



hbz

Wissen. Information. Innovation.

Mit Linked Open Data in die Zukunft

Hessischer Bibliothekartag, Langen 9.5.2011

Dr. Silke Schomburg

1. Open Data | Linked Data

Open Knowledge

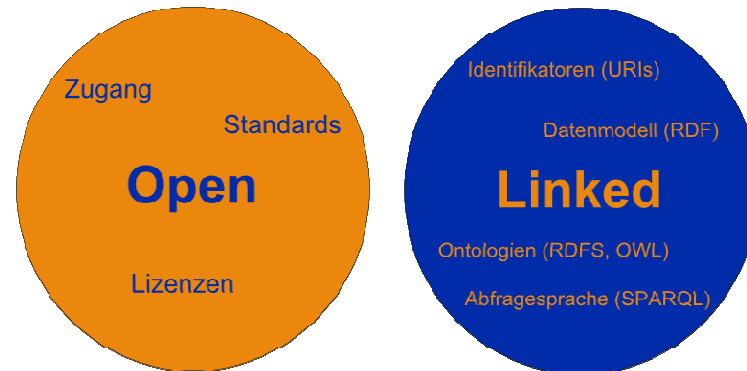
“A piece of knowledge is open if you are free to use, reuse, and redistribute it.”

<http://www.opendefinition.org>

Siehe auch: *Prinzipien zu offenen bibliographischen Daten*,
<http://openbiblio.net/principles/de/>

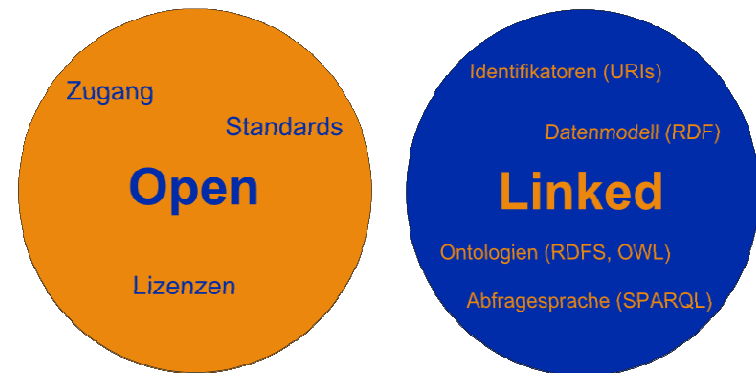
Offenheit ist eine Frage von...

- Zugänglichkeit
- Standards
- **Lizenzen**



- Attribution, Share-alike oder Public-Domain-Waiver
- **Nicht** Non-Commercial

Linked Data ist eine Frage von...



Standards

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

Warum Linked Open Data?

- Linked Open Data: die Zukunft der Datenhaltung und -publikation im Web
- Chance, ein präserter und zuverlässiger Teil des Webs zu werden.
- Internationalität des Ansatzes
- Datenintegration ermöglichen durch Nutzung nicht-bibliotheksspezifischer allgemeiner Standards

Aufgaben

- Vokabularentwicklung
- Veröffentlichung von grundlegenden Daten als RDF-Tripel
 - Titeldaten, Bestandsdaten, Organisations- und Servicebeschreibungen, ...
- Verlinkungen

Hindernisse

- Technischer Aufwand
- Semantic-Web-Technologie ist noch stark in der Entwicklung
 - Vokabulare für viele Zwecke fehlen
 - Tools müssen noch entwickelt werden
 - Provenienzinformation, Versionierung noch nicht geklärt
- Zusammenwirken mit Legacy-Software oft problematisch

Strategie

- LOD als mittel- bis langfristige strategische Perspektive
- Kooperation und Austausch verstärken
- Auch international vernetzen
- Open Data publizieren (auch Rohdaten)
- Linked-Data-Kompatibilität neuer Software (z.B. Ex Libris' Alma) fordern und sicherstellen

2. Linked (Open) Data im hbz

lobid.org

- **Linking Open Bibliographic Data**
- *Titel*daten als Linked Open Data
- URIs und Beschreibungen
bibliothekarischer *Organisationen*
- Verlinkungen beider Dienste über
Bestandsnachweise

Titeldaten: lobid-resources

- RDF-Repräsentation der hbz-Open-Data (bisher 6 Mill. Titeldatensätze)
- Mehr als 90 Millionen RDF-Tripel
- Ca. 6,5 Millionen Links zur GND
- Ca. 9 Millionen Links zu lobid-organisations
- Bsp.: <http://lobid.org/resource/HT002948556>

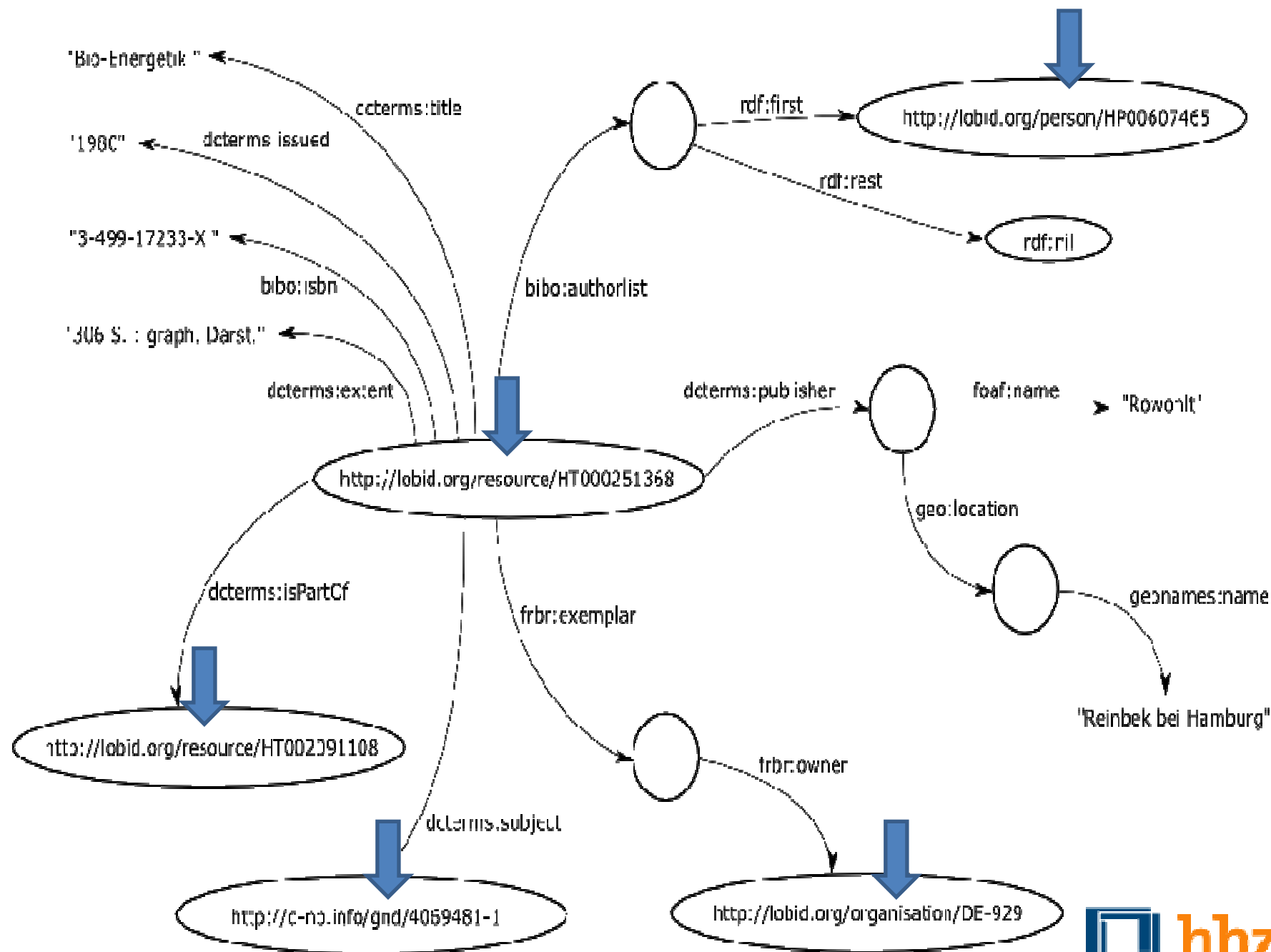
Export und Konvertierung

- Export der Verbunddaten über die Aleph-XML-Schnittstelle
- Umwandlung in verlustfreies MAB-basiertes RDF (RDF/ISO2709)
- Mapping RDF/ISO2709 auf die Bibliographic Ontology
- Konvertierung der Daten nach RDF
- Einspielen in einen Triple Store (4store)

Benutzte Ontologien

- Bibliographic Ontology (Bibo)
- Dublin Core
- FOAF
- FRBR vocab (aber **keine** FRBRisierung der Daten)
- Dokumentation:

<https://wiki1.hbz-nrw.de/display/SEM/Converting+the+Open+Data+from+the+hbz+to+BIBO>



lobid-organisations

- Datenquellen: ISIL-Verzeichnis und MARC Organization Codes der Library of Congress
- Angereichert mit Google Maps
- Problem: (noch) nicht Open Data

Beispiel-URIs

→ Bibliothek des Paul-Ehrlich-Instituts,
Langen:

<http://lobid.org/organisation/DE-F20>

→ Hochschulbibliothekszenrum des
Landes Nordrhein-Westfalen:

<http://lobid.org/organisation/DE-605>

Paul-Ehrlich-Institut, Bundesamt für Sera und Impfstoffe, Bibliothek

ISIL: DE-F20

Address

Street Paul-Ehrlich-Str. 51-59

Postal Code 63225

Place Langen/Hessen

Country

Communication

Phone (06103) 77-0 und 1006

Fax (06103) 77-1264

Email <mailto:pauch@pei.de>

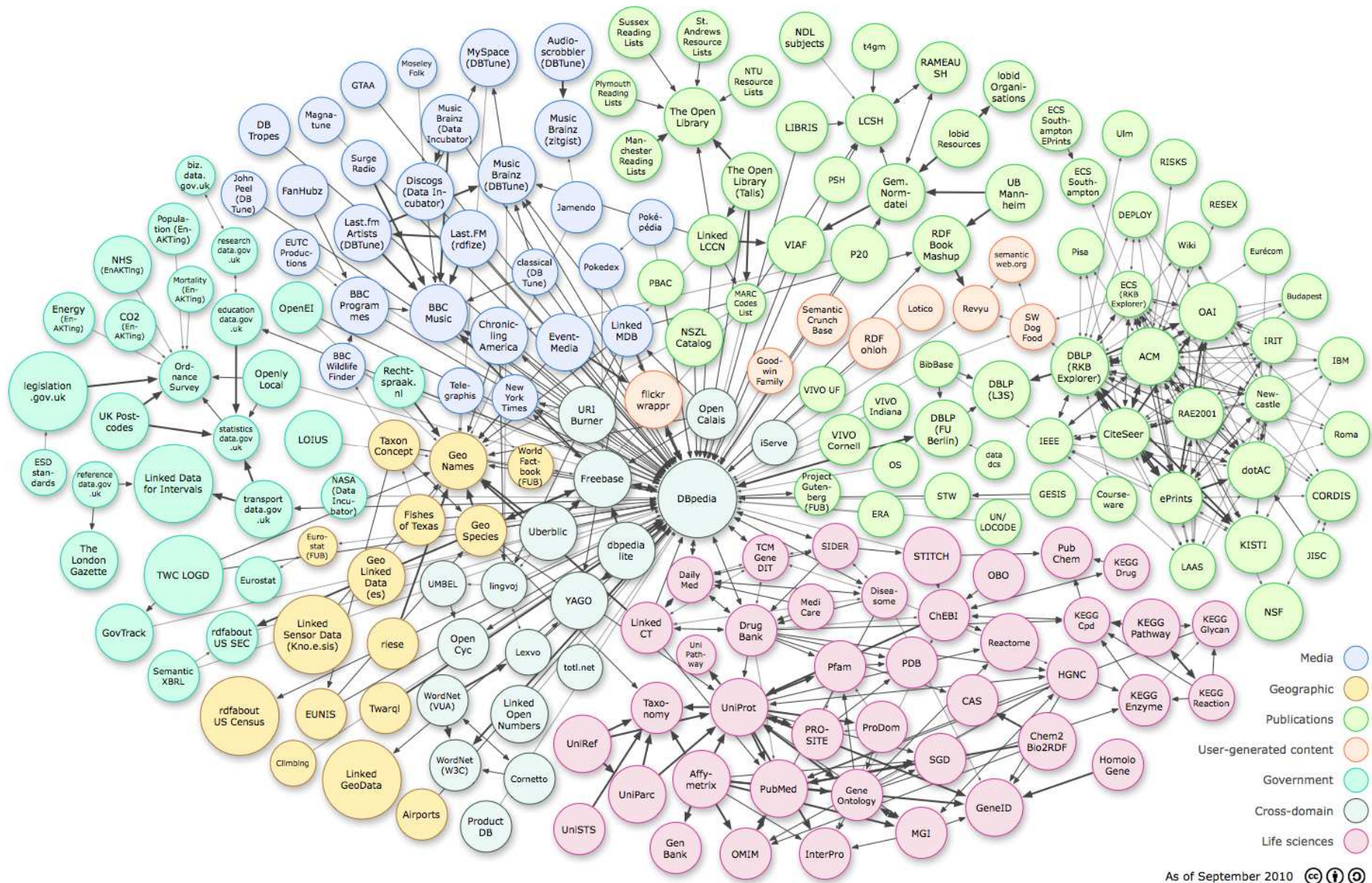
URL <http://www.pei.de>



© 2010 Hochschulbibliothekszentrum | [OPEN](#) [CONTENT](#)

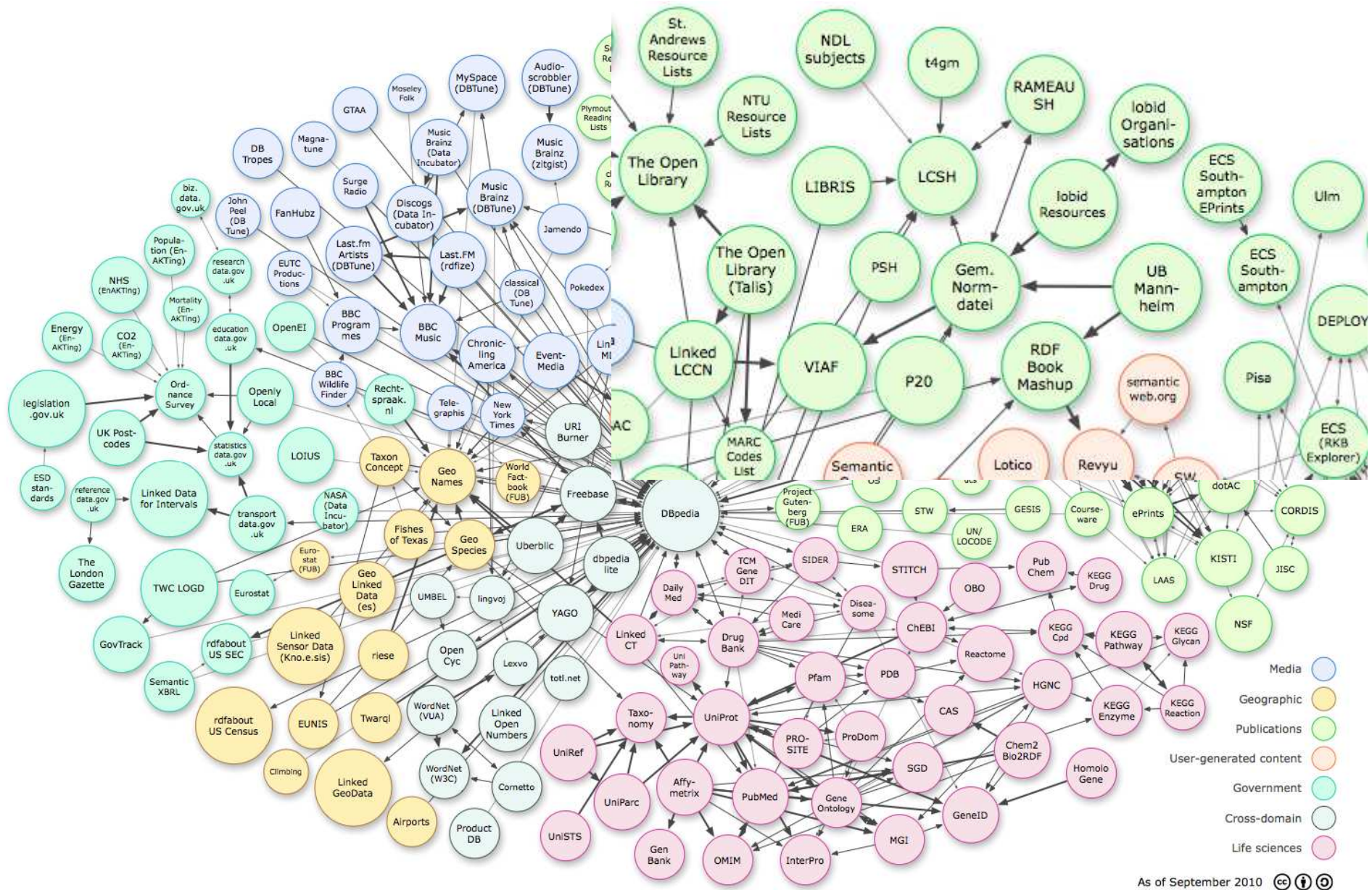
Standards Compliant [XHTML](#)

```
<h1 property="foaf:name">hbz - Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen</h1>
<div rel="hcterms:hasCard" xmlns:hcterms="http://purl.org/uF/hCard/terms/">
  <div typeof="v:VCard" xmlns:v="http://www.w3.org/2006/vcard/ns#">
    <div rel="v:adr">
      <div typeof="v:Address v:Work">
        <span property="v:street-address">Jülicher Str. 6</span>
        <span property="v:postal-code">50674</span>
        <span property="v:locality">Köln</span>
        <span property="v:country-name">Deutschland</span>
      </div>
    </div>
    <div rel="v:tel">
      <div typeof="v:Voice v:Work">
        <span property="rdf:value">(0221) 4 00 75-0</span>
      </div>
    </div>
    <div rel="v:tel">
      <div typeof="v:Fax v:Work">
        <span property="rdf:value">(0221) 4 00 75-180</span>
      </div>
    </div>
    <div rel="v:email">
      <a href="mailto:info-hbz@hbz-nrw.de">info-hbz@hbz-nrw.de</a>
    </div>
    <div rel="v:uri">
      <a href="http://www.hbz-nrw.de">http://www.hbz-nrw.de</a>
    </div>
  </div>
</div>
```

http://de.wikipedia.org/w/index.php?title=Datei:Lod-datasets_2010-09-22_colored.png&filetimestamp=20101022110116





http://de.wikipedia.org/w/index.php?title=Datei:Lod-datasets_2010-09-22_colored.png&filetimestamp=20101022110116



Use Case: Bibliothekslandkarte

- Landkarte mit bibliothekarischen Organisationen
- Filtern nach Verbundzugehörigkeit und Organisationstyp
- Basis: lobid.org/organisations



SPRACHEN

[Deutsch](#)

[Englisch](#)

DIENSTE

RESSOURCEN

[Suchschnittstelle](#)

[SPARQL Endpoint](#)

[SPARQL Webform](#)

[RESTful Beispiel](#)

[Downloadbereich Open](#)

[Data Exporte](#)

ORGANISATIONEN

[Suchschnittstelle](#)

[RESTful HTML-Beispiel](#)

[RESTful RDF-Beispiel](#)

[Verbundlandkarte](#)

META

[Impressum](#)

[Kontakt](#)

[Credits](#)



Use Case: Geobasierte Recherchedienste

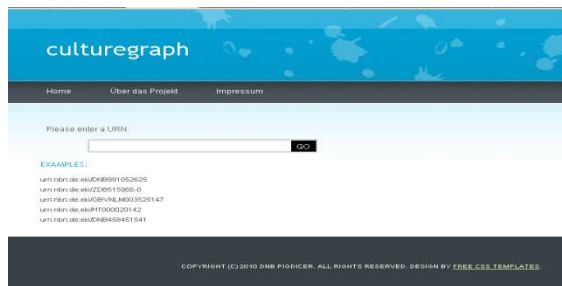
- Titel- und Organisationsdaten sind in einer gemeinsamen RDF-Datenbank (Triple Store)
- Geobasierte Abfragen sind möglich
- Bsp.: Suche nach Buch mit ISBN 978-3-631-58937-3 mit < 5km Entfernung vom hbz

SPARQL-Anfrage

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix v: <http://www.w3.org/2006/vcard/ns#>
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix hcterms: <http://purl.org/uF/hCard/terms/>
prefix dcterms: <http://purl.org/dc/terms/>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
prefix frbr: <http://purl.org/vocab/frbr/core#>
prefix bibo: <http://purl.org/ontology/bibo/>
```

```
SELECT ?lib ( ( (50.9341-xsd:float(?lat))*(50.9341-xsd:float(?lat)) + (6.93549-
    xsd:float(?lon))*(6.93549-xsd:float(?lon))*(0.831939969105-
    (0.00853595*xsd:float(?lat))) ) AS ?d2brgrad) WHERE {
?man bibo:isbn "978-3-631-58937-3" .
?man frbr:exemplar ?ex .
?ex frbr:owner ?lib .
?lib geo:location ?loc .
?loc geo:lat ?lat .
?loc geo:long ?lon .
?uri foaf:name ?nam .
FILTER( (50.9341-xsd:float(?lat))*(50.9341-xsd:float(?lat)) + (6.93549-
    xsd:float(?lon))*(6.93549-xsd:float(?lon))*(0.831939969105-
    (0.00853595*xsd:float(?lat))) < 0.0020219493461806057363177376485694 ) .
} LIMIT 10
```

Resolving- und Lookup-Dienst für bibliothekarische Identifier in culturegraph.org



Ein gemeinsames
Infrastrukturprojekt von DNB und
hbz

Ziele

- Aggregation von Identifiern und anderen relevanten Feldern aus Verbundkatalogen
- Bündelung ähnlicher/gleicher Datensätze anhand einfacher Matchingalgorithmen
- Prägung von URIs für Bündelungen



Lookup und Resolving

- Look Up: Beliebiger bibliographischer Identifier wird zu einheitlichen Ressourcen-URIs (CG-URIs) aufgelöst.
- Resolving: Alle mit einer CG-URI verbundenen Beschreibungen (Titeldatensätze) werden ausgegeben



1. Projektabschnitt

- erste Hälfte 2011
- Aufbau einer ersten Infrastruktur
- Übernahme von ausgewählten Daten
- Prototypische Einbindung von Matching-Algorithmen
- Feedback ausgewählter Verbünde

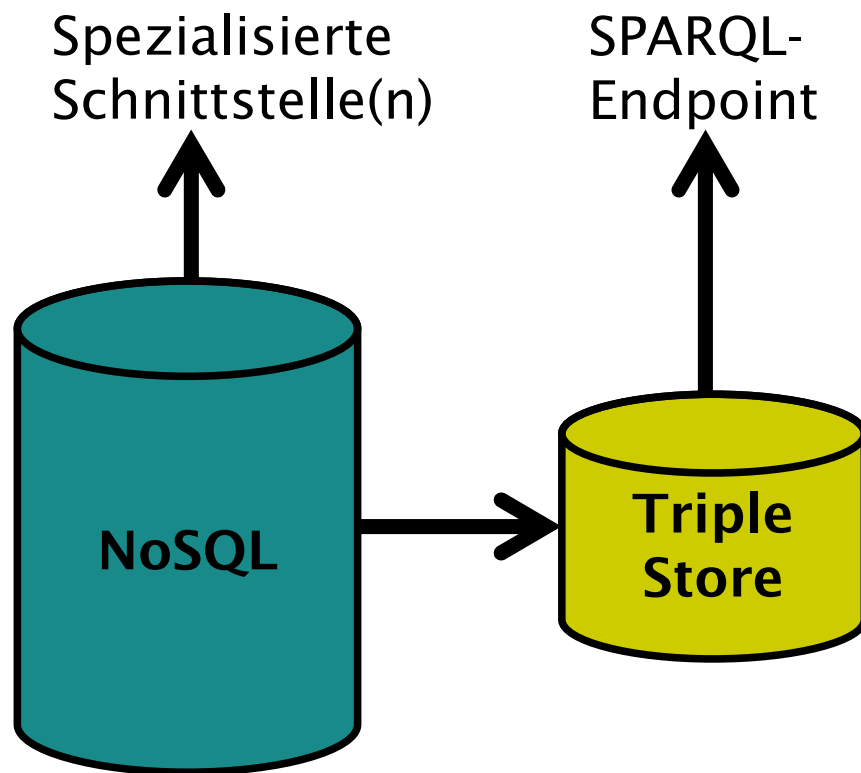


2. Projektabschnitt

- zweite Hälfte 2011
- Verbesserung, Ausbau der Algorithmen
- Test und Nutzung der Plattform durch die Verbundpartner



Architektur



- NoSQL-Datenbank für spezialisierte Anwendungen – z.B. Dublettenerkennung
- Triple Store exponiert die Linked Open Data
- Nur im Triple Store ist LOD

Status

- Relevante Felder für Matchingalgorithmen sind identifiziert
- Die vorgesehene Systemarchitektur existiert
- Importer für die Formate JSON, MAB und MARC sind geschrieben und implementiert
- Datensätze von DNB und hbz wurden testweise importiert und stehen für Tests zur Verfügung



Vielen Dank für Ihre
Aufmerksamkeit.

Fragen?

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

Weiterführende Informationen

Semantic-Web-Bereich im öffentlichen Wiki des hbz:

<https://wiki1.hbz-nrw.de/display/SEM/Home>

Culturegraph-Mailingliste: <http://lists.d-nb.de/mailman/listinfo/culturegraph>

Prinzipien zu offenen bibliographischen Daten:

<http://openbiblio.net/principles/de/>

Adrian Pohl: *Open Data im hbz-Verbund*. Erschienen in ProLibris, 3/2010, S. 109-113. Preprint:

http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2010_open-data.pdf

Adrian Pohl (2011): *Linked Data und die Bibliothekswelt*.

Erscheint im Konferenzband zur ODOK 2010. Preprint:

http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2011_linked_data_ODOK.pdf