

## An Introduction to Linked Open Data

Felix.Ostrowski@googlemail.com (@literarymachine)
Adrian Pohl@hbz-nrw.de (@acka47)

SWIB 2014 Pre-Conference Workshop Monday, December 1st 2014 Bonn

#### Schedule

- Organize in teams
- Introduction: Data Graphs Triples
- → Groupwork
- URIs and Namespaces
- → Groupwork
- Open Data Principles
- → Groupwork
- Identification vs. Description
- Groupwork
- Triple Stores & SPARQL
- Groupwork
- RDF Schema
- Groupwork
- → Summary, Questions & Discussion



## Linked Open Data

- → It's about data ...
- ... more precisely: about open data ...
- ... even more precisely: about linked open data!



### Data, how we know it

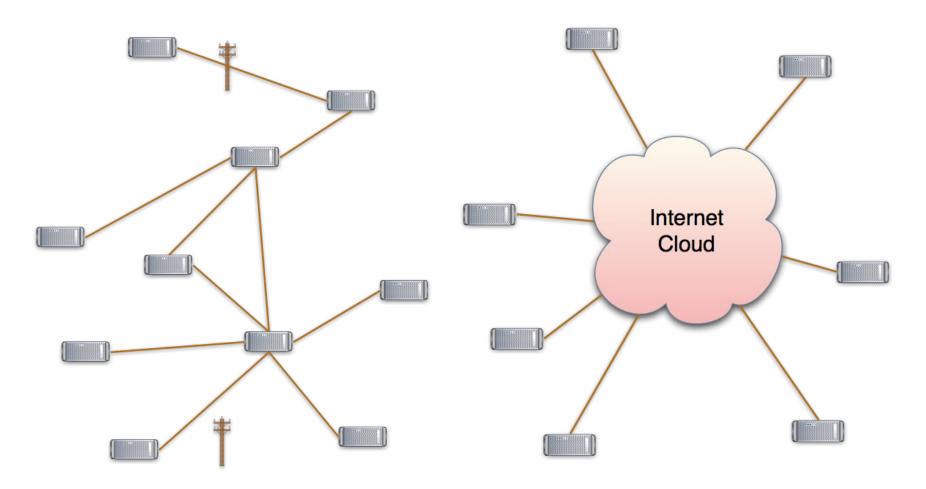


```
LDR
               ----h
FMT
               МН
001
               la HT016905880
               la 20110726
002a
003
               la 20110729
026
               la HBZHT016905880
030
               a|1uc||||17
036a
               la NL
037h
               |a eng
050
               a||||||
051
               m|||f|||
               la 294/61
070
070h
               la 361
080
               la 60
100
               |a Allemang, Dean |9 136636187
104a
               |a Hendler, James A. |9 115664564
331
               |a Semantic web for the working ontologist
335
               |a effective modeling in RDFS and OWL
359
               |a Dean Allemang ; Jim Hendler
403
               la 2. ed.
410
               |a Amsterdam [u.a.]
               la Elsevier MK
412
425a
               |a 2011
433
               |a XIII, 354 S. : graph. Darst.
540a
               la 978-0-12-385965-5
```

(To be honest, we might actually be the only ones knowing such data. And there aren't too many things that one can describe in this way.)



## Along came the Internet





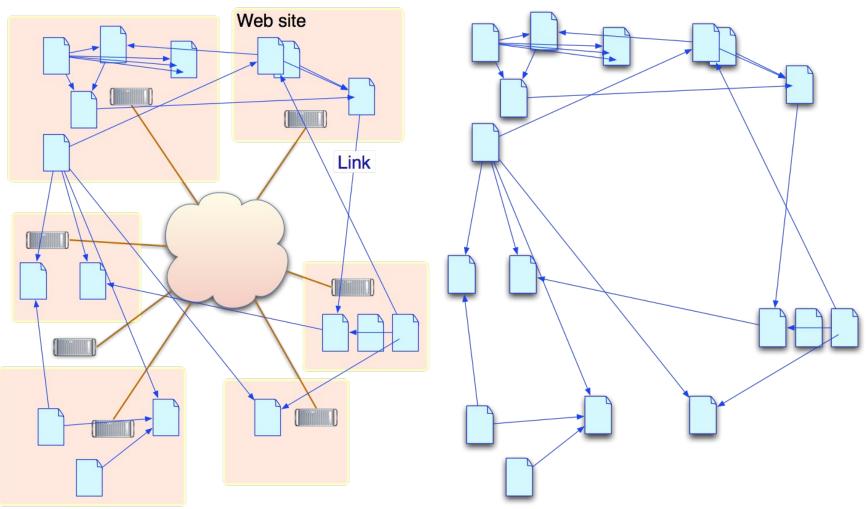
### Data, how others know it

```
<book id="HT016905880">
  <title>Semantic web ... </title>
  <author id="136636187">
    <firstname>Dean</firstname>
    <lastname>Allemang</lastname>
    </author>
  </book>
```

(Of course, "others" does not mean "everybody". But at least you can describe many things this way. Maybe even everything.)

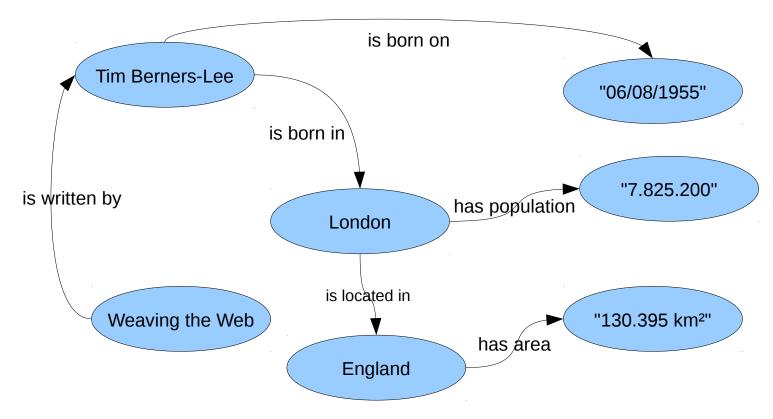


### The World Wide Web





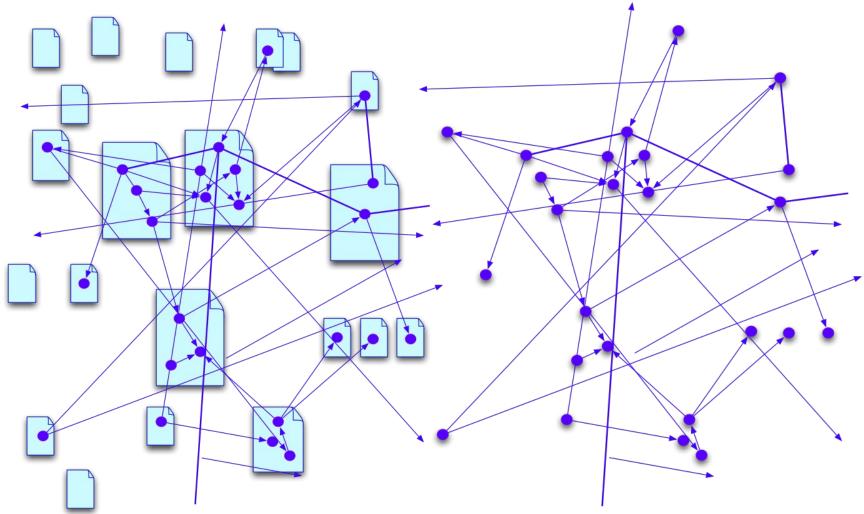
### Data, how the web likes it



(No wonder, it actually looks like a web. Or, if you will, a directed labelled **graph**.)



## The Giant Global Graph





#### Your turn!

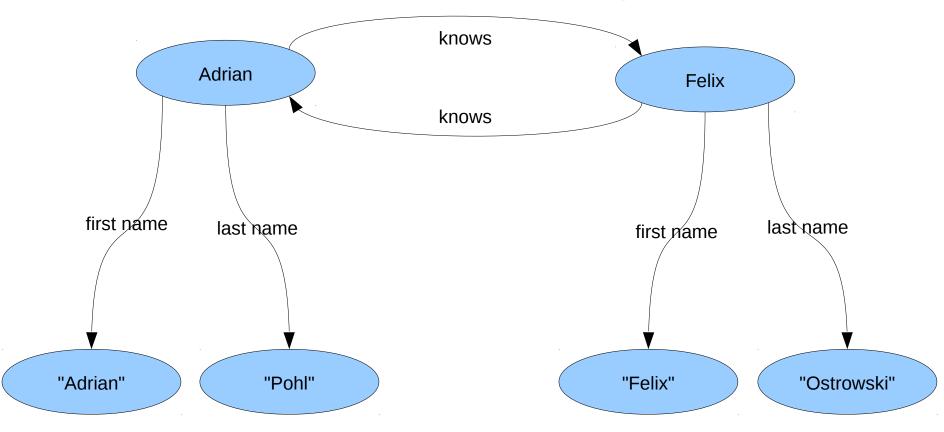


#### Draw a graph of your social network.

(For now, stick with the people on your table)



## A simple social graph





Obviosly a computer will have trouble interpreting such a diagram. The **graph data model** is an **abstract** one, but we can concrete it for the computer.



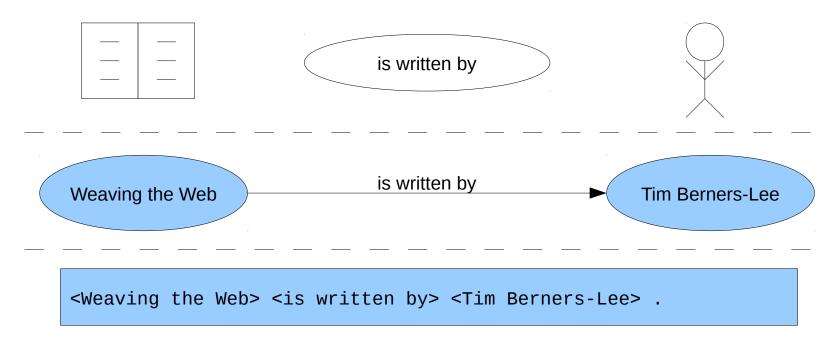
# **Graphs**, (almost) how computers like them

```
<Weaving the Web> <is written by> <Tim Berners-Lee> .
<Tim Berners-Lee> <has first name> "Tim" .
<Tim Berners-Lee> <has last name> "Berners-Lee" .
<Tim Berners-Lee> <is born on> "06/08/1955" .
<Tim Berners-Lee> <is born in> <London> .
<London> <is located in> <England> .
<London> <has population> "7825200" .
<London> <hat Fläche> "130395 km²" .
```

(This notation is called Turtle and it is one of several writing styles for a data model called RDF. RDF stands for "Resource Description Framework"; this is the de-facto standard for publishing Linked Data.

A big advantage of the Turtle notation: humans can actually read it!)

## Basic element: the triple



(A triple is the smallest possible graph. It's components are called **subject**, **predicate** and **object**.)



#### Your turn!



Open the etherpad for your group. In this etherpad, express the graph you have drawn in RDF.



## Simple social graph in RDF

```
<Adrian> <first name> "Adrian" .
  <Adrian> <last name> "Pohl" .
  <Adrian> <knows> <Felix> .
  <Felix> <first name> "Felix" .
  <Felix> <last name> "Ostrowski" .
  <Felix> <knows> <Adrian> .
```



#### What does ...

- →... <Tim Berners-Lee>,
- →... <London> and
- →... <England>

#### stand for, and what does

- <has first name>,
- ><is located in> and
- ><has population>
  mean?



# We need unambigous reference!

Authority files are a good start, but again we'll be the only ones understanding those. On the web, people use **URI**s!

(URI stands for **Uniform Resource Identifier**)



## **URI**

scheme ":" hier-part [ "?" query ] [ "#" fragment ]

(???)



http://de.wikipedia.org/wiki/Uniform\_Resource\_Identifier ftp://ftp.is.co.za/rfc/rfc3986.txt

file:///home/fo/doc/swib13/slides.odp

urn:isbn:978-1608454303



# **Graphs**, how computers really like them

```
<urn:isbn:978-0062515872> <http://purl.org/dc/terms/creator> <http://d-nb.info/gnd/121649091> .
   <http://d-nb.info/gnd/121649091> <http://xmlns.com/foaf/0.1/givenName> "Tim" .
   <http://d-nb.info/gnd/121649091> <http://xmlns.com/foaf/0.1/familyName> "Berners-Lee" .
   <http://d-nb.info/gnd/121649091> <http://xmlns.com/foaf/0.1/birthday> "06/08/1955" .
```

(A pleasant side-effect when using HTTP-URIs – which is what Linked Data is based upon, is that they can be dereferenced. When following such a link, one should get a description of the resource. More on that later.)



# Graphs, (sort of) readable for humans and machines

```
@prefix dc: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix gnd: <http://d-nb.info/gnd/> .

<urn:isbn:978-0062515872> dc:creator gnd:121649091 .
gnd:121649091 foaf:givenName "Tim" .
gnd:121649091 foaf:familyName "Berners-Lee" .
gnd:121649091 foaf:birthday "06/08/1955" .
```

(You can abbreviate URIs using **prefixes**. This also makes it easier to identify the vocabularies you use.)



# But isn't some data we had missing!?

```
<http://d-nb.info/gnd/121649091> <is born in> <London> .
<London> <is located in> <England> .
<London> <has population> "7825200" .
<London> <has area> "130395km²" .
```

(There may not be a URI for everything you want to refer to, neither for entities nor for vocabularies.)



## Don't repeat others, link!

- → Reuse properties from existing vocabularies
- Link to things by simple URI reference
- Think Data-Library (as in Software-Library)



```
@prefix : <#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/terms/> .

:ostrowski foaf:givenName "Felix" .
:ostrowski foaf:familyName "Ostrowski" .
:ostrowski foaf:birthday "28.05.1981" .

<> dc:creator :ostrowski .
```

(When something you want to describe does not have a URI yet, you can use Ids that are relative to the describing document. Since two documents can't be at the same place at the same time, these Ids only have to be unique within that document. "<>" stands for the document itself. You can check here if you are creating valid turtle.)



#### Your turn!



Reformulate your RDF using the FOAF vocabulary. Also, use DC Terms to assert that you are the authors of the describing document. You can also add further metadata about the document if you want.



## Simple social graph using FOAF

```
@prefix : <#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/terms/> .

:adrian foaf:givenName "Adrian" .
:adrian foaf:familyName "Pohl" .
:adrian foaf:knows :felix .
:felix foaf:givenName "Felix" .
:felix foaf:familyName "Ostrowski" .
:felix foaf:knows :adrian .

<> dc:creator <Felix> .
<> dc:creator <Adrian> .
<> dc:created "25.11.2013" .
```



#### Break



### **Open** Data



## Open Definition

"Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."

http://www.opendefinition.org



# Open Data is a question of...

- Licensing
- → Access
- → Formats



# Open Data is a question of...

- → Licensing
- → Access
- → Formats



### Open Data Licenses

- → Attribution
- Attribution-Share-Alike
- → Public-Domain
- → No non-commercial (NC) licenses

http://www.opendefinition.org/licenses/



# Open Data is a question of...

- Licensing
- → Access
- → Formats



#### Access

- ...to the work as a whole
- No more than a reasonable reproduction cost
- Preferably downloading via the Internet without charge



# Open Data is a question of...

- Licensing
- → Access
- → Formats



#### **Formats**

- →Open file format:= "a published specification for storing digital data ... which can ... be used and implemented by anyone"
- Machine-readibility counts!
- → Examples: rdf, json, ods, xls, pdf, docx



Data
vs.
Databases



### Database

"a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means."

From: European Database Directive



#### 'Data'

- A term with different meanings:
  - (1)Content of a database → can be anything
  - (2)Recorded facts (e.g. most catalog records) → aren't copyrightable, only as collection



### Different legal status?

- Legal status of a database and its content may differ
- → Example: a collection of public domain content that as a whole is protected by copyright or related rights

# Opening up data in 8 steps



# 1. Decide what data would be most useful to others

- Your library catalogue & holdings?
- → Special collection data?
- Circulation data?
- Controlled vocabulary?
- **-**



# 2.Getting willing people together



# 3. Clarify potential legal problems

- Check your national legislation
- →Bought data?
- From which vendors?
- →What usage rights & restrictions do contracts give?



# 4. Export the data





### 5. Publish data on the web

#### Index of /download/dumps/DE-605

| Icon  | Name                              | <u>Last modified</u> | <u>S:</u> | <u>ize</u> | Description |
|-------|-----------------------------------|----------------------|-----------|------------|-------------|
| [DTR] | Parent Directory                  |                      |           | _          |             |
|       | DE-605.base 0.rdfmab.tar.bz2      | 09-Aug-2012 00       | 9:19 36   | 65M        |             |
|       | DE-605.base 0.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       |           | 65         |             |
| i i   | DE-605.base 1.rdfmab.tar.bz2      | 09-Aug-2012 00       |           | 70M        |             |
| i i   | DE-605.base 1.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       |           | 65         |             |
| i i   | DE-605.base 2.rdfmab.tar.bz2      | 09-Aug-2012 00       |           | 59M        |             |
| i i   | DE-605.base 2.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       |           | 65         |             |
| į į   | DE-605.base 3.rdfmab.tar.bz2      | 09-Aug-2012 00       |           | 63M        |             |
|       | DE-605.base 3.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       |           | 65         |             |
|       | DE-605.base 4.rdfmab.tar.bz2      | 09-Aug-2012 00       |           | 65M        |             |
|       | DE-605.base 4.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       |           | 65         |             |
| [ ]   | DE-605.base 5.rdfmab.tar.bz2      | 09-Aug-2012 00       |           | 60M        |             |
| [ ]   | DE-605.base 5.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       |           | 65         |             |
| [ ]   | DE-605.base 6.rdfmab.tar.bz2      | 09-Aug-2012 00       |           | 65M        |             |
| [ ]   | DE-605.base 6.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       | 9:21 (    | 65         |             |
| [ ]   | DE-605.base 7.rdfmab.tar.bz2      | 09-Aug-2012 00       | 9:17 3    | 58M        |             |
| [ ]   | DE-605.base 7.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       | 9:21 (    | 65         |             |
| [ ]   | DE-605.base 8.rdfmab.tar.bz2      | 09-Aug-2012 00       | 9:18 3    | 59M        |             |
| [ ]   | DE-605.base 8.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       | 9:21 (    | 65         |             |
| [ ]   | DE-605.base 9.rdfmab.tar.bz2      | 09-Aug-2012 00       | 9:17 3    | 58M        |             |
| [ ]   | DE-605.base 9.rdfmab.tar.bz2.md5  | 15-Aug-2012 09       | 9:21 (    | 65         |             |
| [ ]   | DE-605.base 10.rdfmab.tar.bz2     | 09-Aug-2012 00       | 9:19 36   | 67M        |             |
| [ ]   | DE-605.base 10.rdfmab.tar.bz2.md5 | 15-Aug-2012 09       | 9:21 (    | 66         |             |
| [ ]   | DE-605.base 11.rdfmab.tar.bz2     | 09-Aug-2012 00       | 9:19 36   | 63M        |             |
| [ ]   | DE-605.base 11.rdfmab.tar.bz2.md5 | 15-Aug-2012 09       | 9:21 (    | 66         |             |
| [DIR] | enrich/                           | 03-May-2013 16       | 5:20      | -          |             |
| [ ]   | <pre>example.rdfmab.ttl</pre>     | 12-Mar-2012 16       | 5:08 9    | .7K        |             |
| [ ]   | hbzlod.nt.tar.bz2                 | 23-Aug-2012 19       | 9:17 3    | .8G        |             |
| [ ]   | hbzlod.nt.tar.bz2.md5             | 23-Aug-2012 19       | 9:17 5    | 52         |             |
| [TXT] | <u>license.txt</u>                | 09-Dec-2011 09       |           | 80         |             |
| [DIR] | updates/                          | 16-Jan-2013 15       | 5:56      | -          |             |
| [DIR] | zvdd/                             | 18-Mar-2013 17       | 7:25      | -          |             |



### 6. Apply an open license

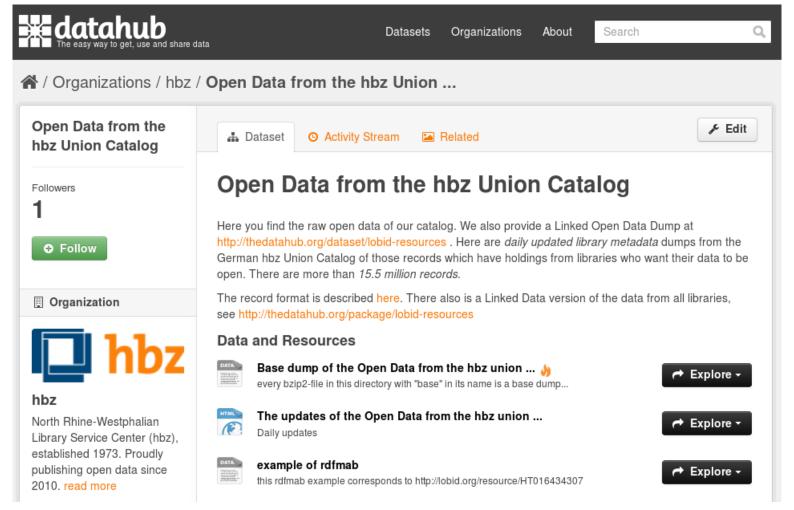
(cc) ZERO

All data which is available for download here is published under a <u>Creative Commons CC0 licence</u>. The data thus is in the Public Domain which means it belongs to everybody and can be used without restrictions for any purpose.

- @prefix cc: <http://creativecommons.org/ns#> .
- <dataset URI> cc:license
- <a href="http://creativecommons.org/publicdomain/zero/1.0/">http://creativecommons.org/publicdomain/zero/1.0/</a>.

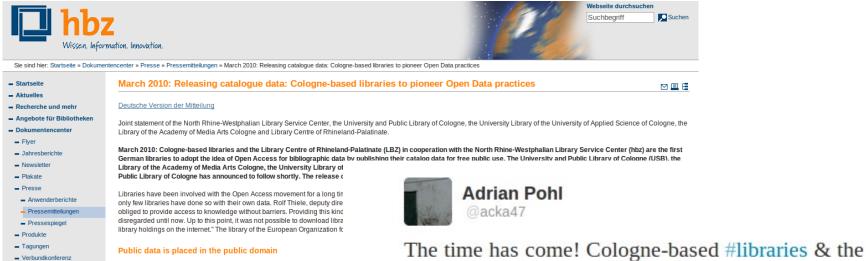


# 7. Register your dataset





### 8. Let others know



The publication of the data enables anybody to download, modify and use i capital, it is important to stick up for the traditional duty of libraries and librar with the lowest restrictions possible," said Silke Schomburg, deputy director to the public without restrictions," she continued.

Cooperation and data exchangie between libraries have been firmly estable enhance cooperation among libraries but enable subsequent use by non-lienhanced by catalog data. The German Wikipedia for example has been er data's half open character, "Schomburg notes.

#### Data for the Semantic Web

- Veröffentlichungen

Impressur

Kontakt

- Projekte

Über uns

- Aus der Region

A kleine Schrift

▲ große Schrift

mittlere Schrift

The North Rhine-Westphalian Library Service Center has recently begun eemerging Semantic Web. The liberalization of bibliographic data provides the discussions with other member libraries of the hbz library network to publish library world. LBZ in cooperation with the #hbz open up their data:  $\label{lbz} http://tr.im/RCnV\ \#opendata\ \#CC0$ 

. . .



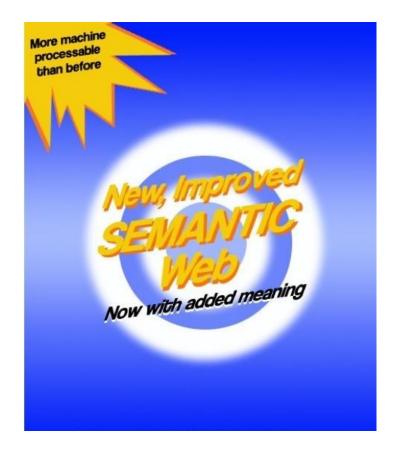
8:16 PM - 12 Mar 10 via TweetDeck - Embed this Tweet

Further information and links to the published datasets are available at <a href="http://www.hbz-nrw.de/projekte/linked\_open\_data">http://www.hbz-nrw.de/projekte/linked\_open\_data</a> (Our new website will be launched on Monday, March 15th.)

For further questions contact



# Now: Improve data & interfaces





#### Your turn!



# Agree on an *open*Creative Commons License within your group and link your document to that license.

(The predicate
<http://creativecommons.org/ns#license>

is well suited for this link, but searching the Web will reveal alternatives.)



# Open licencing

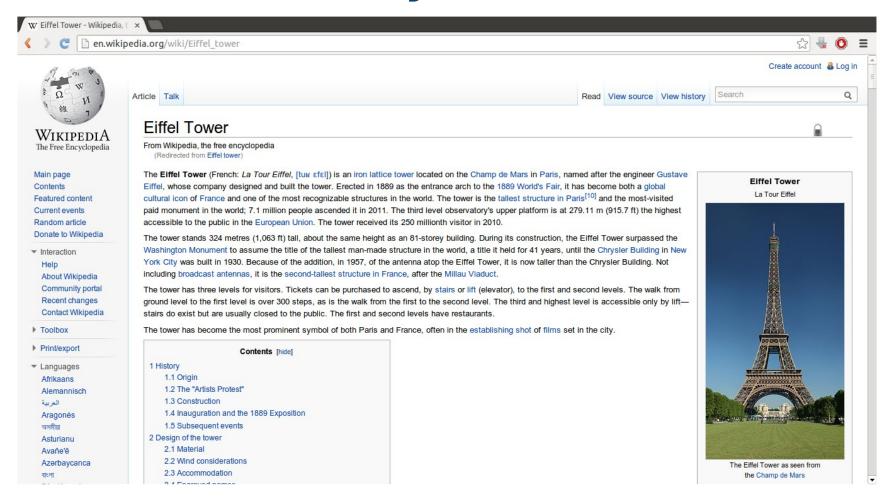
```
@prefix : <#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/terms/> .
:adrian foaf:givenName "Adrian" .
:adrian foaf:familyName "Pohl" .
:adrian foaf:knows :felix .
:felix foaf:givenName "Felix" .
:felix foaf:familyName "Ostrowski" .
:felix foaf:knows :adrian .
<> dc:creator :felix .
<> dc:creator :adrian .
<> dc:created "25.11.2013" .
<> <http://creativecommons.org/ns#license>
      <http://creativecommons.org/publicdomain/zero/1.0/> .
```



#### **Linked Data** in Action



# The Treachery of Documents



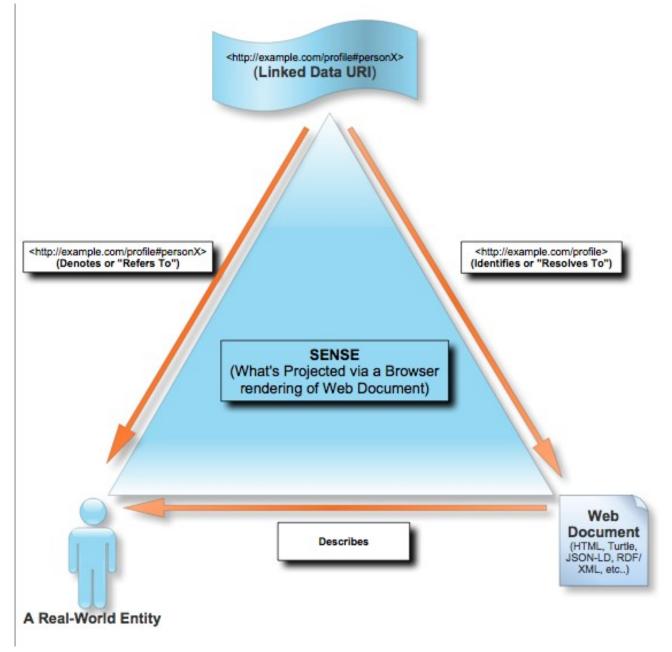
Ceci n'est pas la Tour Eiffel.



Identification and description of a resource ought to be distinguished!

But in the Linked-Data-Paradigm, both are linked.







The description of a resource can be made available in various **formats**. Which format will be delivered can be decided by **Content-Negotiation**.



# JSON-LD

```
"@context": "person.json",
  "@id": "http://ex.org/person/1",
  "name": "Felix",
  "knows": "http://ex.org/person/2"
}
  "@context": "person.json",
  "@id": "http://ex.org/person/2",
  "name": "Adrian",
  "knows": "http://ex.org/person/1"
};
```

```
{
   "name": "http://schema.org/name",
   "knows": {
      "@type": "@id",
      "@id": "http://schema.org/knows",
   }
}
```



#### Your turn!



In your description, link yourself to people from other groups that you know. This doesn't have to be reciprocal.

Also, link (approximately) to the place you live or work. Use DBpedia for this.



#### Break

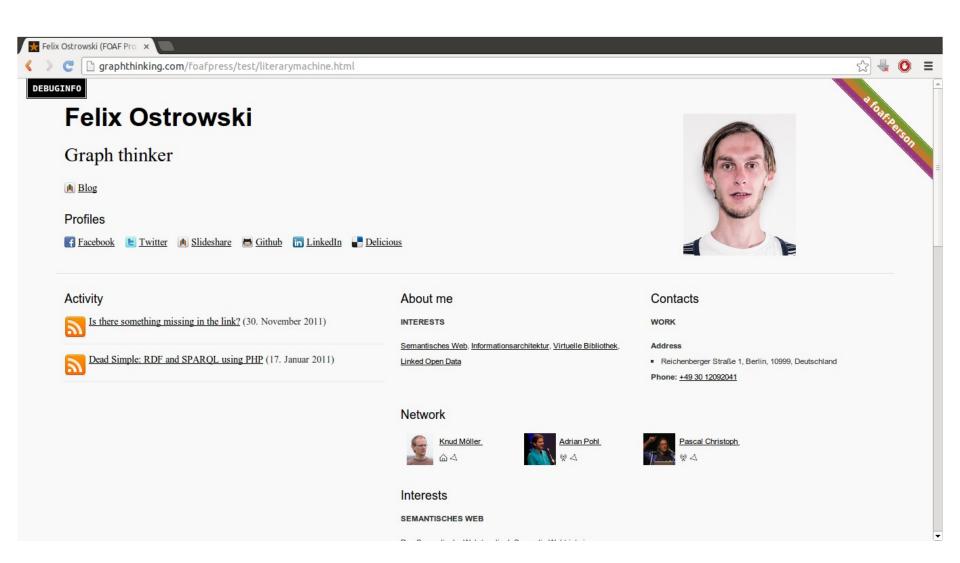


Scattered machine-readable descriptions are useful, but we can do better than that! RDF is a distributed data model that makes it easy to **combine** several descriptions. Furthermore, special databases exist that allow to query RDF data.



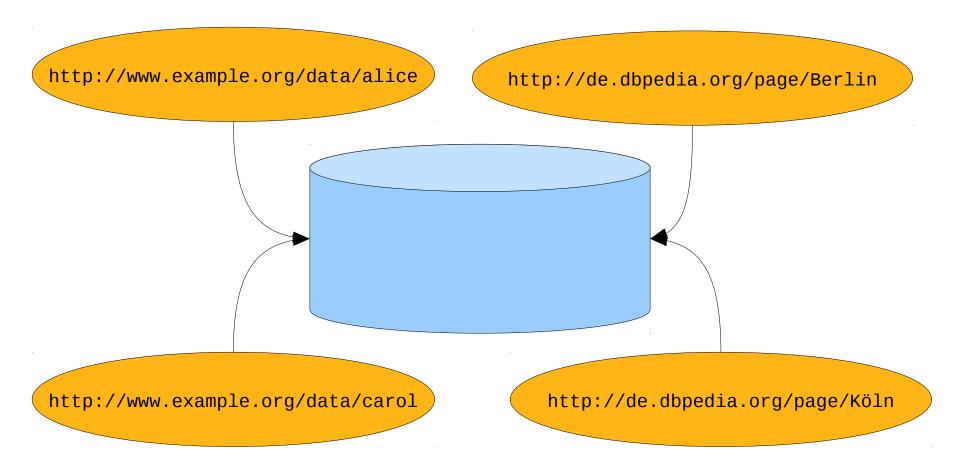
```
<http://ex1.org/adrian> <xmlns.com/foaf/0.1/givenName> "Adrian" .
<http://ex1.org/adrian> <xmlns.com/foaf/0.1/knows> <http://ex2.org/felix> .
<http://ex2.org/felix> <xmlns.com/foaf/0.1/givenName> "Felix" .
<http://ex2.org/felix> <xmlns.com/foaf/0.1/knows> <http://ex1.org/adrian> .
```







### Triple Stores





SPARQL facilitates queries on the data in a triple store. The foundations for this are simply graph patterns. These look almost like triples, the difference being that the contain variables.



```
@prefix ex: <http://example.org/people#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
ex:alice foaf:name "Alice" .
```

```
PREFIX ex: <http://example.org/people#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT * WHERE {
   ex:alice foaf:name ?name .
}
```

```
name
"Alice"
```



```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name1 ?name2 WHERE {
    ?person1 foaf:knows ?person2 .
    ?person1 foaf:name ?name1 .
    ?person2 foaf:name ?name2 .
}
```

| name1   | name2   |
|---------|---------|
| "Alice" | "Bob"   |
| "Bob"   | "Carol" |
| "Carol" | "Alice" |



```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?name ?ortname WHERE {
    ?person1 foaf:knows ?person2 .
    ?person2 foaf:name ?name .
    ?person2 foaf:based_near ?ort .
    ?ort rdfs:label ?ortname .
}
```

| name  | ortname      |
|-------|--------------|
| "Bob" | "Dresden"@de |



### Your turn!



Use SPARQL to analyse your connections. For example you might want to determine who you know directly or indirectly or who comes from the same city as you.



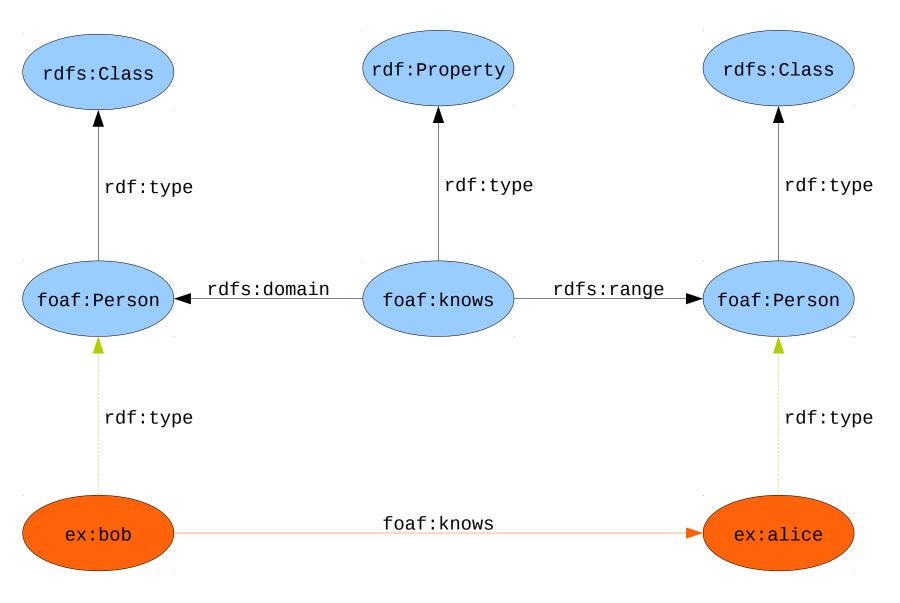
### Break



# Let's put some **Semantic** in the **Web**

The classes and properties being used can be using description languages for vocabularies. The relatively simple RDF Schema (RDFS) is wide spread, but more complex issues can be expressed in the Web Ontology Language (OWL).







```
# Explicit triples
ex:bob foaf:knows ex:alice .
```

```
# Implicit triple, that follow from the schema
ex:bob rdf:type foaf:Person .
ex:alice rdf:type foaf:Person .
```



```
# Explicit triples
ex:bob ex:colleague ex:alice .
```

```
# Implicit triple, that follow from the schema
ex:bob foaf:knows ex:alice .
ex:bob rdf:type foaf:Person .
ex:alice rdf:type foaf:Employee .
ex:alice rdf:type foaf:Employee .
```



### Your turn!



### Create an RDF Schema so that from these assertions

```
@prefix team: <http://example.org/soccer/vocab#> .
@prefix ex: <http://example.org/soccer/resource#> .

ex:team1 team:player ex:bob .
ex:team2 team:player ex:alice .
ex:game1 team:home ex:team1 .
ex:game1 team:away ex:team2 .
```

### the following triples can be inferred.

```
@prefix team: <http://example.org/soccer/vocab#> .
@prefix ex: <http://example.org/soccer/resource#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

ex:team1 rdf:type foaf:Group .
ex:team2 rdf:type foaf:Group .
ex:team1 foaf:member ex:bob .
ex:team2 foaf:member ex:alice .
ex:bob rdf:type foaf:Person .
ex:alice rdf:type foaf:Person .
ex:game1 rdf:type team:Game .
ex:game2 rdf:type team:Game .
```



```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix team: <http://example.org/soccer/vocab#> .
team:player rdf:type rdfs:Property;
            rdfs:subPropertyOf foaf:member ;
            rdfs:domain foaf:Person;
            rdfs:range foaf:Group .
team:home
            rdf:type rdfs:Property;
            rdfs:domain team:Game .
            rdf:type rdfs:Property;
team:away
            rdfs:domain team:Game .
            rdf:type rdfs:Class .
team: Game
```

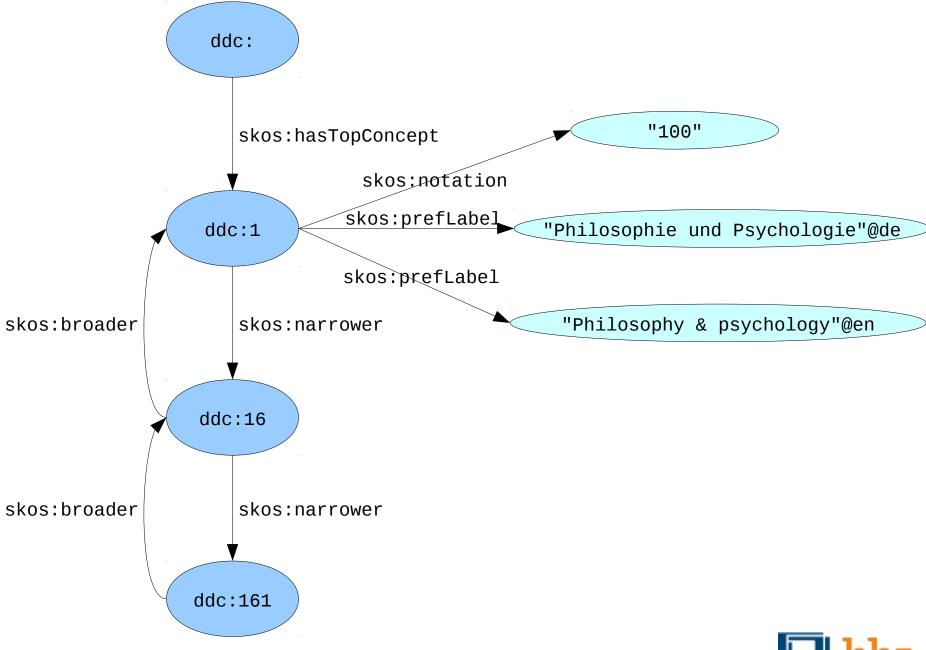


The expressiveness and the possibilities of inference of RDFS and OWL are not always needed.

For controlled vocabularies, the Simple Knowledge Organization System (SKOS) is a simpler alternative that is also based on RDF.

The Dewey Decimal Classification and the

Library of Congress Subject Headings have already found their way into the Linked-Data-world.





# Elements of Linked (Open) Data

#### Using: Mashups

Mashups combine multiple datasets to create a new service. visualisation or information.

#### Using: Search

Linked data search engines allow search across the web of data. Conventional search may present information derived from linked data.

**Using:** Productivity Linked data facilitates data integration for business intelligence or research.

#### Storing and publishing

Linked data can be published in simple flat files on a web server, in databases with a translation layer, or in specialised 'triple stores' built to store and share linked data. Publishing platforms understand requests for linked data & return it formatted as RDF.

#### Querying: SPARQL

SPARQL Protocol and RDF Query Language provides a way to run structured queries over linked data datasets. SPARQL servers expose linked open data to be gueried.

#### Representing: Vocabularies

Vocabularies provide lists (and definitions) of common terms that can be used to describe the things and relationships in a dataset.

#### Representing: Ontologies

Ontologies are vocabularies that record the logical relationships between their terms and support reasoning.

#### Interchanging: RDF

Resource Descriptor Framework (RDF) is a model for representing data as 'triples'. RDF can be serialised into a range of different file formats, including RDF-XML and text-based Turtle or N3 syntax.

#### Identifying: URLs

Using HTTP Uniform Resource Locators (URLs) means that (a) data can be looked up across the Internet; (b) decisions about 'namespaces' for data are managed through the Domain Name System (DNS).

#### Transporting: HTTP (The World Wide Web)

Data is hosted on servers that can talk Hypertext Transfer Protocol (HTTP) to each other and to browsers in order to exchange data across the Internet.

#### Elements of the Linked Open Data Stack (revision 3) - 5th May 2011. CC BY-SA-NC

Draft sketch by Tim Davies (@timdavies / tim@practicalparticipation.co.uk) for IKM Working Paper on Linked Open Data for Development. Comments welcome. Search 'linked open data stack' on http://www.opendataimpacts.net for latest version.

Open data is made available under licenses (or is placed in the public domain) so that others can use and build upon it, free of legal restrictions. Open standards for data files and interchange are used

Licensing: open data



# Thank you!



## License



These slides are published under a Creative Commons license:

http://creativecommons.org/licenses/by/3.0/de/