

Vorträge zur FIS Bildung-Frühjahrstagung vom 22.05.2012 zum Thema

„Linked Open Data“

1. Adrian Pohl (hbz):
Linked Open Data & Bibliotheken - Warum? Was? Wie?
2. Benjamin Zapilko (gesis):
Anwendungsszenarien und Potenziale von Linked Open Data im wissenschaftlichen Forschungsprozess
3. Adrian Pohl (hbz):
Zum Aufbau von Linked-Open-Data-Services im hbz



hbz

Wissen. Information. Innovation.

Linked Open Data & Bibliotheken - Warum? Was? Wie?

FIS Fachtagung, Frankfurt/Main

22. Mai 2012

Adrian Pohl <pohl@hbz-nrw.de>

Re-use data

API services

→ Linked Open Data

In brief

Tech details

Datasets

Credits

Case studies

Hackathons

ThoughtLab

Support for Open Data

Europeana Linked Open Data (LOD)

[Linked Open Data Pilot press release, 17 February 2012](#)



page.

data.europeana.eu currently contains **open** metadata on 2.4 million texts, images, videos and sounds gathered by Europeana. These objects come from data providers who have reacted early and positively to Europeana's initiative of promoting more open data and new **data exchange agreements**. These collections come from 8 direct Europeana providers encompassing over 200 cultural institutions from 15 countries. They cover a great variety of heritage objects, such as this **slovenian version of "O Sole Mio"** from the National Library of Slovenia, or Neil Robson's **memories of the herring business** from the Tyne and Wear Archives & Museums. For more information, see our **datasets**

The data is served according to the Linked Data principles—see the **technical details** and **this animation** for background on our motivation. It is represented in the Europeana Data Model (EDM) and the described resources are **addressable and dereferenceable** by their URIs; for instance, <http://data.europeana.eu/item/92056/BD9D5C6C6B02248F187238E9D7CC09EAF17BEA59> leads either to an HTML page on the Europeana portal for the object it identifies or to raw, machine-processable data on this object. The project has been described in a **paper** presented at the Dublin Core 2011 conference.

Disclaimer: data.europeana.eu is currently in a pilot stage, and can thus be changed at any moment! Your feedback is more than welcome, and may lead to updates in the prototype service.

search



Login or register

Contact us

The Europeana Office
+31 70 3140991
europeana-
lod@googlegroups.com

Dokumentation des Linked Data Services der DNB

📄 14 Hinzugefügt von [IT-Administrator](#), zuletzt bearbeitet von [Hannemann, Jan](#) am 2012-05-02 ([Änderung anzeigen](#))

Deutsche Dokumentation

[Deutsche Dokumentation](#)[↗] (V. 4.1, Stand: 29. Februar 2012. Wird derzeit überarbeitet.) und [FAQ](#)[↗]

Lizenz

Für die Gemeinsame Normdatei (GND) und die Titeldaten im Linked-Data-Service gilt ab sofort die Nutzungslizenz "[Creative Commons Zero](#)[↗]". Die DDC Deutsch steht unter der Lizenz [Creative Commons BY-NC-ND](#)[↗]. Dies betrifft Ressourcen, deren URIs mit <http://d-nb.info/ddc>[↗] beginnen.

Datensets

Deutsche Nationalbibliografie (DNB)

Beispieldatensätze und Informationen zu FTP-Downloads folgen in Kürze.

Gemeinsame Normdatei (GND)

Die Gemeinsame Normdatei ist in der [Visualisierung der Linked-Data-Cloud](#)[↗] (September 2011) unter "GND" (rechts oben) zu finden.

Beispieldatensätze und Informationen zu FTP-Downloads der mit der GND-Ontologie modellierten Normdaten folgen in Kürze.

FTP-Downloads (**RDF-Umsetzung der alten Normdatenformate**):

- Beispiel (Ausschnitt aus der Datenbasis, ca. 500 KB, Stand 04/2010): <http://tinyurl.com/39zy2tw>[↗]
- Personen (PND), Stand 04/2011: [Teil 1](#)[↗] (RDF/XML, ca. 265 MB ZIP File), [Teil 2](#)[↗] (RDF/XML, ca. 5 MB ZIP File)
- Körperschaften (GKD), Stand 04/2011: [Teil 1](#)[↗] (RDF/XML, ca. 95 MB ZIP File), [Teil 2](#)[↗] (RDF/XML, ca. 16 MB ZIP File)
- Sachschlagworte (SWD), Stand 07/2011: [Alle Daten](#)[↗] (RDF/XML, ca. 14 MB ZIP File)

Es besteht die Möglichkeit, auf dieser Seite Fragen und Anregungen zu hinterlassen (siehe "Kommentar hinzufügen" Link am Seitenende). Beantwortete Fragen etc. werden in unregelmäßigen Abständen gelöscht.

English Documentation

[English documentation](#)[↗] (V. 4.1, Status: February 29th 2012. Currently being revised.) and [FAQ](#)[↗]

Licence

Both the authority file of the German-speaking countries - the *Gemeinsame Normdatei (GND)* - and the bibliographic data available through the German National Library's Linked Data Services are released under the [Creative Commons Zero](#)[↗] (CCO) license.

The DDC German is released under the [Lizenz Creative Commons BY-NC-ND](#)[↗] license. This applies to all resources with URIs beginning with "http://d-nb.info/ddc".

Datasets

Deutsche Nationalbibliografie (DNB)

Examples and further information about FTP-downloads will come soon.

Gemeinsame Normdatei (GND)

METADATA SERVICES

DATA SERVICES

1. Linked Open BNB

Background

The British Library is developing a version of the **British National Bibliography** which it is making available as Linked Open Data via a **Talis platform**.

The initial offering includes published books with future releases extending coverage to include serial publications, multipart works, integrating resources (*e.g. loose leaf publications*), kits and forthcoming publications.

BIBLIOGRAPHIC FRAMEWORK TRANSITION INITIATIVE



- [Initiative Home](#)
- [News](#)

More Resources

- > [MARC 21](#)
- > [Library Standards](#)
- > [Working Group on the Future of Bibliographic Control](#)

News and Announcements



A Bibliographic Framework for the Digital Age (October 31, 2011)

Approach

The new bibliographic framework project will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model. The protocols and ideas behind Linked Data are natural exchange mechanisms for the Web that have found substantial resonance even beyond the cultural heritage sector. Likewise, it is expected that the use of RDF and other W3C (World Wide Web Consortium) developments will enable the integration of library data and other cultural heritage data on the Web for more expansive user access to information.

members wrote:

"Recognizing that Z39.2/MARC are no longer fit for the purpose, work with the library and other interested communities to specify and implement a carrier for bibliographic information that is capable of representing the full range of data of interest to libraries, and of facilitating the exchange of such data both within the library community and with related communities." [3]

This quote stems from the report of the Resource Description and Access (RDA)

Was steckt dahinter?

Agenda

1. Warum Linked Open Data?
2. Linked Data
3. Open Data
4. Fazit

1. Warum Linked Open Data?

Motivation

- Nutzung internationaler domänenübergreifender Web-Standards führt zu
 - **Auffindbarkeit** & Verlinkbarkeit,
 - maximaler **Interoperabilität**,
 - einfacher **Mehrfachnutzung**,
 - größerer **Unabhängigkeit von Herstellern**.

Return of Investment

- **Synergieeffekte** durch Vereinheitlichung der Technologie:
 - **Quantitativ**: eingesparte Ressourcen
 - **Qualitativ**: verbesserte Dienstleistungen

2. Linked Data

Das Web: URL, HTTP, HTML

http://example.org/page1

Dokument

text text text text text text
text text text text text text
text text text text text text
text text text [link](#) text text
text text text text text text

http://example.org/page2

Dokument

text text text text text text
text text text text text text
text text text text [link](#) text
text text text text text text
text text text text text text
text [link](#) text text text text
text text text text text text
text text text [link](#) text text
text text text text text text
text text text text text text

Das Web: URL, HTTP, HTML

<http://example.org/page1> $\xrightarrow{\text{link}}$ <http://example.org/page2>

Linked Data = das Web + RDF

- Linked Data ist eine Erweiterung des bestehenden Webs
- HTTP-URIs als Identifikatoren für beliebige Dinge
- Resource Description Framework (RDF) als Beschreibungsmodell
- Verlinkung von Dingen mittels URIs

URIs

http://de.wikipedia.org/wiki/Uniform_Resource_Identifier

<ftp://ftp.is.co.za/rfc/rfc3986.txt>

<file:///home/fo/doc/elag12/slides.odp>

<urn:isbn:978-1608454303>

<http://192.168.0.124>

<mailto:pohl@hbz-nrw.de>

HTTP-URIs für Dinge



HTTP-URIs für Dinge



<<http://viaf.org/viaf/85312226>>

Beschreibung in RDF

@prefix foaf: <http://xmlns.com/foaf/0.1/> .



<http://viaf.org/viaf/85312226> foaf:name "Tim Berners-Lee" .

Beschreibung in RDF

@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@prefix dc: <http://purl.org/dc/terms/> .



<http://viaf.org/viaf/85312226> foaf:name "Tim Berners-Lee" .

<http://lobid.org/resource/HT011263360> dc:title "Weaving the Web" .



Verlinkung von Dingen

@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@prefix dc: <http://purl.org/dc/terms/> .



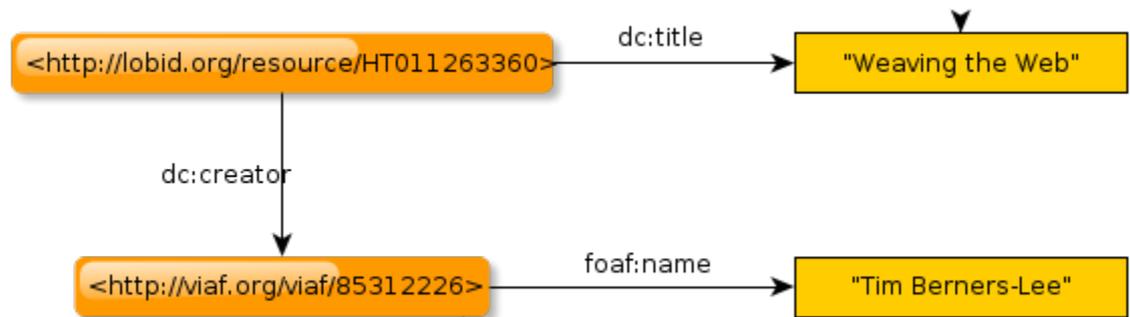
<http://viaf.org/viaf/85312226> foaf:name "Tim Berners-Lee" .

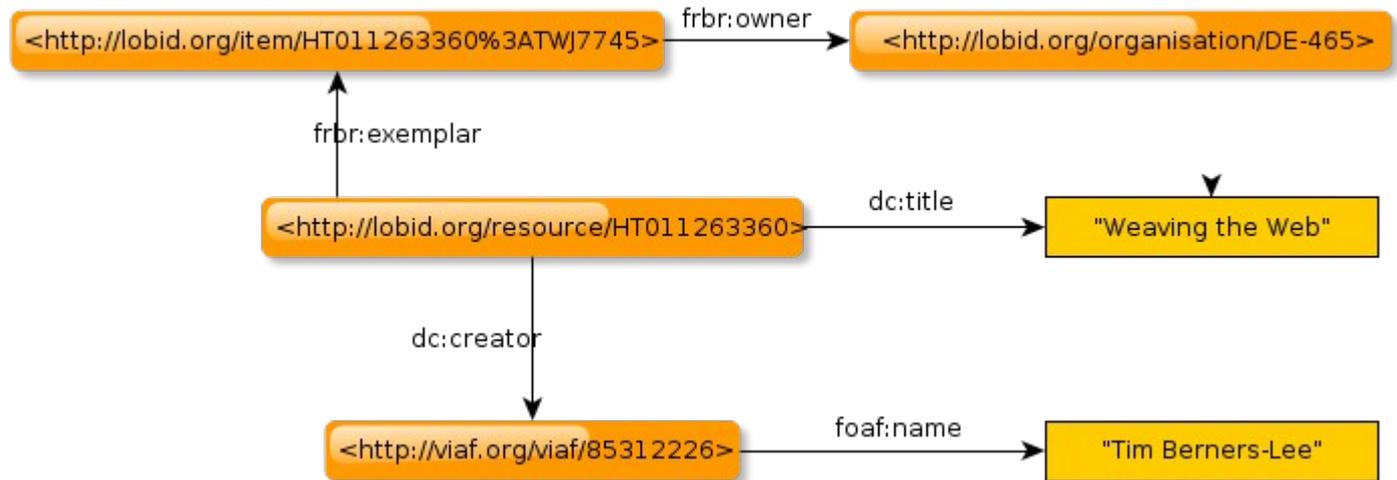


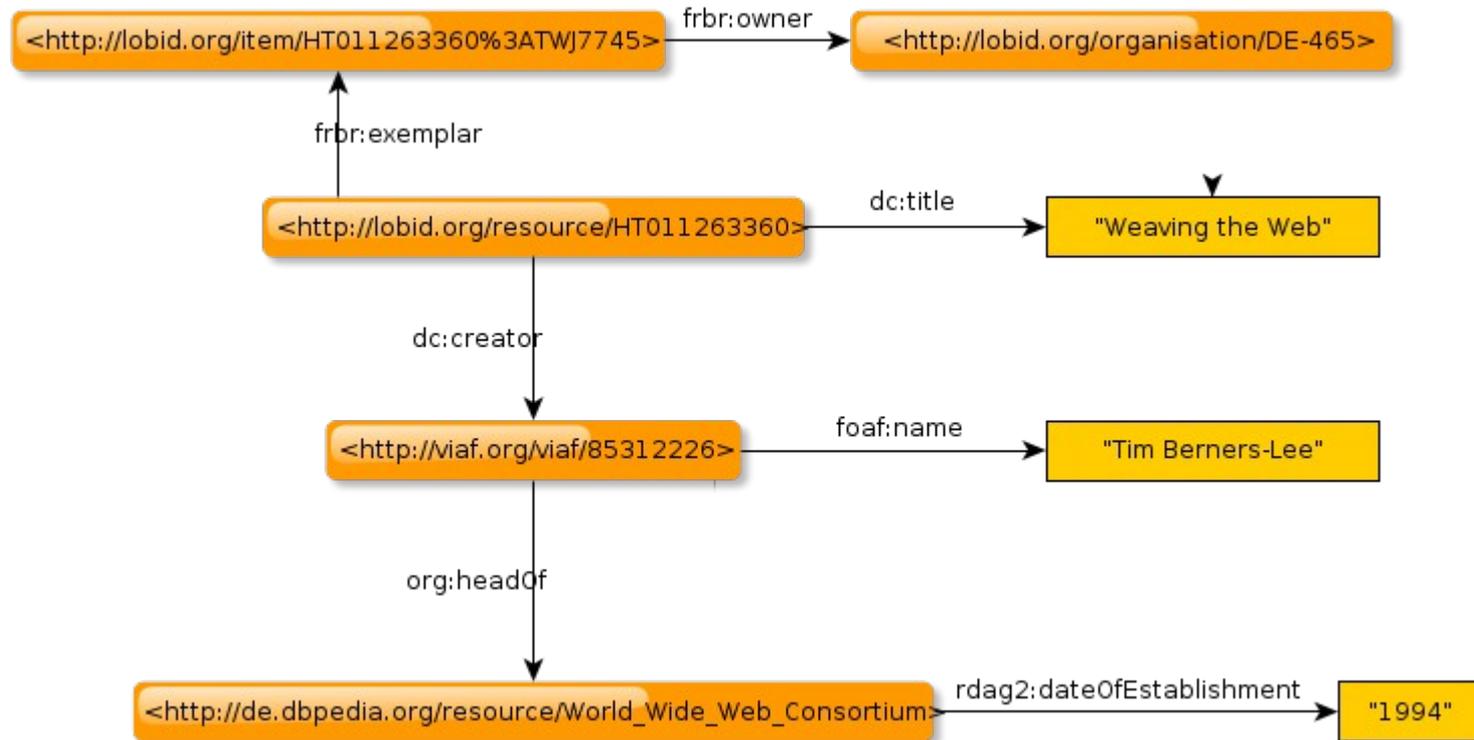
dc:creator

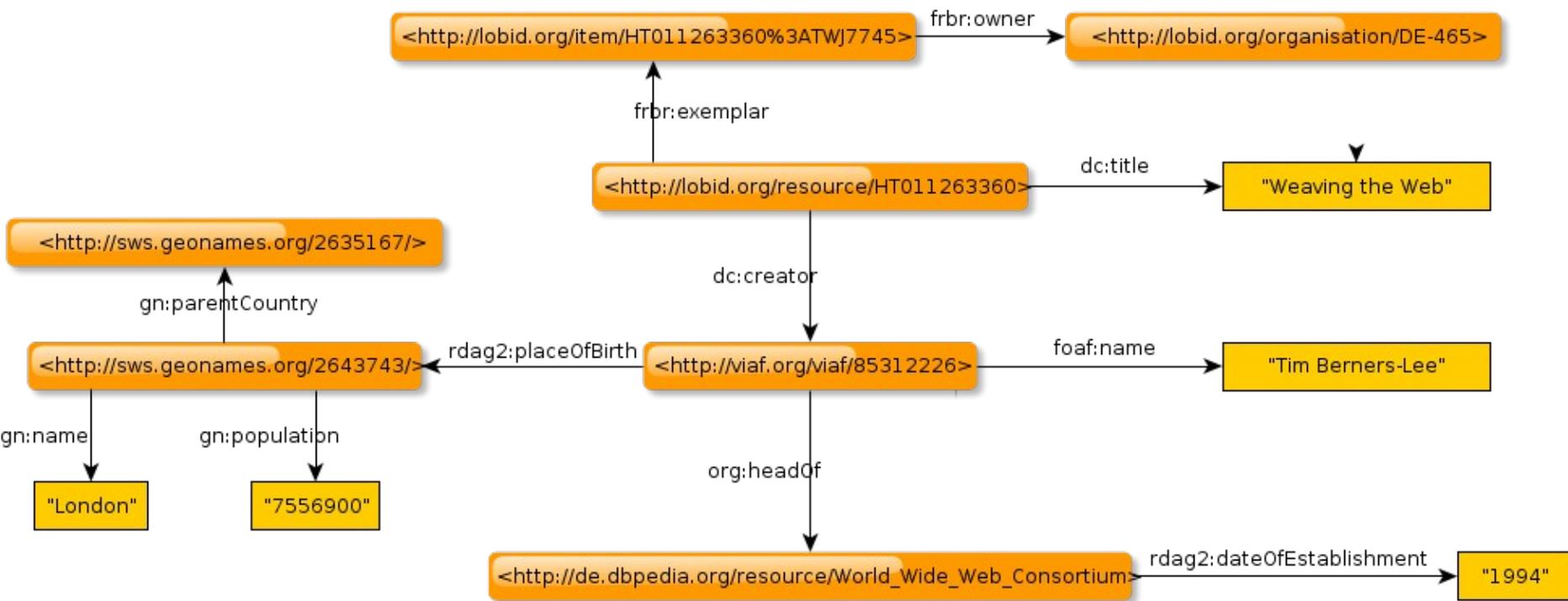
<http://lobid.org/resource/HT011263360> dc:title "Weaving the Web" .











Ontologien und Vokabulare

- Definition von Klassen und Eigenschaften und ihrer Verknüpfungen
- RDF-Schema (RDFS): einfacher Ansatz
- Web Ontology Language (OWL): größere Expressivität

Triple Stores & SPARQL

- RDF-Daten aus verschiedenen Quellen können in Triple Stores abgelegt werden
- Eine Abfrage von Daten in Triple Stores geschieht mit SPARQL

```
@prefix viaf: <http://viaf.org/viaf/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

viaf:85312226 foaf:name „Tim Berners-Lee“ .
```

```
PREFIX viaf: <http://viaf.org/viaf/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT * WHERE {
  viaf:85312226 foaf:name ?name .
}
```

name

„Tim Berners-Lee“

3. Open Data: Rechtliche Standards für ein Web of Data

„Open“

"A piece of knowledge is open if you are free to use, reuse, and redistribute it."

<http://www.opendefinition.org/okd>

Warum?

- Open Access: auch für Daten
- Rechtliche Kompatibilität verteilt vorliegender Daten

Open Data ist eine Frage von...

- Zugang
- Lizenzen
- Standards

Open Data ist eine Frage von...

- **Zugang**
- Lizenzen
- Standards

Zugang

- ...zu den gesamten Daten (API reicht bei großen Datenmengen nicht aus)
- nicht höher als Reproduktionskosten, vorzugsweise gebührenfreier Download im Internet
- Technisch: XMPP > OAI-PMH > http > ftp > scp > Post

Open Data ist eine Frage von...

- Zugang
- **Lizenzen**
- Standards

Geeignete Lizenzen

- Keine Non-Commercial Lizenzen
- Möglichkeiten:
 - Attribution (ODC-BY)
 - Attribution-Share-Alike (ODbL)
 - Public-Domain (CC0, PDDL)

<http://www.opendefinition.org/licenses/#Data>

Open Data ist eine Frage von...

- Zugang
- Lizenzen
- **Formaten**

Formate

→ nicht-proprietär

→ Strukturiertheit &
Maschinenlesbarkeit zählen!

→ Beispiele:

rdf > csv > ods > xls > PDF > docx >
Hardcopy

Daten
vs.
Datenbanken

Datenbank

“a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.”

From: European Database Directive

„Daten“

- Inhalt einer Datenbank → kann alles sein
- Fakten → sind nicht schutzfähig, nur als Sammlung

Unterschiedliche Rechtsgrundlagen

- Der rechtliche Status einer Datenbank und ihres Inhalts kann unterschiedlich sein
- Beispiel: Public-Domain-Inhalte und geschützte Sammlung

4. Fazit

Linked Data ist eine Frage von...

- Web-Standards...
 - für Identifikatoren (HTTP-URIs),
 - für Daten (RDF),
 - für Ontologien (RDFS, OWL),
 - für Abfragen (SPARQL).

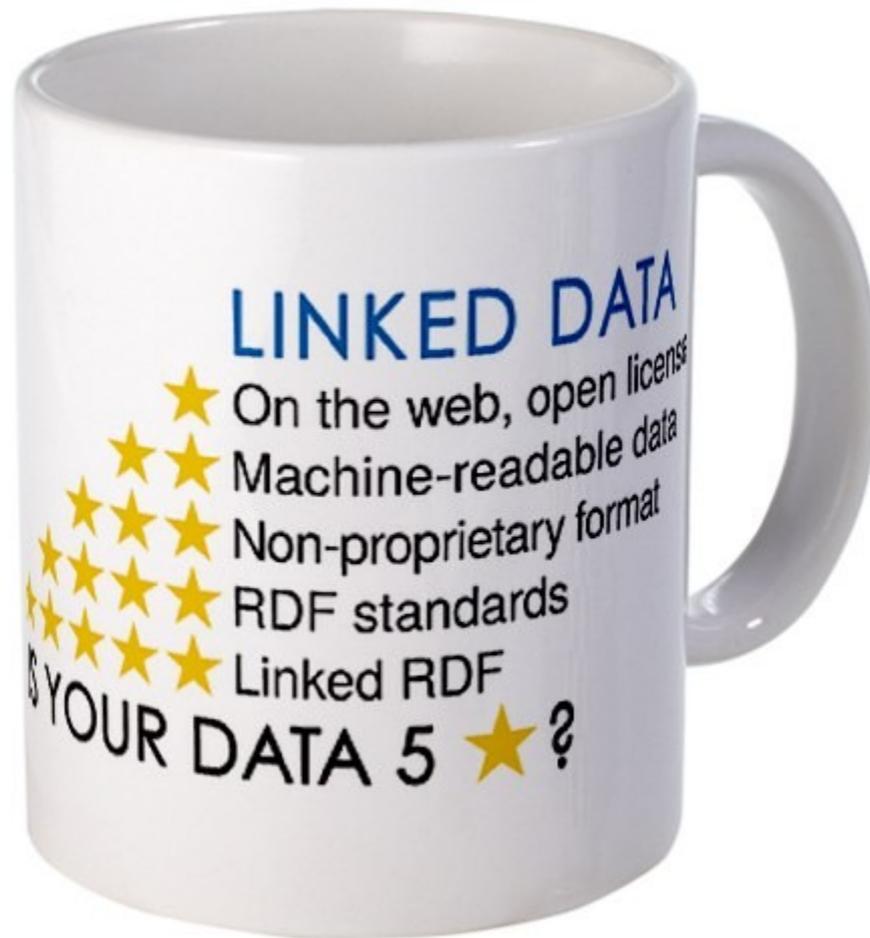
Open Data ist eine Frage von...

- Zugang
- Lizenzen
- Formaten



Linked Open Data





<http://www.w3.org/DesignIssues/LinkedData.html> &
<http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/>

Danke.

Fragen?

Gerne jetzt oder auch später an
semweb@hbz-nrw.de

Lizenz



Diese Folien stehen unter folgender CC-Lizenz:

<http://creativecommons.org/licenses/by/3.0/de/>

Anwendungsszenarien und Potenziale von Linked Open Data im wissenschaftlichen Forschungsprozess

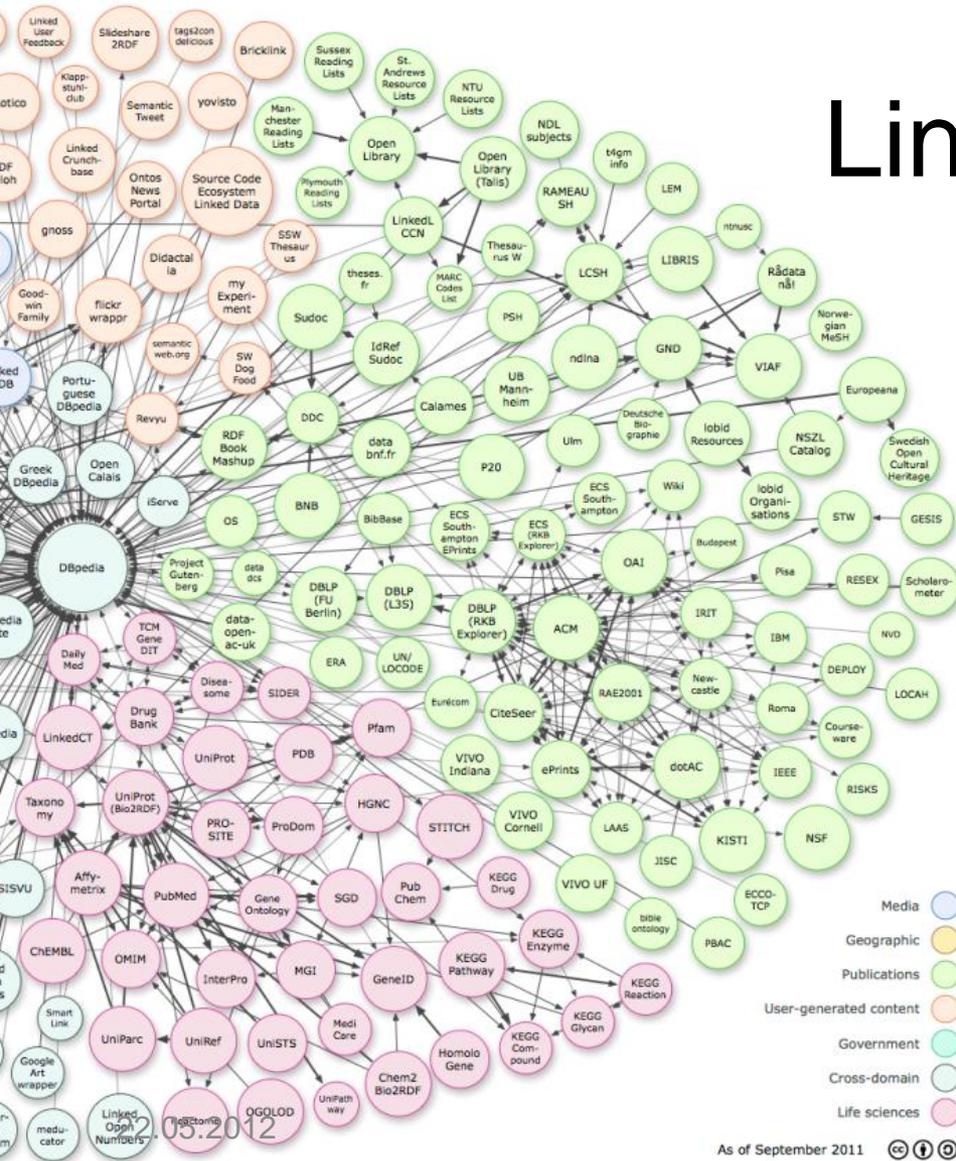
Benjamin Zapilko

FIS Fachtagung 2012, 22.05.2012

Agenda

- Linked Open Data ... und was nun?
- Wissenschaftler als LOD-Nutzer
- Anwendungsszenarien für LOD im Forschungsprozess
 - Forschungsdaten und Informationen recherchieren
 - Forschungsdaten analysieren
- Zusammenfassung

Linked Open Data...



- Immer mehr Daten in der LOD Cloud
- Semantische Technologien immer ausgereifter
- Erste interessante Anwendungen sind da
- „Jeder“ kann seine Daten ins Web bringen

... und was nun?

- Wie gelangen Daten in die LOD Cloud? Wie kann man LOD nutzen?
 - Technologie-Expertise vorhanden
 - Projekte und Initiativen, die Datenanbietern Unterstützung anbieten
- Was kann man mit LOD machen?
 - Erste Anwendungen existieren, aber meist explorierender oder visualisierender Art
- Warum LOD? Für wen? Und wie?
 - abhängig von Daten, Domäne und Nutzern
 - Argumente wie Offenheit und Transparenz als Motivation oft nicht überzeugend genug und politisch oft schwierig



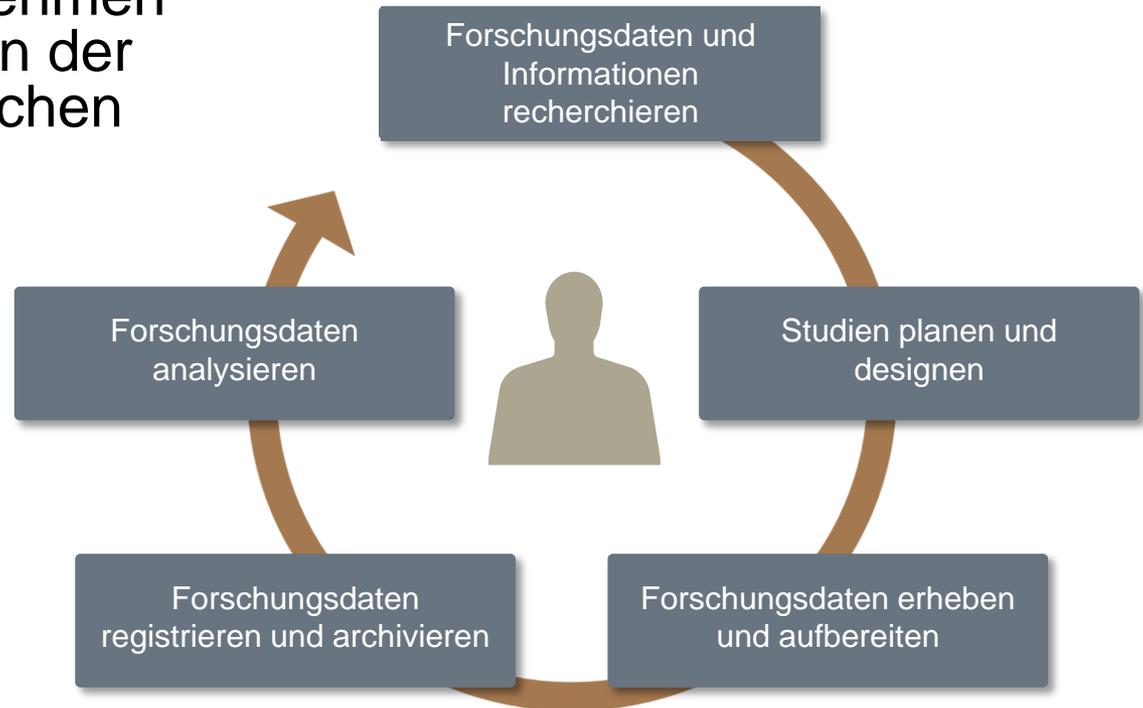
Anreize für Datenanbieter schaffen!
Mehrwerte für Nutzer schaffen!

Nutzergruppe: Wissenschaftler

- Betrachten des Forschungsprozesses eines Wissenschaftlers und darin enthaltene typische Arbeitsschritte
 - von Domäne zu Domäne unterschiedlich
 - verschiedene Schwerpunkte
 - typische Arbeitsabläufe
 - manche Aufgaben sehr zeitintensiv und wiederholen sich immer wieder
- Erkennen von Defiziten, Schwierigkeiten und Problemstellungen
 - Können Teile durch den Einsatz von LOD „aufgewertet“ oder vereinfacht werden?
- Identifizieren von möglichen Mehrwerten und Anreizen durch den Einsatz von LOD

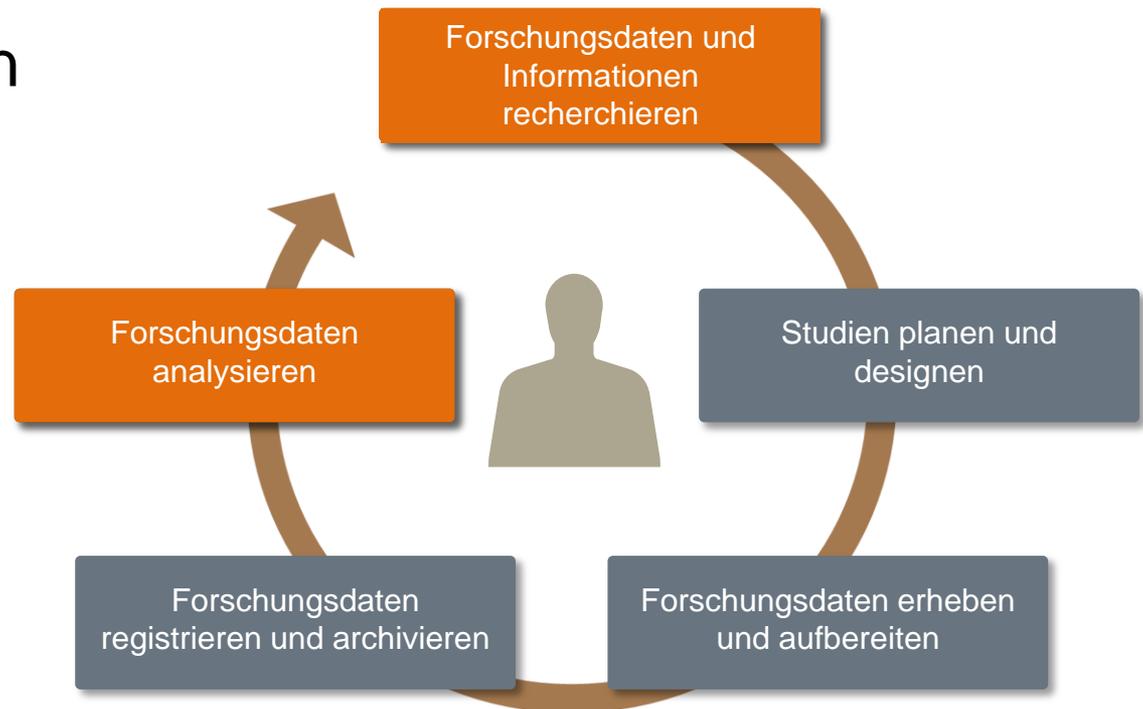
Forschung in den Sozialwissenschaften

- Forschungsdaten nehmen eine zentrale Rolle in der sozialwissenschaftlichen Forschung ein
- 5 Phasen, die innerhalb eines Forschungsprojekts durchlaufen werden



Fokus auf Sekundärforscher

- Forschungsdaten und Informationen recherchieren
- Forschungsdaten analysieren



Forschungsdaten und Informationen recherchieren

Wie ist die allgemeine Meinung zum Thema „Gewalt gegen Kinder in Familien“?

Wurden zu dieser Fragestellung bereits Umfragedaten ausgewertet?

Gibt es zu dieser Fragestellung weitere oder vergleichbare Forschungsdaten?

Welche Forschungsdaten und welche Publikationen helfen mir bei der Beantwortung dieser Fragestellung?

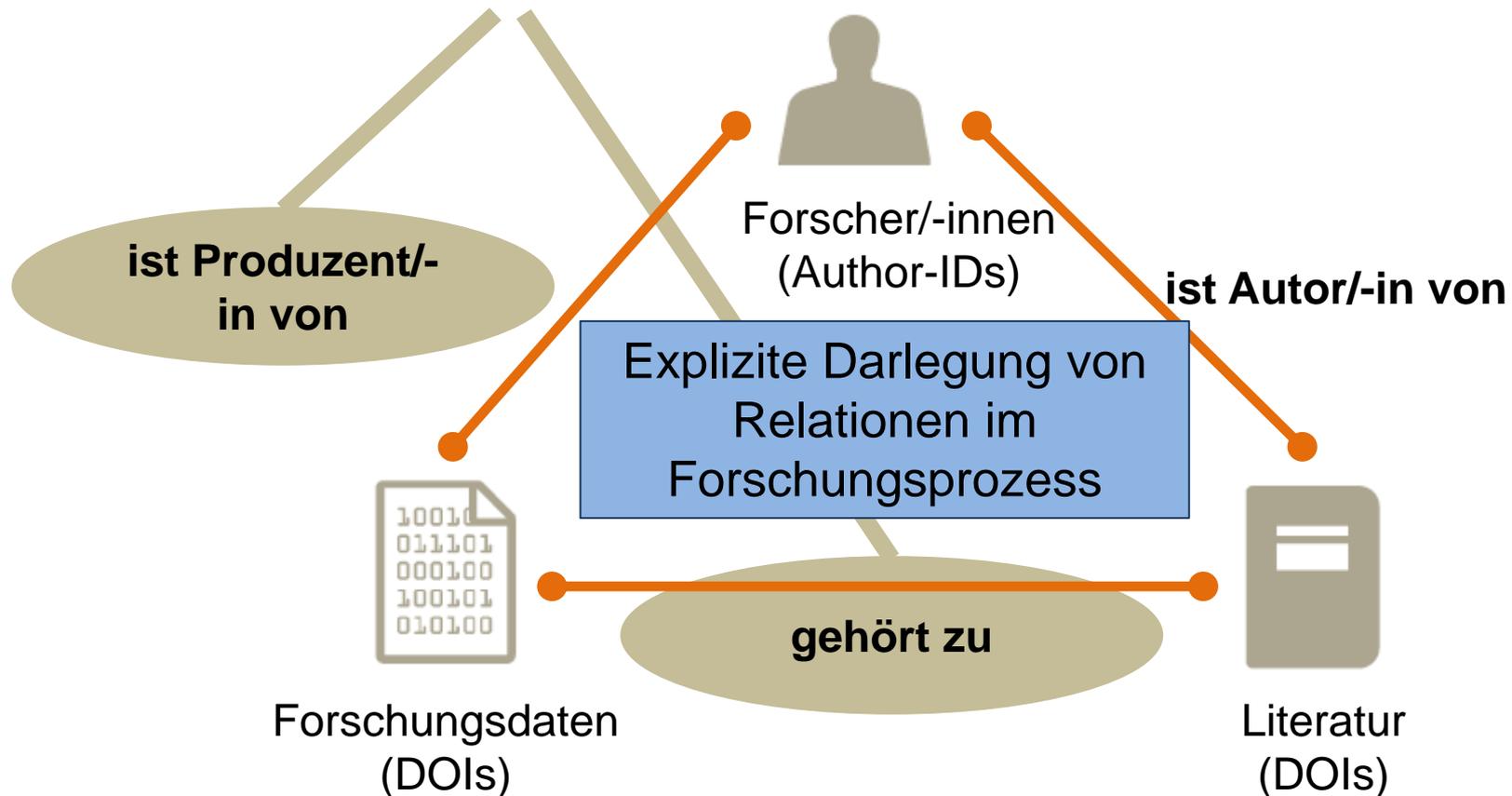
Mögliche Probleme und Hürden

- Recherche über heterogene und meist nicht miteinander verbundene Informationstypen
 - Verschiedene Metadatenstandards
 - Unterschiedlicher Erschließungsgrad
- Recherche gebunden an Institutionen, Datenbanken, Informationsportale etc. oder ganz offen (z.B. Google)
 - Forschungsdaten und Informationen werden dezentral gehalten, gepflegt und aufbereitet (u.a. historisch, rechtlich und disziplinar bedingt)
 - Im Web finden sich relevante Informationen, die von Fachdatenbanken nicht erfasst werden
 - In Fachdatenbanken finden sich relevante Informationen, die durch Websuche nicht gefunden werden



Integrierte Recherchen nur bis zu gewissem Grad möglich!
 Datenintegration und Datenmodellierung äußerst aufwändig!

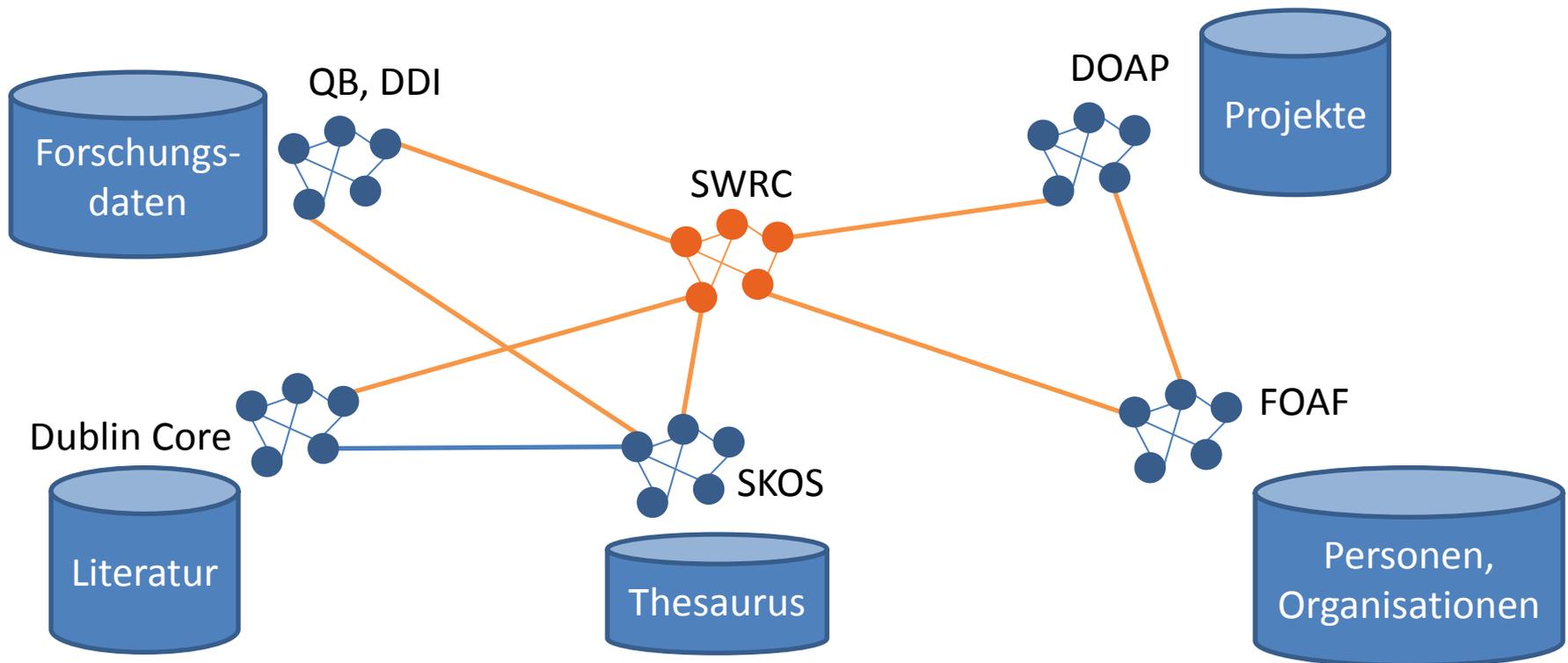
Fehlende Links herstellen



Forschungsprozess darlegen

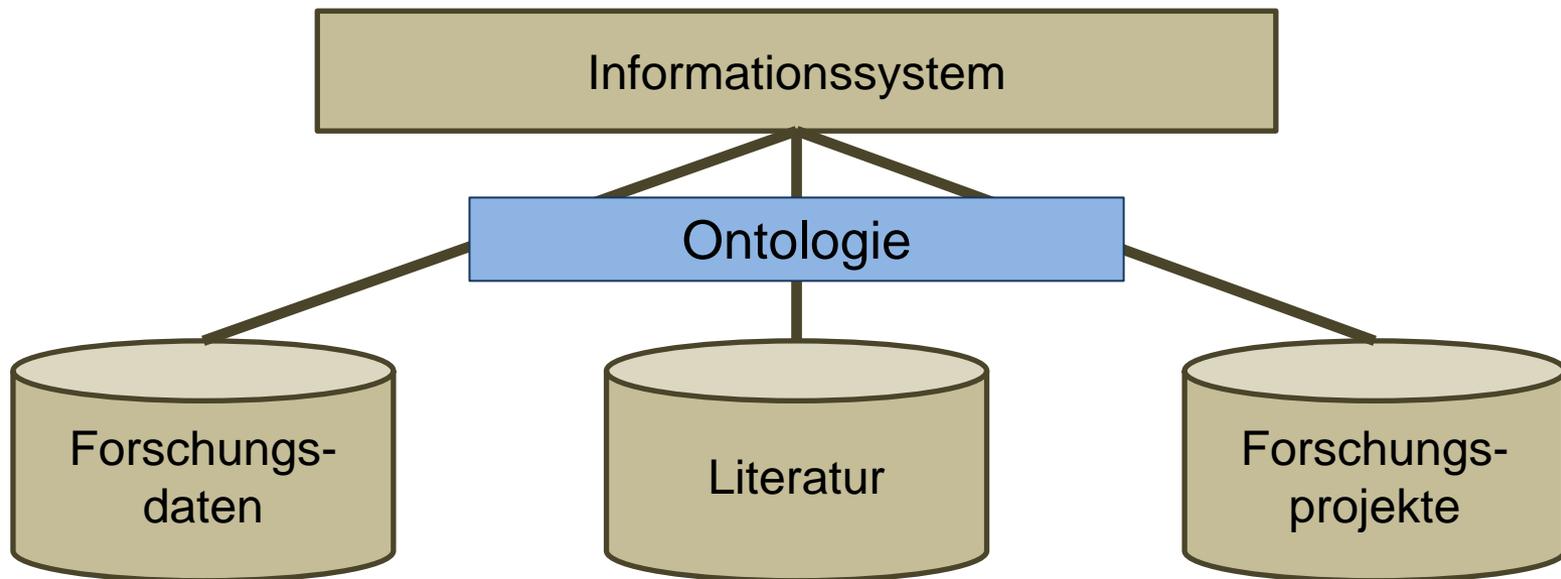
- Einheitliches Datenmodell über heterogene Informationstypen ist aufwändig
- Forschungsprozess zusammenführen durch den Einsatz von Ontologien
 - z.B. die SWRC Ontologie (Semantic Web for Research Communities)
 - über physikalische Grenzen hinweg
 - Datenbestände können unberührt bleiben
 - Gültige Metadatenstandards können beibehalten werden

Forschungsprozess darlegen



Einbettung in bestehende Infrastrukturen

- Ontologien als Modellierungsebene einfügen
 - „schnell“ implementierbar
 - bestehende Infrastrukturen bleiben bestehen



Neue Links herstellen

- DFG-Projekt „InFoLiS – Integration von Forschungsdaten und Literatur in den Sozialwissenschaften“
- gemeinsam mit UB und Uni Mannheim
- Kernziel: Automatisches Erkennen von Studienzitationen in Publikationen
- Output von Links im RDF
 - Diese Link-Information kann in Datenbestände eingebunden werden (z.B. in die Metadaten)

```

<daraDOI1> <isCitedBy> <ssoarURN1>.
<daraDOI2> <isCitedBy> <ssoarURN1>.
<daraDOI3> <isCitedBy> <ssoarURN2>.
<daraDOI4> <isCitedBy> <ssoarURN3>.
      ...
  
```

Neue Links herstellen

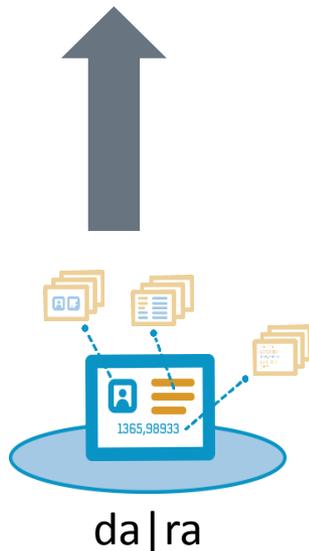
Veröffentlichung:
<ssoarURN1-Link>

Zitierte Studien:
<daraDOI1-Link>
<daraDOI2-Link>



Neue Links herstellen

Veröffentlichung:
<ssoarURN1-Link>



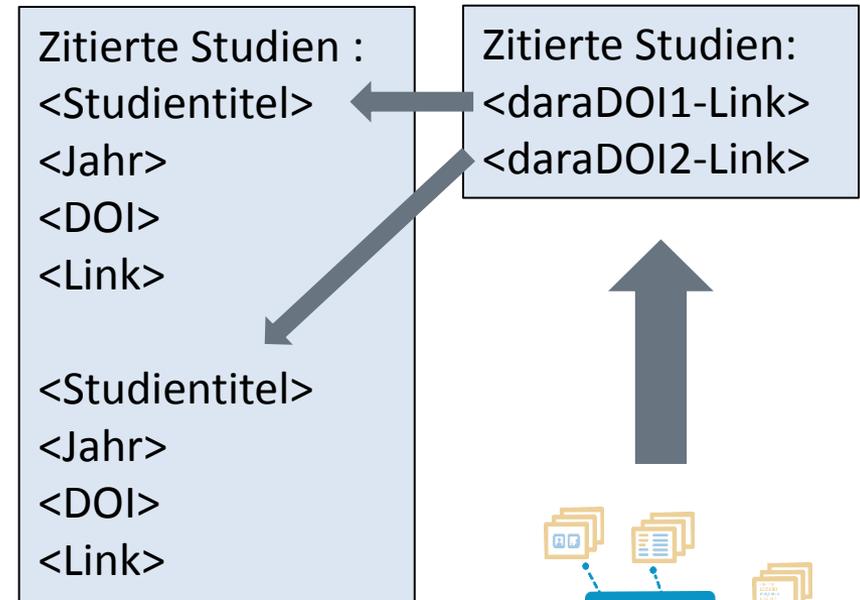
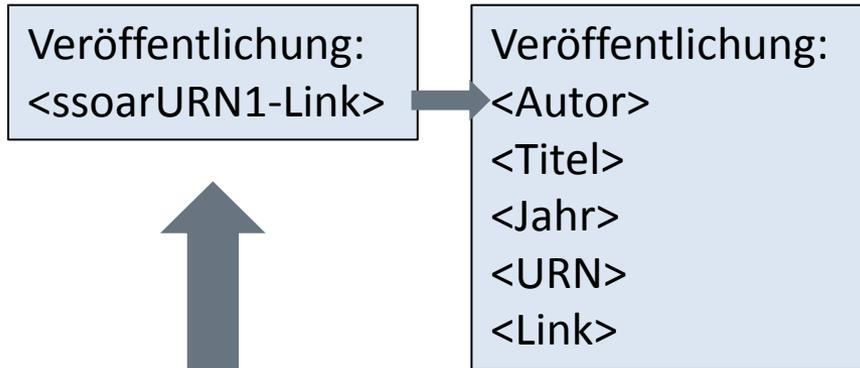
```

<bibo:AcademicArticle rdf:about="http://lod.gesis.org/ssoar/ID/0168-ssoar-123
<dcterms:isPartOf rdf:resource="http://lod.gesis.org/ssoar/">
<dcterms:issued>2009</dcterms:issued>
<dcterms:type>Zeitschriftenartikel</dcterms:type>
<dcterms:creator>Poustka, Fritz</dcterms:creator>
<dcterms:language>en</dcterms:language>
<rdfs:seeAlso>Link to full text: ../files/peer/Peer_125-2010-07-15.txt</rdf
<dcterms:title>Pilot evaluation of the Frankfurt Social Skills Training for
<dcterms:creator>Duketis, Eftichia</dcterms:creator>
<owl:sameAs rdf:resource="http://lod.gesis.org/ssoar/ID/0168-ssoar-123844"/
<bibo:issn>1435-165X</bibo:issn>
<dcterms:license>http://www.ssoar.info/en/home/how-to-deposit-documents/leg
<dcterms:creator>Schmötzer, Gabriele</dcterms:creator>
<foaf:homepage rdf:resource="http://www.ssoar.info/ssoar/View/?resid=12384"
<dcterms:identifier>http://nbn-resolving.de/urn:nbn:de:0168-ssoar-123844</d
<dcterms:creator>Birnkammer, Sabine</dcterms:creator>
<dcterms:extent>S. 327-335</dcterms:extent>
<dcterms:creator>Schlitt, Sabine</dcterms:creator>
<dcterms:source>European Child & Adolescent Psychiatry, Jg. 18, H. 6</d
<bibo:producer rdf:resource="http://lod.gesis.org/ssoar/producer/GESIS"/>
<dcterms:creator>Herbrecht, Evelyn</dcterms:creator>
<dcterms:creator>Bölte, Sven</dcterms:creator>
<dcterms:abstract xml:lang="en">The objective of this pilot study was to ev
</bibo:AcademicArticle>
    
```

Zitierte Studien:
<dataDOI1-Link>
<dataDOI2-Link>



Neue Links herstellen



Neue Links herstellen

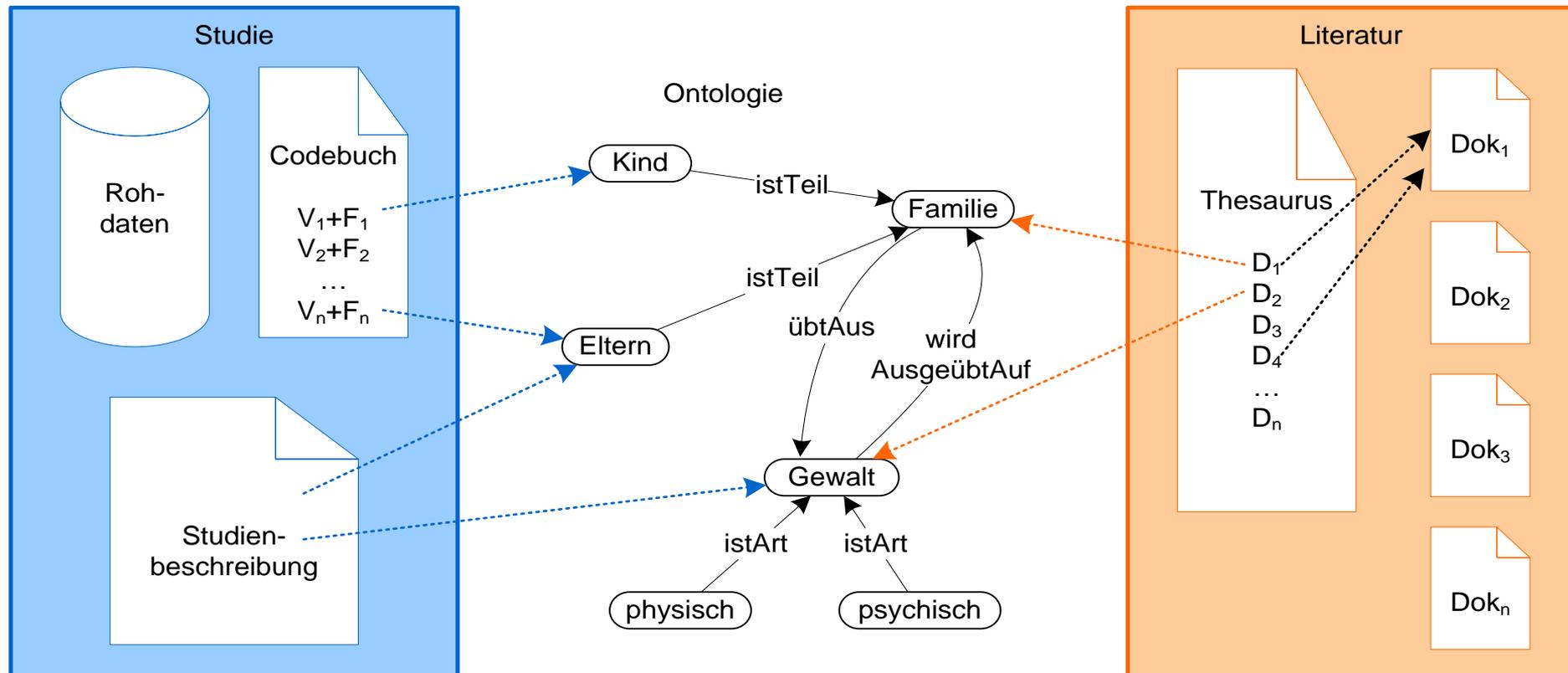
- Anreichern eigener Datenbestände durch externe Datenquellen
 - z.B. DBpedia, Gemeinsame Normdatei (GND), VIAF, Literatur wie UB Mannheim, hbz, Bibsonomy, ...
 - Alternativ: „nur“ Links zu externen Datenbeständen setzen

- Vernetzung von Thesauri, Klassifikationen, etc.
 - z.B. für Search Term Recommendation und weitere IR-Dienste
 - Thesauri können auch bei Vernetzung zu externen Quellen nützlich sein, z.B. Daten mit denselben Keywords in externen Datenbeständen finden

- Vernetzung von Variablen, Indikatoren, Codelisten
 - Suche nach gleichen/ähnlichen Variablen bei unterschiedlichen Forschungsdatenzentren oder verschiedenen Studienserien

Höherer Erschließungsgrad

- Einsatz von Ontologien zur Inhaltserschließung



Mehrwerte für Wissenschaftler

- vernetzte, organisations- und datenbankübergreifende Recherche
- Präzisere Suchergebnisse durch eine semantisch ausdrucksstärkere Inhaltserschließung

Forschungsdaten analysieren

Haben
Arbeitslosen- und
Ausländerquote
Auswirkungen auf
die Angst, den Job
zu verlieren?

Verfügen die Daten
über eine statistische
Qualität, dass ich sie
verwenden kann?

In welchen
Formaten liegen
die recherchierten
Daten vor?

Passen die
Daten überhaupt
zusammen?

Sind die Daten, die
ich gefunden habe,
die die ich
tatsächlich benötige?

Mögliche Probleme und Hürden

- Daten liegen in unterschiedlichen Formaten vor (pdf, html, csv, spss, ...)
 - Konvertierung notwendig
- Datenschemata oder verwendete Codelisten passen nicht zusammen
 - Schema Matching und Datenharmonisierung notwendig
- Daten liegen in verschiedenen Aggregationsebenen vor
- Statistische Qualitätsbeurteilung (Varianzen, Bias, etc.) oft nicht auf Anhieb möglich



Vorverarbeitung von Daten notwendig, um diese beurteilen, analysieren und weiter verarbeiten zu können!

Datenaustausch und -konvertierung

- Forschungsdaten als LOD veröffentlichen
 - Gemeinsame RDF/S oder OWL-Basis trotz unterschiedlicher Metadatenstandards
 - RDF als Austauschformat verwenden
 - Einheitlicher Datenaustausch, -import, -export
- Automatisierung und Vereinheitlichung von Konvertierungsprozessen
- Datenintegration

Schema Matching

Harmonised Unemployment Rate by Gender (Quelle: Eurostat)

	2010-10		2011-01		2011-04	
	Male	Female	Male	Female	Male	Female
Austria	4.4	4.0	4.4	4.6	4.3	4.1
Belgium	7.9	8.3	7.7	7.8	7.5	8.1
Germany	7.2	6.8	6.8	6.4	6.5	6.1
France	9.1	10.3	8.9	10.3	8.6	10.3
...

Schema Matching

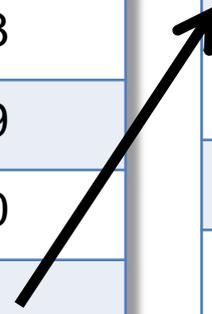
Country	Unemployment Rate by Gender				Year	Month
Albania	2010-10		2011-01		2007	1-01
Austria	Male	Female	Male	Female	2008	February
Belgium	4.4	4.0	4.4	4.6	2009	March
Bulgaria	7.9	8.3	7.7	7.8	2010	...
...	7.2	6.8	6.8	6.4	2011	...
France	9.1	10.3	8.9	10.3	8.6	10.3
...

Schema Matching

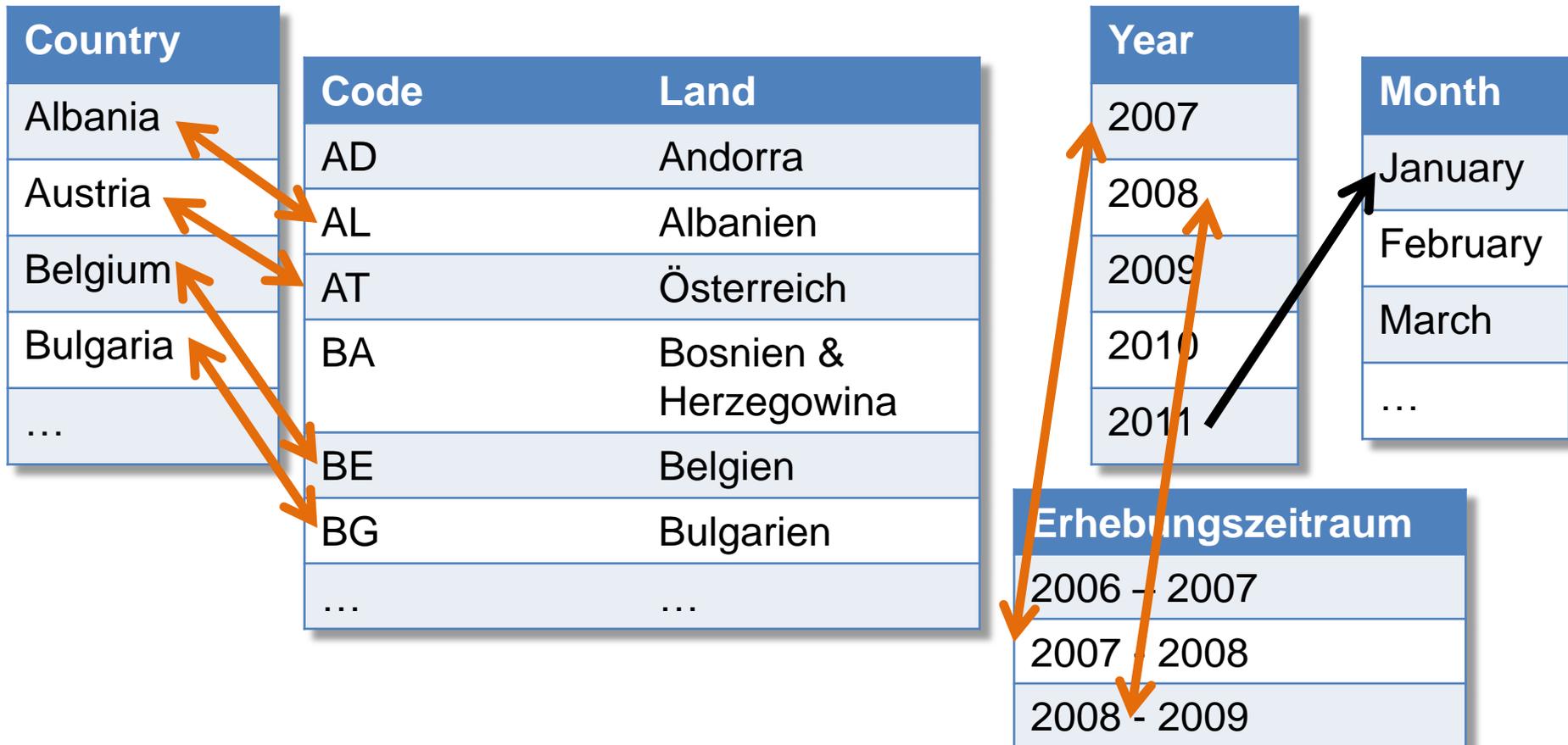
Country
Albania
Austria
Belgium
Bulgaria
...

Year
2007
2008
2009
2010
2011

Month
January
February
March
...

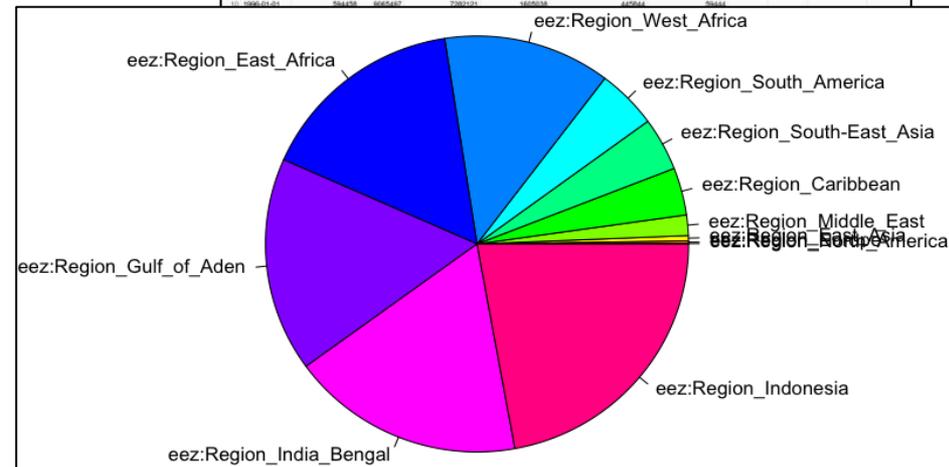
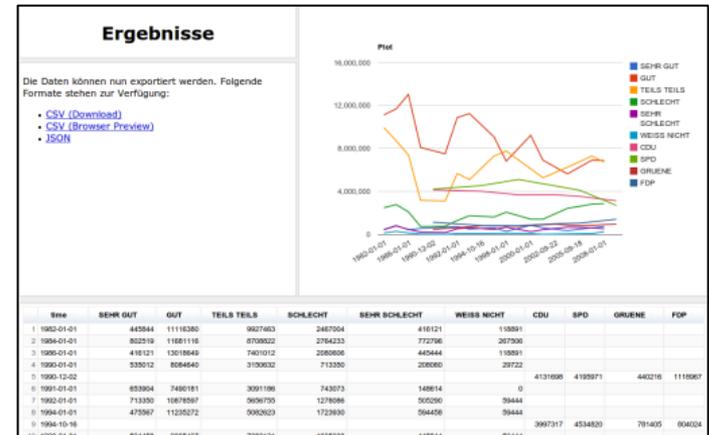


Schema Matching



Weiterverarbeitung von Forschungsdaten

- Anbindung an Analyse-, Grafik- und Statistikprogramme
 - z.B. OLAP-Systeme, Google Visualization API, R Project
 - Unterstützung bei der Beurteilung von Relevanz und statistischer Qualität
 - Ersteinschätzung von Daten (z.B. Trends)



Mehrwerte für Wissenschaftler

- Unterstützung bei der Vorverarbeitung von Forschungsdaten für die Analyse
 - Konvertierung
 - Schema Matching
 - Datenharmonisierung
- Einfacherer Umgang / Interaktion mit heterogenen Forschungsdaten
 - Berechnungen durchführen
 - Visualisierungen erstellen
 - Weiterverarbeitung in Statistikprogrammen

LOD für Datenanbieter

- Interne und externe Vernetzung von Daten
 - Eigene Datenbestände um Daten aus dem Web anreichern
 - Steuerbar: Eigener Scope/Content kann klar abgetrennt werden
 - durch mehr Links nach außen oder
 - einbeziehen zusätzlicher Metadaten
- Präzisere Inhaltserschließung von Daten und Informationen durch semantisch ausdrucksstarke Ontologien
 - Unterstützung durch Suchmaschinen über Mapping zu schema.org
- Anbieten von Mehrwertdiensten basierend auf LOD-Daten
 - z.B. für die Analyse von Forschungsdaten
- Höhere Sichtbarkeit im Web

Zusammenfassung

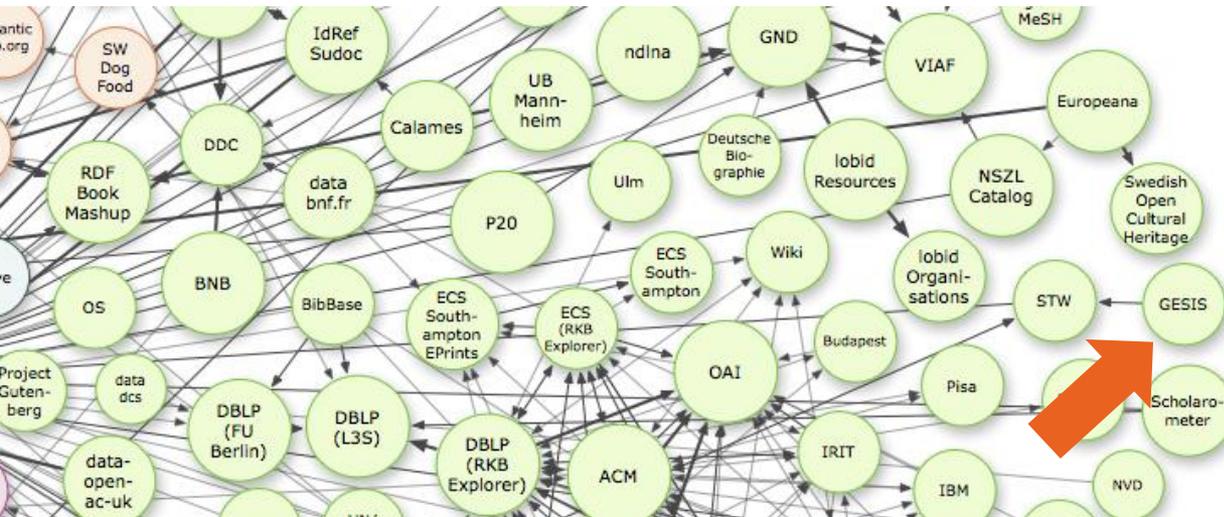
LOD im wissenschaftlichen Forschungsprozess kann

- Forschungsdaten und Informationen enger miteinander vernetzen
- Integrierte Suche über Datenbestände, Organisationsgrenzen hinweg ermöglichen
- Präzisere Suchergebnisse durch fein-granulare Inhaltserschließung liefern
- Typische Schritte der Vor- und Weiterverarbeitung von Forschungsdaten vereinfachen

Offene Punkte

- Offenheit, Datenschutz, lizenzrechtliche Problemstellungen
- Forschung transparenter und nachvollziehbar machen
 - z.B. Forschungsergebnisse veröffentlichen und nachvollziehbar machen
 - Wissenschaftler sind selbst gefragt (stark community-abhängig)
- Anwendungsszenarien von LOD für Primärforscher
 - z.B. Unterstützung bei der Planung und Durchführung von Studien

Vielen Dank!



Benjamin Zapilko

<benjamin.zapilko@gesis.org>

**GESIS – Leibniz-Institut für Sozialwissenschaften
Wissenstechnologien für Sozialwissenschaften (WTS)**



hbz

Wissen. Information. Innovation.

Zum Aufbau von Linked-Open-Data- Services im hbz

FIS Fachtagung, Frankfurt/Main

22. Mai 2012

Adrian Pohl <pohl@hbz-nrw.de>

Agenda

1. Überblick
2. lobid.org
3. culturegraph
4. Perspektiven

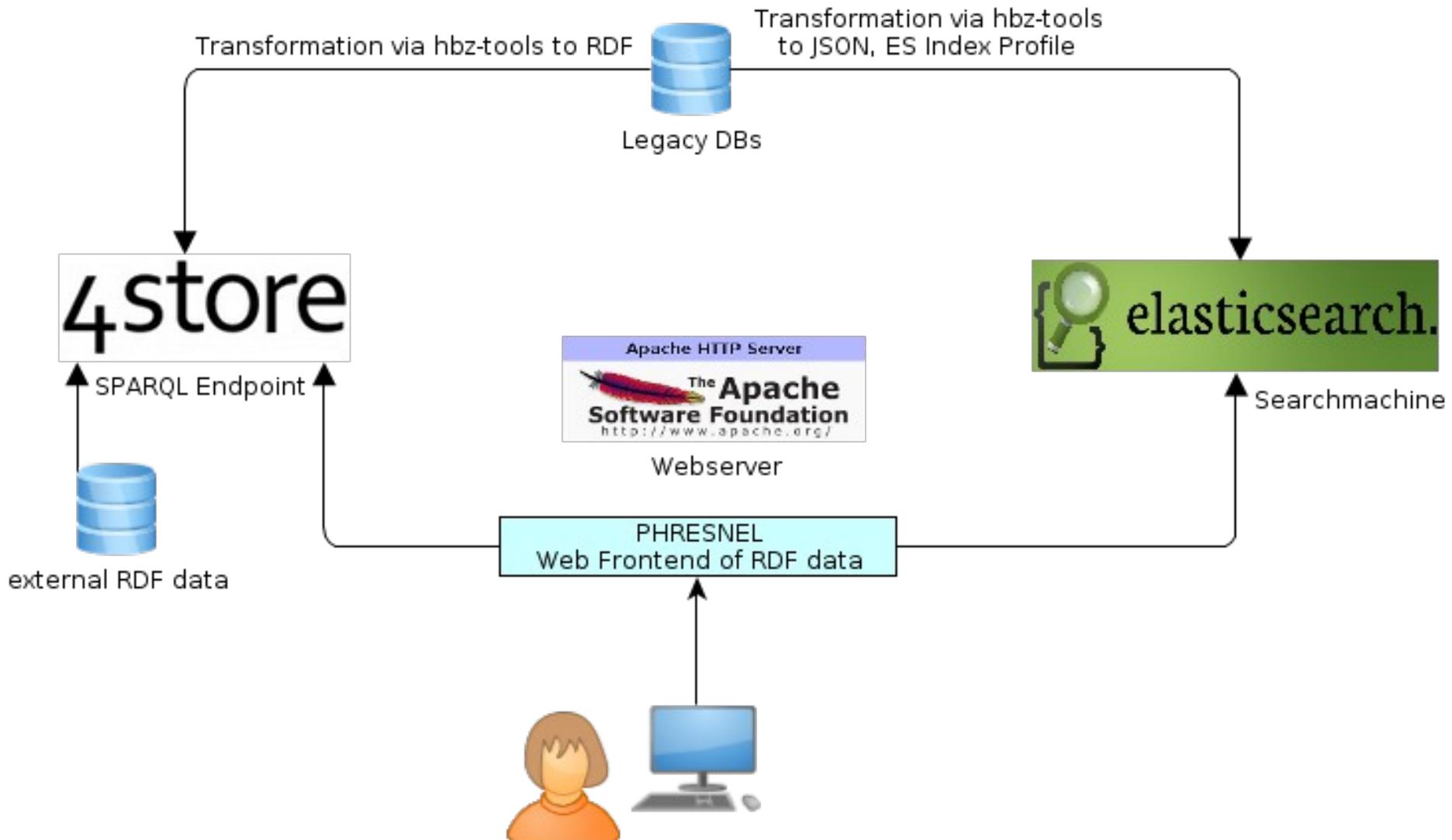
1. Überblick

- Seit 2009 Jährliche SWIB-Konferenz (with ZBW)
- 03/2010 Erste Open-Data-Veröffentlichung
- Seit 06/2010 Koordinierung der OKFN Working Group on Open Bibliographic Data
- 07/2010 lobid.org gelauncht
- 11/2010 Projekt culturegraph.org gestartet
- 01/2011 Prinzipien zu offenen bibliographischen Daten
- 09/2011 Mitbegründung der Igelu Linked Open Data Special Interest Working Group
- 11/2011 Rechlicher Open-Data-Leitfaden veröffentlicht
- DINI-KIM-Empfehlungen zur Öffnung bibliothekarischer Daten

2. lobid.org

Technologie

- Triple Store/SPARQL endpoint: 4store
- Suchmaschine: elasticsearch
- RDF-Browser & Formulargenerator:
Phresnel



All marks mentioned may be trademarks or registered trademarks of their respective owners.

lobid-organisations

- RDF-Beschreibungen für mehr als 40k internationale Organisationen
- Datenquellen: Deutsches ISIL-Verzeichnis & MARC-Organization-Code-Datenbank
- (Noch) nicht offen lizenziert

**DIENSTE**

SPARQL Endpoint

SPARQL Webform

RESSOURCEN

Suche

Downloadbereich Open Data

Exporte

ORGANISATIONEN

Suche

RESTful HTML-Beispiel

META

LOD-Wiki des hbz

Impressum

Kontakt

Credits

<http://lobid.org/organisation/DE-30>

Name	Universitätsbibliothek Johann Christian Senckenberg								
Geokoordinaten									
Anschrift	<table border="1"> <tr> <td>Straße</td> <td>Bockenheimer Landstr. 134/138</td> </tr> <tr> <td>PLZ</td> <td>60325</td> </tr> <tr> <td>Ort</td> <td>Frankfurt/Main</td> </tr> <tr> <td>Land</td> <td>Germany</td> </tr> </table>	Straße	Bockenheimer Landstr. 134/138	PLZ	60325	Ort	Frankfurt/Main	Land	Germany
Straße	Bockenheimer Landstr. 134/138								
PLZ	60325								
Ort	Frankfurt/Main								
Land	Germany								
Telefon	<table border="1"> <tr> <td>Rufnummer</td> <td>(069) 7 98-39205 und 39202</td> </tr> </table>	Rufnummer	(069) 7 98-39205 und 39202						
Rufnummer	(069) 7 98-39205 und 39202								
Telefon	<table border="1"> <tr> <td>Rufnummer</td> <td>(069) 7 98-39380</td> </tr> </table>	Rufnummer	(069) 7 98-39380						
Rufnummer	(069) 7 98-39380								
ISIL	DE-30								
Homepage	http://www.ub.uni-frankfurt.de								
E-Mail	mailto:Auskunft@ub.uni-frankfurt.de								
Kontakt-QR									
Verbundsystem	http://lobid.org/organisation/DE-603								
Weitere Information	http://dispatch.opac.d-nb.de/DB=1.2/CMD?ACT=SRCHA&IKT=8529&TRM=DE-30								
Sitz	http://sws.geonames.org/6553153/								
Wikipedia-Link	http://de.wikipedia.org/wiki/Universitätsbibliothek_Johann_Christian_Senckenberg								
DBpedia-Link	http://de.dbpedia.org/resource/Universitätsbibliothek_Johann_Christian_Senckenberg								

lobid-resources

- LOD-Schnittstelle für offene Daten aus dem hbz-Verbundkatalog
- Lizenziert mit CC0
- Momentan gut 10M Einträge
- Links in LOD-Cloud (GND, ZDB, DBpedia, dewey.info, Projekt Gutenberg...)
- Siehe Beschreibung auf *the Data Hub*

<http://thedatahub.org/dataset/lobid-resources>

**DIENSTE**

SPARQL Endpoint

SPARQL Webform

RESSOURCEN

Suche

Downloadbereich Open

Data Exporte

ORGANISATIONEN

Suchschnittstelle

RESTful HTML-Beispiel

META

LOD-Wiki des hbz

Impressum

Kontakt

Credits

http://lobid.org/resource/HT002948556	
Titel	With reference to reference
Autor	http://d-nb.info/gnd/135539897
Erscheinungsjahr	1983
Typ	< http://purl.org/ontology/bibo/Book >
Typ	< http://purl.org/vocab/frbr/core#Manifestation >
Sprache	http://id.loc.gov/vocabulary/iso639-2/eng
Erscheinungsort	Indianapolis u.a.
Verlag	Hackett
Format	print
Umfang	VIII, 200 S.
Schlagwort	Referenz <Linguistik>
Schlagwort	Semantik
ISBN-10	0915145529
ISBN-10	0915145537
ISBN-13	9780915145539
ISBN-13	9780915145522
Exemplar	http://lobid.org/item/HT002948556%3AAE1772
Besitzer	Rheinisch-Westfälische TH, Hochschulbibliothek
Signatur	AE1772
Exemplar	http://lobid.org/item/HT002948556%3A1948.ELG.E.1983
Besitzer	Institut für Philosophie der Universität, Bibliothek
Signatur	1948.ELG.E.1983
Exemplar	http://lobid.org/item/HT002948556%3AHLK260
Besitzer	Universitätsbibliothek Duisburg-Essen, Campus Essen
Signatur	HLK260
Exemplar	http://lobid.org/item/HT002948556%3A33+i+E+296
Besitzer	Institut für Ethik, Geschichte und Theorie der Medizin, Bibliothek

Katalogisierung in RDF

- Phresnel: RDF-Browser zur Generierung von HTML-Seiten via SPARQL
- Basiert auf dem Fresnel Display Vocabulary for RDF
- Kann auch zur Generierung von Formularen zur Datenproduktion verwendet werden.

Probleme

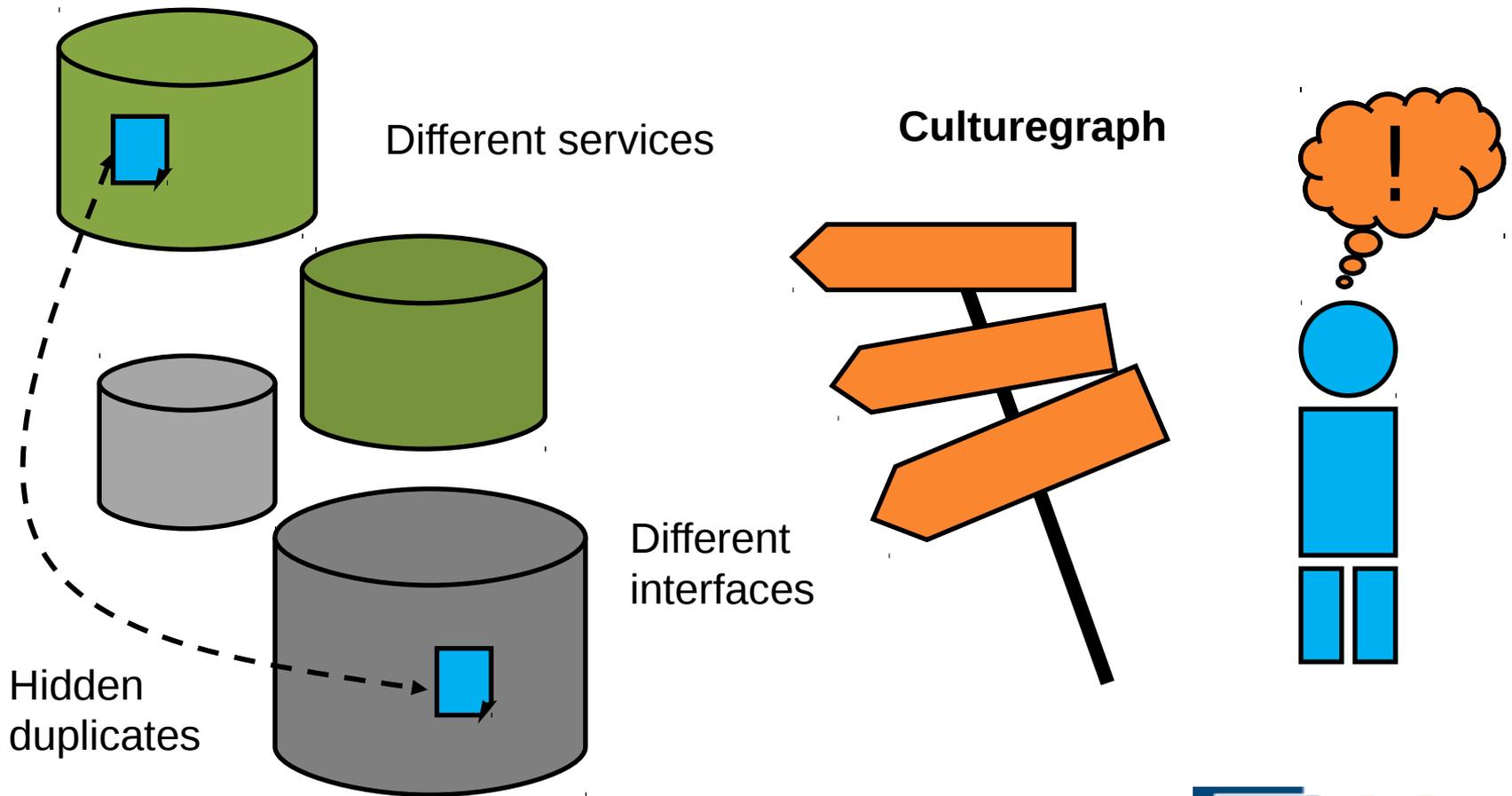
- Datenproduktion in Legacy-System & Updates mittels Vollabzügen
- Daten sind nicht aktuell
- Daten können nicht verbessert & an die Quelle zurückgeführt werden
- Fehlende Versionierung
- Fehlende Update-Schnittstelle

3. culturegraph

Offene Plattform zur Datenverlinkung

- Persistente Identifizierung von Entitäten
- Verwaltung und Veröffentlichung von Mappings
- Berechnung von Ähnlichkeiten und Verknüpfungen
- Offene Daten / Freie Software

culturegraph als Wergweiser



Erstes Projekt: Resolving & Lookup

- **Input:** Deutsche Verbundkataloge
- **Ziel:** Bündelung von Titeln
- **Service:**
 - Publikation der Bündel
 - Prägen von URIs für Bündel
 - Suche von Bündeln auf Basis bekannter Identifikatoren

Erreichter Stand

- Projekt erfolgreich abgeschlossen
- Skalierbare, performante Infrastruktur wurde aufgebaut
- Aktueller Datenbestand:
 - Monografien seit 1945
 - aktuell ca. 88 Mio. Datensätze im System
 - Verarbeitung (Import/Analyse/Clustering) aller Daten in ca. 2 Std.
- Beta-Dienst online:
<http://www.culturegraph.org/demonstrator>

4. Perspektiven

Ergänzung weiterer Informationen in lobid.org

- Dienstleistungen und Öffnungszeiten
- Sammlungsinformationen
- Verfügbarkeitsinformationen
- Kundeninformationen (geschützt)
- Links zu:
 - Open Library
 - Geonames (Verlagsorte)
 - ...

Erforschung zukünftiger Katalogisierungssysteme

Zukünftige web-basierte
Katalogisierungsdienste

- basieren auf LOD-Standards,
- pushen Änderungen in Echtzeit,
- unterstützen Versionierung,
- sind Teil des Giant Global Graph.

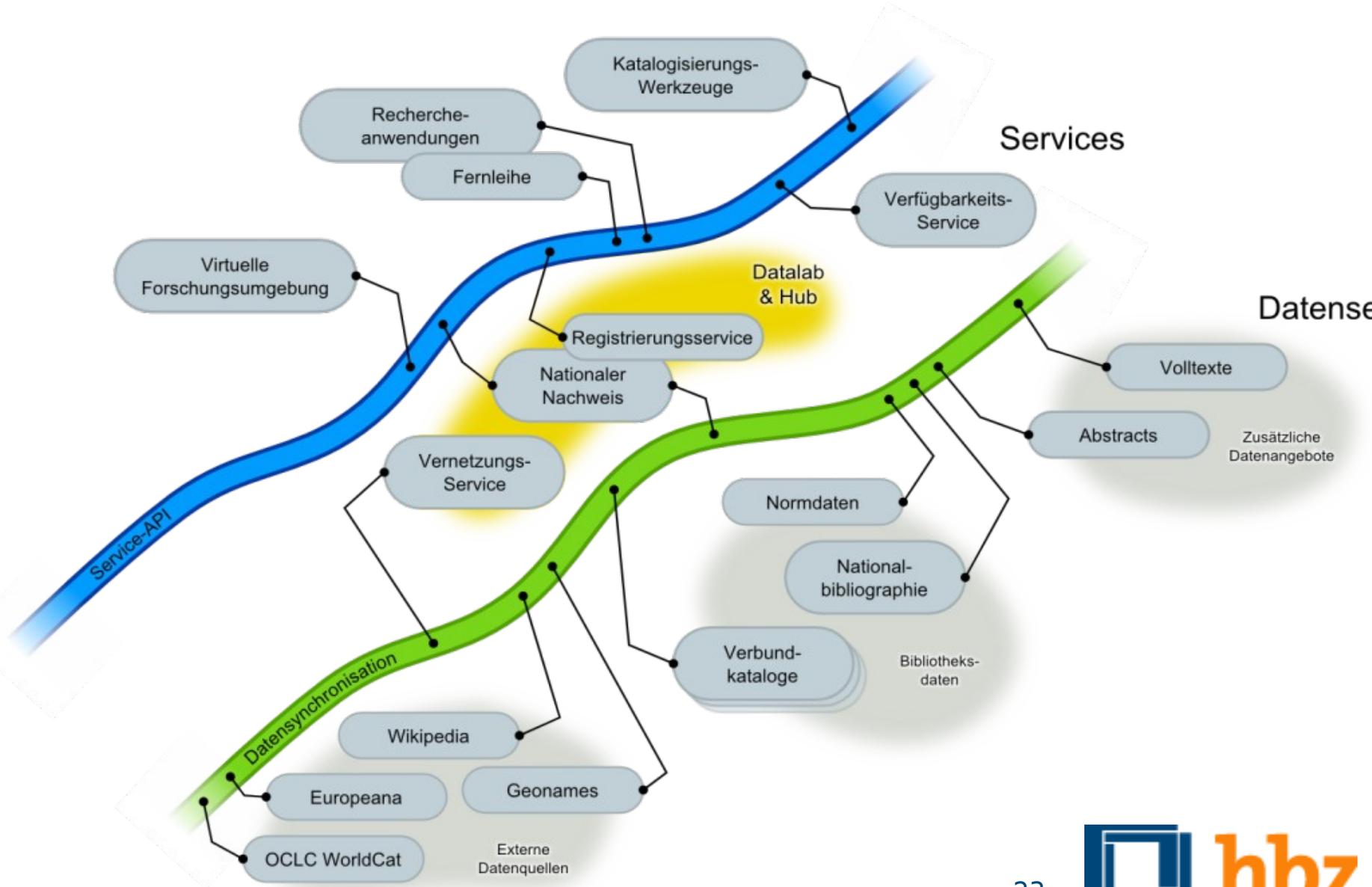
Aufbau einer neuen Bibliotheksinfrastruktur

- Basierend auf Linked-Open-Data-Prinzipien
- Datensynchronisation in Echtzeit
- Web Services

Anwendungen

Services

Datensets



SWIB 2012

- **"Semantic Web in Libraries"** (SWIB)
hosted by hbz and ZBW - Leibniz Information
Centre for Economics
- Cologne, **27. - 29.11.2012**
- <http://swib.org/>

Danke.

Fragen?

Gerne jetzt oder auch später an
semweb@hbz-nrw.de

Weitere Informationen

→ [hbz LOD blog](#) (Nachrichten rund um lobid.org)

→ Phresnel source code on github:

<https://github.com/lobid/PhresnelApp>

Lizenz



Diese Folien stehen unter folgender CC-Lizenz:
<http://creativecommons.org/licenses/by/3.0/de/>