

# Kolloquium Wissensinfrastruktur

25. Januar 2013

## Linked Data für Bibliotheken

Cord Wiljes

# Überblick

Motivation

Linked Data

Bibliographische Vokabulare für Linked Data

schema.org

Aktueller Stand + Ausblick

# Motivation

The screenshot shows a web browser window titled "UB Bielefeld: Bibliothekskatalog - Mozilla Firefox". The address bar contains the URL "katalog.ub.uni-bielefeld.de/cgi-bin/search.cgi?sprache=GER&art=f". The browser's toolbar includes various icons and a search engine dropdown set to "latex kursiv".

The main content area features the logo of the "Universität Bielefeld" and the "Universitätsbibliothek" with the tagline "INFORMATION plus!". Below this is a green header with the text "Bibliothekskatalog" and a link "Zur Startseite der Bibliothek".

The search interface includes several tabs: "Einfache Suche", "Erweiterte Suche" (selected), "Export", "Systematik", "Ausleihkonto", "Fernleihe", "Hilfe", and "Beenden".

The search form contains the following elements:

- A "Freitext" search input field.
- Fields for "Titel (-stichworte)" and "Autor", each with an "Index" button.
- Buttons for "Suchvorschläge", "Suchen", and "Felder leeren".
- A checkbox for "Treffer nach Erscheinungsjahr sortieren".
- An "Optionen" section with checkboxes for "nur Zeitschriftentitel, Zeitungen", "nur Online-Publikationen", and "Verfügbarkeit anzeigen" (checked).
- A "Treffer pro Seite" dropdown set to "10" and a "Sortierung" dropdown.
- A link for "Freitext: Tipps zur Suche".

# Anfrage

Gib mir alle Romane von Nobelpreisträgerinnen!

- ▶ Bibliothekskatalog
- ▶ Wikipedia

# Anwendungsfall aus der Chemie

## Glasübergangstemperaturen von organischen Aerosolen



Prof. Thomas Koop

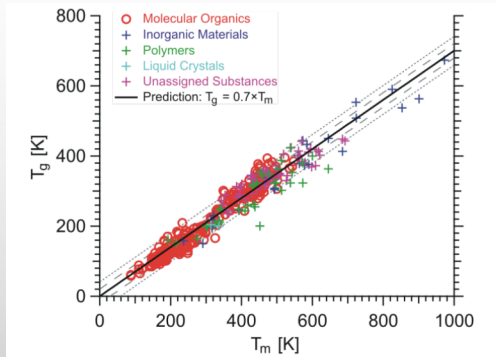


# Aufgabe

## **Aufgabe: Finde alle Glasübergangstemperaturen von organischen Verbindungen!**

- ▶ Es existiert keine zentrale Datenbank, die die gesuchten Informationen enthält.
- ▶ Die Daten sind über zahlreiche Publikationen verstreut.
- ▶ Quellen mit unterschiedlichem Grad an Zuverlässigkeit
- ▶ Zeitaufwand für die Literaturrecherche: > 100 h

# Ergebnis



Beziehung von Glasübergangstemperatur  $T_g$  und Schmelzpunkt  $T_m$

# Was tun?

## Lösungsvorschläge:

- ▶ Volltextsuche?
- ▶ eine gigantische Datenbank?
- ▶ semantische Analyse von natürlicher Sprache?
- ▶ Linked Data!

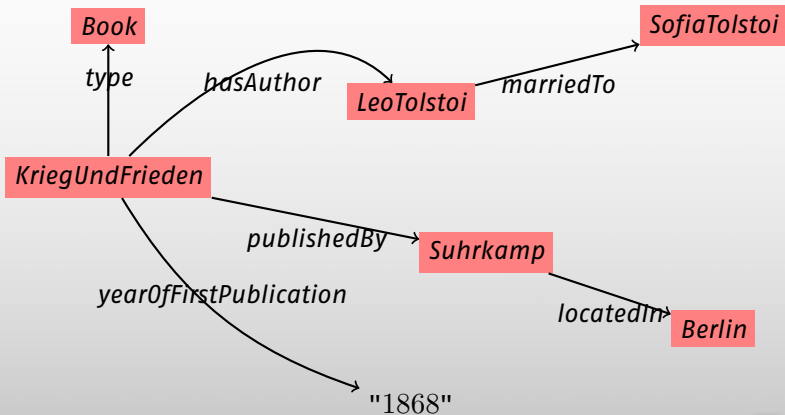


# Linked Data

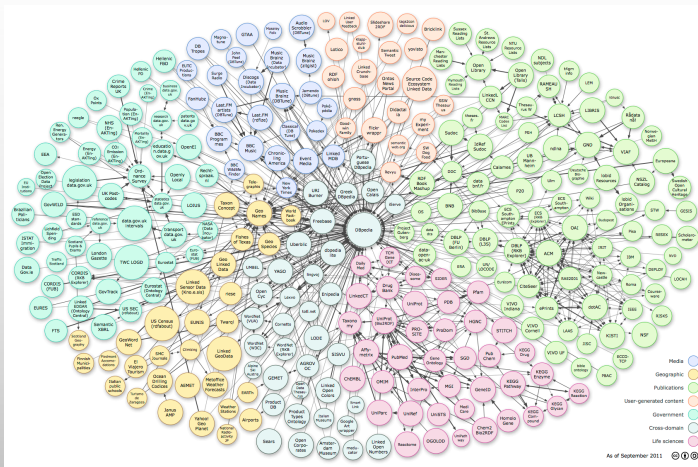


Tim Berners-Lee

# Wie funktioniert Linked Data?



# Linked Data Cloud



# RDF

Subjekt Prädikat Objekt .

KriegUndFrieden hatAutor LeoTolstoi .

```
<http://dbpedia.org/resource/Krieg_und_Frieden>  
  <http://purl.org/dc/elements/1.1/creator>  
    <http://dbpedia.org/resource/Leo_Tolstoi> .
```

```
dbpedia:Krieg_und_Frieden dc:creator dbpedia:Leo_Tolstoi .
```

# Linked Data

- ▶ Datenmodell: RDF = gerichteter Graph
- ▶ "*Anyone can say anything about anything.*"
- ▶ Vom "*Web of Documents*" zum "*Web of Data*"

# RDF

## Serialisierungen

- ▶ stand-alone:
  - ▶ N-Triples, Turtle
  - ▶ RDF/XML
  - ▶ JSON-LD
- ▶ in HTML
  - ▶ Microformats
  - ▶ RDFa
  - ▶ Microdata

# Auflösung

## Gib mir alle Bücher von Nobelpreisträgerinnen!

```
PREFIX yago: <http://dbpedia.org/class/yago/>  
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
PREFIX prop: <http://dbpedia.org/property/>  
PREFIX dbo: <http://dbpedia.org/ontology/>  
PREFIX res: <http://dbpedia.org/resource/>  
SELECT DISTINCT ?name ?booktitle  
WHERE {  
    ?uri rdf:type yago:WomenNobellaureates .  
    ?uri prop:name ?name .  
    ?book dbo:author ?uri .  
    ?book prop:name ?booktitle .  
}
```

# Linked Data: Nutzen

- ▶ Datenintegration
- ▶ Verknüpfung mit externen Datenquellen
- ▶ Suchmaschinenoptimierung



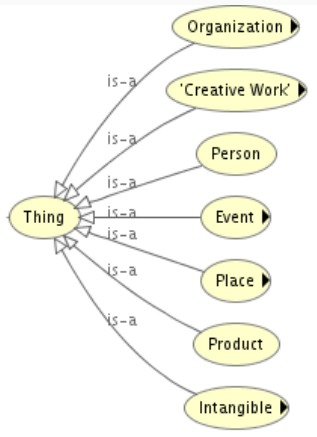
# Bibliographische Vokabulare für Linked Data

- ▶ **Dublin Core:** metadata elements for cataloging library items
- ▶ **BIBO** (Bibliographic Ontology)
- ▶ **SKOS** (Simple Knowledge Organizational System)
- ▶ **MADS/RDF** (Metadata Authority Description Schema in RDF)
- ▶ **FRBR** (Functional Requirements for Bibliographic Records): bibliothekswissenschaftliches Datenmodell für bibliographische Metadaten.

# schema.org

- ▶ gestartet am 2. Juni 2011
- ▶ gemeinsames Vokabular
- ▶ Google, Bing, Yahoo! und Yandex
- ▶ Microdata-Serialisierung

## schema.org



# schema.org

## Thing > CreativeWork

Article Blog, Book, Comment, ItemList, Map, MediaObject, Movie, MusicPlaylist, MusicRecording, Painting, Photograph, Recipe, Review, Sculpture, SoftwareApplication, TVEpisode, TVSeason, TVSeries, WebPage, WebPageElement,

# schema.org

## Thing > CreativeWork

about, accountablePerson, aggregateRating, alternativeHeadline, associatedMedia, audio, author, award, awards, comment, contentLocation, contentRating, contributor, copyrightHolder, copyrightYear, creator, dateCreated, dateModified, datePublished, discussionUrl, editor, encoding, encodings, genre, headline, inLanguage, interactionCount, isFamilyFriendly, keywords, mentions, offers, provider, publisher, publishingPrinciples, review, reviews, sourceOrganization, text, thumbnailUrl, version, video

## Thing > CreativeWork > Book

bookEdition, bookFormat, illustrator, isbn, numberOfPages

## Thing > CreativeWork > Article > ScholarlyArticle

articleBody, articleSection, wordCount

# Microdata

```
<div>The Catcher in the Rye – Mass Market Paperback</div>  
<br>by <a href="/author/jd_salinger.html">J.D. Salinger</a>  
<br>Product details  
<br>224 pages  
<br>Publisher: Little, Brown, and Company – May 1, 1991  
<br>Language: English  
<br>ISBN–10: 0316769487
```

# Microdata

```
<div itemscope itemtype="http://schema.org/Book">
  <span itemprop="name">The Catcher in the Rye</span> -
  <link itemprop="bookFormat"> href="http://schema.org/Paperback"
  by <a itemprop="author"> itemscope itemtype="http://schema.org/Person"
  <link itemprop="http://purl.org/dc/elements/1.1/creator" href="http://
  Product details
  <span itemprop="numberOfPages">224</span> pages
  Publisher: <span itemprop="publisher">Little, Brown, and Company</span>
  <meta itemprop="datePublished" content="1991-05-01">May 1, 1991
  Language: <span itemprop="inLanguage">English</span>
  ISBN -10: <span itemprop="isbn">0316769487</span>
</div>
```

# Microdata to Turtle

```
@prefix md: <http://www.w3.org/ns/md#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix schema: <http://schema.org/> .  
@prefix dct: <http://purl.org/dc/elements/1.1/> .
```

```
<> md:item ( [ a schema:Book;  
              schema:author [ a schema:Person;  
                              schema:name "J.D. Salinger"  
              dct:creator <http://dbpedia.org/resource/J._...  
              schema:bookFormat schema:Paperback;  
              schema:datePublished "1991-05-01";  
              schema:inLanguage "English";  
              schema:isbn "0316769487";  
              schema:name "The Catcher in the Rye";  
              schema:numberOfPages "224";  
              schema:publisher "Little, Brown, and Company
```



# Projekte

- ▶ Library Linked Data Incubator Group des W3C
- ▶ OpenAIREplus (Open Access Infrastructure for Research in Europe)
- ▶ Linked Data Service der UB Mannheim
- ▶ Hochschulbibliothek der RWTH Aachen
- ▶ Europeana
- ▶ Linked Data Initiative der Library of Congress
- ▶ British Library

# Herausforderungen

- ▶ Standards ✓ RDF, SPARQL, RDFS, OWL, ...
- ▶ Anwendungen
- ▶ gemeinsame Vokabulare
- ▶ Kulturwandel: Open Science
- ▶ neue Arbeitsabläufe
- ▶ Vertrauenswürdigkeit der Daten

# Rolle der Bibliotheken

- ▶ Aufbau und Betrieb einer Infrastruktur für Forschungsdaten
- ▶ Experten für die Organisation großer Datenmengen (Taxonomien, Vokabulare,...)
- ▶ Lokaler Service für Wissenschaftler
- ▶ Garanten für qualitativ hochwertige und inhaltlich vertrauenswürdige Daten: Bibliotheken als Data Hubs im Web of Linked Data

# Linked Data in Bielefeld

- ▶ CITEC Open Research / Toolkit
- ▶ OpenAIREplus FP7 Pilotprojekt: Linking Data and Publications
- ▶ CMS + Blogs

## nächste Schritte

- ▶ Anreicherung der PUB-Daten mit schema.org-Metadaten
- ▶ Umsetzung ausgewählter Referenzprojekte in Linked Data