



hbz

Wissen. Information. Innovation.

An Introduction to Linked Open Data

Felix.Ostrowski@googlemail.com (@literarymachine)

Pascal Christoph@hbz-nrw.de (@dr0ide)

Adrian Pohl@hbz-nrw.de (@acka47)

#swib12 Pre-Conference Workshop

Monday, November 26th 2012

Cologne

Schedule

- Organize in teams
- Introduction: Data – Graphs – Triples
- Groupwork
- URIs and Namespaces
- Groupwork
- Open Data
- Groupwork
- Identification vs. Description
- Content-Negotiation
- 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          -----M2.01200024-----h
FMT          MH
001          |a HT016905880
002a         |a 20110726
003          |a 20110729
026          |a HBZHT016905880
030          a|luc|||||17
036a         |a NL
037b         |a eng
050          a|||||17
051          m||f|||
070          |a 294/61
070b         |a 361
080          |a 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          |a 2. ed.
410          |a Amsterdam [u.a.]
412          |a Elsevier MK
425a         |a 2011
433          |a XIII, 354 S. : graph. Darst.
540a         |a 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.)

Data, how others know it

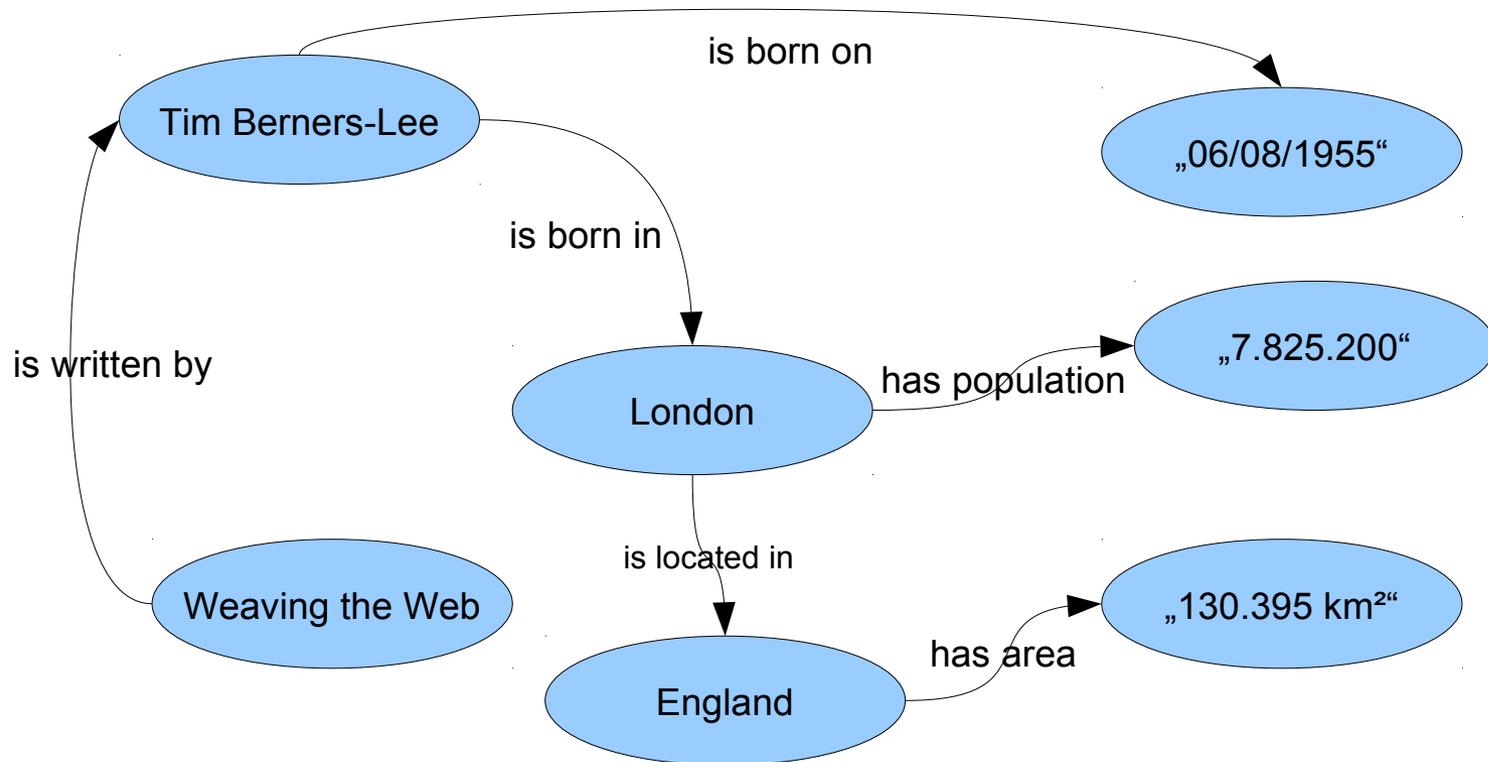
```
+-----+-----+-----+-----+
| id      | firstname | lastname | birthday |
+-----+-----+-----+-----+
| 136636187 | Dean      | Allemang | NULL      |
+-----+-----+-----+-----+

+-----+-----+-----+-----+
| id      | title                                     | author |
+-----+-----+-----+-----+
| HT016905880 | Semantic web for the working ontologist | 136636187 |
+-----+-----+-----+-----+
```

```
<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.)

Data, how the web likes it



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

Obviously 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.

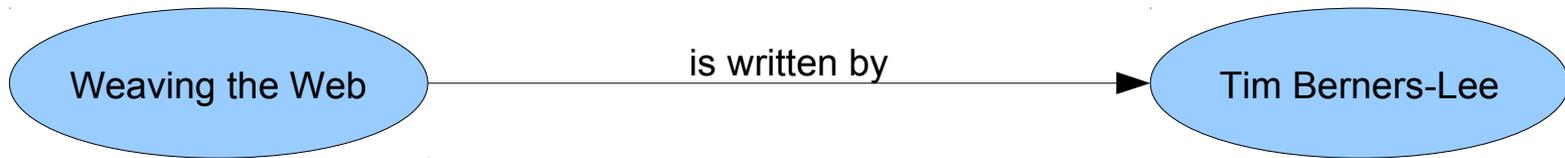
Graphs, (almost) how computers like them

```
<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**



```
<Weaving the Web> <is written by> <Tim Berners-Lee> .
```

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

Your turn!

Create an etherpad for your group at
<http://etherpad.lobid.org/>

In this pad, describe the members of your group using RDF expressed in Turtle. You can just make up the predicates for now.

(If you'd rather do so, go ahead and describe a fictional character.)

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 **unambiguous reference!**

Controlled vocabularies are a good start, but again we'll be the only ones understanding those. On the web, people use **URIs!**

(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/elag12/slides.odp>

<urn:isbn:978-1608454303>

(URIs don't necessarily need domain names; IP-adresses also work: <http://192.168.0.124>)

Graphs, how computers really like them

```
<urn:isbn:978-0062515872> <http://purl.org/dc/terms/creator> <http://viaf.org/viaf/85312226> .  
<http://viaf.org/viaf/85312226> <http://xmlns.com/foaf/0.1/givenName> „Tim” .  
<http://viaf.org/viaf/85312226> <http://xmlns.com/foaf/0.1/familyName> „Berners-Lee” .  
<http://viaf.org/viaf/85312226> <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 viaf: <http://viaf.org/viaf/> .  
  
<urn:isbn:978-0062515872> dc:creator gnd:121649091 .  
viaf:121649091 foaf:givenName „Tim“ .  
viaf:121649091 foaf:familyName „Berners-Lee“ .  
viaf:121649091 foaf:birthday „06/08/1955“ .
```

But isn't some data we had missing!?

```
<http://viaf.org/viaf/85312226> <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. When looking for URIs for entities, **DBpedia** is a good place to start. It's a bit tougher for vocabularies, but **LOV** (Linked Open Vocabularies) is worth a try.

A couple of tricks

```
@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.

```
@prefix :          <#> .
@prefix foaf:      <http://xmlns.com/foaf/0.1/> .
@prefix dbpedia:  <http://de.dbpedia.org/resource/> .

:ostrowski foaf:givenName „Felix“ ;
           foaf:familyName „Ostrowski“ ;
           foaf:birthday „28.05.1981“ ;
           foaf:based_near dbpedia:Berlin ,
                           dbpedia:Berlin-Kreuzberg .
```

(When making several statements about the same resource, you don't have to repeat its identifier. The predicate-object pairs are separated by a semicolon. When listing several values for the same predicate of a resource, that predicate also does not have to be repeated. Simply separate the values by comma. You can check [here](#) whether you are creating valid turtle.)

Your turn!

Within your group, agree on an identifier for each member and describe them using the **FOAF** vocabulary. Don't forget to state that you know each other! Also, use **DC Terms** to assert that you are the authors of the describing document. You can also add further metadata about the document.

Break

Open Data

„Open Data“ - Definition

Open Knowledge

“A piece of knowledge is open if you are free to use, reuse, and redistribute it.”

<http://www.opendefinition.org/okd/>

Open Data is a question of...

- Access
- Licenses
- Formats

Open Data is a question of...

- Access
- Licenses
- Formats

Access

- ...to the whole data
- No more than a reasonable reproduction cost
- Preferably downloading via the Internet without charge
- Technically: XMPP > OAI-PMH > http > ftp > scp > paper-based

Open Data is a question of...

- Access
- Licenses
- Formats

Open Data Licenses

- Attribution (ODC-BY)
- Attribution-Share-Alike (OdbL)
- Public-Domain (CC0, PDDL)
- CC-BY, CC-BY-SA for some uses
- **No** non-commercial licenses

<http://www.opendefinition.org/licenses/#Data>

Open Data is a question of...

- Access
- Licenses
- Formats

Formats

→ Machine-readability counts!

→ Examples:

rdf > csv > ods > xls > PDF >
docx > Hardcopy

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

- Content of a database → can be anything
- Recorded facts → aren't copyrightable, only as collection

Different legal status?

- Legal status of database and its contents may differ
- Example: Public Domain content and copyrighted collection

Opening up data in 7 steps

1. Getting willing people together



2. Clarify potential legal problems

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

3. Export the data



4. Publish data on the web

Index of /download/dumps/DE-605

 Name	Last modified	Size	Description
 Parent Directory		-	
 DE-605.base_0.rdfmab.tar.bz2	13-Mar-2012 19:33	235M	
 DE-605.base_0.rdfmab.tar.bz2.md5	14-Mar-2012 13:17	63	
 DE-605.base_1.rdfmab.tar.bz2	13-Mar-2012 19:33	232M	
 DE-605.base_1.rdfmab.tar.bz2.md5	14-Mar-2012 13:17	63	
 DE-605.base_2.rdfmab.tar.bz2	13-Mar-2012 19:32	232M	
 DE-605.base_2.rdfmab.tar.bz2.md5	14-Mar-2012 13:17	63	
 DE-605.base_3.rdfmab.tar.bz2	13-Mar-2012 19:33	235M	
 DE-605.base_3.rdfmab.tar.bz2.md5	14-Mar-2012 13:17	63	
 DE-605.base_4.rdfmab.tar.bz2	13-Mar-2012 19:32	233M	
 DE-605.base_4.rdfmab.tar.bz2.md5	14-Mar-2012 13:17	63	
 DE-605.base_5.rdfmab.tar.bz2	13-Mar-2012 19:33	237M	

5. Apply an open license



All data which is available for download here is published under a [Creative Commons CC0 licence](#). 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

<http://creativecommons.org/publicdomain/zero/1.0/> .

6. Register your dataset

the Data Hub — The easy way to get, use and share data acka47 Logout

ckan [Add a dataset](#) [Search](#) [Groups](#) [About](#)

Open Data from the hbz Union Catalog

[View](#) [Resources \(3\)](#) [Related \(0\)](#) | [History](#) [Settings](#) [Authorization](#)

There are *daily updated library metadata* dumps from the hbz Union Catalog, those records which have holdings from libraries who want their data to be open. There are more than *10 million records*.

Views: 293

License: [Creative Commons CCZero](#)

[OPEN DATA](#)

Tags

[bibliographic](#) [library](#) [lld](#) [lod](#)

Groups

[Bibliographic Data](#)

Resources [\(edit\)](#)

[Base dump of the Open Data from the hbz union catalog](#) 25 [tar.bz2](#)

[The updates of the Open Data from the hbz union catalog](#) 21 [tar.bz2](#)

[example of rdfmab](#) 7 [text/turtle](#)

Additional Information [\(settings\)](#)

Field	Value
Source	https://wiki1.hbz-nrw.de/display/SEM/Recently published Open Data exports
Author	North Rhine-Westphalian Library Service Center
Maintainer	North Rhine-Westphalian Library Service Center
Version	2012-03-08
State	active

7. Let others know

- Startseite
- Aktuelles
- Recherche und mehr
- Angebote für Bibliotheken
- Dokumentencenter
 - Flyer
 - Jahresberichte
 - Newsletter
 - Plakate
 - Presse
 - Anwenderberichte
 - Pressemitteilungen
 - Pressespiegel
 - Produkte
 - Tagungen
 - Verbundkonferenz
 - Veröffentlichungen
- Impressum
- Kontakt
- Projekte
- Über uns
- Aus der Region

-  kleine Schrift
-  mittlere Schrift
-  große Schrift

March 2010: Releasing catalogue data: Cologne-based libraries to pioneer Open Data practices



[Deutsche Version der Mitteilung](#)

Joint statement of the North Rhine-Westphalian Library Service Center, the University and Public Library of Cologne, the University Library of the University of Applied Science of Cologne, the Library of the Academy of Media Arts Cologne and Library Centre of Rhineland-Palatinate.

March 2010: Cologne-based libraries and the Library Centre of Rhineland-Palatinate (LBZ) in cooperation with the North Rhine-Westphalian Library Service Center (hbz) are the first German libraries to adopt the idea of Open Access for bibliographic data by publishing their catalog data for free public use. The University and Public Library of Cologne (USB), the Library of the Academy of Media Arts Cologne, the University Library of Public Library of Cologne has announced to follow shortly. The release c

Libraries have been involved with the Open Access movement for a long time, but only a few libraries have done so with their own data. Rolf Thiele, deputy director, is obliged to provide access to knowledge without barriers. Providing this kind of access was disregarded until now. Up to this point, it was not possible to download library holdings on the internet. "The library of the European Organization for

Public data is placed in the public domain

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

Cooperation and data exchange between libraries have been firmly established and enhanced cooperation among libraries but enable subsequent use by non-librarians enhanced by catalog data. "The German Wikipedia for example has been enhanced by data's half open character," Schomburg notes.

Data for the Semantic Web

The North Rhine-Westphalian Library Service Center has recently begun to publish emerging Semantic Web. The liberalization of bibliographic data provides it with discussions with other member libraries of the hbz library network to publish library world.

Further information and links to the published datasets are available at http://www.hbz-nrw.de/projekte/linked_open_data. (Our new website will be launched on Monday, March 15th.)

For further questions contact:



Adrian Pohl

@acka47

The time has come! Cologne-based [#libraries](#) & the [LBZ](#) in cooperation with the [#hbz](#) open up their data: <http://tr.im/RCnV> [#opendata](#) [#CC0](#)

 Reply  Delete  Favorite

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

Your turn!

Agree on an open 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.)

Linked Data in Action

Identification and **description** of a resource ought to be distinguished!
But in the Linked Data paradigm, there are currently two common variants to link both together.

Hash URIs

`http://www.example.org/people#alice`

The server ignores the **fragment identifier** and delivers the description

`http://www.example.org/people`

HTTP 303 Redirects

`http://www.example.org/people/alice`

`http://www.example.org/data/alice`

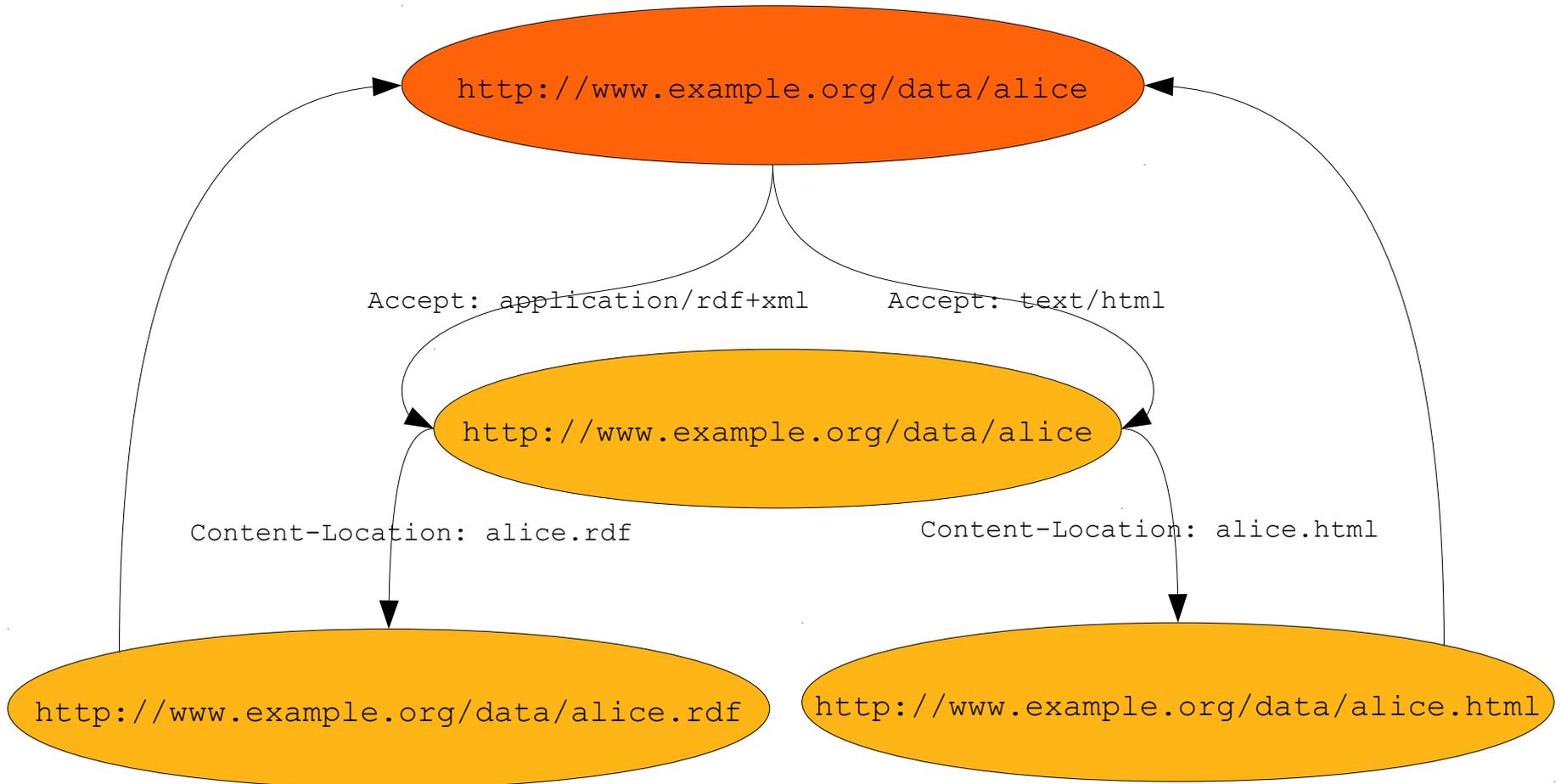
The server **redirects**
to the URL of the
description

Location: `http://www.example.org/data/alice`

`http://www.example.org/data/alice`

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

Content Negotiation



The easiest way to publish Linked Data is to serve **static RDF-files** that use **Hash-URIs** from a webserver that is set up to use **Multiviews** for Content-Negotiation.

```
# /var/www/people.ttl

@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix      : <#> .

:ostrowski
  foaf:givenName "Felix" ;
  foaf:familyName "Ostrowski" .

:christoph
  foaf:givenName "Pascal" ;
  foaf:familyName "Christoph" .
```

```
# /var/www/.htaccess
AddType text/turtle .ttl
AddType application/rdf+xml .rdf
Options +MultiViews
```

With this simple setup, the **URIs of the people** are:
<http://localhost/personen#ostrowski>
<http://localhost/personen#christoph>
The **URL of the description** is:
<http://localhost/personen>
Content-Negotiation for Turtle and RDF/XML is activated.

```
# /var/www/personen.rdf

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://localhost/lodws/test#">
  <rdf:Description rdf:about="http://localhost/lodws/test#christoph">
    <foaf:familyName>Christoph</foaf:familyName>
    <foaf:givenName>Pascal</foaf:givenName>
  </rdf:Description>
  <rdf:Description rdf:about="http://localhost/lodws/test#ostrowski">
    <foaf:familyName>Ostrowski</foaf:familyName>
    <foaf:givenName>Felix</foaf:givenName>
  </rdf:Description>
</rdf:RDF>
```

Demo: Consuming Linked Data with cURL

Your turn!

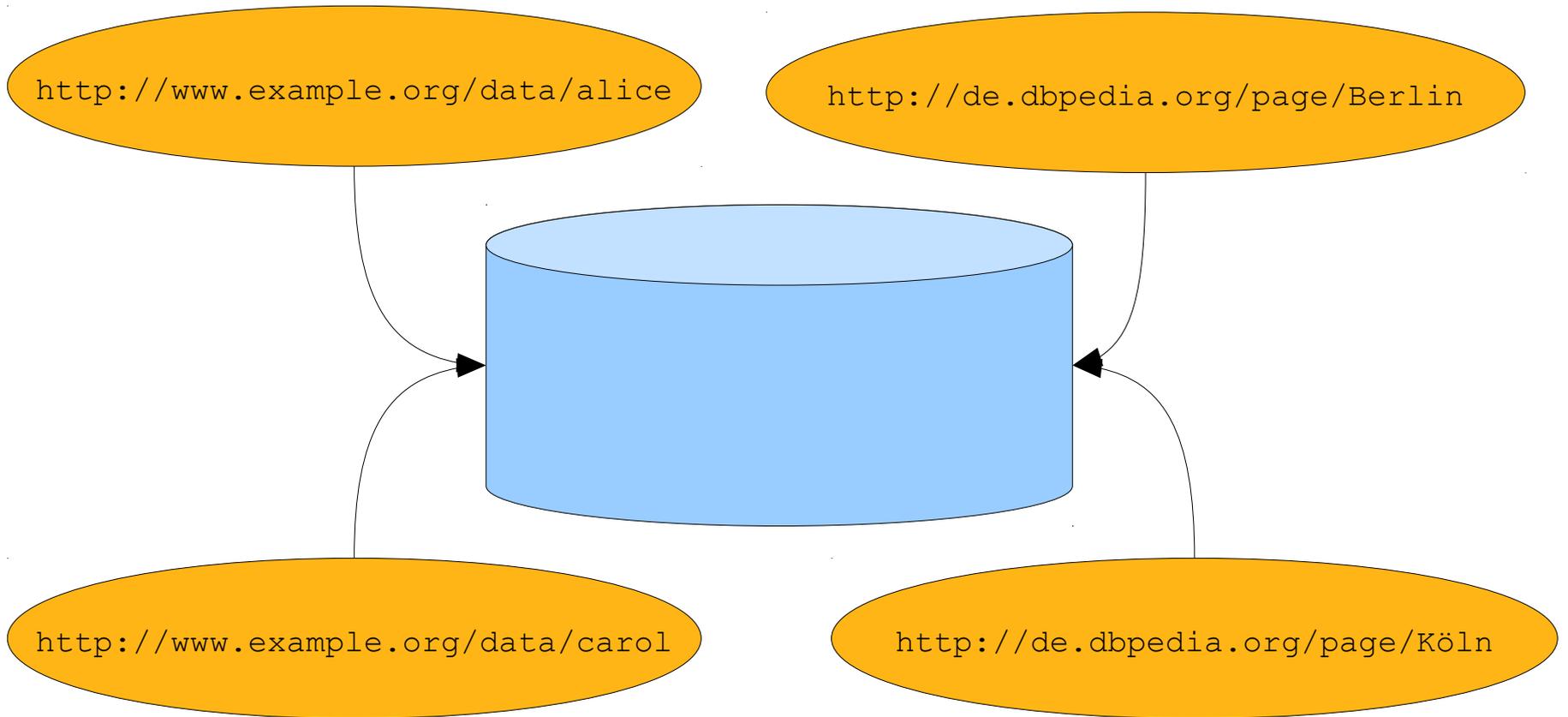
In your description, link yourself to people from other groups that you know. This doesn't have to be reciprocal.

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.

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 they 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 ex: <http://example.org/people#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

ex:alice foaf:name „Alice” ;
         foaf:knows ex:bob .
ex:bob foaf:name „Bob” ;
       foaf:knows ex:carol .
ex:carol foaf:name „Carol” ;
         foaf:knows ex: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      ex: <http://example.org/people#> .
@prefix      foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dbpedia: <http://de.dbpedia.org/resource/> .

ex:alice foaf:name „Alice” ;
         foaf:knows ex:bob ;
         foaf:based_near dbpedia:Berlin .
ex:bob   foaf:name „Bob” ;
         foaf:knows ex:carol ;
         foaf:based_near dbpedia:Dresden .
```

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

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

name	placename
„Bob”	„Dresden”@de

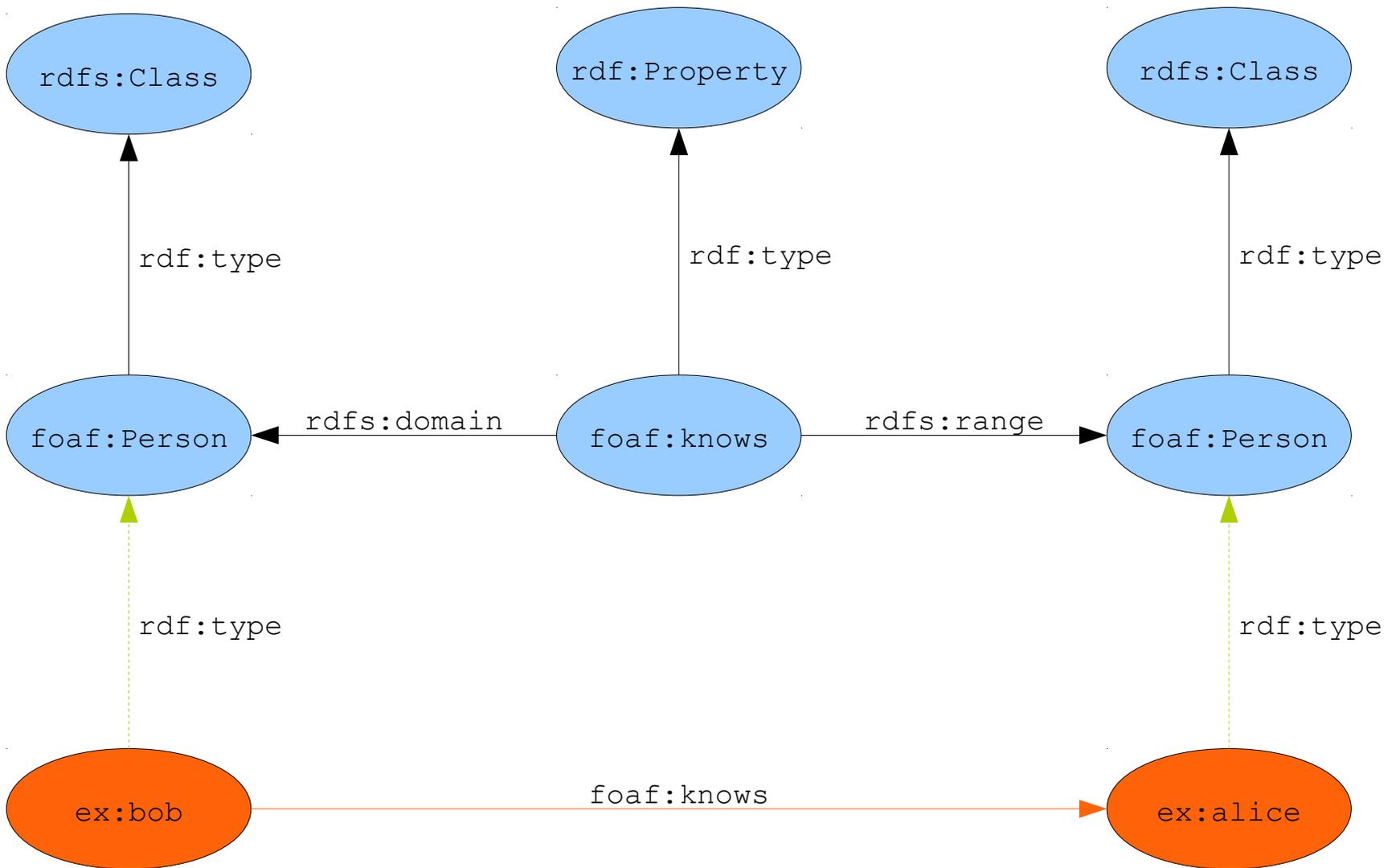
Your turn!

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 .
```

```
# RDF Schema
foaf:knows rdf:type rdfs:Property ;
           rdfs:range foaf:Person ;
           rdfs:domain foaf:Person .
foaf:Person rdf:type rdfs:Class .
```

```
# 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 .
```

```
# RDF Schema as a „bridge“ across vocabularies
ex:colleague rdfs:subPropertyOf foaf:knows ;
              rdfs:domain        ex:Employee ;
              rdfs:range         ex:Employee .
ex:Employee  rdf:type           rdfs:Class ;
              rdfs:subClassOf   foaf:Person .
```

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

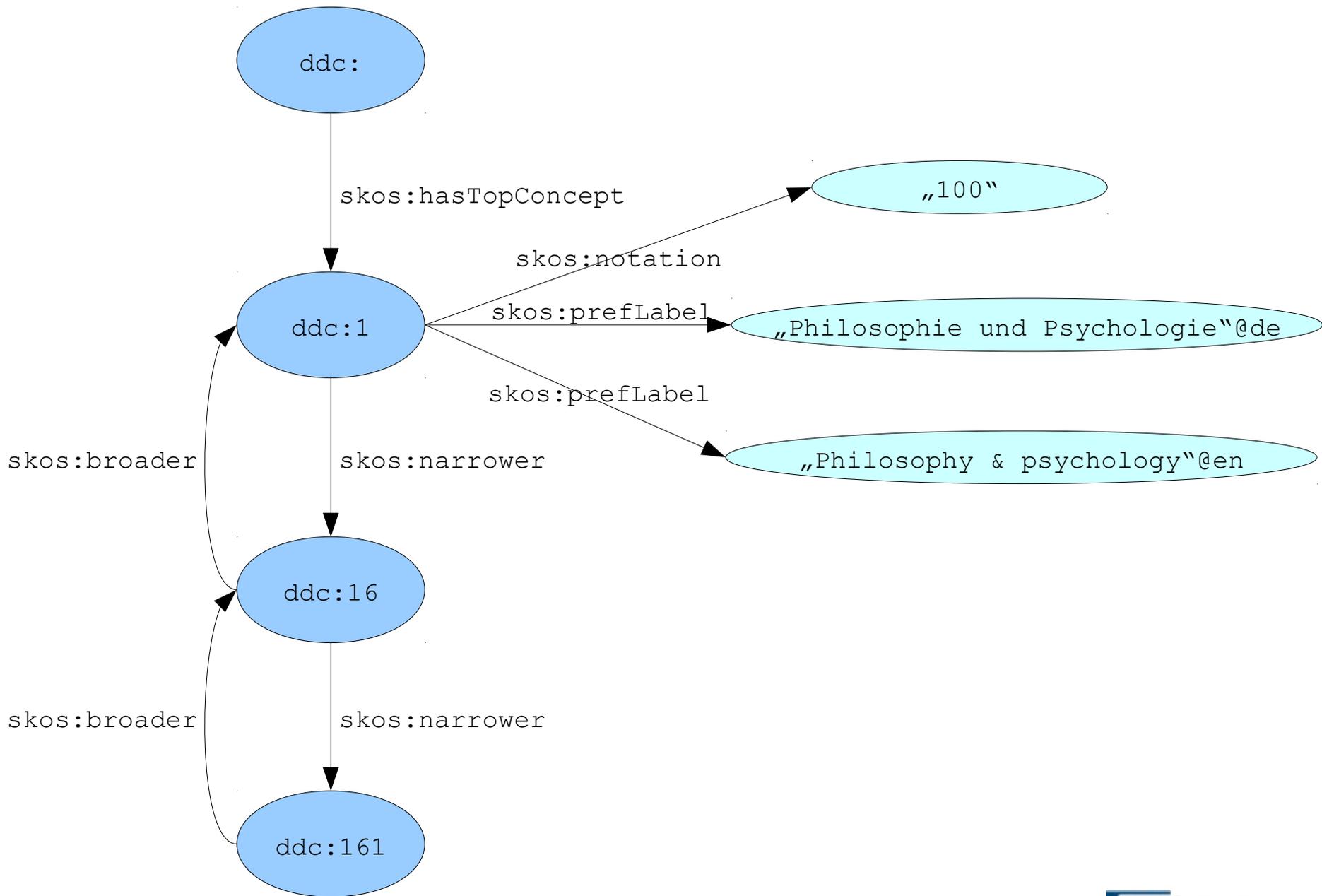
team:home      rdf:type    rdfs:Property ;
                rdfs:domain    team:Game .

team:away      rdf:type    rdfs:Property ;
                rdfs:domain    team:Game .

team:Game      rdf:type    rdfs:Class .
```

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.



Linked Data Principles

- 1) Use URIs as names for things.
- 2) Use HTTP URIs so that people can look up those names.
- 3) When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
- 4) Include links to other URIs. So that they can discover more things.

Thanks.

Questions?

Here and now or also later to
semweb@hbz-nrw.de

Credits

„Shunting“ by John Spooner CC-BY

<http://www.flickr.com/photos/johnspooner/2199685678/>

Exit arrow by ccarlstead CC-BY <http://www.flickr.com/photos/cristic/553170147/>

License



These slides are published under a Creative Commons license:

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