

Seminar Gesellschaftliche Strukturen im digitalen Wandel

Vortrag Linked Data in der Universitätsbibliothek Leipzig

Franz Teichmann¹

¹Universität Leipzig

04.11.2014

1 Dateninfrastruktur der Universitätsbibliothek Leipzig

- Datentypen
- Ausgangszustand
- Problemstellung

2 praktische Lösung

- Begriffserklärungen
- Linked Data Prinzipien
- Datentransformationen und Integration
- neues System

3 Diskussion

- Bibliotheken seit jeher Quelle akademischen Wissens
- historisch gewachsene Informationssammlung

- Bibliotheken seit jeher Quelle akademischen Wissens
- historisch gewachsene Informationssammlung
 - stetige Wartung, Aufbereitung
 - dadurch sehr hohe Qualität
 - vollständig indizierte Sammlung verschiedener Medien
 - lokal verfügbar (Bestands- und Leihexemplare)
- langsame Erweiterung auf digitales Angebot

- Bibliotheken seit jeher Quelle akademischen Wissens
- historisch gewachsene Informationssammlung
 - stetige Wartung, Aufbereitung
 - dadurch sehr hohe Qualität
 - vollständig indizierte Sammlung verschiedener Medien
 - lokal verfügbar (Bestands- und Leihexemplare)
- langsame Erweiterung auf digitales Angebot
 - meist nur Online-Katalog
 - beschränkt auf einfache Suchen, indirekter Zugang
- notwendig: Ausbau der Informations-Dienstleitung mit neuen technischen Möglichkeiten

- Titeldaten
 - Titel
 - Autor
 - ISBN, PPN
- Exemplardaten
 - Barcode
 - Signatur
 - RFID
- Stammdaten
 - Standorte, Öffnungszeiten
- Normdaten
 - z.B. abweichende Schreibweisen

- Titeldaten im Format MARC21
 - Machine-Readable Cataloging
 - feldorientiertes Format, ISO 2709
 - digitale Repräsentation von Karteikarten durch Binärverkettung
 - Wegweisersymbole trennen Datenfelder und ermöglichen Navigation

- Titeldaten im Format MARC21
 - Machine-Readable Cataloging
 - feldorientiertes Format, ISO 2709
 - digitale Repräsentation von Karteikarten durch Binärverkettung
 - Wegweisersymbole trennen Datenfelder und ermöglichen Navigation
 - Standard im Datenaustausch zwischen Bibliotheken
 - in den 60er Jahren von der Library of Congress entwickelt

010 tag	Library of Congress Control Number (LCCN)
020 tag	(ISBN)
100 tag	Name des Autors
245 tag	Informationen zum Titel
260 tag	zusätzliche Informationen zur Publikation
300 tag	Beschreibung des Buches
650 tag	Fachgebiet

Tabelle: meistverwendete MARC21 Tags

- Exemplardaten
 - Daten eingefügt in Libero-Server-System
 - interne Datenhaltung nicht transparent
 - Funktionen unbekannt, Mitarbeiter entwickelten über Jahre Schnittstelle nach außen
 - Abfrage einzelner Datensätze über resource_id möglich
 - Indikator für vorhandene Exemplare in der ÜB Leipzig in Subfeld der Titeldaten

- Exemplardaten
 - Daten eingefügt in Libero-Server-System
 - interne Datenhaltung nicht transparent
 - Funktionen unbekannt, Mitarbeiter entwickelten über Jahre Schnittstelle nach außen
 - Abfrage einzelner Datensätze über resource_id möglich
 - Indikator für vorhandene Exemplare in der UB Leipzig in Subfeld der Titeldaten
- Stammdaten
 - von Mitarbeitern der UB Leipzig in XML getippt
 - beeindruckend fehlerfrei und problemlos in viele Formate konvertierbar
- Normdaten
 - von der deutschen Nationalbibliothek (GND) in RDF angeboten

- Finden eines "gemeinsamen Nenners"
- dynamisches Überführen aller Daten in einheitliches Format
 - Erweiterbarkeit
 - effiziente Suche erforderlich
- Konzeption einer nutzerfreundlichen Schnittstelle nach außen
- Tests und Rollout des Systems

- Beibehalten der grundlegenden Infrastruktur
- RDF als gewähltes einheitliches Format
- Linked Data als Technologie der Wahl
 - Daten konzeptionell kompatibel, auch zu anderen Quellen
 - (vorgegeben)
 - gewünschte Funktionalität
 - gute Referenzen
- Entwurf, Planung einer neuen Präsentationsschicht
- Tests und Rollout des Systems

- XML
 - Extensible Markup Language
 - hierarchisch strukturierte Daten in Textform
 - menschenlesbar, lizenzfrei, plattformunabhängig

- XML

- Extensible Markup Language
- hierarchisch strukturierte Daten in Textform
- menschenlesbar, lizenzfrei, plattformunabhängig

- RDF

- Resource Description Framework
- Datensätze in Form simpler Aussagen
- Subjekt-Prädikat-Objekt "Bob knows Fred"
- Verkettung zu Wissensgraphen, darstellbar in vielen Formaten (Serialisierungen), z.B. XML

- XML
 - Extensible Markup Language
 - hierarchisch strukturierte Daten in Textform
 - menschenlesbar, lizenzfrei, plattformunabhängig
- RDF
 - Resource Description Framework
 - Datensätze in Form simpler Aussagen
 - Subjekt-Prädikat-Objekt "Bob knows Fred"
 - Verkettung zu Wissensgraphen, darstellbar in vielen Formaten (Serialisierungen), z.B. XML
- URI
 - Uniform Resource Identifier: Zeichenkette zur eindeutigen Identifizierung von Ressourcen
 - Untermenge: Uniform Resource Locator (URL)
 - Benennung über primären Zugriffsmechanismus (http,ftp)

- XML

- <http://www.loc.gov/standards/marcxml/Sandburg/sandburg.xml>

- RDF

```
@prefix dbr: <http://dbpedia.org/resource/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema> .
@prefix dbp: <http://dbpedia.org/property/> .
dbr:Leipzig rdfs:label "Leipzig"@de .
dbr:Leipzig dbp:hasMayor dbr:Burkhard_Jung .
dbr:Leipzig dbp:locatedIn dbr:Saxony,
dbr:Germany.
```

- nach Tim Berners Lee, 2009:
<http://www.w3.org/DesignIssues/LinkedData.html>
- Verwendung von URIs als Bezeichner für alle Ressourcen
 - ermöglicht Unterscheidung und eindeutige Bestimmung von Entitäten
 - gilt auch für Prädikate

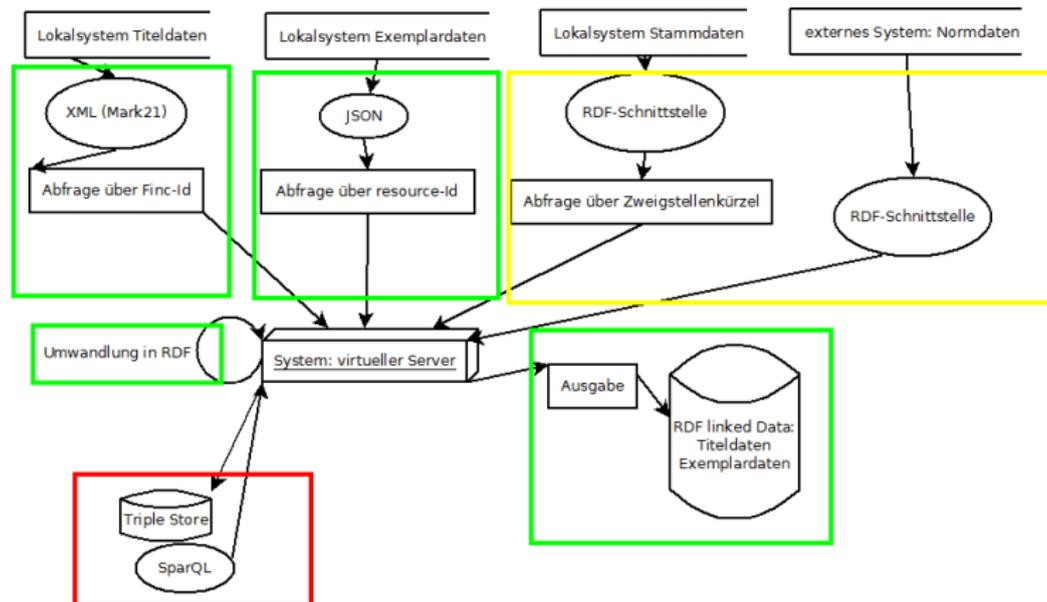
- nach Tim Berners Lee, 2009:
<http://www.w3.org/DesignIssues/LinkedData.html>
- Verwendung von URIs als Bezeichner für alle Ressourcen
 - ermöglicht Unterscheidung und eindeutige Bestimmung von Entitäten
 - gilt auch für Prädikate
- Verwendung von HTTP als Schema
 - die URI ist damit eine URL und kann in jedem Webbrowser eingegeben werden
 - Dereferenzierung
 - Unterscheidung zwischen Ressource selbst und Website über Ressource

- nach Tim Berners Lee, 2009:
<http://www.w3.org/DesignIssues/LinkedData.html>
- Verwendung von URIs als Bezeichner für alle Ressourcen
 - ermöglicht Unterscheidung und eindeutige Bestimmung von Entitäten
 - gilt auch für Prädikate
- Verwendung von HTTP als Schema
 - die URI ist damit eine URL und kann in jedem Webbrowser eingegeben werden
 - Dereferenzierung
 - Unterscheidung zwischen Ressource selbst und Website über Ressource
- Einhaltung gewisser Standards bei der Dereferenzierung von HTTP-URIs
 - gemeint ist Rückgabe von RDF/XML beim Nachschlagen einer Ressource
 - sinnvolle Informationen über die Ressource

- nach Tim Berners Lee, 2009:
<http://www.w3.org/DesignIssues/LinkedData.html>
- Verwendung von URIs als Bezeichner für alle Ressourcen
 - ermöglicht Unterscheidung und eindeutige Bestimmung von Entitäten
 - gilt auch für Prädikate
- Verwendung von HTTP als Schema
 - die URI ist damit eine URL und kann in jedem Webbrowser eingegeben werden
 - Dereferenzierung
 - Unterscheidung zwischen Ressource selbst und Website über Ressource
- Einhaltung gewisser Standards bei der Dereferenzierung von HTTP-URIs
 - gemeint ist Rückgabe von RDF/XML beim Nachschlagen einer Ressource
 - sinnvolle Informationen über die Ressource
- Darstellung von Zusammenhängen zwischen Ressourcen
 - sinnvoller Grad der Vernetzung innerhalb der Wissensbasis
 - Vernetzung mit anderen Wissensbasen z.B. mit owl:sameAs

- Titeldaten
 - extrahieren des ursprünglichen MARC21-Blob aus der Datenbank
 - parsing nach XML mithilfe von regulären Ausdrücken (Leitfaden GND)
 - Transformation nach RDF mit Script der Library of Congress
- Exemplardaten
 - Abfrage und Transformation (kompatible Formate)
 - Verlinkung mit Titeldaten

Stand der Umsetzung



- Würden Sie das neue Angebot der UB Leipzig nutzen?
- Wie schätzen Sie die praktische Relevanz der Funktionalitätserweiterung ein?
- Welche Funktionalität fehlt bisher?

- Bibliothek der Moderne und modernes Urheberrecht
 - Ansichtsexemplare
 - digitale Leseräume mit Zugangsbeschränkung
 - Unterscheidung zwischen digitaler und physischer Präsenz in der Bibliothek