UNIVERSITÄT BERN

Vorlesung Open Data:

Linked Data und Semantic Web

Termin 11, 10. Mai 2017

Dr. Matthias Stürmer und Prof. Dr. Thomas Myrach

Forschungsstelle Digitale Nachhaltigkeit
Institut für Wirtschaftsinformatik
Universität Bern

Terminübersicht Vorlesung

1. 23.02.2017: Informationen zur Vorlesung und Einführung OD und OG
2. 02.03.2017: Open Government Data und Impact Measuring Framework
3. 09.03.2017: Digitale Nachhaltigkeit
4. 16.03.2017: Entstehung und Anwendung des Öffentlichkeitsgesetz
5. 23.03.2017: Geistiges Eigentum und die Open-Bewegung
6. 30.03.2017: Open Data Journalism
7. 06.04.2017: Open Finance and Participatory Budgeting
8. 13.04.2017: Aid Transparency und Open Aid

   20.04.2017: Osterferien
9. 27.04.2017: Open Geodata
10. 04.05.2017: Open Procurement
11. 11.05.2017: Linked Data und Semantic Web
12. 18.05.2017: Open Corporate Data

    25.05.2017: Auffahrt
13. **01.06.2017: Abschlusspräsentationen Open Data Apps**

19. Mai 2016

Vorlesung

1. Linked Data und Semantic Web
2. Gastreferat von Adrian Gschwend, Zazuko GmbH: Technische Aspekte von Linked Data

Übung

1. Informationen zu den Abschlusspräsentationen
2. Programming Coaching

Agenda

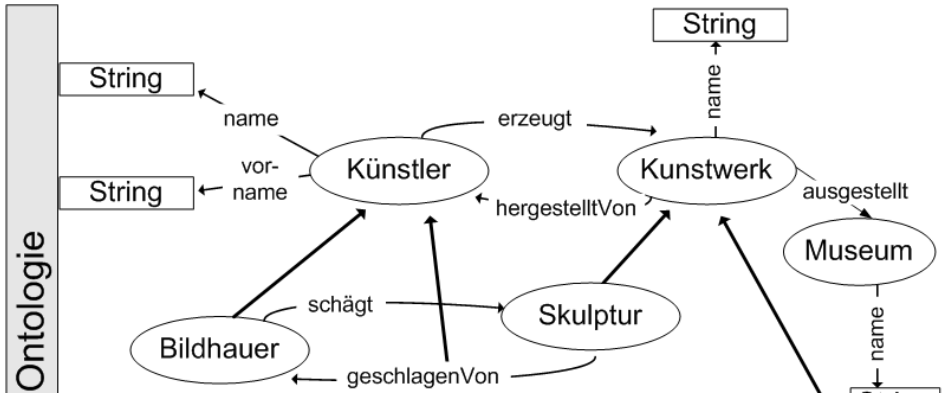
1. **Einführung Linked Data**
2. Beispiele von Linked Data Anwendungen

TED Referat von Tim Berners-Lee, 2009

Quelle: http://www.ted.com/talks/tim_berners_lee_on_the_next_web?language=de

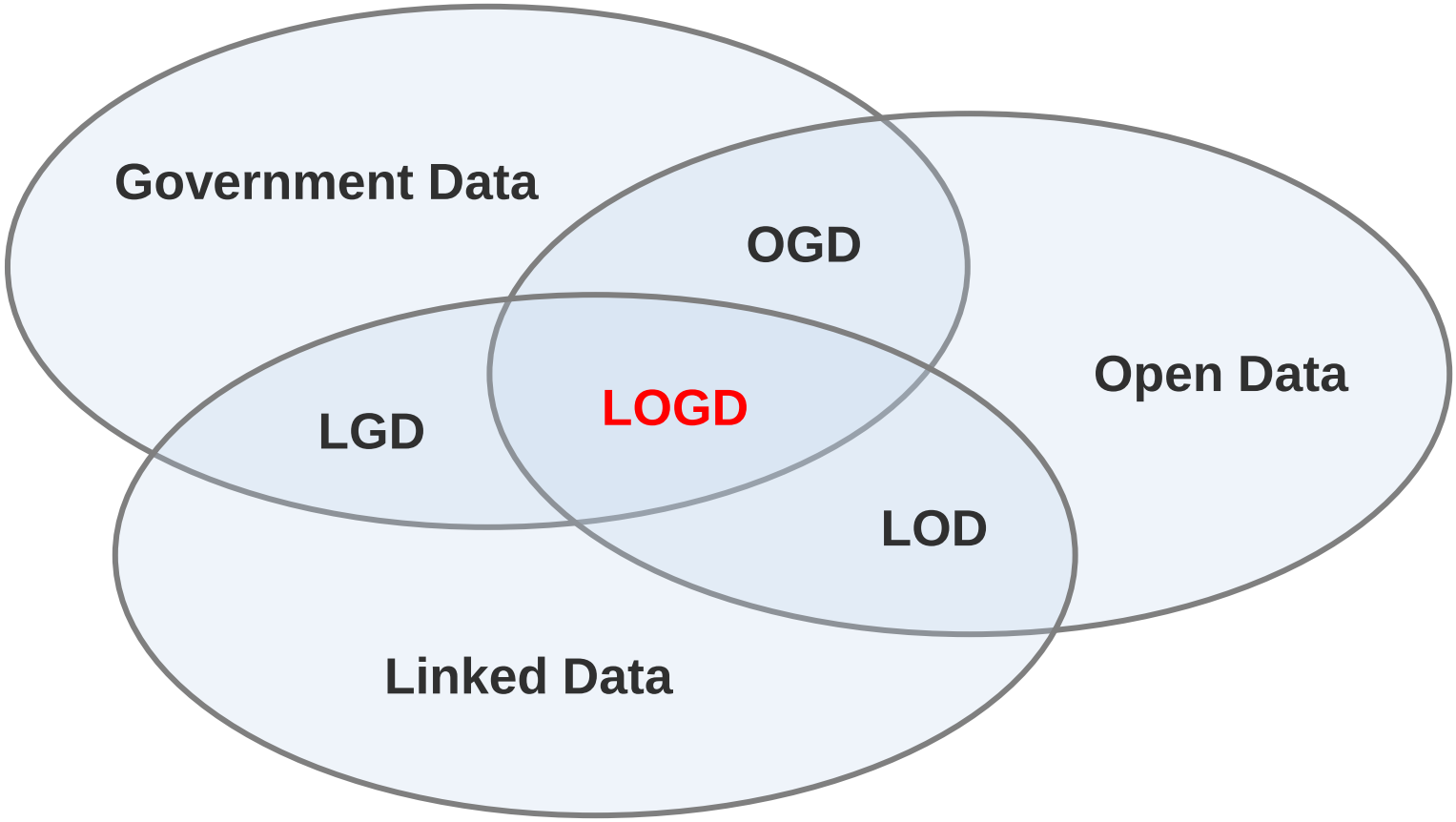
Begrifflichkeiten

- **Syntax:** Zeichenfolgen (bspw. Orthographie und Grammatik)
- **Semantik:** inhaltliche Bedeutung (bspw. Aussage eines Textes)
- **Semantic Web:** maschinenlesbare Verknüpfungen von Inhalten
- **Metadaten:** maschinenlesbare Informationen über Daten
- **Linked Data:** Daten, die zueinander in Beziehung stehen
- **Ontologie:** logische Beziehungen von Begriffen

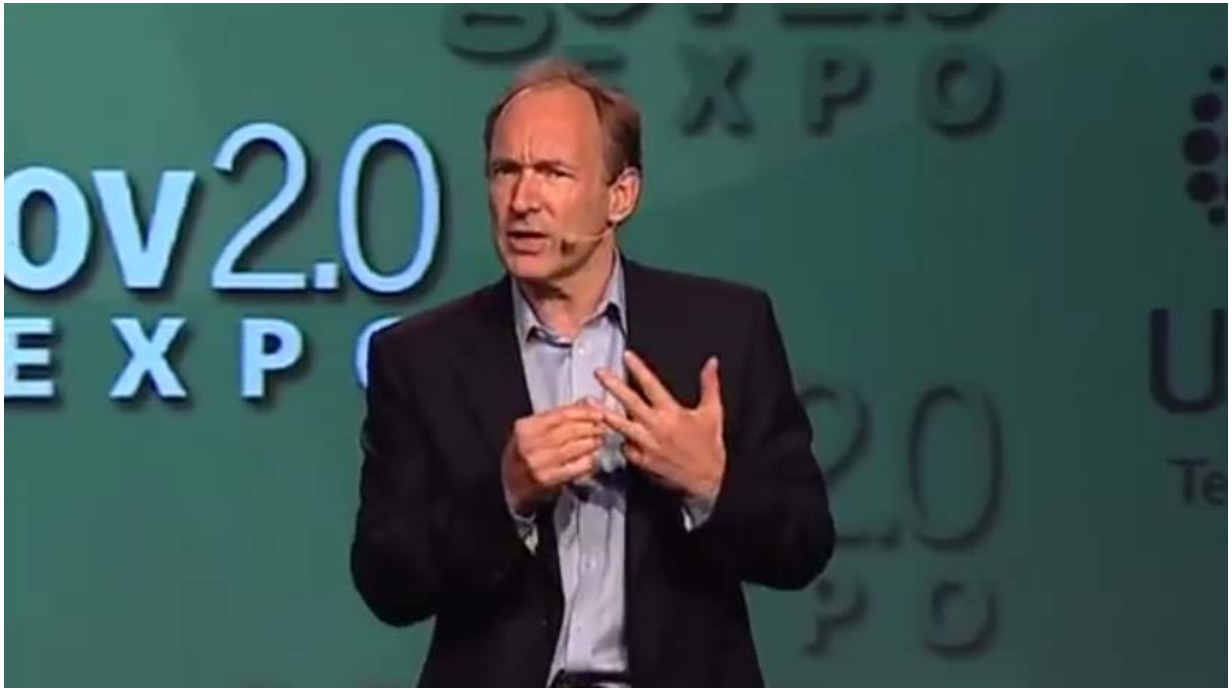


Quelle: http://de.wikipedia.org/wiki/Ontologie_(Informatik)

Linked Open Government Data



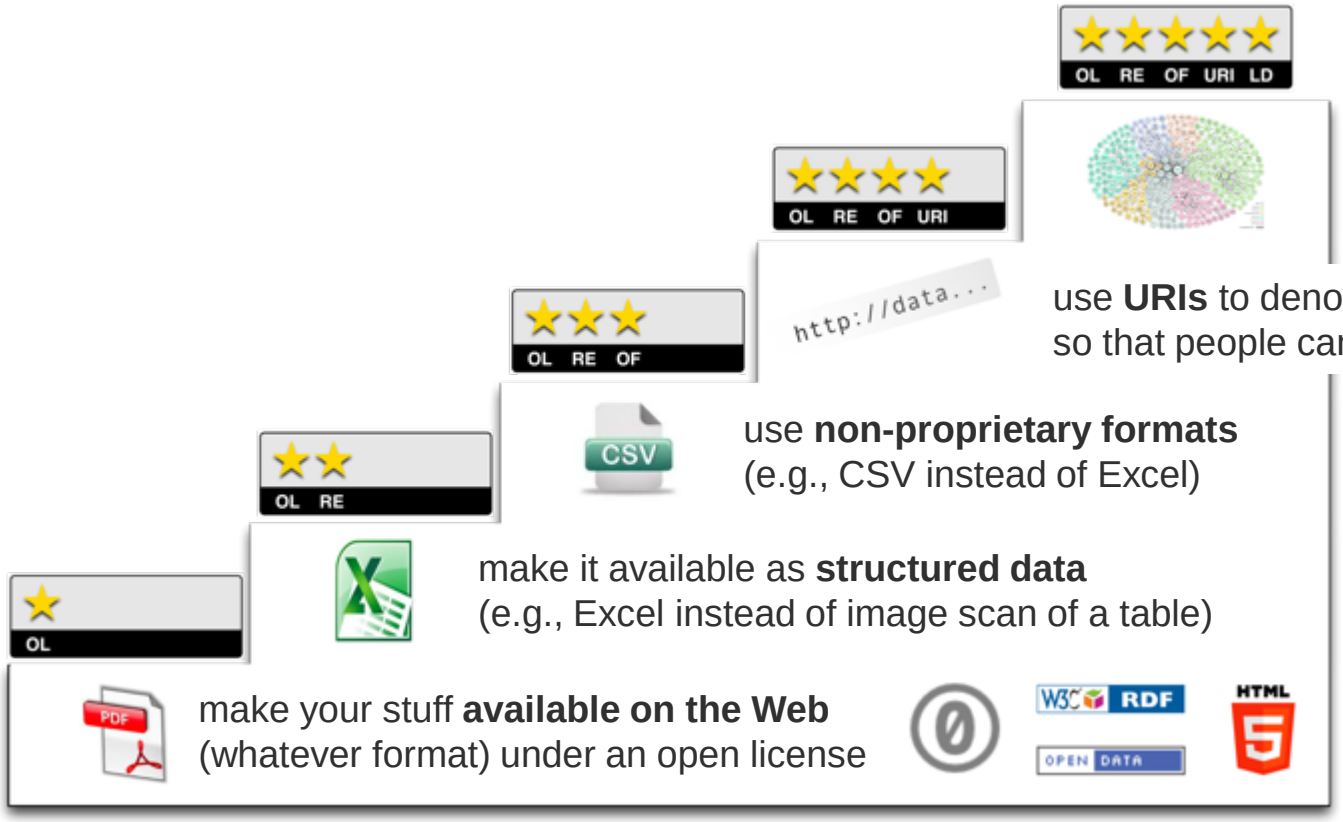
5 Star Open Data



Gov 2.0 Expo 2010: Tim Berners-Lee,
"Open, Linked Data for a Global Community"

Quelle: http://www.youtube.com/watch?v=ga1aSJXCFe0

5 Star Open Data



link your data to other data to provide context

use **URIs** to denote things, so that people can point at your stuff

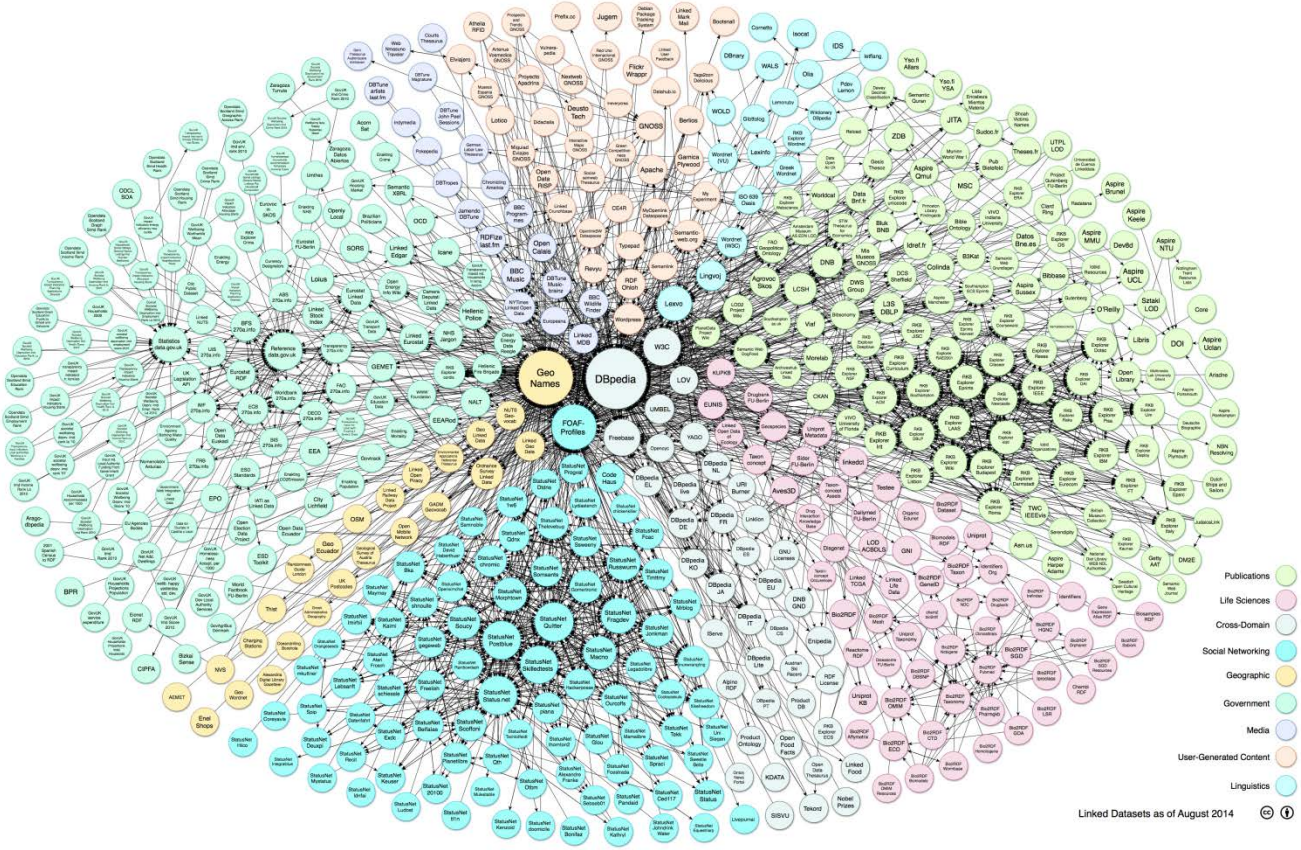
use **non-proprietary formats** (e.g., CSV instead of Excel)

make it available as **structured data** (e.g., Excel instead of image scan of a table)

make your stuff **available on the Web** (whatever format) under an open license

Quelle: http://5stardata.info

Linked Open Data Cloud Diagram



Source: "Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net

Linked Open Data Cloud Diagram

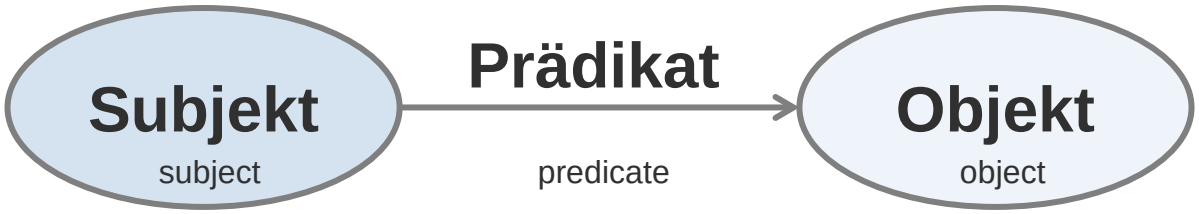
Amount of datasets per topic:

Topic	Datasets	%
Government	183	18.05%
Publications	96	9.47%
Life sciences	83	8.19%
User-generated content	48	4.73%
Cross-domain	41	4.04%
Media	22	2.17%
Geographic	21	2.07%
Social web	520	51.28%
Total	1014	

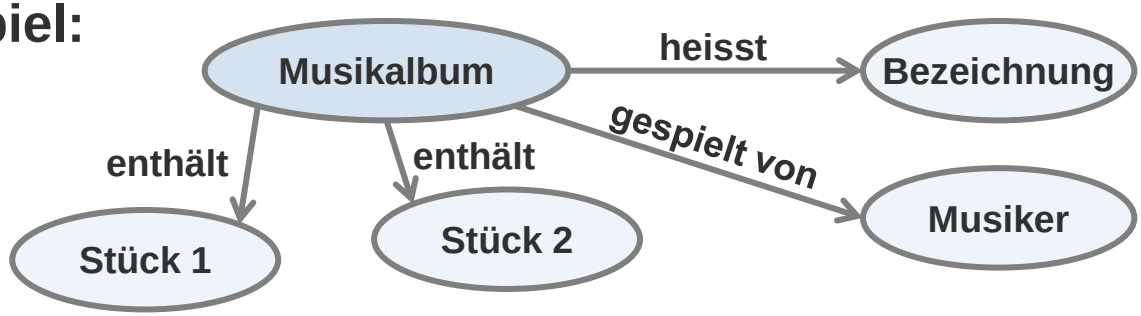
Source: "Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net

Resource Description Framework RDF

- Offener **World Wide Web Consortium W3C** Standard
- Formale Definition eines **gerichteten Graphen**
- Grundsätzlich aufgebaut als **Tripel:**



- **Beispiel:**



Quelle: http://www.devx.com/semantic/Article/34816

Resource Description Framework RDF

```
<rdf:Description rdf:about="subject">
    <predicate rdf:resource="object" />
    <predicate>literal value</predicate>
<rdf:Description>
```

```
<?xml version="1.0" encoding="UTF-8"?>

<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:region="http://www.country-regions.fake/">

    <rdf:Description rdf:about="http://en.wikipedia.org/wiki/Oxford">
        <dc:title>Oxford</dc:title>
        <dc:coverage>Oxfordshire</dc:coverage>
        <dc:publisher>Wikipedia</dc:publisher>
        <region:population>10000</region:population>
        <region:principaltown rdf:resource="http://www.country-regions.fake/oxford"/>
    </rdf:Description>

</rdf:RDF>
```

Quelle: http://www.linkeddatatools.com/introducing-rdf-part-2

Dublin Core

- Dublin Core Metadata Initiative (DCMI), gegründet 1994
- Ziel ist einheitliche Beschreibung von Metadaten
- Suchmaschinen sollen Dokumente besser auffinden können
- Beispiele:
  - ***title:*** Titel des Dokuments
  - ***subject:*** Thema des Inhalts
  - ***coverage:*** Eingrenzung inhaltlich abgedeckte Bereich
  - ***description:*** Kurzzusammenfassung des Inhalts des Dokuments
  - ***publisher:*** Name der veröffentlichenden Instanz
- Anwendung:
  - `<dc:title>Oxford</dc:title>`
  - `<dc:publisher>Wikipedia</dc:publisher>`

Quelle: https://de.wikipedia.org/wiki/Dublin_Core

Schema.org

schema.org

UNIVERSITÄT BERN

- Gegründet 2011 von Bing, Google und Yahoo!
- Grosse Sammlung von Metadaten Schemas um Web-Inhalte einheitlich von Suchmaschinen auffindbar zu machen
- Zur Zeit 571 Types und 831 Properties
- Beispiel **«City»** https://schema.org/City
  - ***address:*** Physical address of the item
  - ***geo:*** The geo coordinates of the place
  - ***photo:*** A photograph of this place
  - ***containedInPlace:*** The basic containment relation between a place and one that contains it.

City

Thing > Place > AdministrativeArea > City

A city or town.

Usage: Between 1000 and 10,000 domains

- Weitere Types: **Product, Offer, Vehicle, Organisation, Event etc.**

Quelle: http://seopressor.com/blog/dublin-core-vs-schemaorg-metadata-comparison/

DUBLIN CORE	VS	SCHEMA.ORG
1995	Year Founded	2011
DCMI, OCLC & NCSA	Founders	Bing, Google & Yahoo
DCMI	Maintained By	Open Community
Physical & Web Resources	Use For	Only Web Resources
Less SEO Friendly	SEO Friendly	Highly SEO Friendly
Abstract & Broad	Data Type	Highly Specific
Wide	Compatibility	Only With Web Resources
Here To Stay	Future Prospect	Rapid Growth
Despite its slowing growth, it is far from being obsolete. If compatibility is vital, DC is what you want to cover.	Verdict	Arguably the future of meta data. If online presence is what you want, Schema is where you want to head for.

Quelle: http://seopressor.com/blog/dublin-core-vs-schemaorg-metadata-comparison/

JSON Linked Data Serialisierung

UNIVERSITÄT BERN

JSON for Linking Data

Data is messy and disconnected. JSON-LD organizes and connects it, creating a better Web.

Linked Data

Linked Data empowers people that publish and use information on the Web. It is a way to create a network of standards-based, machine-readable data across Web sites. It allows an application to start at one piece of Linked Data, and follow embedded links to other pieces of Linked Data that are hosted on different sites across the Web.

A Simple Example

```
{
  "@context": "http://json-ld.org/contexts/person.jsonld",
  "@id": "http://dbpedia.org/resource/John_Lennon",
  "name": "John Lennon",
  "born": "1940-10-09",
  "spouse": "http://dbpedia.org/resource/Cynthia_Lennon"
}
```

JSON-LD

JSON-LD is a lightweight Linked Data format. It is easy for humans to read and write. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. JSON-LD is an ideal data format for programming environments, REST Web services, and unstructured databases such as CouchDB and MongoDB.

Quelle: http://json-ld.org

Agenda

1. Einführung Linked Data
2. **Beispiele von Linked Data Anwendungen**

DBpedia

DBpedia: Anmeldung

Änderungen | Kommentare | Index | weitere ▾ | Suche

About

Note We are moving the content of this website to our new page currently located here, we will switch within the next days (written 29th of Aprli)

DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link the different data sets on the Web to Wikipedia data. We hope that this work will make it easier for the huge amount of information in Wikipedia to be used in some new interesting ways. Furthermore, it might inspire new mechanisms for navigating, linking, and improving the encyclopedia itself.

Wiki Contents

This Wiki provides information about the DBpedia community project:

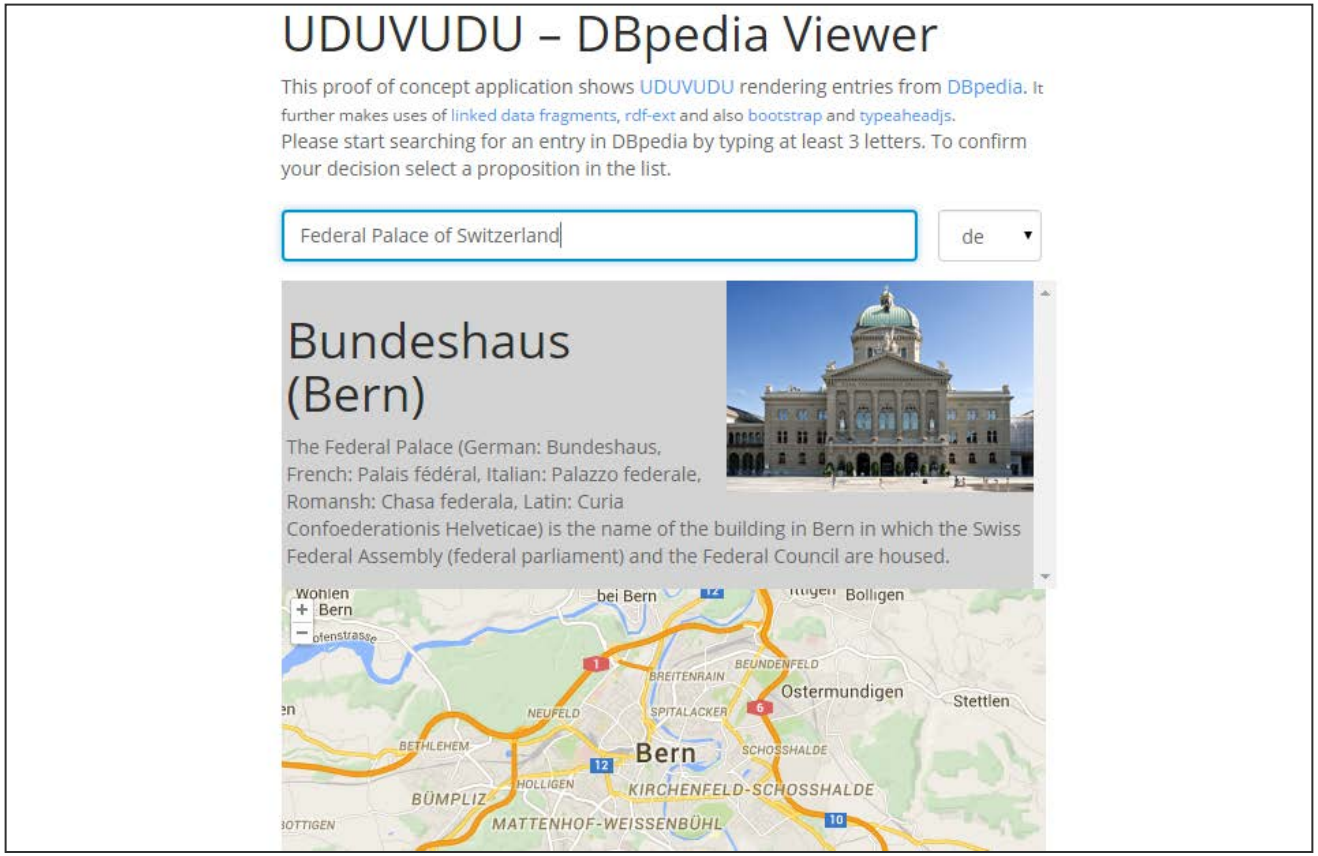
- Datasets gives an overview about the DBpedia knowledge base.
- Ontology gives an overview about the DBpedia ontology.
- Online Access describes how the data set can be accessed via a SPARQL endpoint and as Linked Data.
- Downloads provides the DBpedia data sets for download.
- Interlinking describes how the DBpedia data set is interlinked with various other datasets on the Web.
- Use Cases lists different use cases for the DBpedia data set.
- Extraction Framework describes the DBpedia information extraction framework.
- Data Provision Architecture paints a picture of the software and protocols used to serve DBpedia on the Web.
- Community explains how the DBpedia community collaborates and how people can contribute to the DBpedia effort.
- DBpedia Mapping Wiki containing the mappings used by the DBpedia extraction.
- DBpedia Internationalization Effort working towards providing multiple language-specific versions of DBpedia.
- DBpedia-Live presents the new DBpedia-Live framework.
- DBpedia Spotlight presents the DBpedia Spotlight tool for the semantic annotation of textual content.
- Credits lists the people and institutions that have contributed to DBpedia so far.
- Change Log lists the DBpedia releases and gives an overview about the changes for earch release.
- Next steps describes ideas and future plans for the DBpedia project.

The DBpedia Knowledge Base

Knowledge bases are playing an increasingly important role in enhancing the intelligence of Web and enterprise search and in supporting information integration. Today, most knowledge bases cover only specific domains, are created by relatively small groups of knowledge engineers, and are very cost intensive to keep up-to-date as domains change. At the same time, Wikipedia has grown into one of the central knowledge sources of mankind, maintained by thousands of contributors.

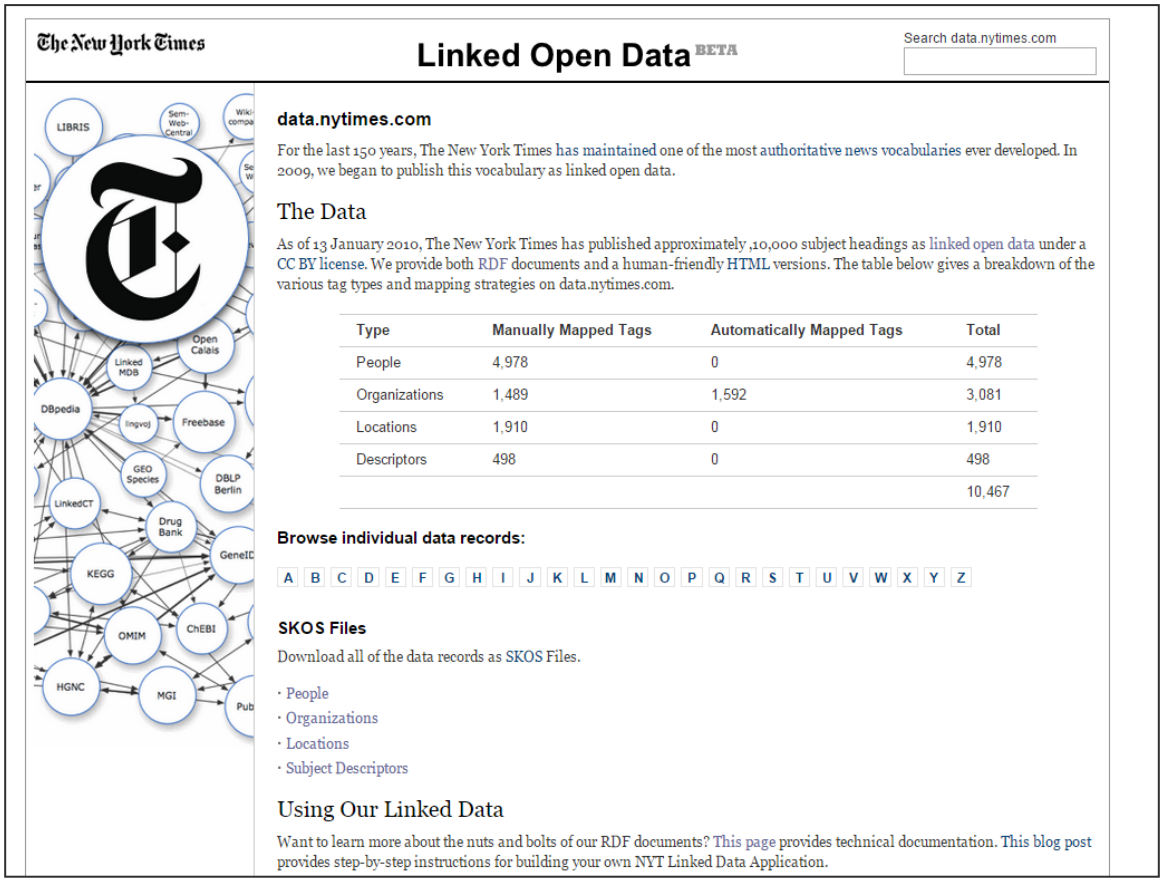
Quelle: http://wiki.dbpedia.org

Inhalte von DBpedia anzeigen



Quelle: http://dbpedia.exascale.info/

Linked Open Data von New York Times



The New York Times

Linked Open Data BETA

Search data.nytimes.com

data.nytimes.com

For the last 150 years, The New York Times has maintained one of the most authoritative news vocabularies ever developed. In 2009, we began to publish this vocabulary as linked open data.

The Data

As of 13 January 2010, The New York Times has published approximately ,10,000 subject headings as linked open data under a CC BY license. We provide both RDF documents and a human-friendly HTML versions. The table below gives a breakdown of the various tag types and mapping strategies on data.nytimes.com.

Type	Manually Mapped Tags	Automatically Mapped Tags	Total
People	4,978	0	4,978
Organizations	1,489	1,592	3,081
Locations	1,910	0	1,910
Descriptors	498	0	498
			10,467

Browse individual data records:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

SKOS Files

Download all of the data records as SKOS Files.

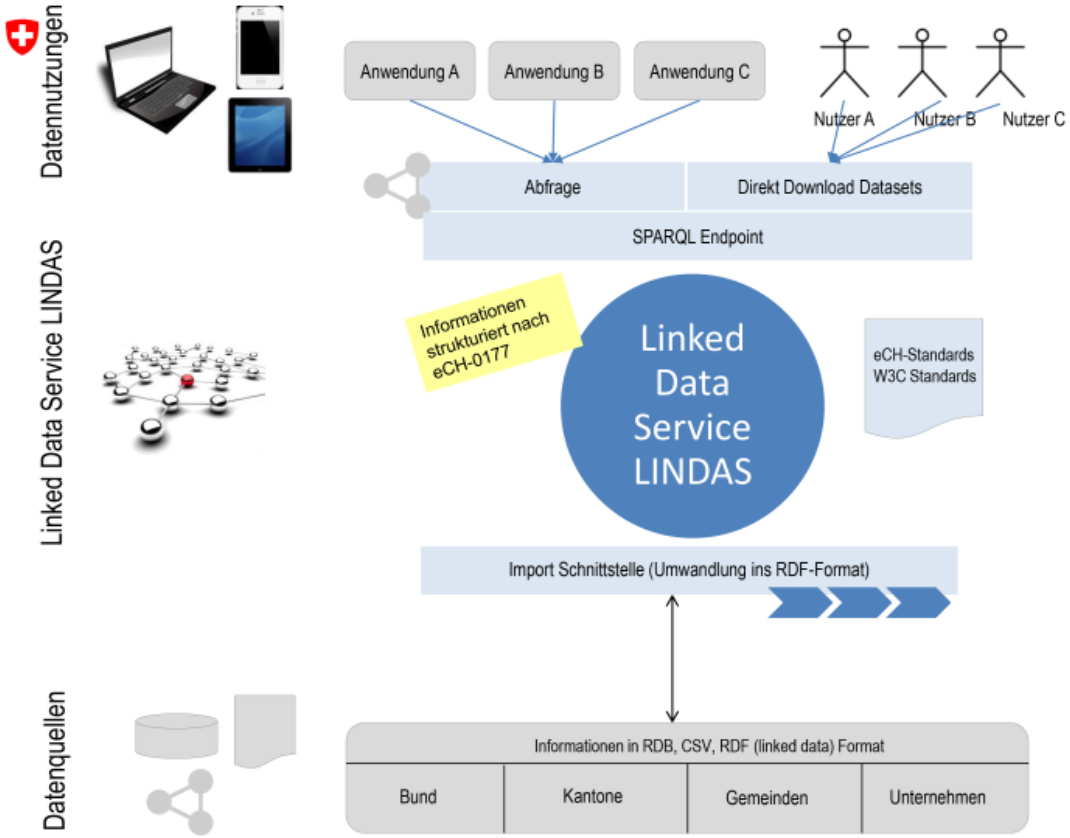
- People
- Organizations
- Locations
- Subject Descriptors

Using Our Linked Data

Want to learn more about the nuts and bolts of our RDF documents? This page provides technical documentation. This blog post provides step-by-step instructions for building your own NYT Linked Data Application.

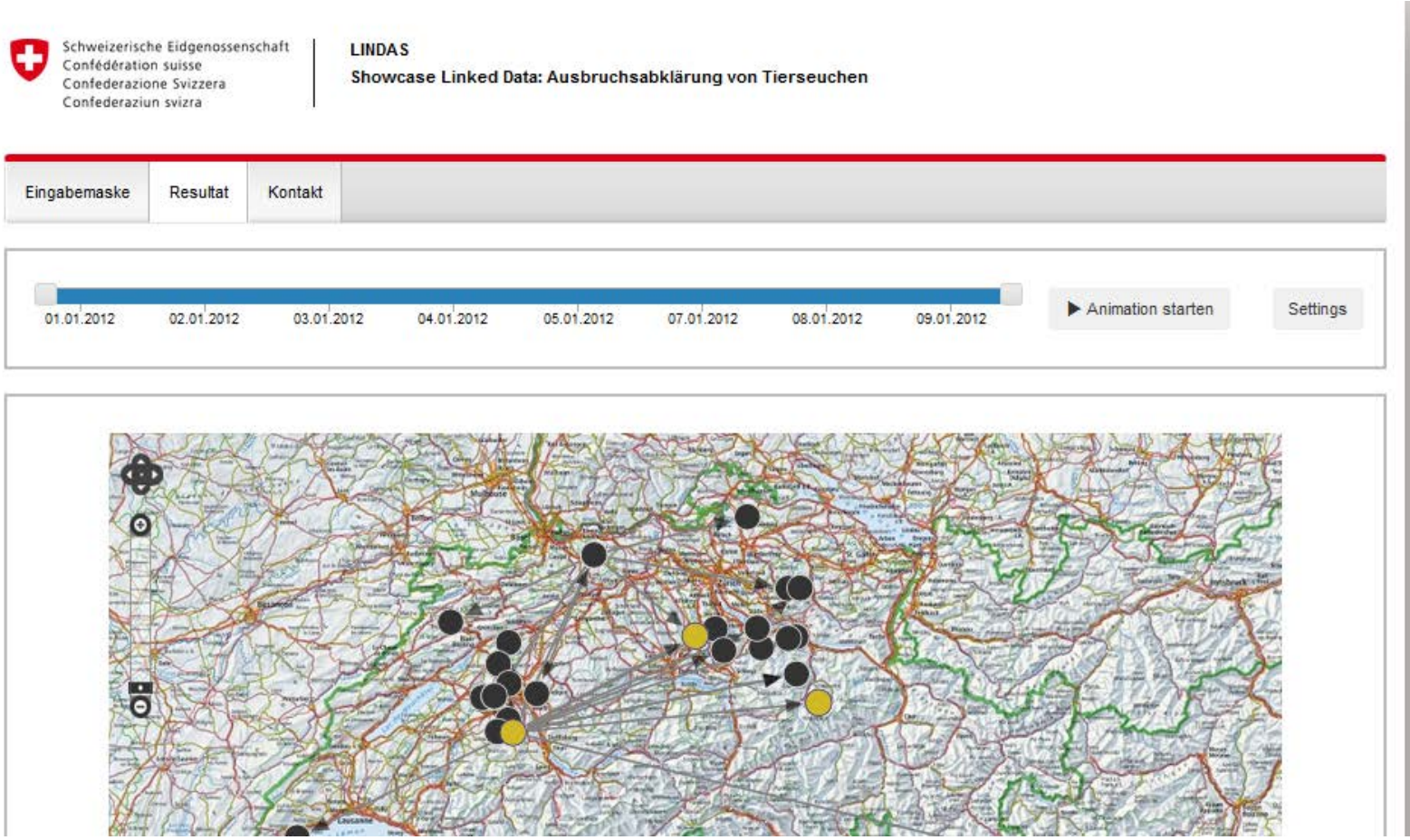
Quelle: http://data.nytimes.com

LINDAS vom Seco



Quelle: http://lindas-data.ch

Use Case Seuchenausbreitung



Quelle: http://outbreak-tool2.fdn.iwi.unibe.ch