickeln

*Oder auf bereits vorhandene Standards zurückgreifen!*

## Beispiel: MARC

- ▶ Entwickelt 1966 auf Basis von Karteikarten
- ▶ MARC must die! (Tennant 2002)
- ▶ Noch immer üblichster Standard ( $\Rightarrow$  BIBFRAME?)

*Wo liegt das Problem?*

# Beispiel: 900er-Felder in MARC

- ▶ Grundsätzlich Teil des MARC-Standard
- ▶ Inhaltlich nicht Teil des MARC-Standard

*eigene Erweiterungen  $\neq$  gemeinsame Struktur*

## Beispiel: unAPI & SRU

- ▶ Verschiedene Schnittstellen an der VZG
- ▶ Viele, meist einheitliche Formate
- ▶ Konvertierung on-the-fly

PICA → MARC21 → MARCXML → MODS ...

<http://unapi.gbv.de/>



## Beispiel: JSON und XML

```
{  
  "author": "Emma Goldman",  
  "year": "2014",  
  "isbn": "978-3-89401-810-8"  
}
```

```
<record>  
  <author>Goldman, Emma</author>  
  <year>2014</year>  
  <isbn>978-3-89401-810-9</isbn>  
</record>
```

## Beispiel: JSON & JSON

```
{  
  "author": "Emma Goldman",  
  "year": "2014",  
  "isbn": "978-3-89401-810-8"  
}
```

```
{  
  "author": [ "Goldman, Emma" ],  
  "date": "2014",  
  "isbn": "9783894018108"  
}
```

## Beispiel: Creator-Feld in Dublin Core

- ▶ Nachname, Vorname
- ▶ Vorname Nachname
- ▶ Vorname Nachname (Sprache: Original)
- ▶ ...

*Metadaten-Strukturen wachsen wie Unkraut*

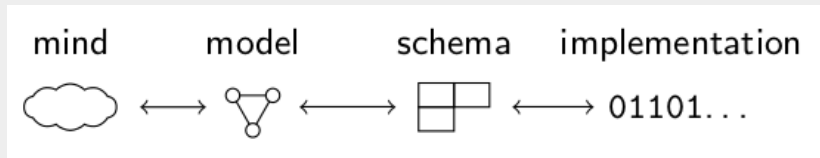
# Metadaten-Strukturen

# Metadaten-Strukturen?

- ▶ Wesentlich für Metadaten ist das Vorhandensein einer einheitlichen Struktur. Diese kann u.A. als Schema (Kategorienkatalog, Datendefinitionssprache), Profil, Regelwerk, Datenformat oder Modell (Ontologiesprache) vorliegen.

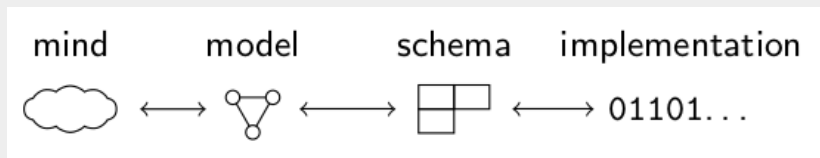
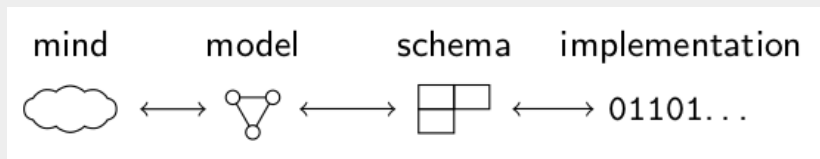
Aus meinem Eintrag “Metadaten” im Lexikon der Bibliotheks- und Informationswissenschaft (LBI) von 2011

# Wie entstehen Daten(-Strukturen)?



Datenmodellierung

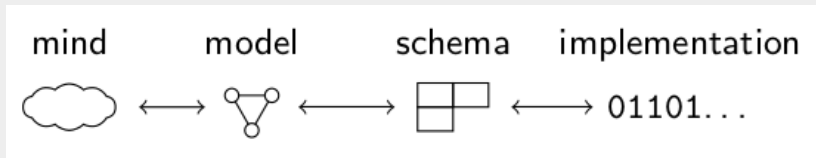
# Schnittmengen zwischen Metadaten-Strukturen



# Metadateninteroperabilität durch Mappings



# Was kann gemappt werden?



Ebene	Form des Mappings
Mind	Begriffsklärung
Model	Ontology oder Terminology Mapping
Schema	Ontology oder Schema Mapping
Implementation	Schema Mapping oder Konvertierung

# Beispiele

- ▶ Model Mapping
  - ▶ *Wir müssen reden...*
- ▶ Terminology Mapping
  - ▶ RVK AN 65800  $\approx$  MeSH D007998
- ▶ Schema/Ontology Mapping
  - ▶ Datenfelder ( $\Rightarrow$  JSON/XML-Beispiel)
- ▶ Konvertierung
  - ▶ Zeichenkodierungen

# Sprachgebrauch uneinheitlich!

## Strukturen

- ▶ Model
- ▶ Schema
- ▶ Format
- ▶ Ontologie
- ▶ Terminologie
- ▶ ...

## aufeinander abbilden

- ▶ Mapping
- ▶ Matching
- ▶ Konkordanz
- ▶ Crosswalk
- ▶ Konvertierung
- ▶ ...

# Mapping-Praxis

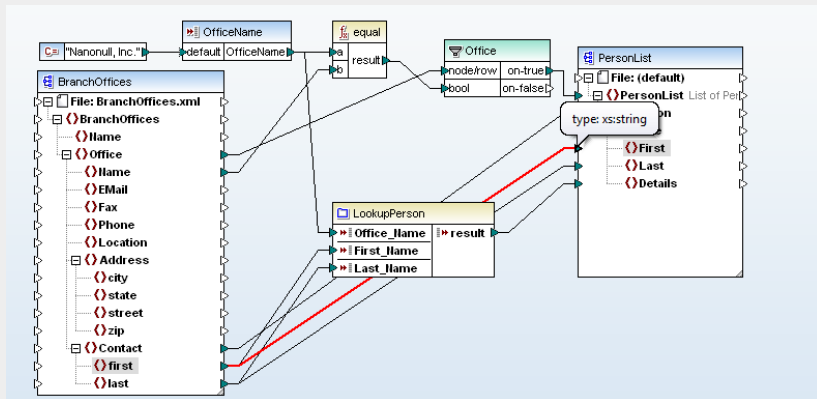
# Wie kommen wir zu Mappings?

- ▶ Intellektuelle Vergleiche
  - ▶ Konkordanztabellen erstellen
  - ▶ Konvertierung programmieren
- ▶ Automatische Verfahren und Heuristiken
  - ▶ Label-basiert (“author”=“authors”)
  - ▶ Strukturbasiert (z.B. Datentypen)
  - ▶ Instanzbasiert

# Mapping-Verfahren und -Tools

- ▶ Viele Forschungssysteme
  - ▶ <http://oaei.ontologymatching.org/> (OAEI)
  - ▶ <http://ontologymatching.org/>
- ▶ Bestandteil einiger ETL- und BD-Tools
- ▶ Einige erfolgreiche kommerzielle Systeme
  - ▶ Eher spezialisiert
  - ▶ Mehr Konvertierung statt Mapping statt Matching

# Beispiel: Altanova MapForce



<https://www.altova.com/de/mapforce.html>

# Beispiel: Projekt coli-conc

<http://coli-conc.gbv.de/>

- ▶ Sammlung von Terminologien und Konkordanzen  
⇒ zusammen mit [BARTOC.org](http://bartoc.org)
- ▶ Bereitstellung von Konkordanzen
- ▶ Tool zur Erstellung und Bewertung von Konkordanzen

*DDC, RVK, GND, BK...*



# Heuristik: Instanz-basiertes Matching

- ▶ Idee
  - ▶ Gleiche oder ähnliche Werte(verteilungen)
- ▶ Annahmen
  - ▶ Beide Schemas müssen mit Werten gefüllt sein
  - ▶ Beide Datenbasen müssen Duplikate enthalten
  - ▶ Duplikate müssen gleiche Attribute enthalten
- ▶ Beispiel: coli-conc Mapping-Algorithmus
  - ▶ Titeldaten im GBV-Katalog hat Notationen mehrerer KOS
  - ▶ Kookkurenz  $\Rightarrow$  Semantische Ähnlichkeit

# Zusammenfassung

# Metadaten

- ▶ Beschreibende Daten mit einheitlicher Struktur
- ▶ auf verschiedene Ebenen (Modell, Schema, Format...)
- ▶ mal explizit (Standard, Regeln...) mal implizit

# Interoperabilität

- ▶ Unterschiedliche Systeme  
⇒ Unterschiedliche (Metadaten)strukturen
- ▶ Schnittmenge von Gemeinsamkeiten
- ▶ Mapping auf einer passenden Ebene
- ▶ Heuristiken zum Finden von Mappings

# Weitere Quellen

- ▶ Quellen dieses Vortrags (CC BY-SA) <https://github.com/jakobib/metadatainteroperability2017>
- ▶ Vorlesungsfolien zur semantischen Datenintegration <https://github.com/hshdb/MWM-317-02>
- ▶ Jakob Voss: Describing Data Patterns (2013) <http://aboutdata.org>