

Eine Neuordnung im Bibliotheks- und Informationssektor?

Ein Beitrag über zukünftige Nachhaltigkeit, Innovation und Strukturbildung in der Informationsversorgung

4. Kongress für Information und Bibliothek, Leipzig 2010

Reinhard Altenhöner, DNB / Dr. Silke Schomburg, hbz

16. März 2010

Agenda

- Warum eigentlich „Neuordnung“: Der Rahmen
- Treiber für Veränderungen der Informationslandschaft
- Antworten der DNB & des hbz

Rahmen

- Evaluierungsaktivitäten WR, DFG-Projekt, KII
- Veränderte Nutzer-/Nutzungsanforderungen: Wissen ist vernetzt, Nutzer arbeiten verteilt, Services werden vernetzt, User interagieren aktiv (mit)
- Technisch relevante Entwicklungen
 - Semantic Web, LOD
 - SOA/WEB/Cloud/Schichtenarchitektur
- Bibliotheksumfeld: Neues Katalogisierungsparadigma RDA?!
- Neue Geschäftsmodelle, neue Partnerschaften (GRID, Forschungsdaten, LZA)
- → Sichtbarkeit von Bibliotheken?

Prüfsteine für zukünftige Dienste

- Unterstützung von Nutzerumgebungen
- Integrationsfähigkeit / Flexibilität von Daten und Diensten
- Modulare Software-Architekturen
- Transparente wohl definierte Schnittstellen, WebAPIs

Stand heute

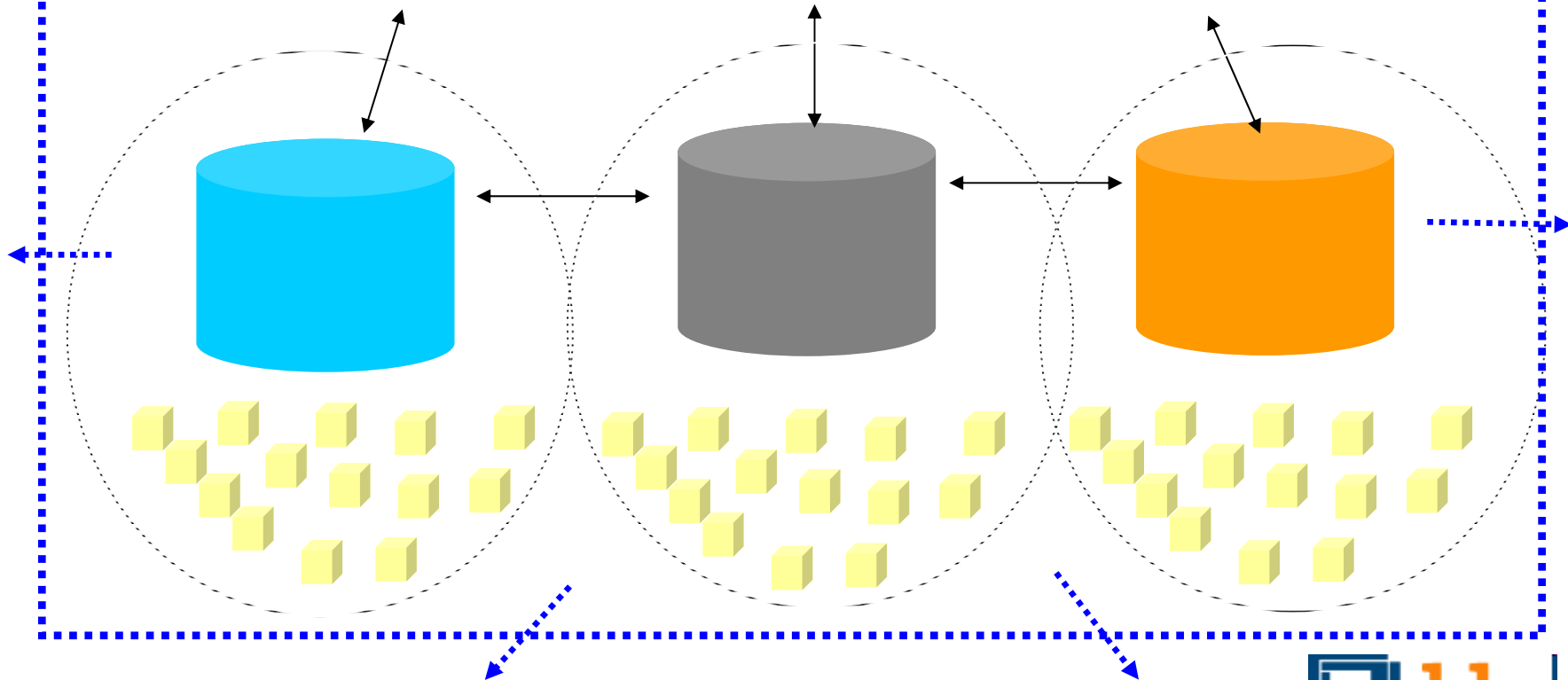
- Daten, die man erzeugt oder auch benötigte Fremddaten, werden „lokal“ gespeichert. Das ist aufwändig, Synchronisierungsprozesse beschäftigen die Community intensiv.
- Dies Daten werden über eigene Zugangssysteme zur Verfügung gestellt.
- DNB Stufe 1: Neue Verteilverfahren (OAI, SRW/SRU-Update)
- DNB Stufe 2: Vorbereitung für die verteilte Datenhaltung. Voraussetzung: konsequente Verwendung von Identifiern, die alle Daten zitierfähig halten
- Einsatz von Workflow-Engines

„geschlossene Welten“

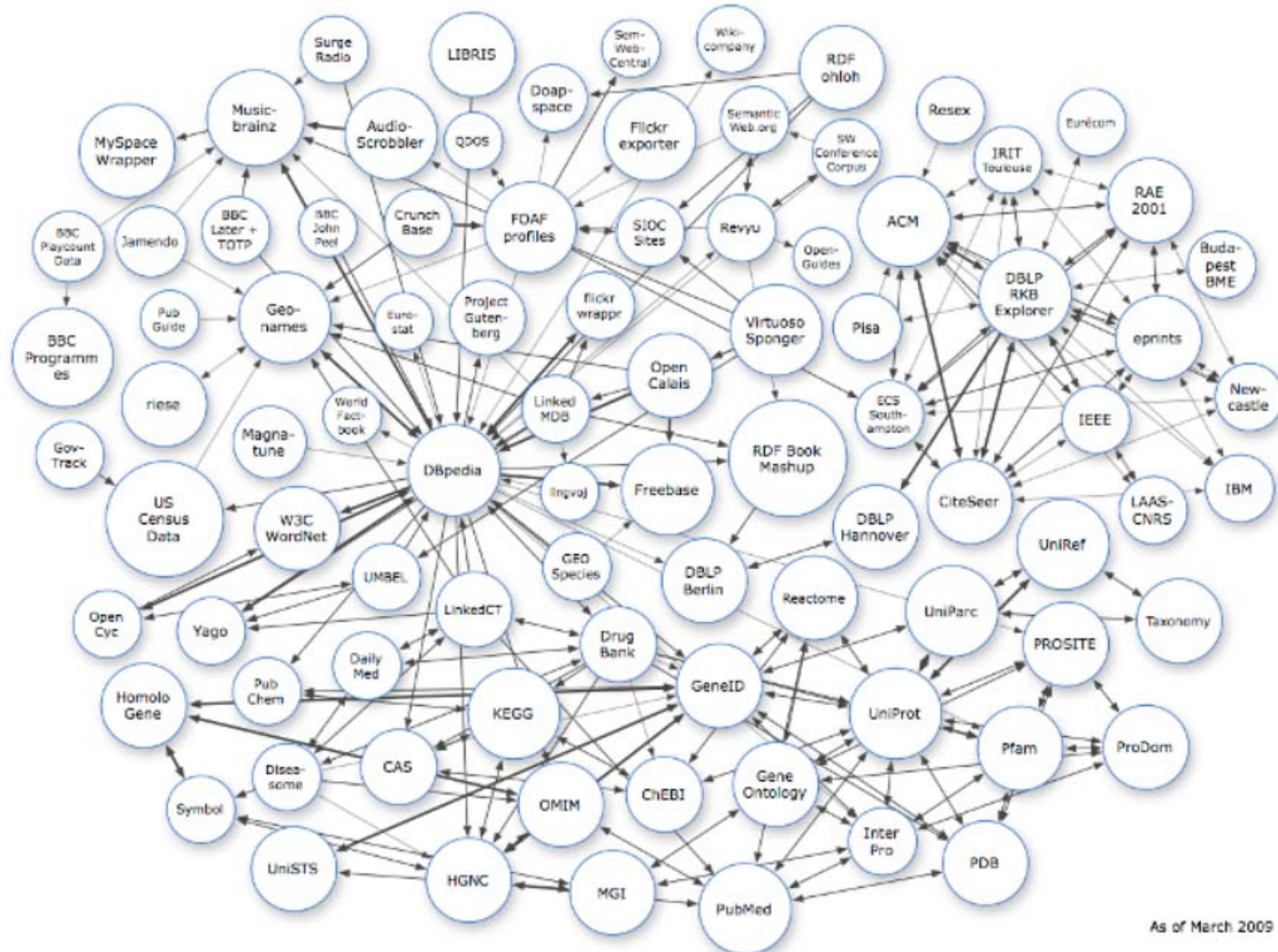
Proprietäre Schnittstellen

Proprietäre Schnittstellen

Zentrale Datenbanken (Normdaten, ZDB)



Linked Data und das Semantic Web



As of March 2009

Arbeitsschwerpunkt Linked Data

- Problem im Web: Was und wo sind zuverlässige und vertrauenswürdige Daten?
→ Bibliotheken sind eine solche Quelle
- Neuer Ansatz
 - Nachnutzbarkeit der Daten (freie Zugänglichkeit, Technik)
 - Reiche Daten (strukturierte Information)
 - Vermeidung von Redundanzen

Linked Data – Linking open data

- (Abstrakte) Ressourcen (“things”) werden verlinkt; ihnen wird eine Eigenschaft / ein Wert (Attribut) zugewiesen. Jede Eigenschaft einer Sache kann man als eine Beziehung (Relation) zu einer anderen Sache formulieren
- Datenmodell: RDF (Subjekt-Prädikat-Objekt-Ausdrücke)
- Technik: konsequenter Einsatz von URIs für die Identifikation von Entitäten, Beziehungen und kontrolliertem Vokabular
- Atomisierung von Sachverhalten, maschinenlesbare Abbildung

Aktionsfeld Linked Data - DNB

- Bereitstellung von Normdaten + bibliographischen Daten als linked data
- Laufende Ontologieentwicklung: Anknüpfung an vorhandene, Verfeinerungen; internationale Kooperation
- RDA-Evaluierung
- Geschäftsmodell: Grundsätzlich sind die Daten für die nicht-kommerzielle Nutzung frei, bei kommerzieller Nutzung wollen wir gefragt werden.
- Ein definitives Lizenzmodell gibt es noch nicht. Creative Commons BY-NC im Prinzip geeignet, aber nur für das Gesamtset. Aber: der einzelne Datensatz?

Linked Data Aktivitäten

- 1. Prototyp im begrenzten Zugriff online seit 15.02.2010 im Approvalsysteem der DNB: PND
 - Verwendete Vokabulare: RDA Group 2 Elements (Namenbezogene Elemente), RDA Relationships for Persons / Corporate Bodies / Families, Relationship Vocabulary, GND (Gemeinsame Normdatei) – (noch) nicht registriert
- 2. Prototyp: Online im Produktivsystem am 30.03.2010
 - Verbesserte PND-Umsetzung, erste Umsetzung SWD, Vokabulare: SKOS / Dublin Core / (gnd?), Verknüpfungen zwischen SWD und PND, erste Links zu externen Datenquellen (DBpedia)
- Probleme / Weiterführende Aspekte:
 - Performantes SPARQL-Interface
 - Datenhaltung / Datenpflege in RDF?
 - Neue Erschließungsumgebung

Erster Prototyp: Beispieldatensatz

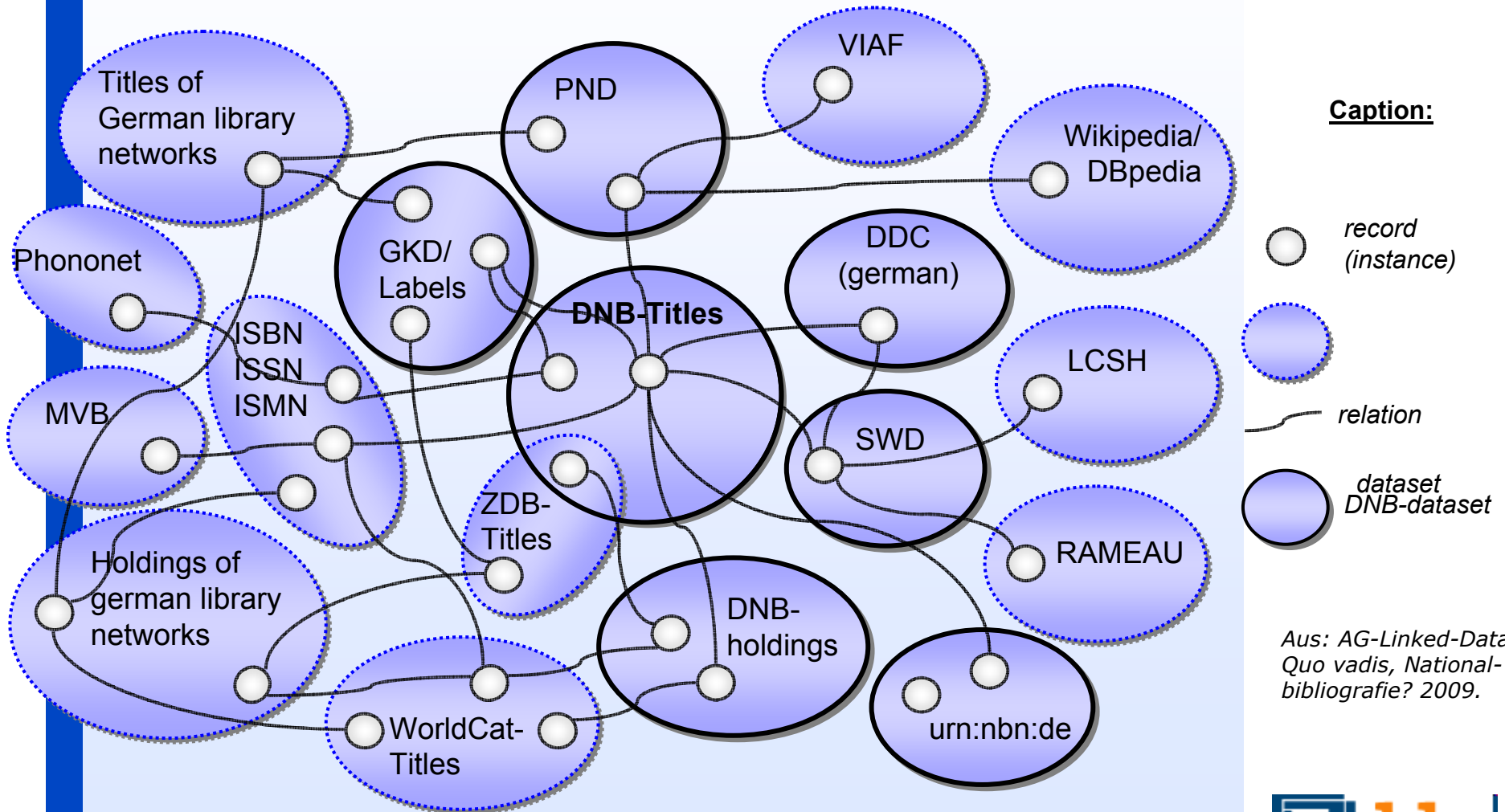
```

<rdf:RDF>
<rdf:Description rdf:about="http://d-nb.info/gnd/130477141">
<gnd:academicTitle rdf:datatype="http://www.w3.org/2001/
XMLSchema#string">Dr. med.</gnd:academicTitle>
<rdaGr2:professionOrOccupation xml:lang="de">Schauspielerin</rdaGr2:professionOrOccupation>
<rdaGr2:identifierForThePerson rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
(DE-588)130477141</rdaGr2:identifierForThePerson>
<rdaGr2:dateOfBirth rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
1966</rdaGr2:dateOfBirth>
<rdaGr2:gender>http://RDVocab.info/termList/gender/1003</rdaGr2:gender>
<gnd:variantNameForThePerson rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
Furtwängler-Burda, Maria</gnd:variantNameForThePerson>
<rdaGr2:placeOfBirth rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
München</rdaGr2:placeOfBirth>
<gnd:preferredNameForThePerson rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
Furtwängler, Maria</gnd:preferredNameForThePerson>
<rdaGr2:biographicalInformation xml:lang="de">
Dt. Schauspielerin, Sprecherin und Ärztin</rdaGr2:biographicalInformation>
<gnd:invalidIdentifier rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
(DE-588a)114984972</gnd:invalidIdentifier>
</rdf:Description>
</rdf:RDF>
  
```

Zweiter Prototyp: Beispieldatensatz

```
<rdf:Description rdf:about="http://d-nb.info/gnd/4003359-4">
  <skos:scopeNote>Für Übungen, die allgemein dem Wohlbefinden im phys. u. psych. Sinnen dienen sollen, sowie für psychotherapeutische Verfahren benutze Atemtherapie</skos:scopeNote>
  <skos:definition>benutzt für die med. Therapie u. Prophylaxe von Erkrankungen</skos:definition>
  <skos:related>Atemtherapie</skos:related>
  <skos:broader>Krankengymnastik</skos:broader>
  <skos:altLabel>Atemübung</skos:altLabel>
  <skos:altLabel>Atemtraining</skos:altLabel>
  <skos:altLabel>Atemschulung</skos:altLabel>
  <skos:prefLabel>Atemgymnastik</skos:prefLabel>
  <dcterms:identifier>(DE-588c)4003359-4</dcterms:identifier>
</rdf:Description>
```

Wo wollen wir damit hin?



Caption:

Aus: AG-Linked-Data. Quo vadis, Nationalbibliografie? 2009.

Vision der DNB

- Verlässlicher Anbieter und Produzent von relevanten Daten im Semantic Web
- Anknüpfungspunkt / Aggregationsknoten / "Enabler" für weitere Verknüpfungen / Datenanreicherungen
- Kein isolierter "Vollanbieter", sondern ein (wichtiger) Knoten von vielen im Netz
- Schwerpunkt unterstützende Dienste:
 - Automatische Indexierungsverfahren
 - Semantische Anreicherung/Verknüpfung von Daten
 - Langzeitarchivierung
 - Standards

Perspektiven der Verbände

- **Kooperation – gemeinsam in die Zukunft**
- Kooperationen und Vernetzung der Verbände werden gelebt und müssen zukünftig weiter ausgebaut werden
- Grundsätzlicher Softwarewandel bzw. -wechsel der Verbunddatenbanken steht mittelfristig an
- Zukunftsweisende Dienste müssen aufgebaut werden, dabei Fragen beantworten wie “make or buy”?
 Beispiele: LZA, Digitalisierung, Authentifizierung und Semantic Web

Aktivitäten des hbz: Perspektiven

- Kundenumfrage: Volldienstleister oder Einzeldienstleister?
- Überarbeitung des hbz-Produktportfolios
- Erstellung eines Perspektivpapiers für die Arbeiten des hbz in den nächsten fünf Jahren
- Auslaufen der Zielvereinbarungen zwischen Ministerium und Hochschulen zum 31.12.2010
- Neue regionale Aufgaben wie das Digitale Archiv NRW als Kooperationsprojekt zwischen Bibliotheken, Museen und Archiven geplant

Freigabe der Katalogdaten

- Notwendige Bedingung, um am Linked-Open-Data-Web mitzuweben
- Freigabe ist politisch-rechtlicher Schritt, der die Wichtigkeit gemeinfreier Daten verdeutlicht und die Datenfreigabe zum Diskussionsthema macht
 -> Wunsch zur Folge weiterer Bibliotheken und Verbände
- Pionierarbeit, um die Bibliothekskataloge ins Semantic Web zu migrieren

Aktivitäten des hbz: Linked Open Data

- Gemeinsame Erklärung von vier Kölner Bibliotheken und des LBZ zur Freigabe ihrer Titeldaten
- Regelmäßige Treffen einer offenen Linked-Open-Bibliographic-Data-Arbeitsgruppe. Themen:
 - Ontologien für bibliographische Daten im Semantic Web (RDA, Bibliographic Ontology, ...),
 - Rechtlichen Bedingungen (Freigabe der Daten mit einer Public-Domain-Lizenzierung)
 - Konversion der Daten nach RDF

Nächste Schritte I

- **Etablierung einer Open-Data-Praxis**
- Schaffung einer internationalen Arbeitsumgebung für
 - die Freigabe bibliographischer Daten,
 - ihre Beschreibung,
 - Verzeichnung,
 - Aktualisierung und
 - Versionierung.

Siehe hierzu <http://radar.oreilly.com/2010/03/truly-open-data.html>.

Nächste Schritte II

- **Migration der Daten ins Linked-Data-Web**
- Gemeinsame Entwicklung von Vokabularen (auf der Basis von RDA!?)
- Kollaborative Konversion bibliographischer Daten nach RDF
- Open Source: Teilen des Quellcodes von Konvertierungswerkzeugen

Nächste Schritte III

- **Linked-Data-Katalogisierung**
- Nicht nur Altdaten sollen ins Linked-Data-Web migriert werden
- Wir brauchen Webanwendungen zur direkten Katalogisierung von Linked Open Bibliographic Data
- Selber Open-Source-Lösungen entwickeln?
- Was machen die Bibliothekssystemanbieter?