



## **Linked Open Projects**

-

### **Nachnutzung von Projektergebnissen im Semantic Web**

Magnus Pfeffer (Dipl.-Inf., M.A. LIS)  
[magnus.pfeffer@bib.uni-mannheim.de](mailto:magnus.pfeffer@bib.uni-mannheim.de)

Kai Eckert (Dipl.-Inf., Dipl.-Wirt.-Inf.)  
[kai.eckert@bib.uni-mannheim.de](mailto:kai.eckert@bib.uni-mannheim.de)



## Überblick

- Projektergebnisse UB Mannheim
- Daten und Formate
- Linked Data



## Projekt: Automatische Klassifikation nach RVK

- Start 2005
- Datenbasis
  - SWB und HeBIS Katalogdaten
  - Umfang: 22,6 Mio Einträge
- Ergebnisse
  - >1 RVK-Notation für jeden Eintrag
  - Gewichtung der Notation(en)



## Konkordanz HeBIS-SWD auf Titelebene

- Start 2010
- Spezialfall des RVK-Projekts
- Datenbasis
  - SWB und HeBIS Katalogdaten
  - Umfang: 22,6 Mio Einträge
- Ergebnisse
  - Aussagen der Form  $PPN_a$  „sehr ähnlich zu“  $PPN_b$
  - Übernahme der RVK-Notation(en) von  $PPN_a$  für  $PPN_b$



## Konkordanz RVK-SWD

- Start 2009
- Datenbasis
  - SWB Katalogdaten
  - Umfang: 13,3 Mio Einträge
- Ergebnis
  - Korrelation SWD-ID zu RVK-Notation



## Nachnutzung der Ergebnisse: Anforderungen

- Datenformat
  - Einheitliche Syntax
  - Maschinenlesbar
- Ressourcen
  - Verknüpfung mit den Ursprungsdaten
- Vokabular
  - Nachvollziehbare Semantik
  - Projektübergreifend
- Zugang
  - Einheitlicher Zugriff
  - Möglichkeit der Selektion



## Nachnutzung der Ergebnisse:

## Bislang

### ■ Datenformat

- Einheitliche Syntax
- Maschinenlesbar

CSV, XML

### ■ Ressourcen

- Verknüpfung mit den Ursprungsdaten

Identnummer

### ■ Vokabular

- Nachvollziehbare Semantik
- Projektübergreifend

Spaltenüberschrift  
Feldbeschreibung

### ■ Zugang

- Einheitlicher Zugriff
- Möglichkeit der Selektion

Archive im Web  
Archivebene



## Nachnutzung der Ergebnisse:

## Bislang

- |   |  |                                |
|---|--|--------------------------------|
| <ul style="list-style-type: none"> <li>■ Datenformat                             <ul style="list-style-type: none"> <li>■ Einheitliche Syntax</li> <li>■ Maschinenlesbar</li> </ul> </li> </ul>         | <p>CSV, XML</p>                                | <p>← OK</p>                    |
| <ul style="list-style-type: none"> <li>■ Ressourcen                             <ul style="list-style-type: none"> <li>■ Verknüpfung mit den Ursprungsdaten</li> </ul> </li> </ul>                      | <p>Identnummer</p>                             | <p>← Nicht eindeutig</p>       |
| <ul style="list-style-type: none"> <li>■ Vokabular                             <ul style="list-style-type: none"> <li>■ Nachvollziehbare Semantik</li> <li>■ Projektübergreifend</li> </ul> </li> </ul> | <p>Spaltenüberschrift<br/>Feldbeschreibung</p> | <p>← Uneinheitlich</p>         |
| <ul style="list-style-type: none"> <li>■ Zugang                             <ul style="list-style-type: none"> <li>■ Einheitlicher Zugriff</li> <li>■ Möglichkeit der Selektion</li> </ul> </li> </ul>  | <p>Archive im Web<br/>Archivebene</p>          | <p>← Unhandlich<br/>← Grob</p> |



## Nachnutzung der Ergebnisse:

## Bislang

## Linked Data

### ■ Datenformat

- Einheitliche Syntax
- Maschinenlesbar

CSV, XML

RDF/XML

### ■ Ressourcen

- Verknüpfung mit den Ursprungsdaten

Identnummer

URI

### ■ Vokabular

- Nachvollziehbare Semantik
- Projektübergreifend

Spaltenüberschrift  
Feldbeschreibung

externes Vokabular

### ■ Zugang

- Einheitlicher Zugriff
- Möglichkeit der Selektion

Archive im Web  
Archivebene

URI per HTTP  
Sparql Abfrage



## Datenformat: RDF

- Tripel, die Aussagen (Statements) darstellen:
  - **Subjekt**, **Prädikat**, **Objekt**
  - Beispiel: **Deutschland** **hat als Hauptstadt** **Berlin**.
- Subjekt und das Prädikat als URI immer eindeutig bezeichnet
- Objekt als URIs oder Datenwert (Literal)
- Beispiele:
  - **ex:locations/germany** **ex:vocab/capital** **ex:locations/berlin**
  - **ex:locations/germany** **ex:vocab/name** „**Deutschland**“
  - **ex:locations/berlin** **ex:vocab/name** „**Berlin**“



## Ressourcen: URIs

- Eindeutige und dauerhafte Beschreibung einer Resource
  - <http://dblp.uni-trier.de/rec/pers/e/Eckert:Kai/>
- Eine Ressource durch verschiedene URIs beschreibbar
  - <http://dblp.uni-trier.de/rec/pers/e/Eckert:Kai/>
  - [http://ki.informatik.uni-mannheim.de/people/kai\\_eckert.html](http://ki.informatik.uni-mannheim.de/people/kai_eckert.html)
  - <http://de.linkedin.com/in/kaieckert>
  - → **Keine Eineindeutigkeit!**
- Äquivalenz von URIs als RDF Statement
  - <http://de.linkedin.com/in/kaieckert> `owl:sameAs` <http://dblp.uni-trier.de/rec/pers/e/Eckert:Kai/> ↵
  - kann jederzeit und von jedem definiert werden
  - Auch Prädikate (Vokabular) gleichsetzbar
  - Transparente Verknüpfung von Datenbeständen → Linked Data





## Exemplarische Umsetzung

- Bibliografische Ressourcen des SWB
  - URIs für jede verwendete PPN
  - <http://data.bib.uni-mannheim.de/data/swb/00446737X>
  - URIs für die RVK-Notationen
  - <http://data.bib.uni-mannheim.de/data/rvk/MR+4000>
  - URIs für Schlagworte bereits als Linked Data
  - <http://d-nb.info/gnd/4029577-1>
  
- Projektbezogene Vokabularien
  - Dublin Core / ISBD für Metadaten
  - SKOS für SWD und RVK
  - Eigenes Prädikat für Aussage „sehr ähnlich zu“



## Public Beta Website

- Website / Dokumentation
  - <http://data.bib.uni-mannheim.de/>
- SparQL Endpoint
  - <http://data.bib.uni-mannheim.de/sparql/>
- Exploration Interface
  - <http://data.bib.uni-mannheim.de/isparql/>



Bildnachweis: Ramzi Hashisho, <http://www.sxc.hu/profile/bigevil600>

## Fragen / Diskussion



23.07.2010

Linked Open Projects