

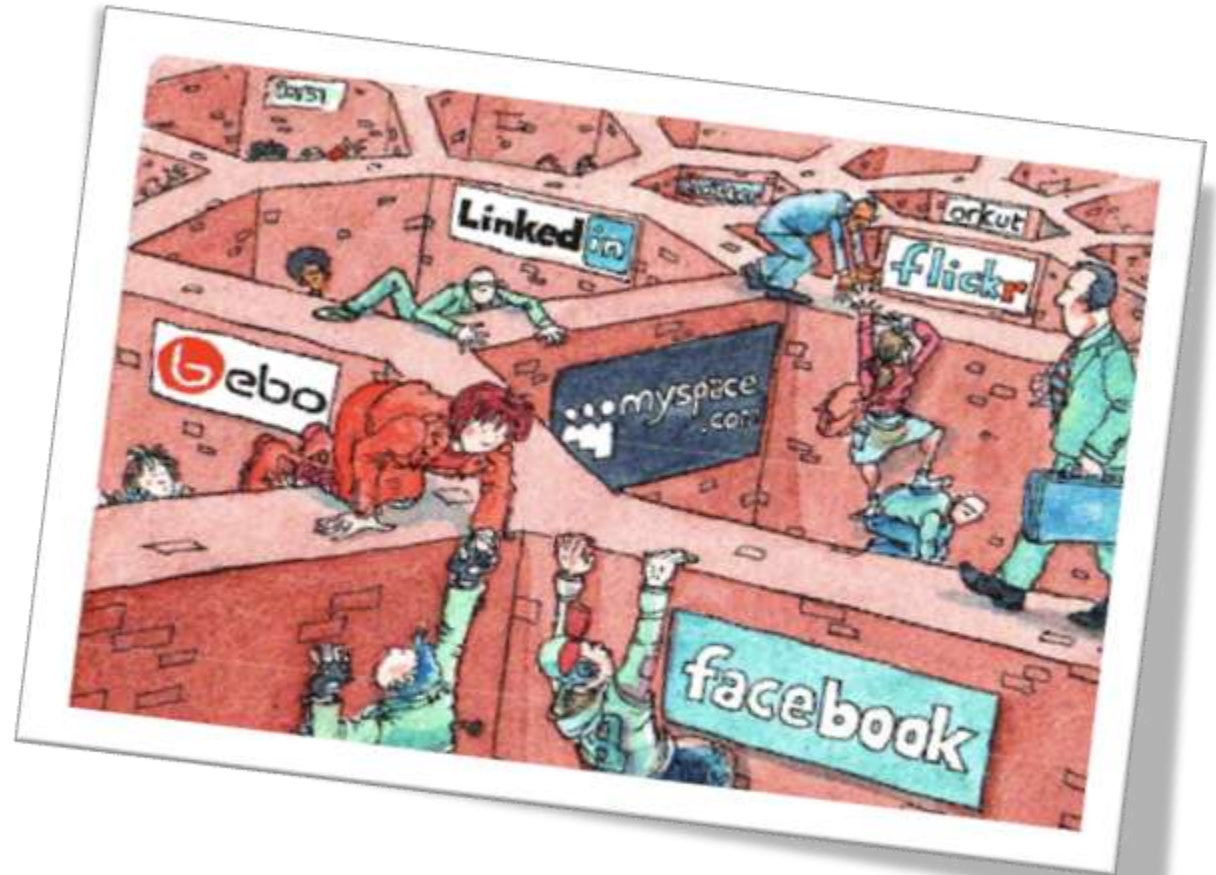


Linked Data, Microformats, RDFa, Microdata

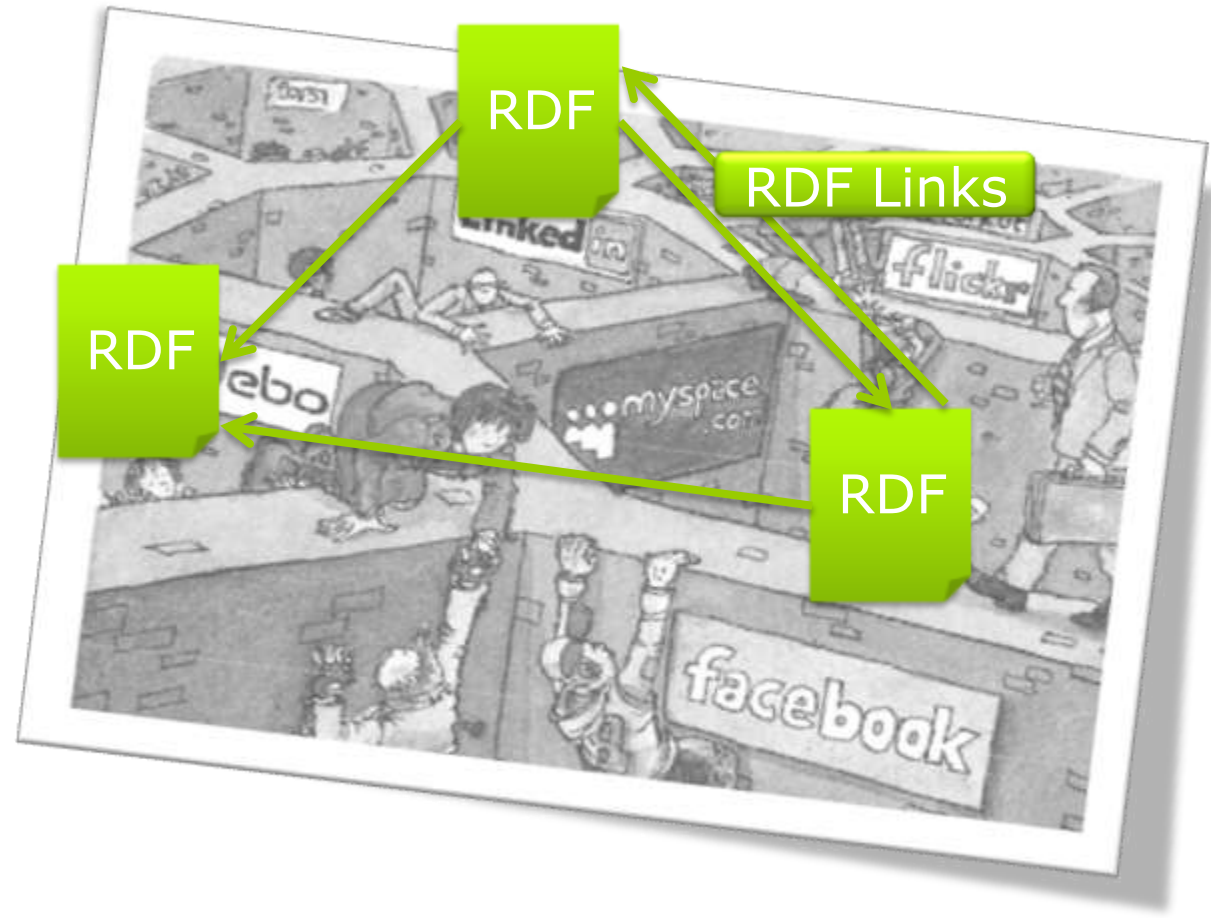
Markus Luczak-Rösch
Freie Universität Berlin
Institut für Informatik
Netzbasierende Informationssysteme
markus.luczak-roesch@fu-berlin.de

Probleme?

→ Data Silos



Web of Linked Data



Linked Data Prinzipien

1. URIs als Namen für alle "Dinge"

`http://dbpedia.org/resource/Berlin`

2. `http://` URIs damit man im Web auf diese Namen zugreifen kann

Content Negotiation



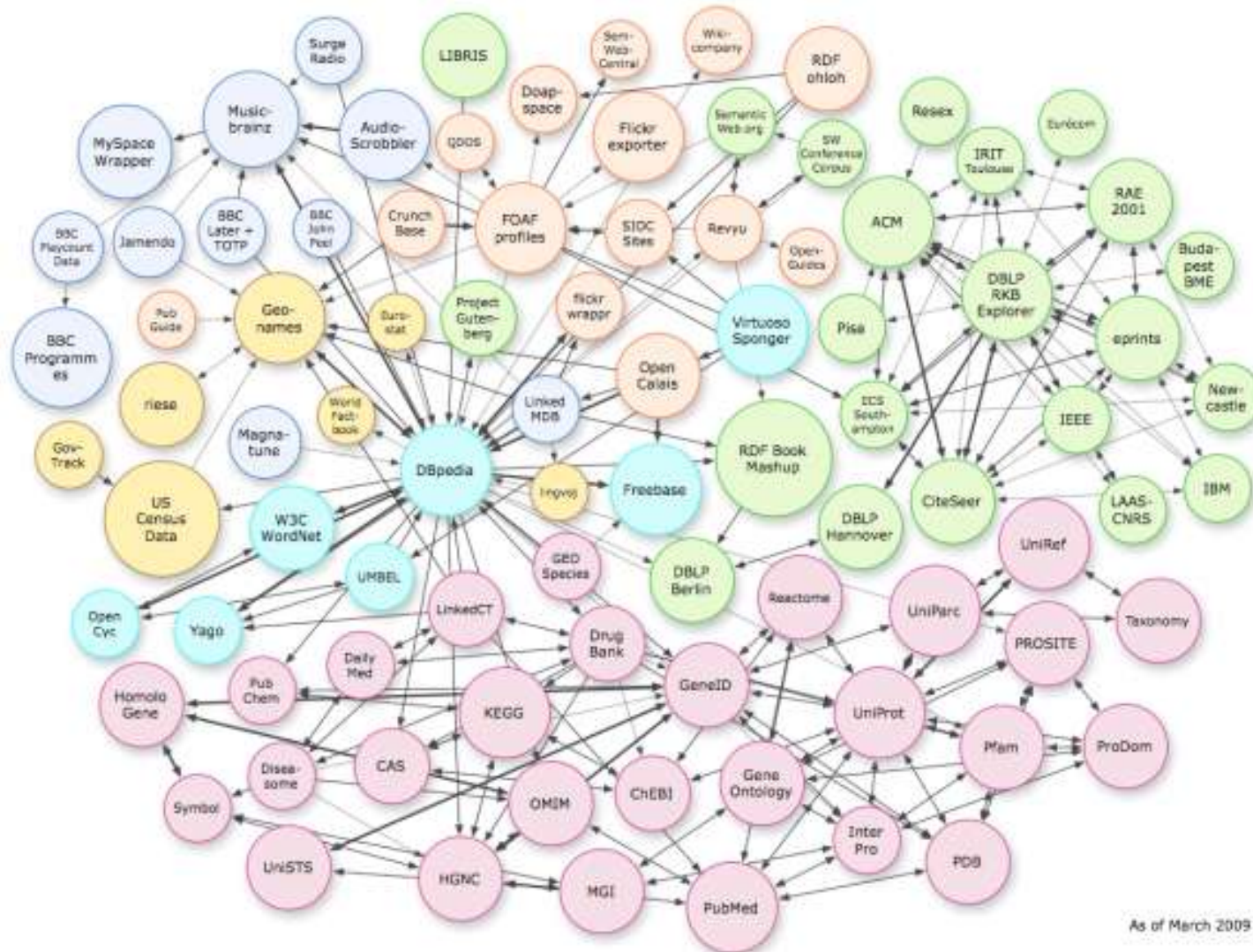
3. Wenn eine URI aufgerufen wird sollen sinnvolle Informationen entsprechend der Standards (RDF, SPARQL) geliefert werden

`http://dbpedia.org/page/Berlin`
`http://dbpedia.org/data/Berlin`

4. Links zu anderen URIs, damit Nutzer mehr "Dinge" finden können

`yago-res:Berlin` S
 owl:sameAs P
 dbpedia:Berlin O

Linking Open Data Cloud



Linked Data erzeugen

- Domäne festlegen
- Identifier
- Vokabular
- RDF-Links

Domäne festlegen, Daten verstehen

- Personen
- Geo
- Ämter
- ...
- Omas Kuchen

- **Klaus Wowereit**

- ist: regierender
Bürgermeister von
Berlin

- **Berlin**

- lat...
- long...
- ...

- **Thing**

- Mayor
- City
- ...

- Klaus Wowereit

- Mayor

http:// URIs

http://...

http://...

- Klaus Wowereit

- Mayor

eigener Namensraum

<http://dbpedia.org/...>

<http://dbpedia.org/...>

- Klaus Wowereit

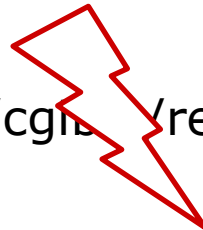
- Mayor

implementierungsunabhängig

<http://dbpedia.org/...>

<http://dbpedia.org/...>

http://www4.wiwiss.fu-berlin.de:2020/demos/dbpedia/cgi-bin/resources.php?id=Klaus_Wowereit



- Klaus Wowereit

- Mayor

Slash oder Hash?

http://dbpedia.org/daten.rdf#Klaus_Wowereit

<http://dbpedia.org/konzepte.owl#Mayor>

http://dbpedia.org/Klaus_Wowereit

<http://dbpedia.org/Mayor>

- Klaus Wowereit

- Mayor

Faktenwissen oder Konzeptwissen?

[http://dbpedia.org/resource/
Klaus_Wowereit](http://dbpedia.org/resource/Klaus_Wowereit)

[http://dbpedia.org/ontology/
Mayor](http://dbpedia.org/ontology/Mayor)

- Klaus Wowereit

http://dbpedia.org/resource/Klaus_Wowereit	← Ressource
http://dbpedia.org/page/Klaus_Wowereit	← HTML-Seite
http://dbpedia.org/data/Klaus_Wowereit	← Daten

- Klaus Wowereit

- http://dbpedia.org/resource/Klaus_Wowereit ← Ressource
- http://dbpedia.org/page/Klaus_Wowereit ← HTML-Seite
- http://dbpedia.org/data/Klaus_Wowereit ← Daten
- http://dbpedia.org/data/Klaus_Wowereit.rdf
- http://dbpedia.org/data/Klaus_Wowereit.ntriples

• Wiederverwenden

- Geo
- **FOAF**
- GoodRelations
- SIOC
- DOAP
- ...

[http://xmlns.com/foaf/0.1/
Person](http://xmlns.com/foaf/0.1/Person)

• Entwickeln

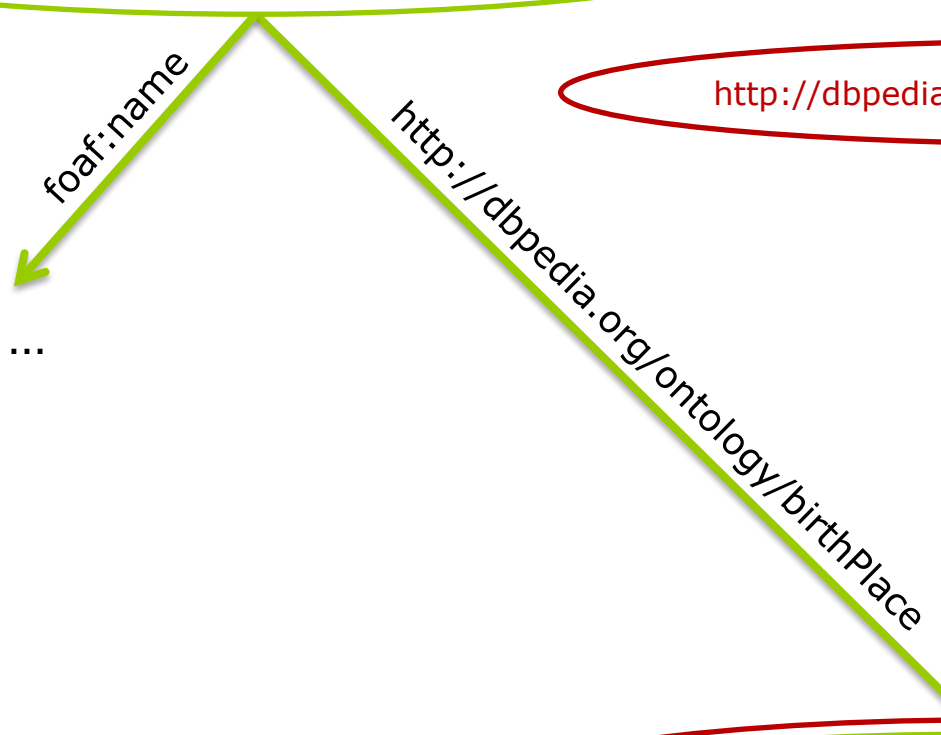
- Thing
 - **Person**
 - **OfficeHolder**
 - ...
 - ...

[http://dbpedia.org/ontology/
Person](http://dbpedia.org/ontology/Person)

[http://dbpedia.org/ontology/
OfficeHolder](http://dbpedia.org/ontology/OfficeHolder)

- Wiederverwenden nicht wiedererfinden!
- Mischen!
 - Geo <http://xmlns.com/foaf/0.1/Person>
 - FOAF
 - Dublin Core http://www.w3.org/2003/01/geo/wgs84_pos#lat
 - DBpedia Ontology <http://dbpedia.org/ontology/leader>
 - ... <http://dbpedia.org/ontology/City>

<http://www.markus-luczak.de/me>



http://dbpedia.org/resource/Klaus_Wowereit



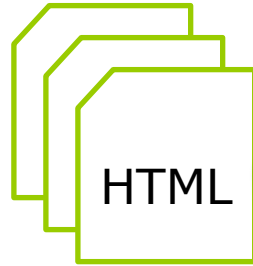
<http://dbpedia.org/resource/Berlin>

- owl:sameAs
- ...
- foaf:homepage
- foaf:topic
- foaf:based_near
- foaf:maker/foaf:made
- foaf:depiction
- rdfs:seeAlso
- ...

HTTP GET



`http://dbpedia.org/resource/Klaus_Wowereit`



HTML



RDF

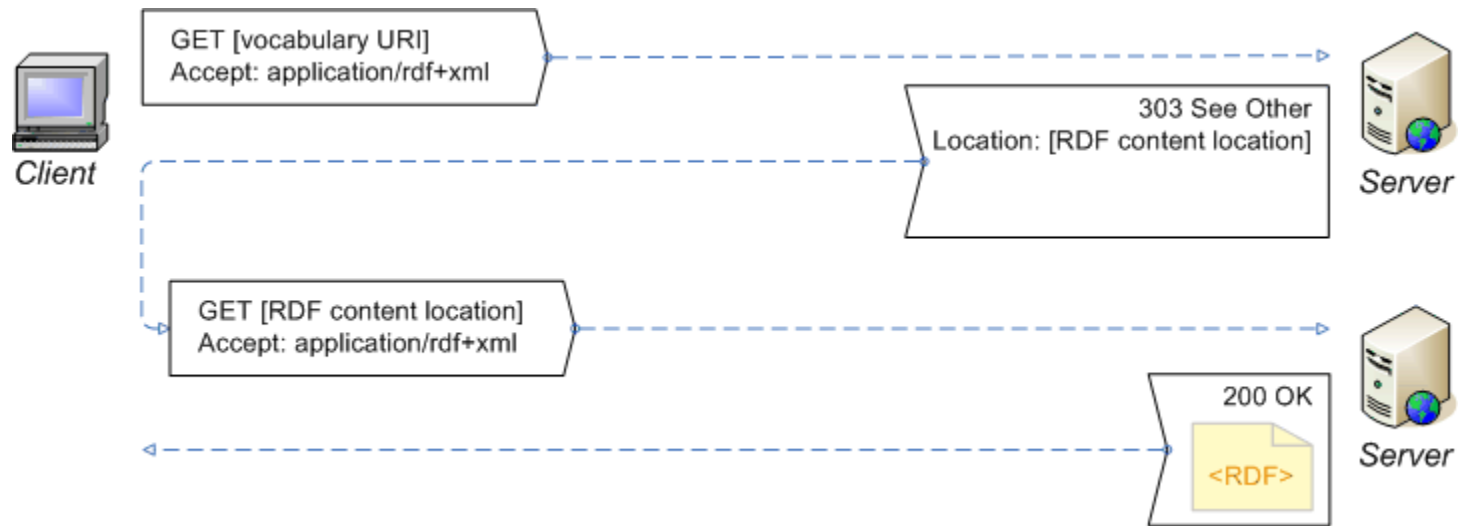
`http://dbpedia.org/page/Klaus_Wowereit`

`http://dbpedia.org/data/Klaus_Wowereit`

Linked Data Server
Infrastructure



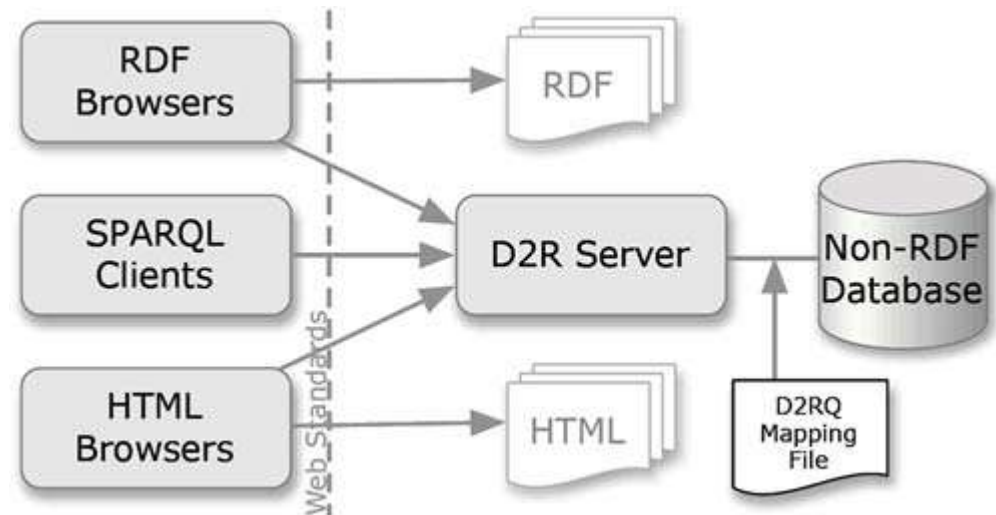
Data Source



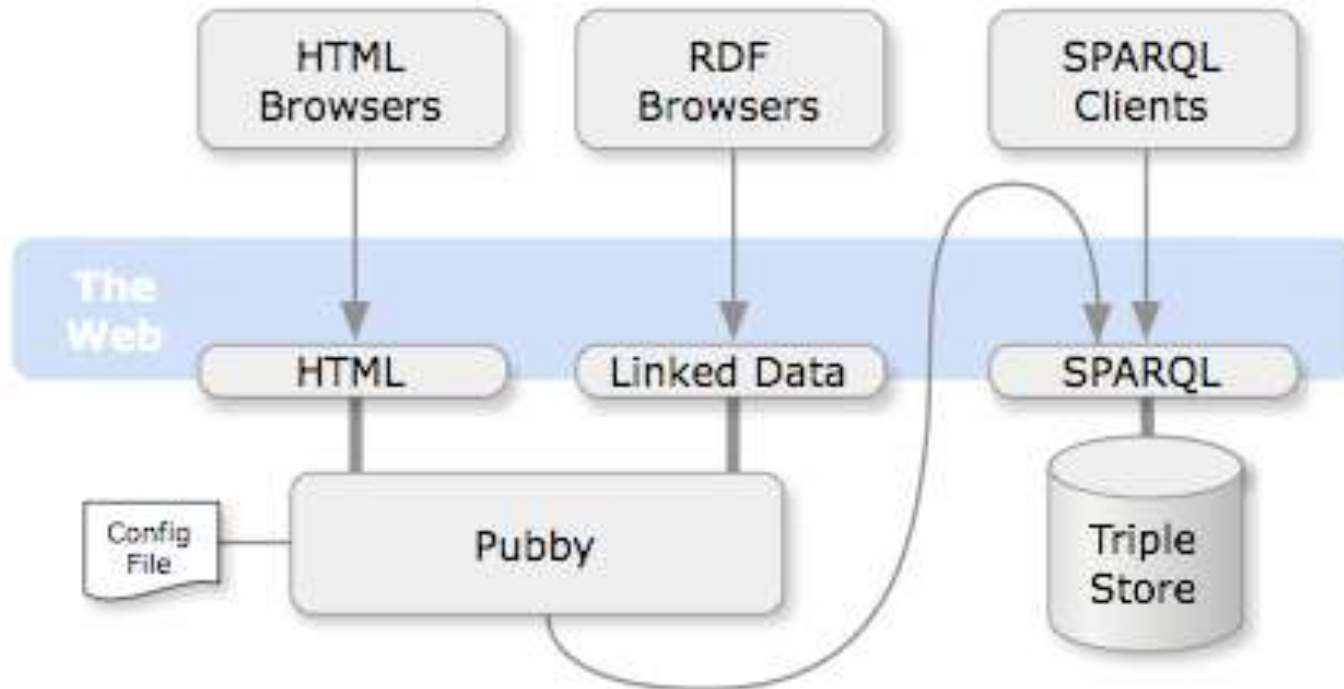
Linked Data anbieten/serven

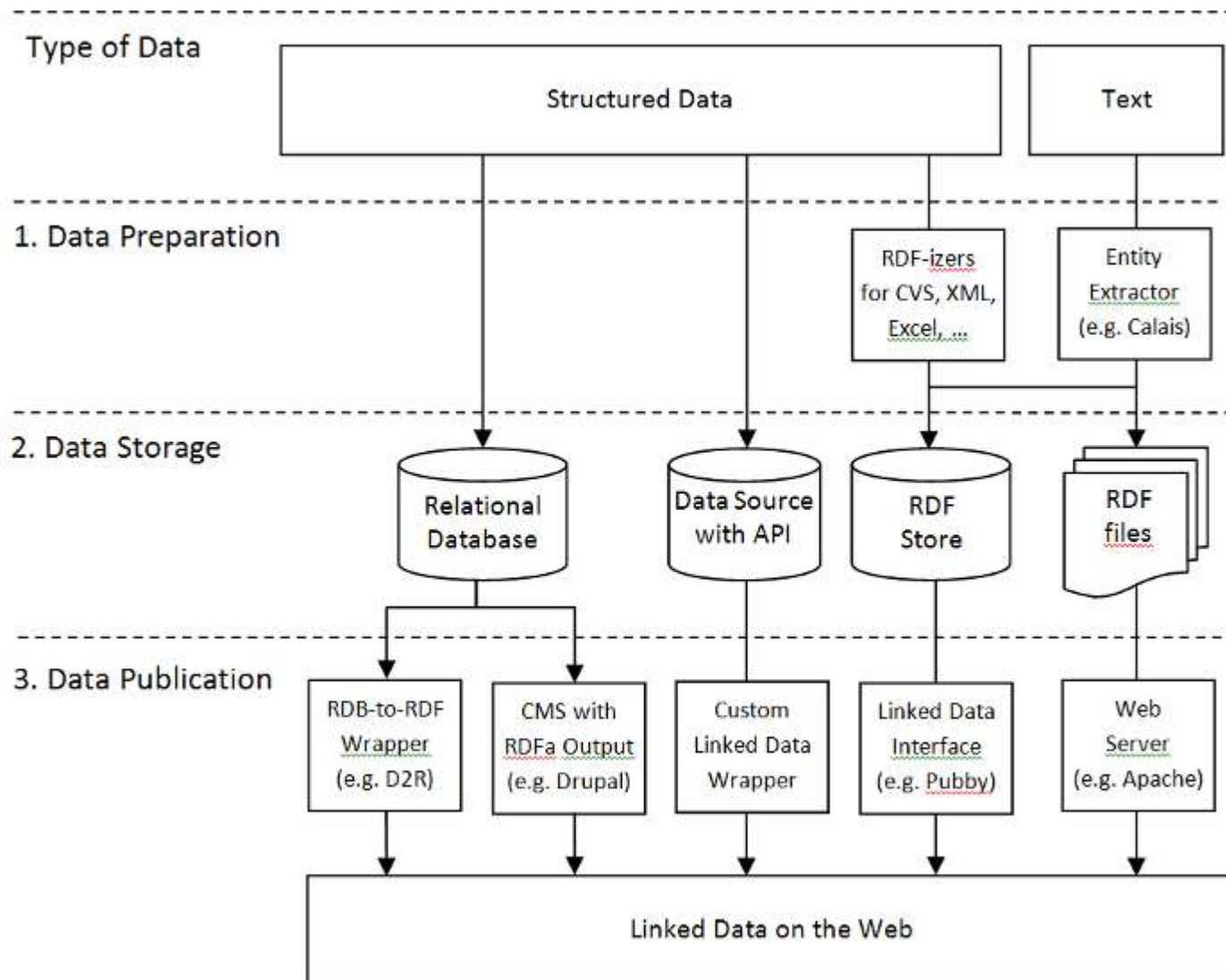
- statische RDF-Dateien
- relationale Datenbanken
- Linked Data Server
- API-Wrapper
- RDFa

- Mappen eines relationalen Schemas auf RDF (-Klassen und -Beziehungen)
 - deklarative Mapping-Sprache (D2RQ)
 - SPARQL-Endpoint
 - Linked Data Server



Linked Data Server (Bsp.: Pubby)





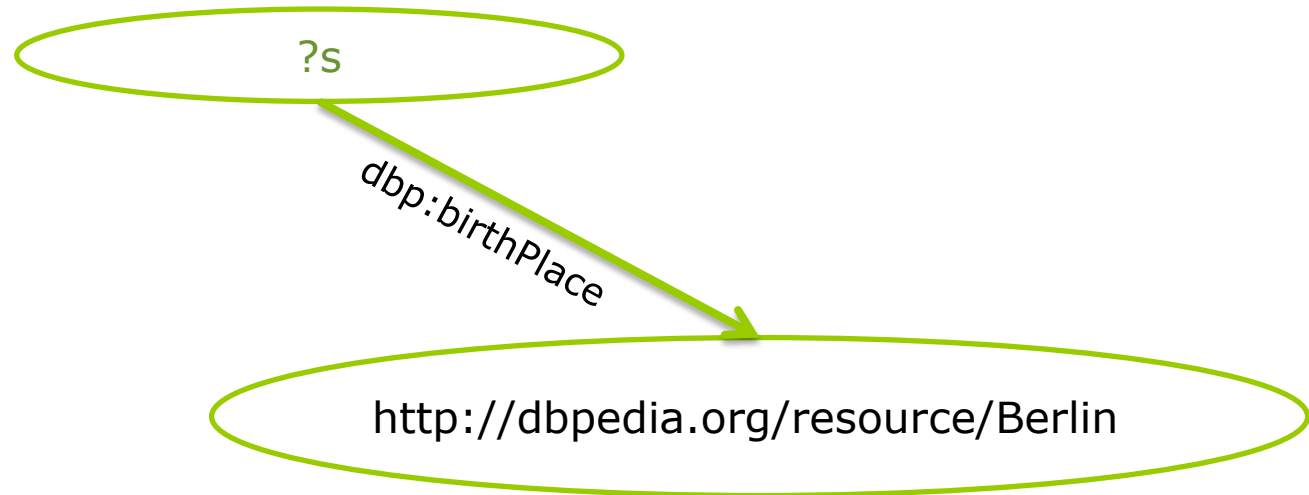
- URIs finden
- Zusätzliche Daten finden
- SPARQL-Endpoints finden

- Gibt es (bereits) einen Identifier für das, was mich interessiert?
 - Lookup-Services für Datensets
 - [http:// lookup.dbpedia.org](http://lookup.dbpedia.org)
 - <http://rkbexplorer.com>
 - ...
 - Web of Data Search Engines
 - <http://sindice.com>
 - <http://ws.nju.edu.cn/falcons/objectsearch/index.jsp>

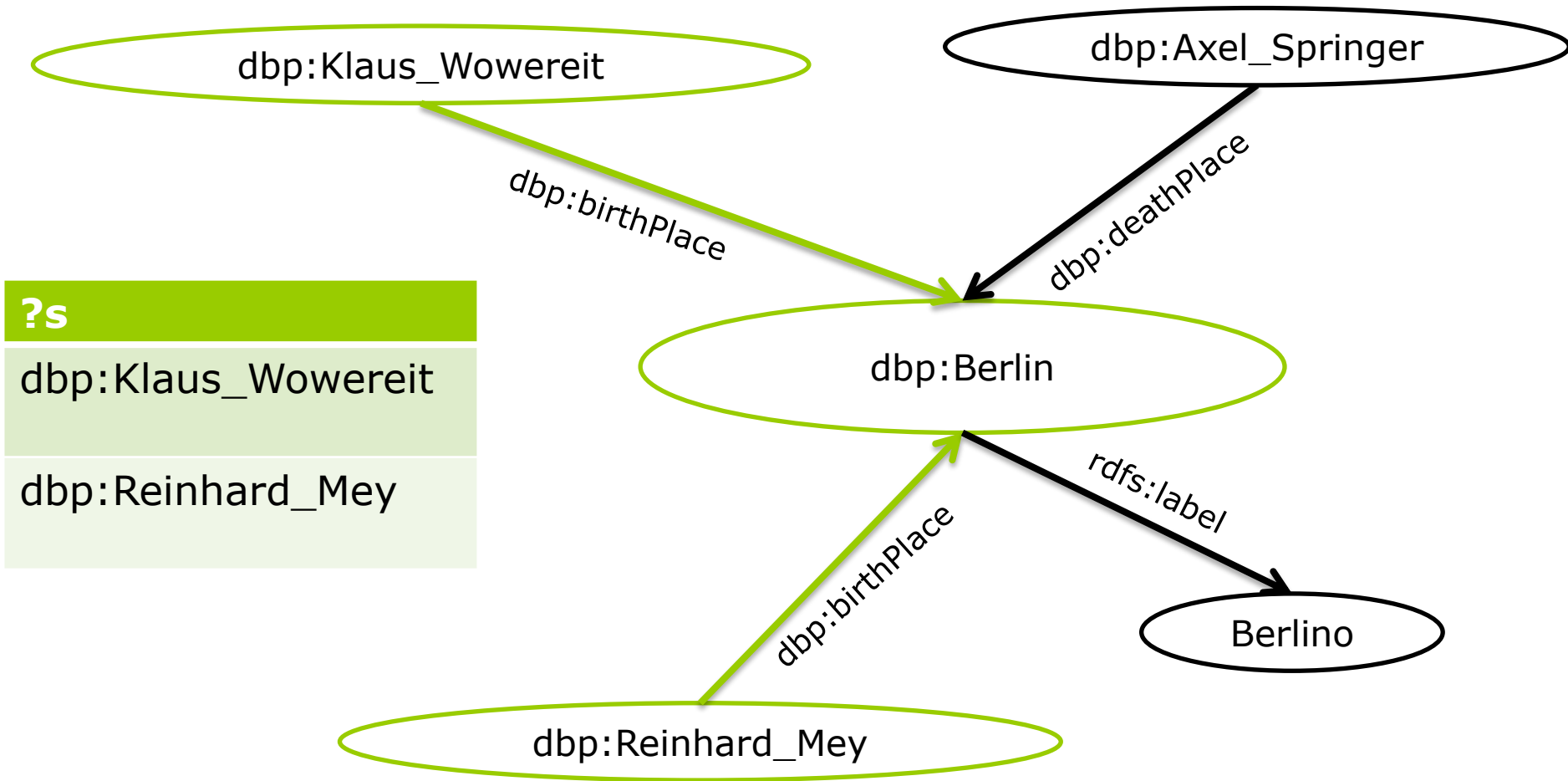
- Wo finde ich weitere Daten für die URI, die ich zur Hand habe?
 - Links verfolgen
 - `rdfs:seeAlso`
 - `owl:sameAs`
 - Co-Referenz-Dienst verwenden
 - <http://sameas.org>
 - Web of Data Search Engines

- <http://www.w3.org/wiki/SparqlEndpoints> (depr.)
- <http://thedatahub.org/group/lodcloud>

- SPARQL: Anfragesprache für RDF-Daten
- Grundsatz: Pattern-Matching
 - beschreibe Graphpattern
 - frage RDF-Graph mit diesem Pattern an
 - Subgraphen, die Pattern matchen kommen in die Ergebnismenge



SPARQL-Anfragen über Linked Data



- Anfrage via RESTful Service mit Parameter *query*
GET /sparql?query=PREFIX+rdf... HTTP/1.1
Host: dbpedia.org
- Antwort liefert ein SPARQL-Result-Format (XML oder JSON)

```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
  <head>
    <variable name="x"/>
    <variable name="hpage"/>
  </head>

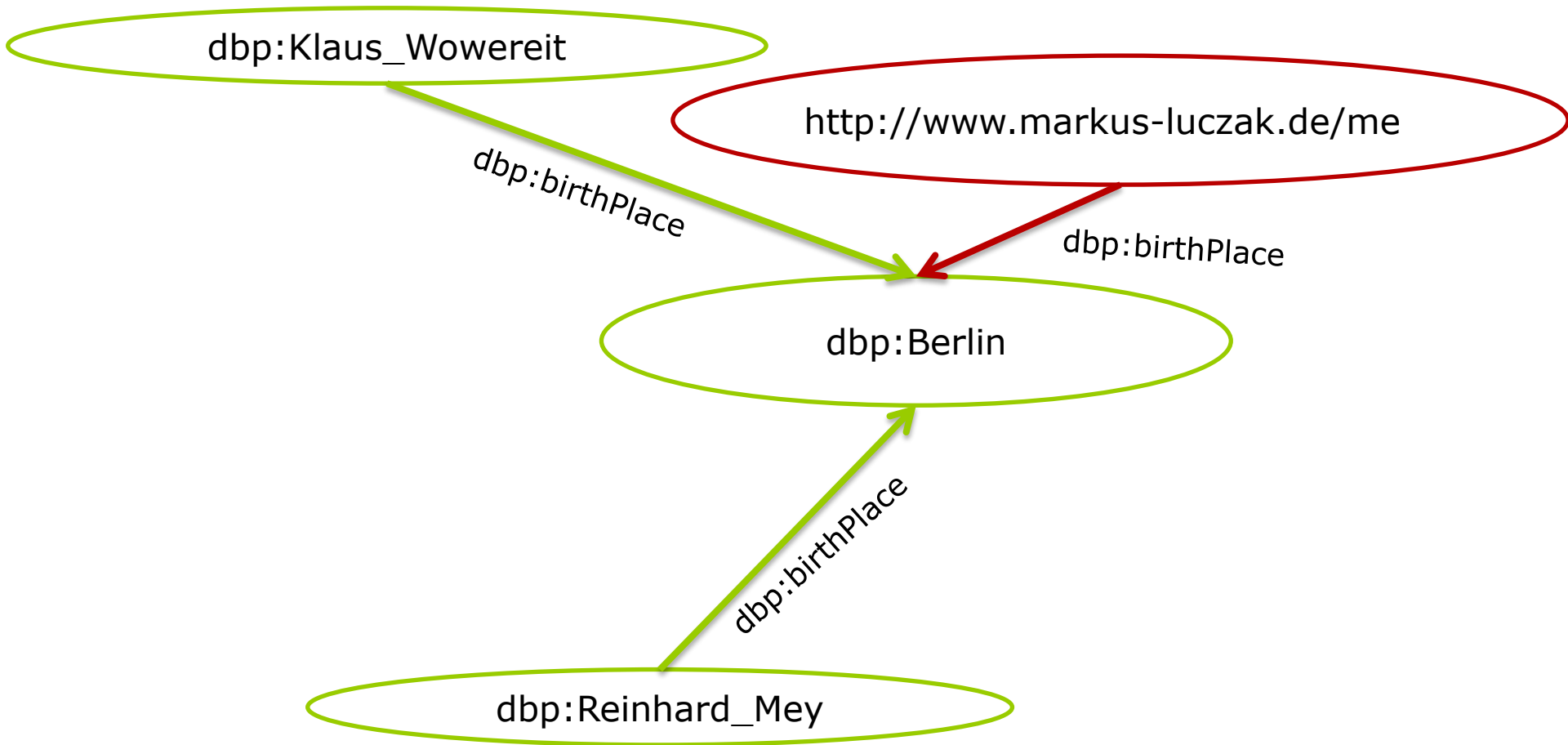
  <results>
    <result>
      <binding name="x"> ... </binding>
      <binding name="hpage"> ... </binding>
    </result>

    <result>
      <binding name="x"> ... </binding>
      <binding name="hpage"> ... </binding>
    </result>
    ...
  </results>
</sparql>
```

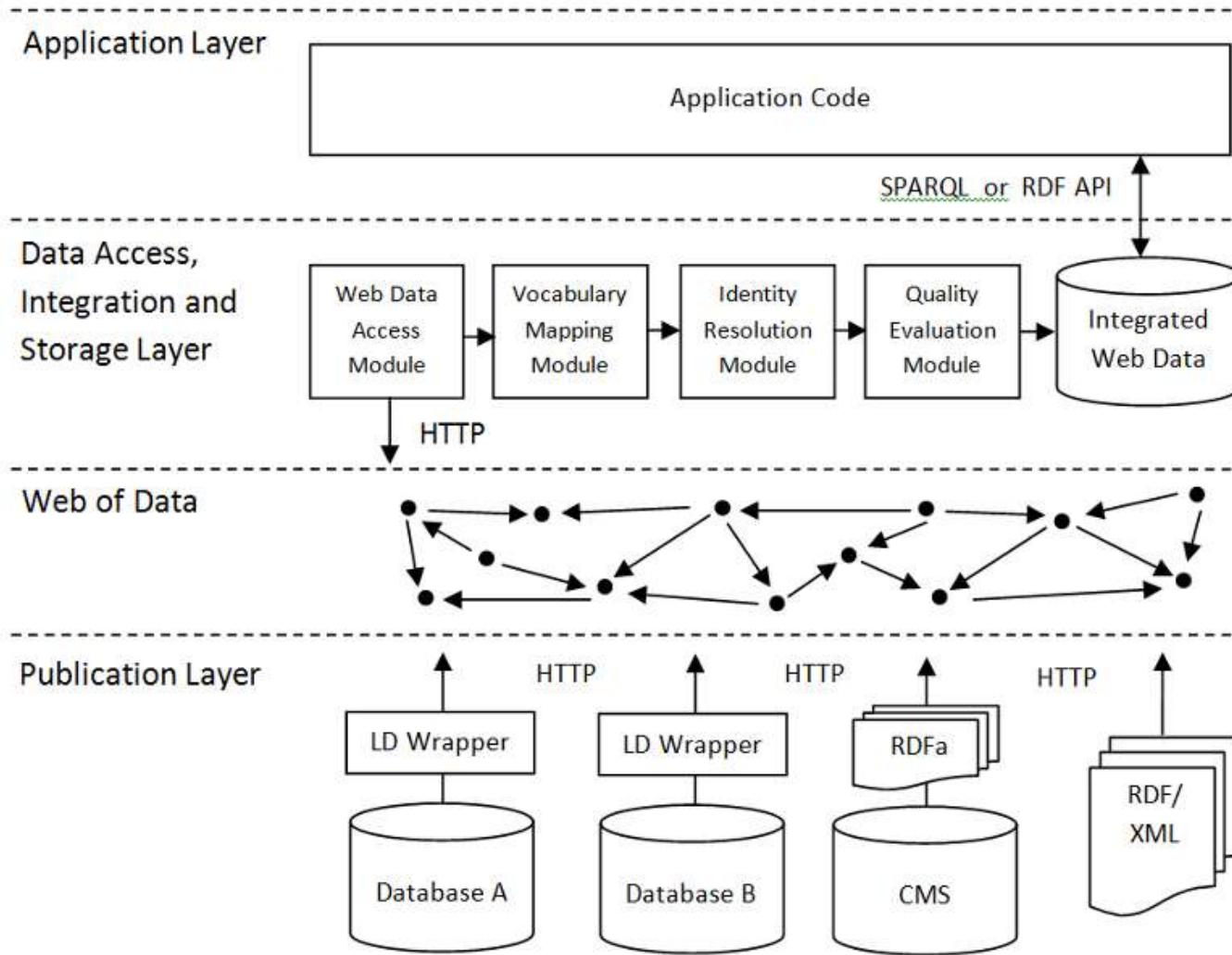
<http://www.w3.org/TR/rdf-sparql-XMLres/>

```
{
  "head": { "vars": [ "book", "title" ]
  },
  "results": {
    "bindings": [
      {
        "book": { "type": "uri", "value": "http://example.org/book/book6" },
        "title": { "type": "literal", "value": "Harry Potter and the Half-Blood Prince" }
      },
      {
        "book": { "type": "uri", "value": "http://example.org/book/book5" },
        "title": { "type": "literal", "value": "Harry Potter and the Order of the Phoenix" }
      },
      {
        "book": { "type": "uri", "value": "http://example.org/book/book4" },
        "title": { "type": "literal", "value": "Harry Potter and the Goblet of Fire" }
      },
      {
        "book": { "type": "uri", "value": "http://example.org/book/book3" },
        "title": { "type": "literal", "value": "Harry Potter and the Prisoner Of Azkaban" }
      },
      {
        "book": { "type": "uri", "value": "http://example.org/book/book2" },
        "title": { "type": "literal", "value": "Harry Potter and the Chamber of Secrets" }
      },
      {
        "book": { "type": "uri", "value": "http://example.org/book/book1" },
        "title": { "type": "literal", "value": "Harry Potter and the Philosopher's Stone" }
      }
    ]
  }
}
```

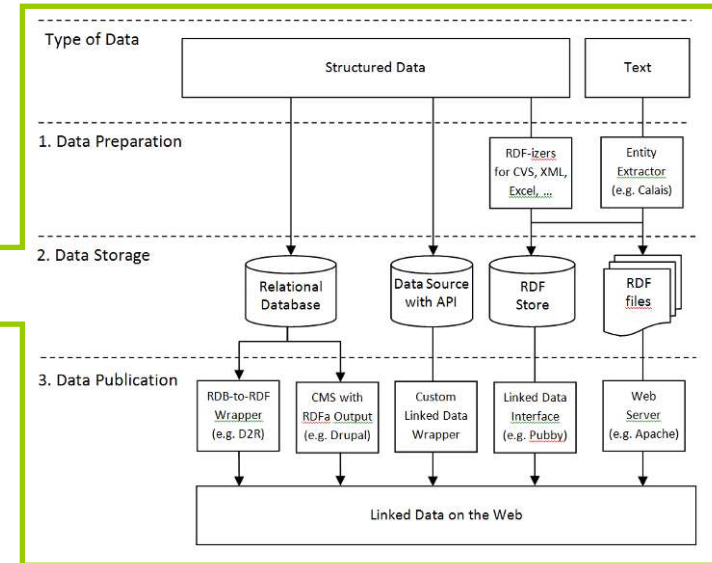
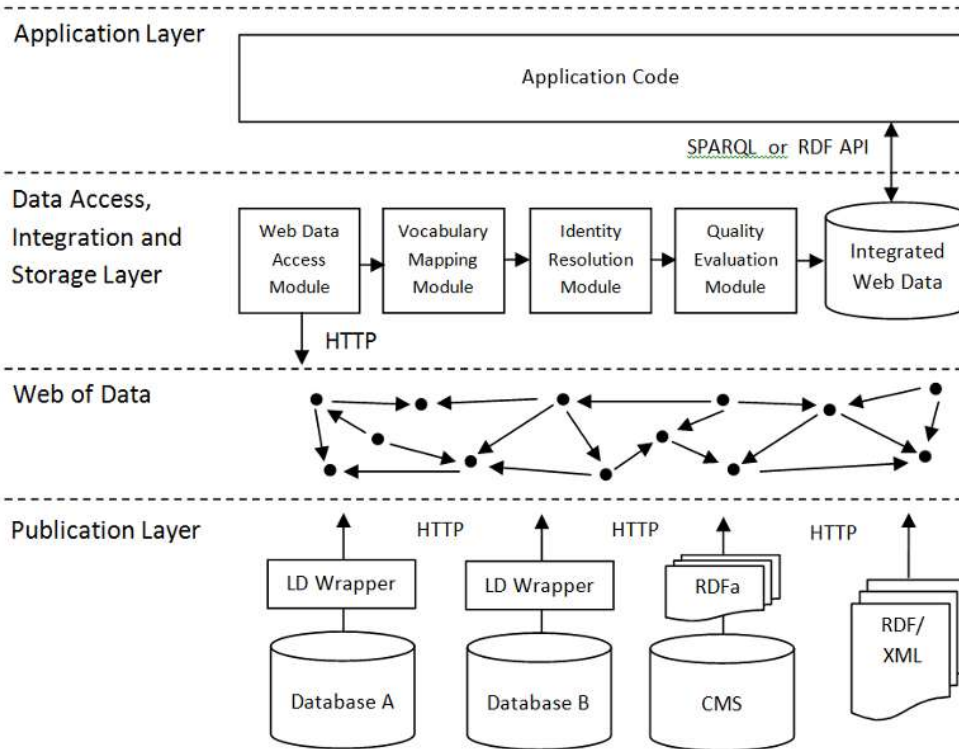
<http://www.w3.org/TR/rdf-sparql-json-res/>

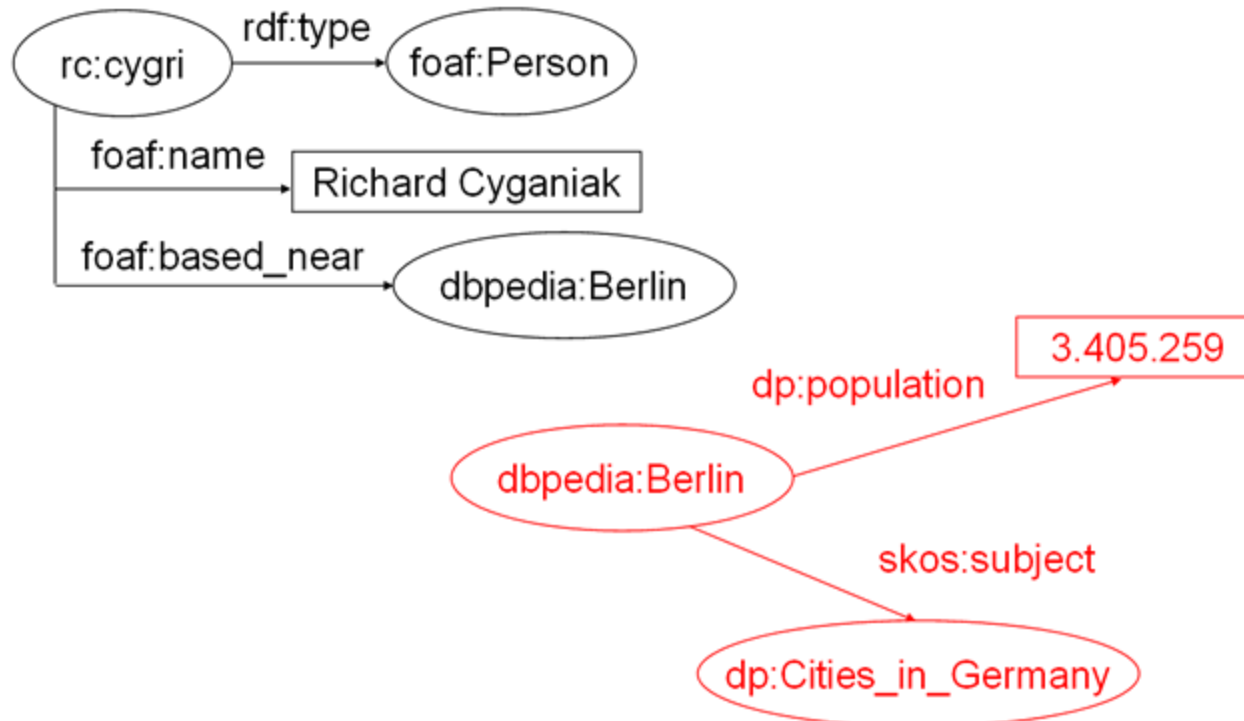


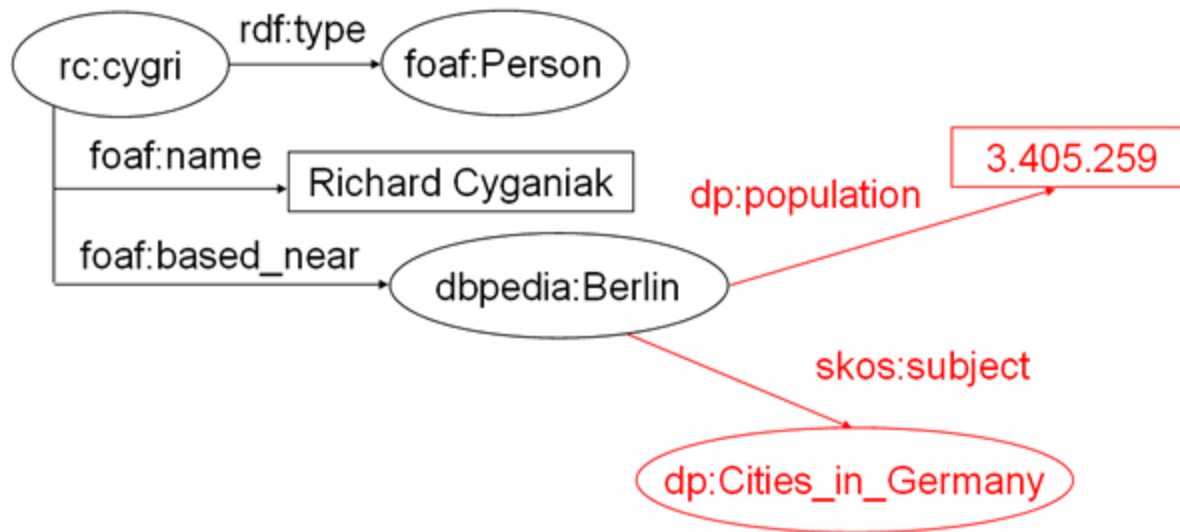
- verteilte Datenhaltung erzeugt Herausforderungen für Anfragen
- **Anfrageansätze**
 - follow-up Anfragen ← anwendungsspezifisch, proprietär
 - zentrale Datensammlung anfragen ← eher trivial
 - föderierte Anfrage ← interessanter
 - Idee: Mediator anfragen, der Subanfragen an relevante Quellen verteilt und Ergebnis integriert
 - Link-Traversierung zur Anfragezeit ← sehr interessant
 - Idee: verfolge Links in gefundenen Ergebnissen, um dynamisch den Datenbestand zu vergrößern und beziehe diese zusätzlichen Daten dann in die Anfrageauswertung mit ein

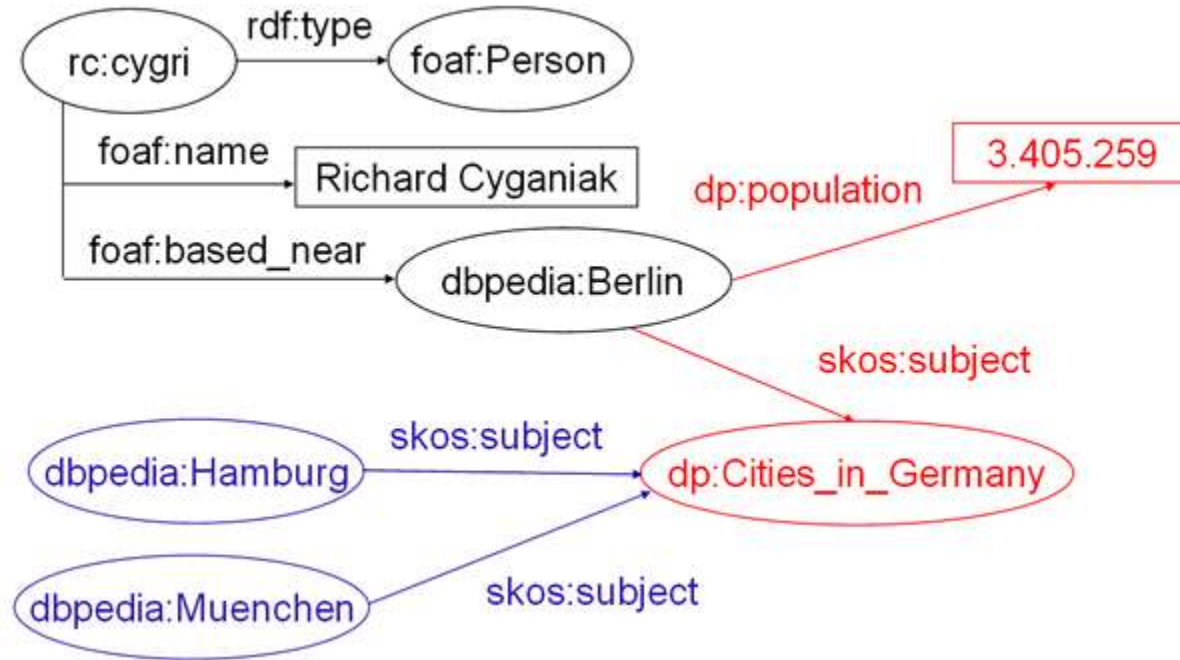


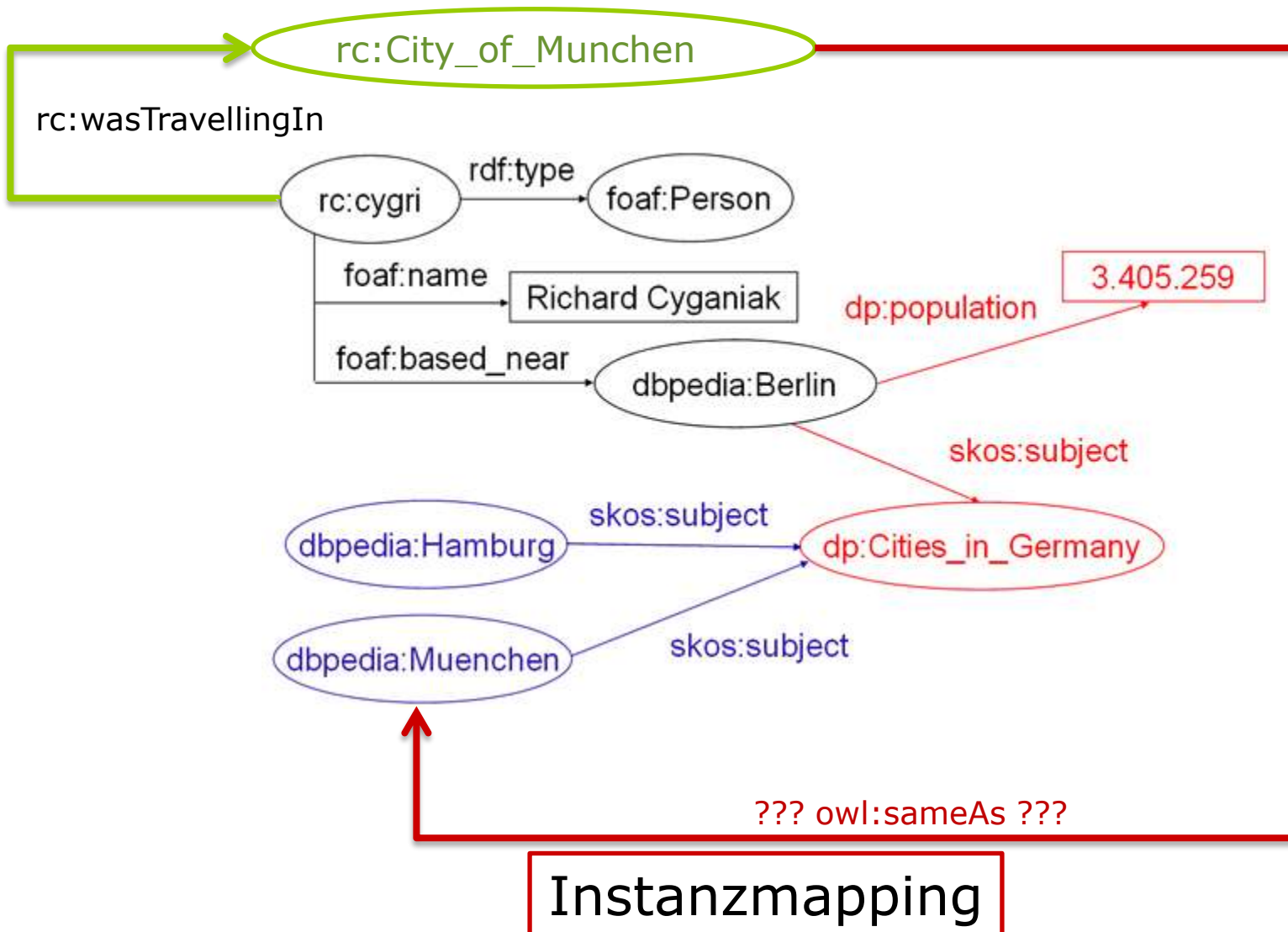
Linked Data Anwendungen











<http://xmlns.com/foaf/0.1/Person> owl:sameAs  <http://dbpedia.org/ontology/Person>

- owl:sameAs
- owl:equivalentClass
- owl:equivalentProperty

Schemamapping

Shopping-Ergebnisse für **galaxy tab**



- [Samsung Galaxy Tab 10.1 16 GB - Android 3.0 \(Honeycomb ...](#)
★★★★★ 10 Erfahrungsberichte - 354 € - 59 Anbieter
- [Samsung Galaxy Tab WiFi 16 GB - Android 2.2 1 GHz](#)
★★★★★ 18 Erfahrungsberichte - 279 € - 100 Anbieter
- [Samsung Galaxy Tab 16 GB - Android 2.2 1 GHz](#)
★★★★★ 33 Erfahrungsberichte - 222 € - 94 Anbieter

oliver kahn birthday

About 222,000 results (0.43 seconds)

June 15, 1969

Oliver Kahn, Date of birth
[Show details](#)

[Oliver Kahn \(Eine Vuvuzela zum Geburtstag \) "Happy Birthday Olli ...](#)



www.youtube.com/watch?v=kEdvFUoPJdo

15 Jun 2010 - 2 min - Uploaded by Hallodeutschland10

Oliver Kahn ist heute(15.06.2010) . 41 geworden und die

- strukturierte Daten in **auf einfachste Weise** in HTML einbetten
 - zwei einfache Einheiten
 - Entities
 - Properties
 - drei wichtige HTML-Attribute
 - class
 - rel
 - rev
 - vordefinierte Standardvokabulare (Erweiterung durch poshformats; microformats \subset poshformats, POSH = Plain Old Semantic HTML)

- Robert Tolksdorf und Markus Luczak-Rösch haben die Arbeitsgruppen-Web-Site erstellt.

```
<div>
```

```
<a href=„http://www.robert-tolksdorf.de/Robert“>  
    Robert Tolksdorf
```

```
</a>
```

```
und
```

```
<a href="http://www.markus-luczak.de/#me">  
    Markus Luczak-Rösch
```

```
</a>
```

```
haben die Arbeitsgruppen-Web-Site erstellt.
```

```
</div>
```

```
<div>
```

```
<a href="http://www.robert-tolksdorf.de/Robert"
  rel="author">
  Robert Tolksdorf
```

```
</a>
```

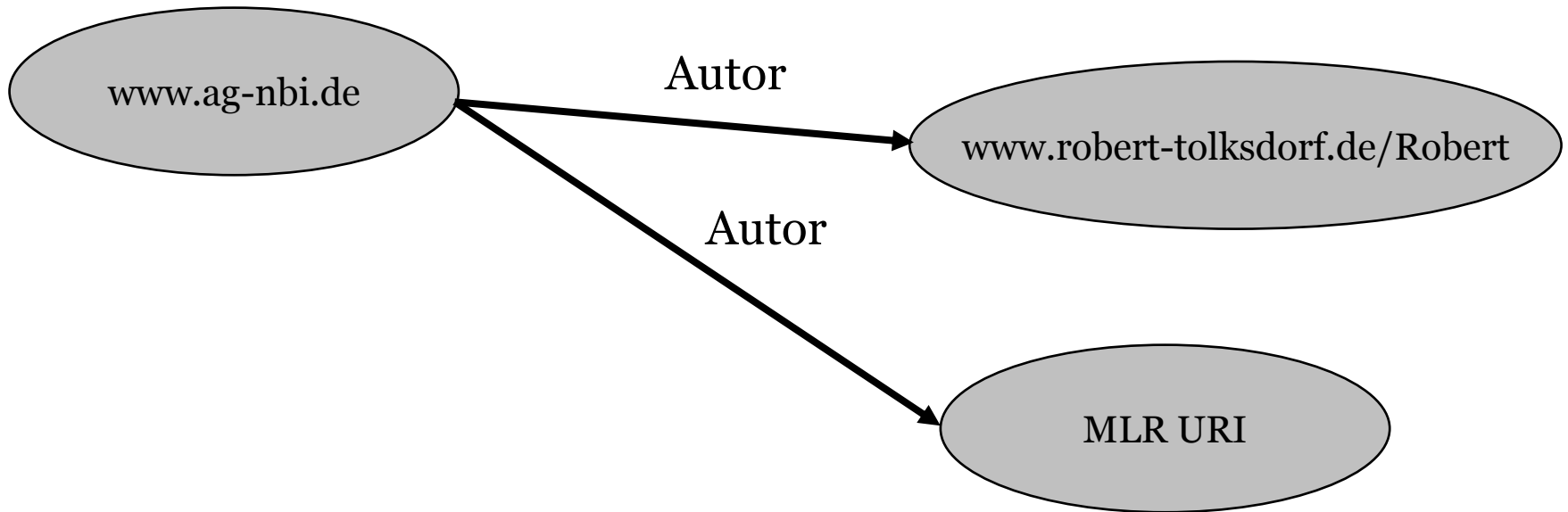
```
und
```

```
<a href="http://www.markus-luczak.de/"
  rel="author">
  Markus Luczak-Rösch
```

```
</a>
```

```
haben die Arbeitsgruppen-Web-Site erstellt.
```

```
</div>
```



```
<div>  
  <a href="http://www.markus-luczak.de/" >  
    Markus Luczak-Rösch  
  </a>  
  arbeitet für die  
  <a href="http://www.fu-berlin.de/" >  
    FU Berlin  
  </a>.  
</div>
```

```
<div><span class="vcard">  
  <a class="fn url"  
    href="http://www.markus-luczak.de/" >  
    Markus Luczak-Rösch  
</a></span>  
arbeitet für die  
  <span class="vcard">  
    <a class="fn org url"  
      href="http://www.fu-berlin.de/" >  
      FU Berlin  
    </a>  
  </span>.  
</div>
```

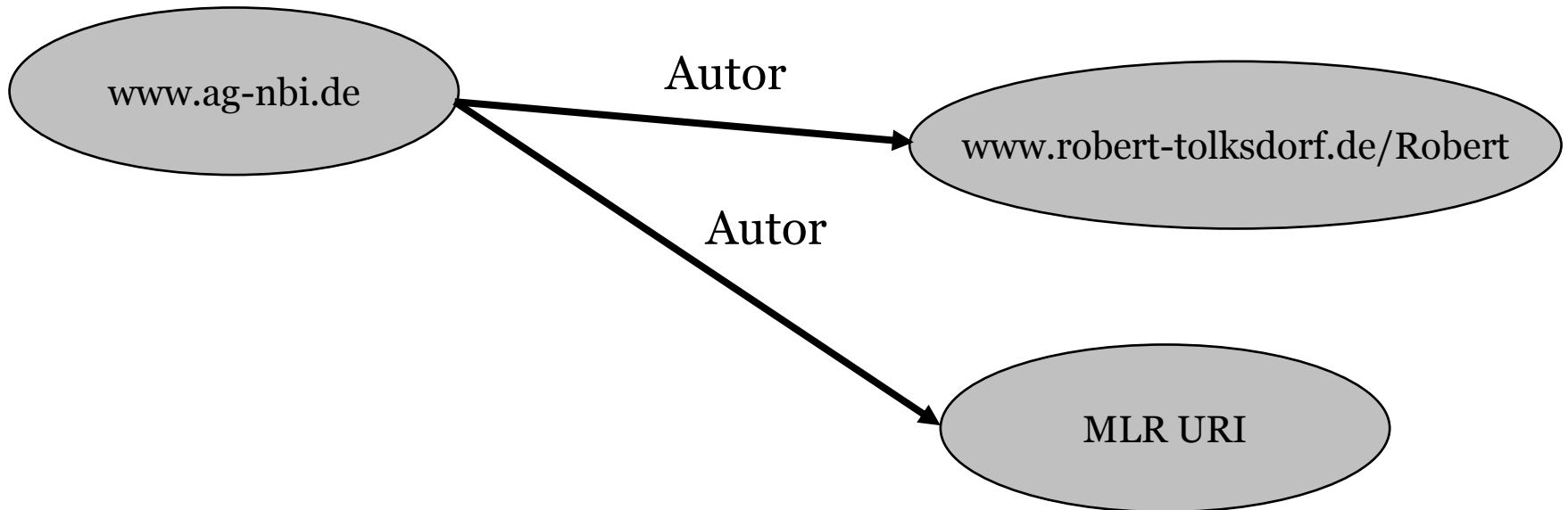
```
<div class="vcard">  
  <span class="fn" >  
    Markus Luczak-Rösch  
  </span>  
  arbeitet für die  
  <span class="org" >  
    FU Berlin  
  </span>.  
</div>
```

```
<div class="vcard">
  <a class="fn url"
    href="http://www.markus-luczak.de/" >
    Markus Luczak-Rösch
  </a>
  arbeitet für die
  <span class="org vcard" >
    <a class="url fn org"
      href="http://www.fu-berlin.de" >
      FU Berlin</a>
    </span>.
</div>
```


- **RDF Daten** in HTML einbetten
 - vollständiges RDF Datenmodell
 - Mischung von Schemata möglich
 - wichtige HTML-Attribute
 - about
 - property
 - rel
 - rev

```
<div about="" rel="dcterms:author">  
  <a href="http://www.robert-tolksdorf.de/Robert">  
    Robert Tolksdorf  
  </a>  
  und  
  <a href="http://www.markus-luczak.de/#me">  
    Markus Luczak-Rösch  
  </a>  
  haben die Arbeitsgruppen-Web-Site erstellt.  
</div>
```

<> dcterms:author < http://www.robert-tolksdorf.de/Robert > ,
< http://www.markus-luczak.de/#me > .



- einfache Art, um strukturierte Daten („maschinenlesbare Labels“) in HTML einzubetten
 - Standard ab HTML 5
 - itemprop
 - itemref
 - content
 - itemscope
 - itemtype
 - verschachtelte Name-Wert-Paare
 - Gruppe von Name-Wert-Paaren → Item
 - atomares Name-Wert-Paar → Property
 - Zugriff via Microdata DOM API

<div>

Robert Tolksdorf

und

Markus Luczak-Rösch

haben die Arbeitsgruppen-Web-Site erstellt.

</div>

```
<div><span itemscope>  
  <a itemprop="url"  
    href="http://www.robert-tolksdorf.de/Robert">  
    <span itemprop="name">  
      Robert Tolksdorf </span>  
  </a> </span>  
und <span itemscope>  
  <a itemprop="url"  
    href="http://www.markus-luczak.de/#me">  
    <span itemprop="name">  
      Markus Luczak-Rösch </span>  
  </a> </span>  
haben die Arbeitsgruppen-Web-Site erstellt.  
</div>
```

```
<div><span itemscope
  itemtype="http://example.org/people/professor">
  <span itemprop="name">
    Robert Tolksdorf
  </span> </span>
und
<span itemscope
  itemtype="http://example.org/people/lecturer">
  <span itemprop="name">
    Markus Luczak-Rösch</span> </span>
haben die Arbeitsgruppen-Web-Site erstellt.
</div>
```

```
<div><span itemscope
  itemtype="http://example.org/people/professor,,
  itemid= "http://www.robert-tolksdorf.de/Robert">
  <a itemprop="name"
    href="http://www.robert-tolksdorf.de/Robert">
    Robert Tolksdorf
  </a> </span>
  und
  ...
</div>
```



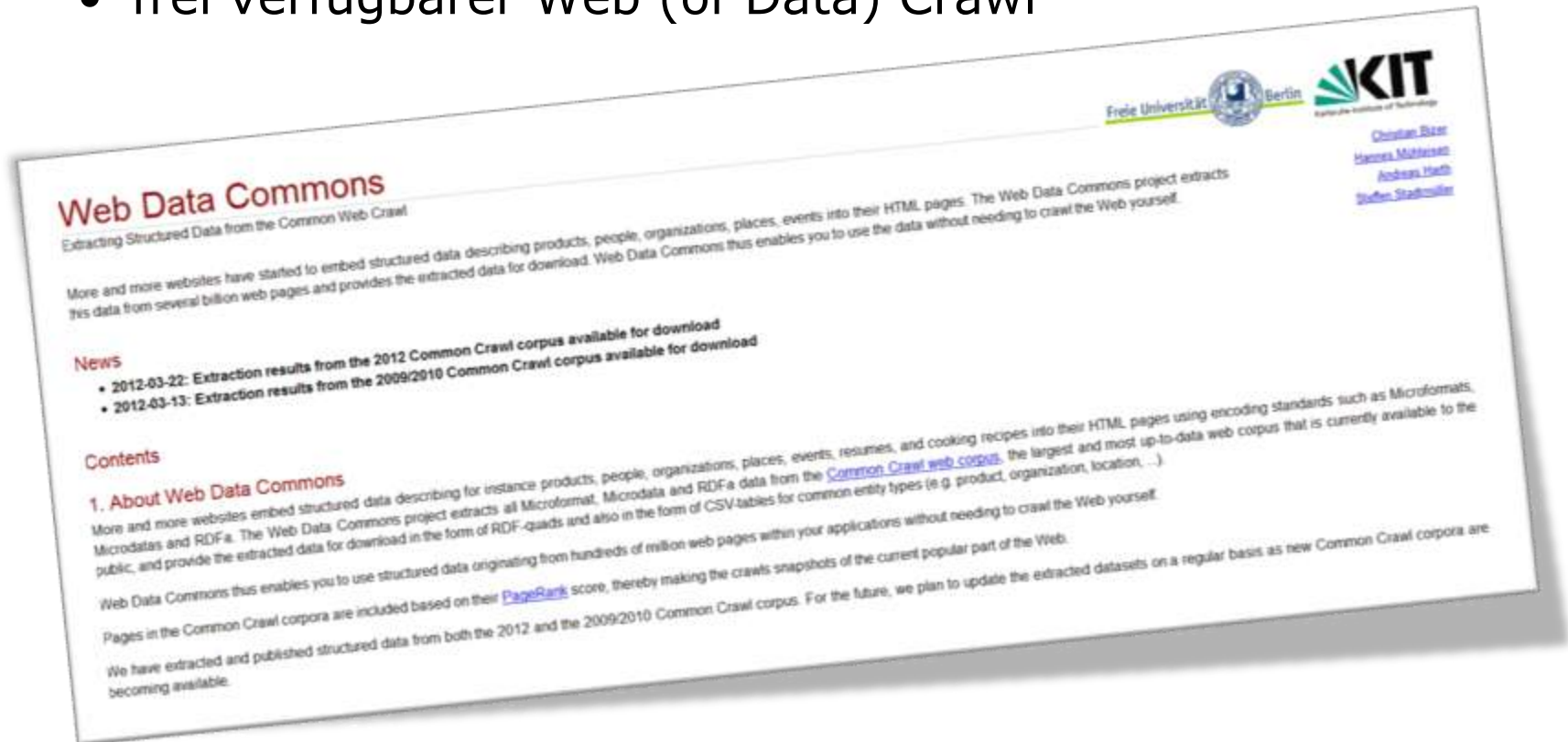
```
<div><span itemscope
  itemtype="http://example.org/people/professor,,
  itemid= "http://www.robert-tolksdorf.de/Robert">
  <span itemprop="name">
    Robert Tolksdorf
  </span>
  <meta itemprop="knows"
    content="Markus Luczak-Rösch" />
  </span>
  und
  ...
</div>
```

```
<div><span itemscope
  itemtype="http://example.org/people/professor,,
  itemid= "http://www.robert-tolksdorf.de/Robert">
  <span itemprop="name">
    itemref="meta-knows"
    Robert Tolksdorf
  </span>
  <meta id="meta-knows" itemprop="knows"
    content="Markus Luczak-Rösch" />
  </span>
  und
  ...
</div>
```

- Microdata Vokabular für die (meisten) prominenten Anwendungsfälle im Web
 - erstellt und betrieben von Google, Yahoo und Microsoft



- Extraktion strukturierter Daten aus dem Common Crawl Korpus
 - frei verfügbarer Web (of Data) Crawl



Microdata, RDFa, Microformats

Feature	RDFa 1.1	Microdata 1.0	Microformats 1.0
Relative Complexity	High	Medium	Low
Data Model	Graph	Tree	Tree
Item optionally identified by IRI	Yes	Yes	No
Item type optionally specified by IRI	Yes	Yes	No
Item properties specified by IRI	Yes	Yes	No
Multiple objects per page	Yes	Yes	Yes
Overlapping objects	Yes	Yes	No
Plain Text properties	Yes	Yes	Yes
IRI properties	Yes	Yes*	No
Typed Literal properties	Yes	No	No
XML Literal properties	Yes	No	No
Language tagging	Yes	Yes	Inconsistent

<http://manu.sporny.org/2011/uber-comparison-rdfa-md-uf/>

Microdata, RDFa, Microformats

Feature	RDFa 1.1	Microdata 1.0	Microformats 1.0
Override text and IRI content	Yes	No	Text only
Clear mapping to RDF	Yes	Problematic	No
Target Languages	8 (XHTML1, HTML4, HTML5, XHTML5, XML, SVG, ePub, OpenDocument)	2 (HTML5, XHTML5)	4 (XHTML1, HTML4, HTML5, XHTML5)
New Attributes	8 about, datatype, profile, prefix, property, resource, typeof, vocab	5 itemid, itemprop, itemref, itemscope, itemtype	0
Re-used Attributes	5 content, href, rel, rev, src	5 content, src, href, data, datetime	4 class, title, rel, href
Multiple IRI types per object	Yes	RDF only	No
Multiple statements per element	Yes	No	Yes
“Locally scoped” vocabulary terms	Yes, via vocab	Yes, via itemscope	No
Item Chaining	Yes	Basic	No

Microdata, RDFa, Microformats

Feature	RDFa 1.1	Microdata 1.0	Microformats 1.0
Transclusion	No	Yes	Yes, via include pattern
Compact IRIs	Yes	No	No
Prefix rebinding	Yes	No	No
Vocabulary Mashups	Yes	No	No
HTML5 time element support	Not yet	Yes	No
Different attributes for different property types	Yes property for text, rel/rev for URLs, resource/content for overrides	No	Yes class for text and rel for URLs
Transform to JSON	Yes (RDFa API)	Yes (Parser and Microdata DOM API)	No
DOM API	Yes	Yes	No
Unified Parser	Yes	Yes	No

- Web of Data = Linked Data + Microformats + RDFa + Microdata
- Einfachheit (Microformats, Microdata) vs. Flexibilität (Linked Data, RDFa)

- Linked Data Prinzipien
- URIs
- Content Negotiation
- SPARQL-Basics
- vergleich der Datenmodelle (Graph vs. Baum)

- Tutorium: Ihre Fragen
- Vorlesungstermin: Projektauftrakt