

Vorlesung Semantic Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vorlesung im Wintersemester 2011/2012

Dr. Heiko Paulheim

Fachgebiet Knowledge Engineering

Linked Open Data

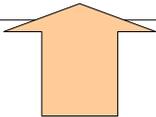
- Was wir bisher kennen gelernt haben:
 - RDF als universelle Beschreibungssprache
 - RDF Schema zur Definition von Konzepten und Vokabularen
- Linked Open Data
 - nutzt diese Techniken
 - stellt Daten im Web zur Verfügung
- Die Linked Open Data Cloud
 - hat nichts mit Cloud Computing zu tun
 - ist eine große, freie Informationssammlung

Linked Open Data

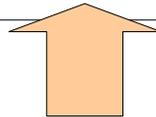
– was heißt hier "Linked"?



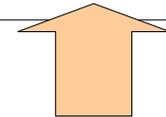
```
:p a :Physician .
:p :hasDegree "Dr." .
:p :hasName "Mark Smith" .
:p :hasAddress :a .
:a :street "Main Street" .
:a :number "14"^^xsd:int .
:a :city "Smalltown" .
:p :hasOpeningHours [
  a rdf:Bag ;
  [ :day :Monday;
    :from "9"^^xsd:int;
    :to "11"^^xsd:int;
  ]
  ...
]
```



```
:s a :City .
:s :name "Smalltown" .
:s :lat "49.86"^^xsd:double .
:s :long "8.65"^^xsd:double .
:s :district "Birmingham" .
...
```



```
:d a :District .
:d :name "Birmingham" .
:d :pop "347891"^^xsd:int .
:d :locatedIn "England" .
...
```



Linked Open Data

– was heißt hier Linked?

- Information ist im Web verstreut
- Bei HTML gibt es schon ein Konzept dazu:
 - den Hyperlink
 - Mehr Infos beim `W3C`
- Dieses Konzept macht sich auch Linked Open Data zu nutze!

Linked Open Data

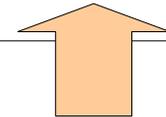
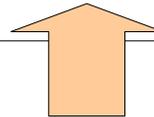
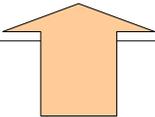
– was heißt hier "Linked"?



```
:p a :Physician .
:p :hasDegree "Dr." .
:p :hasName "Mark Smith" .
:p :hasAddress :a .
:a :street "Main Street" .
:a :number "14"^^xsd:int .
:a :city
  <http://.../smalltown> .
:p :hasOpeningHours [
  a rdf:Bag ;
  [ :day :Monday;
    :from "9"^^xsd:int;
    :to "11"^^xsd:int;
  ]
  ...
```

```
:s a :City .
:s :name "Smalltown" .
:s :lat "49.86"^^xsd:double .
:s :long "8.65"^^xsd:double .
:s :district
  <http://.../birmingham> .
  ...
```

```
:d a :District .
:d :name "Birmingham" .
:d :pop "347891"^^xsd:int .
:d :locatedIn "England" .
  ...
```



Linked Open Data

– was heißt hier "Linked"?

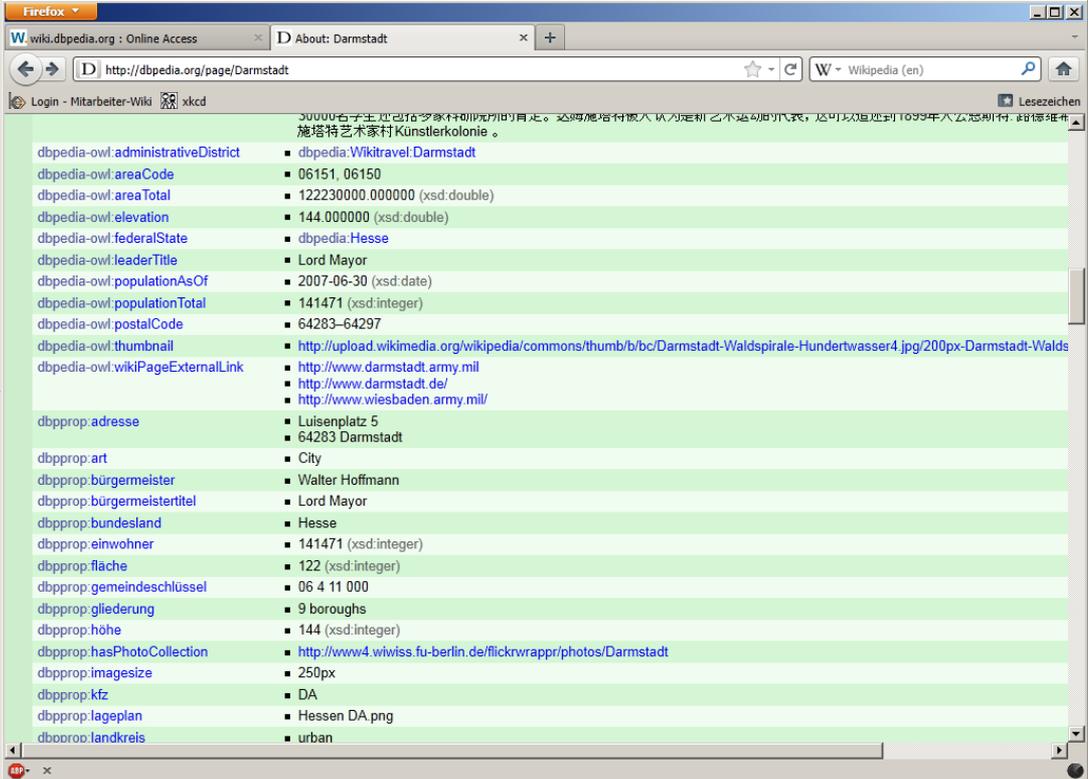


- Linked Open Data sind RDF-Daten
 - verteilt vorgehalten
- URIs
 - haben wir bis jetzt einfach als Identifier genutzt
 - sind in Linked Open Data Verweise auf andere Daten
 - die sich auflösen lassen!
 - "dereferenzierbare URIs" (URLs)
 - Für parallele Nutzung mit HTML Content Negotiation oder RDFa einsetzen

Linked Open Data – was heißt hier "Linked"?

▪ Beispiel:

- `<#Heiko> :worksIn <http://dbpedia.org/resource/Darmstadt> .`



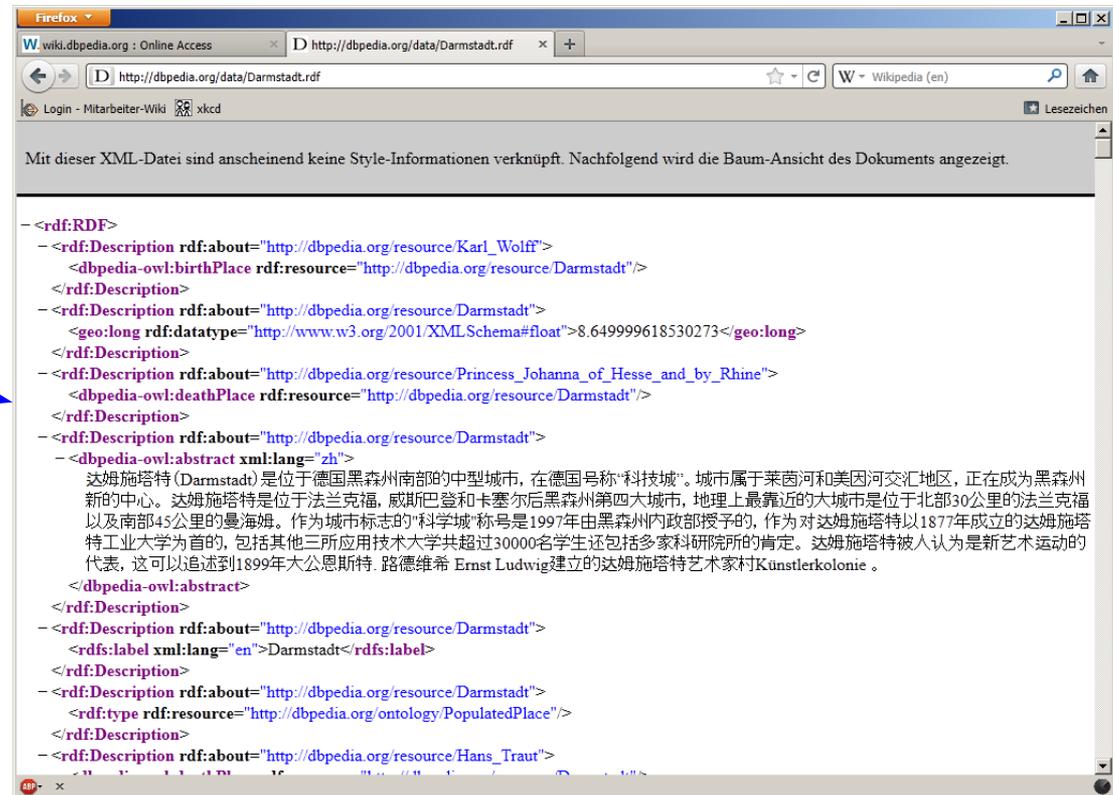
The screenshot shows a Firefox browser window with the address bar containing `http://dbpedia.org/page/Darmstadt`. The main content area displays a list of properties and their values for Darmstadt, including administrative district, area code, population, and postal code. A blue arrow points from the URL `http://dbpedia.org/resource/Darmstadt` in the list item above to the address bar.

dbpedia-owl:administrativeDistrict	dbpedia:Wikitravel:Darmstadt
dbpedia-owl:areaCode	06151, 06150
dbpedia-owl:areaTotal	122230000.000000 (xsd:double)
dbpedia-owl:elevation	144.000000 (xsd:double)
dbpedia-owl:federalState	dbpedia:Hesse
dbpedia-owl:leaderTitle	Lord Mayor
dbpedia-owl:populationAsOf	2007-06-30 (xsd:date)
dbpedia-owl:populationTotal	141471 (xsd:integer)
dbpedia-owl:postalCode	64283–64297
dbpedia-owl:thumbnail	http://upload.wikimedia.org/wikipedia/commons/thumb/b/bc/Darmstadt-Waldspirale-Hundertwasser4.jpg/200px-Darmstadt-Walds
dbpedia-owl:wikiPageExternalLink	http://www.darmstadt.army.mil http://www.darmstadt.de/ http://www.wiesbaden.army.mil/
dbpprop:adresse	Luisenplatz 5 64283 Darmstadt
dbpprop:art	City
dbpprop:bürgermeister	Walter Hoffmann
dbpprop:bürgermeistertitel	Lord Mayor
dbpprop:bundesland	Hesse
dbpprop:einwohner	141471 (xsd:integer)
dbpprop:fläche	122 (xsd:integer)
dbpprop:gemeindeschlüssel	06 4 11 000
dbpprop:gliederung	9 boroughs
dbpprop:höhe	144 (xsd:integer)
dbpprop:hasPhotoCollection	http://www4.wiwiss.fu-berlin.de/flickrwrappr/photos/Darmstadt
dbpprop:imagesize	250px
dbpprop:kfz	DA
dbpprop:lageplan	Hessen DA.png
dbpprop:landkreis	urban

Linked Open Data – was heißt hier "Linked"?

- Beispiel:

- `<#Heiko> :worksIn <http://dbpedia.org/resource/Darmstadt>`.



Links in HTML und in Linked Open Data

- Vergleiche

Heiko arbeitet in `Darmstadt`.

und

`:Heiko :worksIn <http://dbpedia.org/resource/Darmstadt> .`

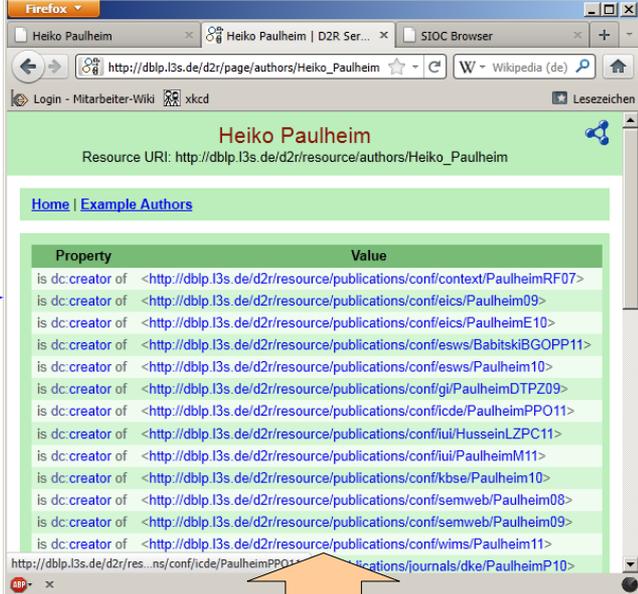
- Merke:

- In Linked Open Data sind Links immer *explizit* typisiert
- Der Typ des Links ist damit dem Computer unmittelbar verständlich

Linked Open Data: Mehrere Datensets zu einer Ressource

- Wichtiger Spezialfall: owl:sameAs*

```
:Heiko
  owl:sameAs
  <http://dblp.13s.de/d2r/page/
  authors/Heiko_Paulheim>
```



Heiko Paulheim
Resource URI: http://dblp.13s.de/d2r/resource/authors/Heiko_Paulheim

Home | Example Authors

Property	Value
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/context/PaulheimRF07>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/eics/Paulheim09>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/eics/PaulheimE10>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/esws/BabitskiBGOPP11>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/esws/Paulheim10>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/gi/PaulheimDTPZ09>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/icde/PaulheimPPO11>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/iii/HusseinLZPC11>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/iii/PaulheimM11>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/kbse/Paulheim10>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/semweb/Paulheim08>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/semweb/Paulheim09>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/conf/wims/Paulheim11>
is dc.creator of	<http://dblp.13s.de/d2r/resource/publications/journals/dke/PaulheimP10>



* OWL bekommen wir erst noch – das muss uns jetzt aber nicht stören...

Linked Open Data: Mehrere Datensets zu einer Ressource



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Wichtiger Spezialfall: `owl:sameAs*`
- Verlinkt zwei identische Ressourcen
 - Notwendig wegen Non-Unique Naming Assumption
- Eines der am häufigsten falsch verstandenen Konzepte
- Use:
 - Zwei Datensätze über die gleiche Person
- Abuse:
 - Ein Datensatz über eine Person und die Homepage dieser Person
 - Das Starbucks am Luisenplatz und die Firma Starbucks
 - Die Stadt Hamburg und das Bundesland Hamburg
 - Die Institution Kanzleramt und das Gebäude Kanzleramt

* OWL bekommen wir erst noch – das muss uns jetzt aber nicht stören...



Linked Open Data: Mehrere Datensets zu einer Ressource



TECHNISCHE
UNIVERSITÄT
DARMSTADT

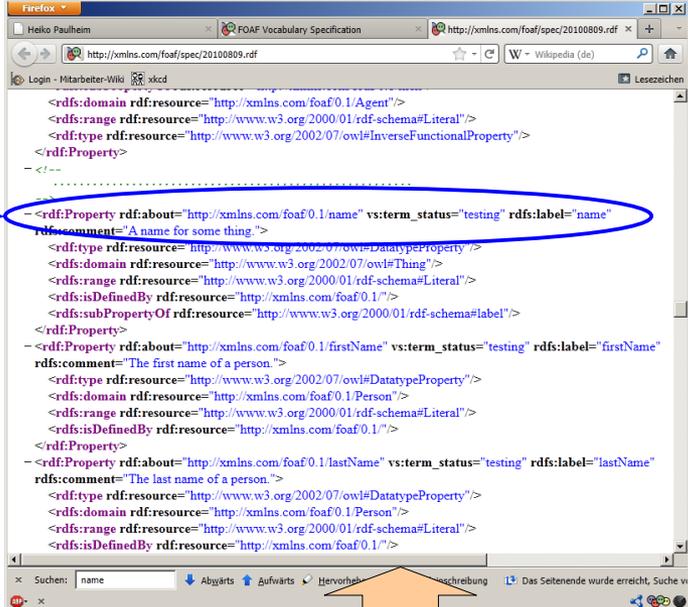
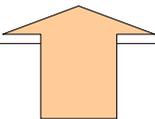
- Alternativen zu `owl:sameAs`*
 - Allgemeiner Verweis auf verwandte Information:
`rdfs:seeAlso`
 - Verweis auf (HTML-)Homepage:
z.B. `foaf:homepage`

* OWL bekommen wir erst noch – das muss uns jetzt aber nicht stören...

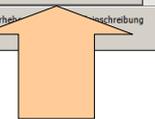
Linked Open Data: Links auf das Schema

- noch ein wichtiger Spezialfall für HTTP-URIs:
 - Auflösen von Schema-Links

```
:Heiko  
<http://xmlns.com/foaf/0.1/name>  
"Heiko Paulheim" .
```



```
<rdf:Property rdf:about="http://xmlns.com/foaf/0.1/name" vs:term_status="testing" rdfs:label="name"  
  rdfs:comment="A name for some thing" >  
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty" />  
  <rdf:domain rdf:resource="http://www.w3.org/2002/07/owl#Thing" />  
  <rdf:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal" />  
  <rdf:isDefinedBy rdf:resource="http://xmlns.com/foaf/0.1/" />  
  <rdf:subPropertyOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#label" />  
</rdf:Property>  
- <rdf:Property rdf:about="http://xmlns.com/foaf/0.1/firstName" vs:term_status="testing" rdfs:label="firstName"  
  rdfs:comment="The first name of a person." >  
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty" />  
  <rdf:domain rdf:resource="http://xmlns.com/foaf/0.1/Person" />  
  <rdf:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal" />  
  <rdf:isDefinedBy rdf:resource="http://xmlns.com/foaf/0.1/" />  
</rdf:Property>  
- <rdf:Property rdf:about="http://xmlns.com/foaf/0.1/lastName" vs:term_status="testing" rdfs:label="lastName"  
  rdfs:comment="The last name of a person." >  
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty" />  
  <rdf:domain rdf:resource="http://xmlns.com/foaf/0.1/Person" />  
  <rdf:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal" />  
  <rdf:isDefinedBy rdf:resource="http://xmlns.com/foaf/0.1/" />  
</rdf:Property>
```

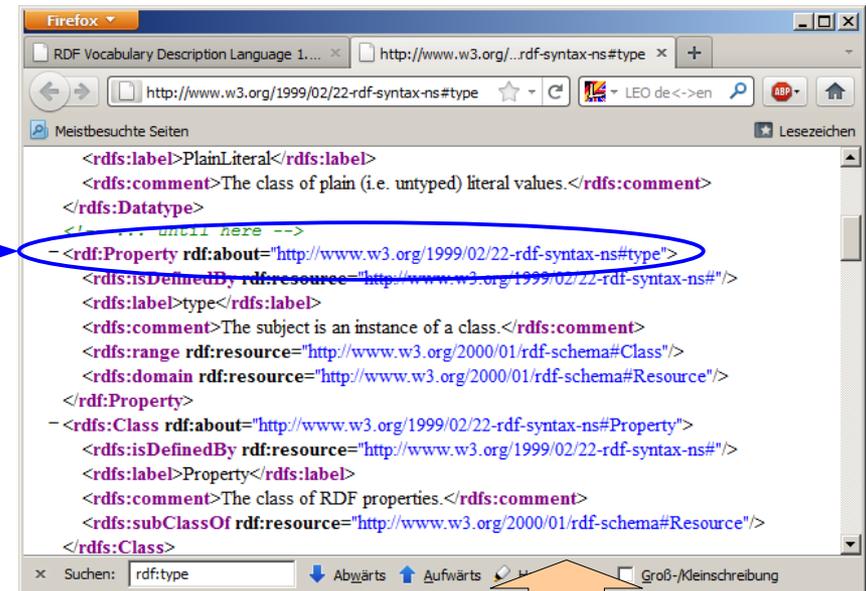


Linked Open Data: Links auf das Schema

- Das funktioniert auch mit den "eingebauten" Schemata

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

```
:Heiko rdf:type Person .  
...
```



```
<rdf:label>PlainLiteral</rdf:label>  
<rdf:comment>The class of plain (i.e. untyped) literal values.</rdf:comment>  
</rdf:Datatype>  
<!-- ... until here -->  
<rdf:Property rdf:about='http://www.w3.org/1999/02/22-rdf-syntax-ns#type'>  
<rdf:isDefinedBy rdf:resource='http://www.w3.org/1999/02/22-rdf-syntax-ns#'/>  
<rdf:label>type</rdf:label>  
<rdf:comment>The subject is an instance of a class.</rdf:comment>  
<rdf:range rdf:resource='http://www.w3.org/2000/01/rdf-schema#Class'/>  
<rdf:domain rdf:resource='http://www.w3.org/2000/01/rdf-schema#Resource'/>  
</rdf:Property>  
<rdf:Class rdf:about='http://www.w3.org/1999/02/22-rdf-syntax-ns#Property'>  
<rdf:isDefinedBy rdf:resource='http://www.w3.org/1999/02/22-rdf-syntax-ns#'/>  
<rdf:label>Property</rdf:label>  
<rdf:comment>The class of RDF properties.</rdf:comment>  
<rdf:subClassOf rdf:resource='http://www.w3.org/2000/01/rdf-schema#Resource'/>  
</rdf:Class>
```



Linked Open Data: Die vier Grundregeln

- Vier Grundregeln,
vorgeschlagen von Tim Berners-Lee (mal wieder)
1. Identifiziere Dinge mit URIs
 2. Verwende auflösbare HTTP-URIs
 3. Hinterlege an diesen URIs nützliche Informationen,
verwende dabei Standards
 4. Füge Links zu anderen URIs hinzu

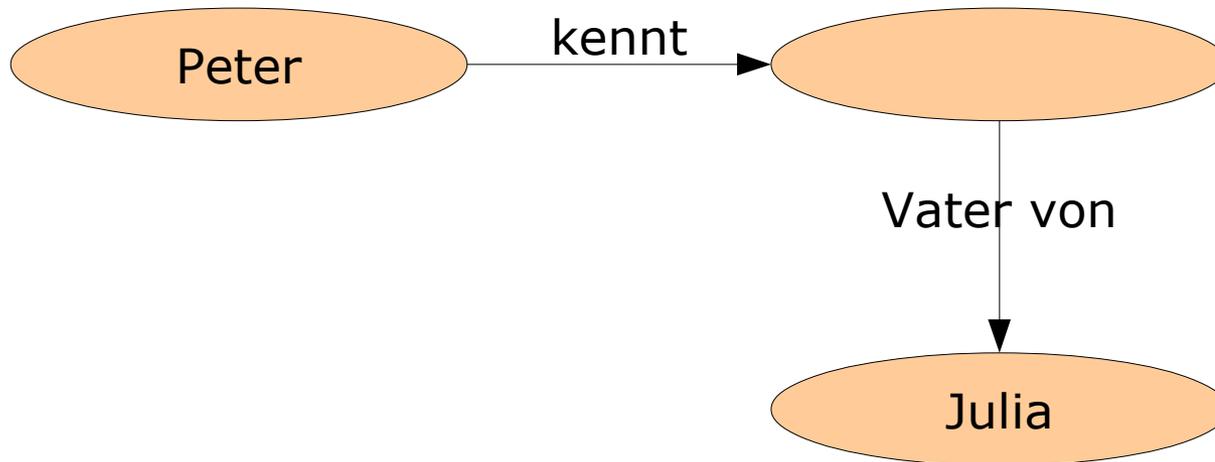
Linked Open Data: Was hinterlegt man an einem URI?



TECHNISCHE
UNIVERSITÄT
DARMSTADT

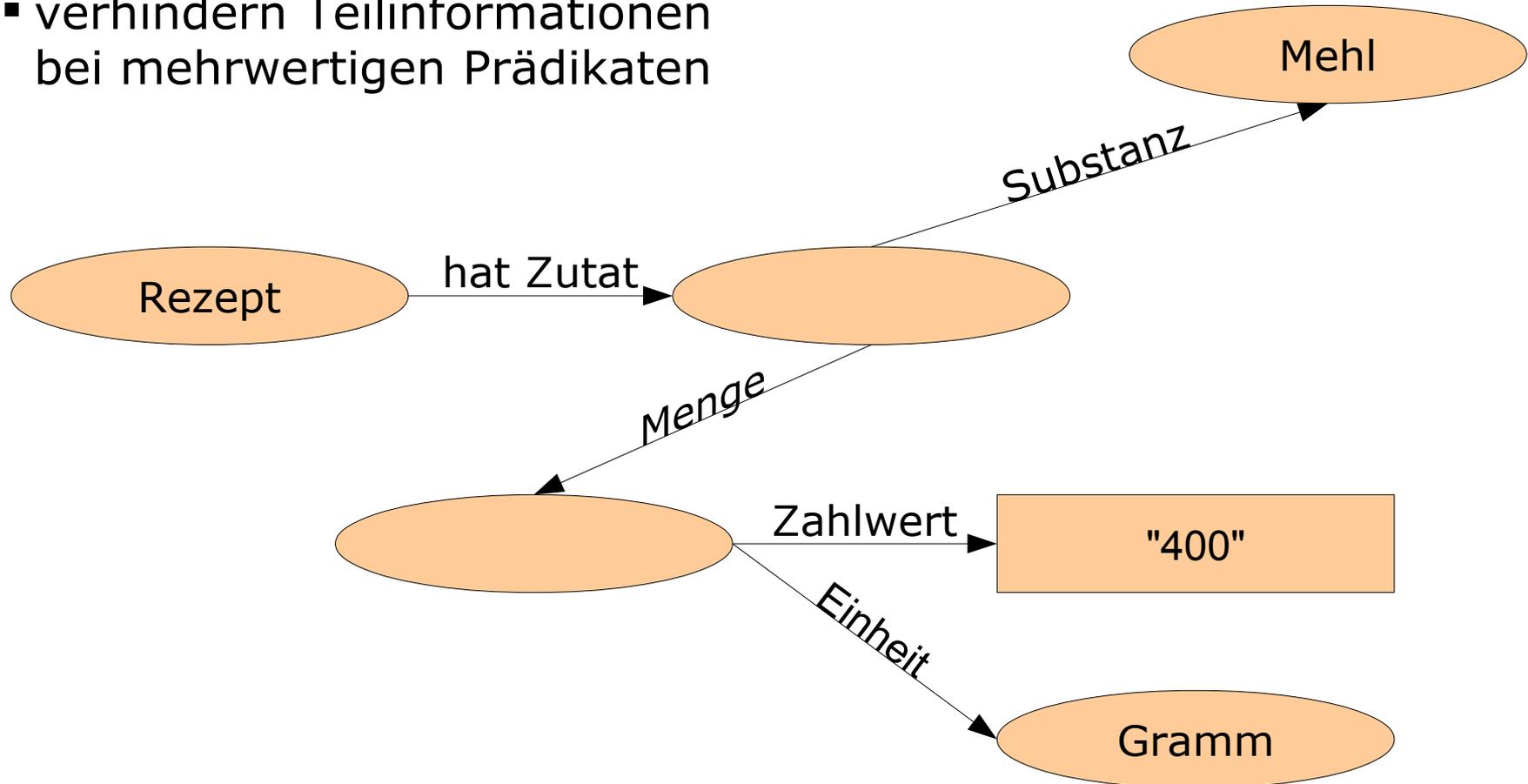
- Grundprinzip: hinterlege ein komplettes RDF-Molekül zu der Ressource
- Was ist ein RDF-Molekül?
 - Alle Tripel, die die Ressource als Subjekt oder Objekt enthalten
 - Alle Blank Nodes haben mindestens zwei ein- oder ausgehende Kanten

- Verhindern "Sackgassen" beim Browsen



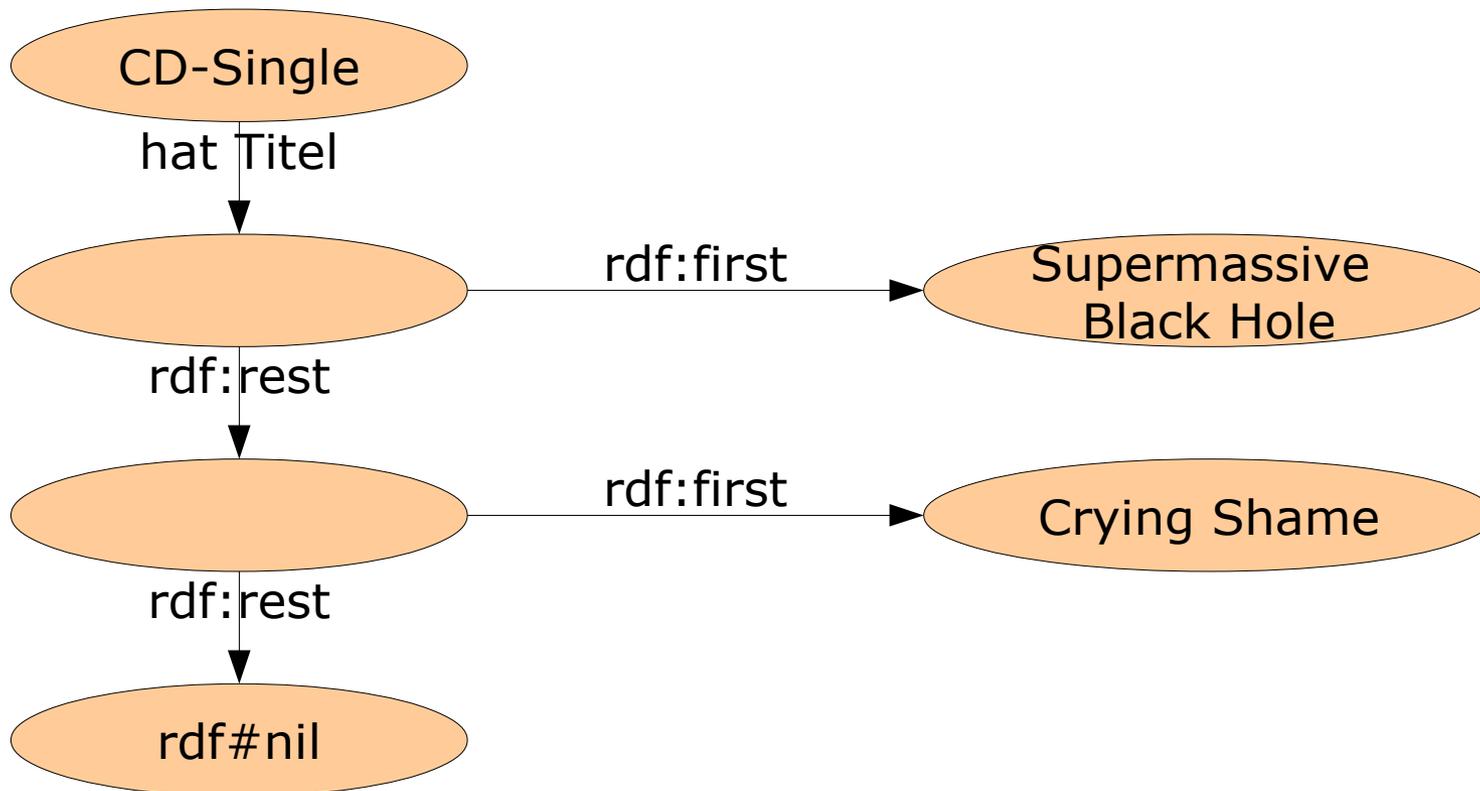
RDF-Moleküle

- verhindern Teilinformationen bei mehrwertigen Prädikaten



RDF-Moleküle

- verhindern unvollständige Listen

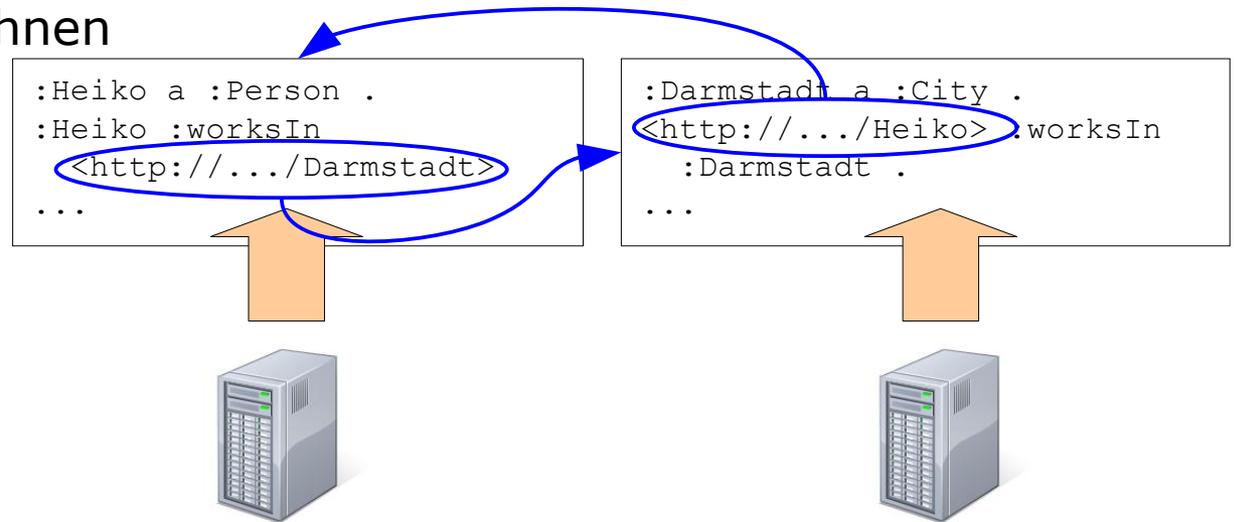




- Was ist ein RDF-Molekül?
 - Alle Tripel, die die Ressource als Subjekt oder Objekt enthalten
 - Alle Blank Nodes haben mindestens zwei ein- oder ausgehende Kanten
- Folgen:
 - Tripel werden u.U. dupliziert (Redundanzen!)
 - Moleküle können sehr groß werden

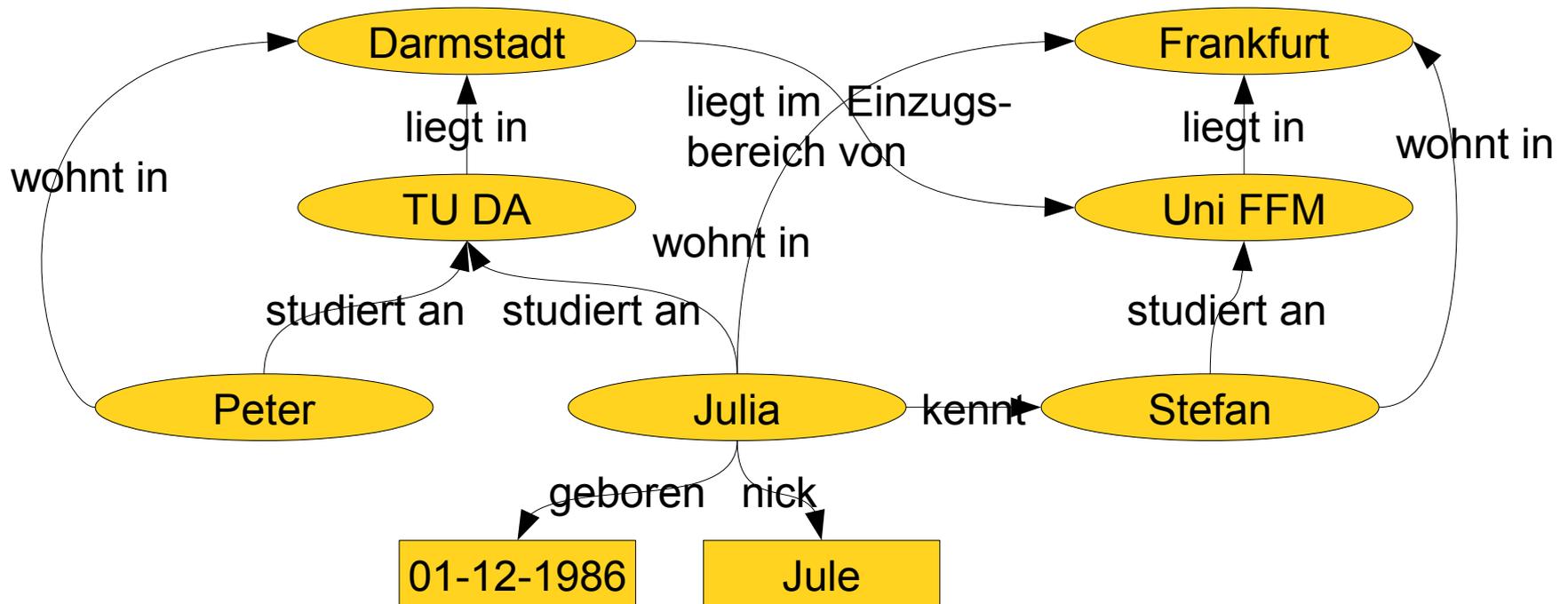
RDF-Moleküle: Theorie und Praxis

- Nach der reinen Lehre müssten alle Tripel repliziert werden
- Pragmatischer Ansatz
 - Welche Informationen (Tripel) sind für den Aufrufer interessant?
 - z.B. der Wohnort einer Person
 - aber nicht alle Personen, die an einem Ort wohnen



RDF-Moleküle: Pragmatischer Ansatz

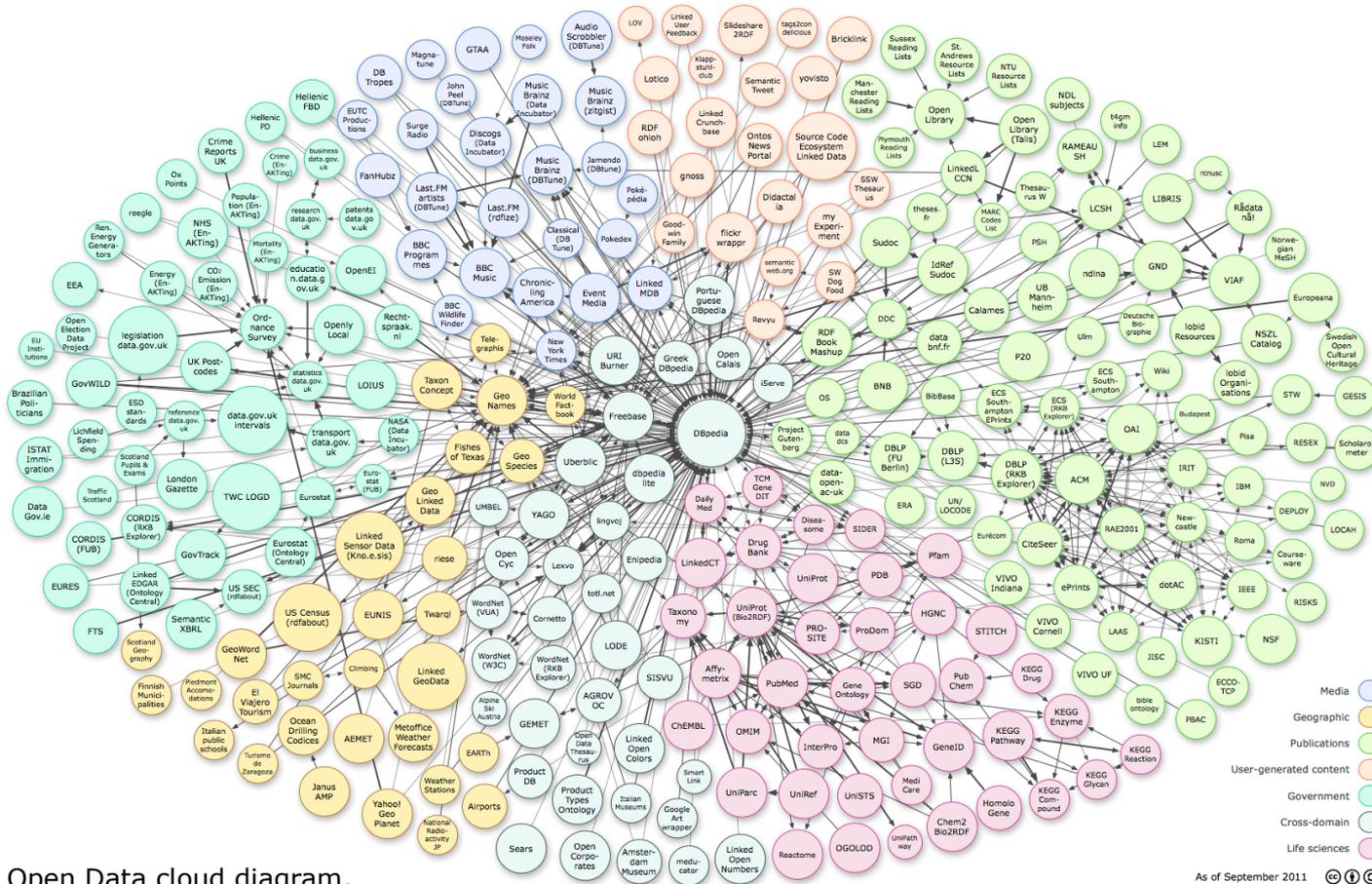
▪ Beispiel-Graph



RDF-Moleküle: Pragmatischer Ansatz

- Wie könnten hier sinnvolle Moleküle aussehen
 - ...für eine Stadt?
 - ...für eine Universität?
 - ...für eine Person?

Die Linked Open Data Cloud



Linking Open Data cloud diagram,
by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>



Die Linked Open Data Cloud

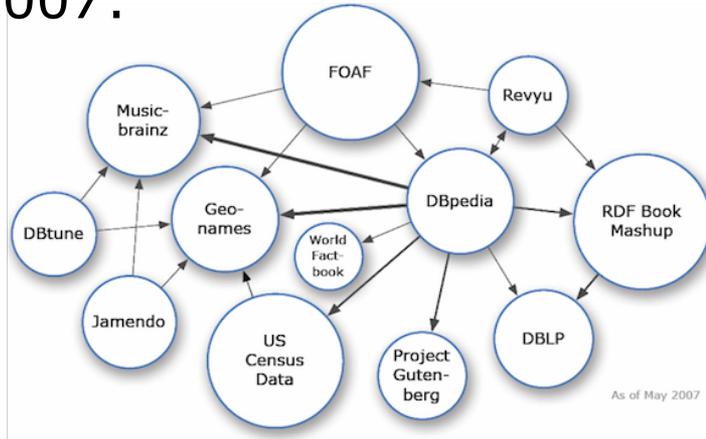
- In Zahlen:
 - 295 Datensets
 - ~32 Mrd. Tripel
 - ~500 Mio. Links zwischen Datensets

- In Themen:
 - Medien
 - Geodaten
 - Wissenschaftliche Publikationen
 - Regierungsdaten
 - Biologie
 - User Generated Content
 - übergreifende Themen

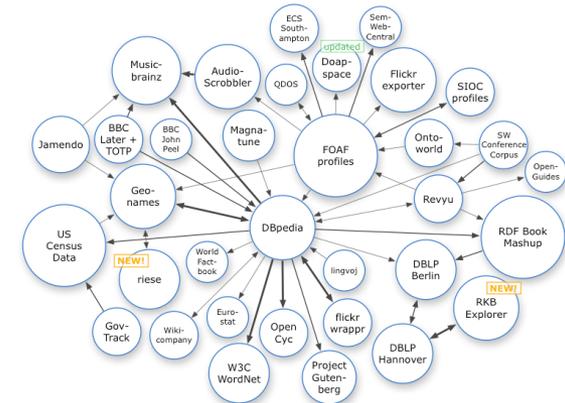
Quelle: <http://www4.wiwiss.fu-berlin.de/lodcloud/>

Die Linked Open Data Cloud: Geschichte

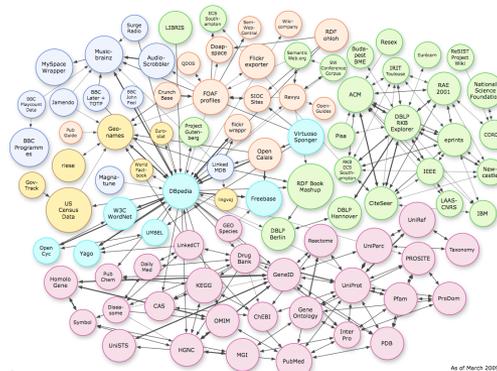
■ Mai 2007:



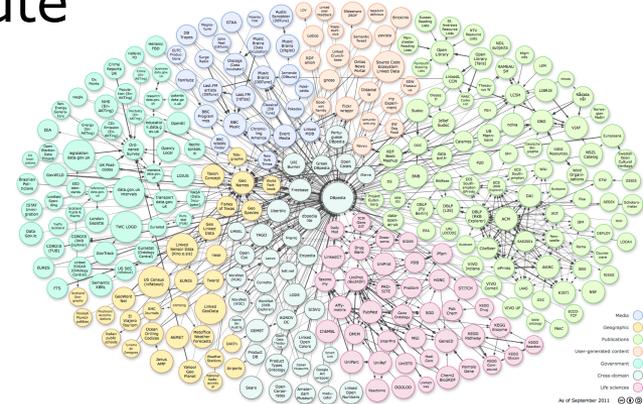
■ März 2008:



■ März 2009



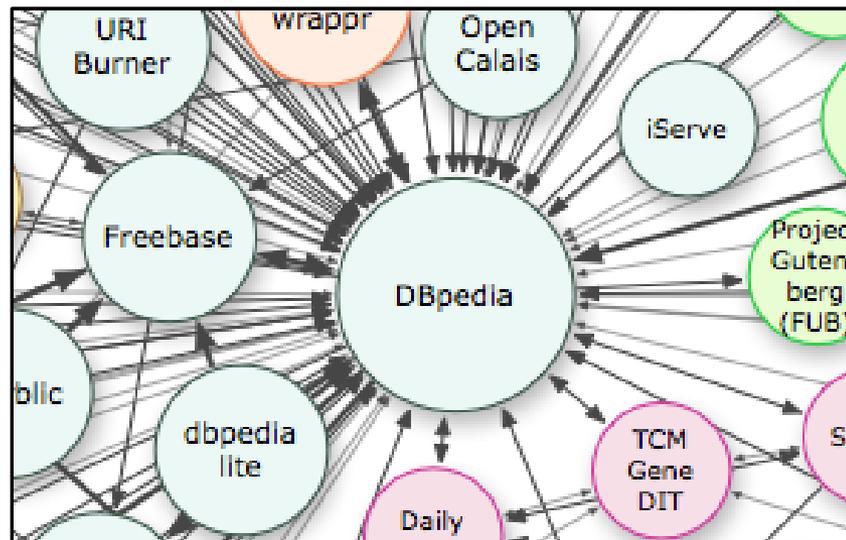
■ heute



Linking Open Data cloud diagram,
by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

Das "Herz" der LOD Cloud: DBPedia

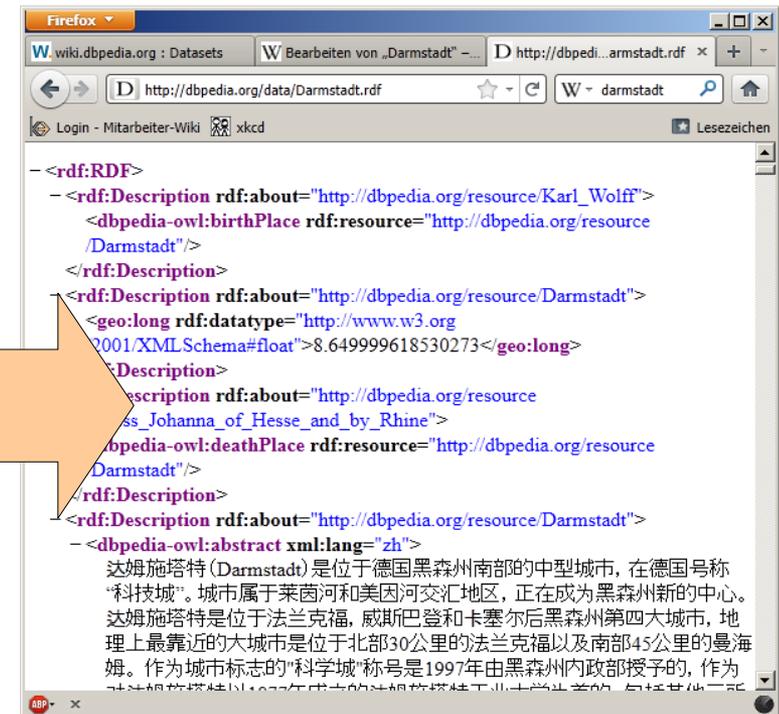
- Allgemeinwissen über 3,5 Millionen Entitäten
- 500 Millionen Tripel
- Mit 100 weiteren Datensets verknüpft
 - Das am stärksten vernetzte Datenset



Linking Open Data cloud diagram,
by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

DBPedia: Funktionsweise

Wappen	Deutschlandkarte
	
<pre>{{Infobox Gemeinde in Deutschland Art = Stadt Regierungsbezirk = Darmstadt Bundesland = Hessen Breitengrad = 49/52//N Längengrad = 08/39//E Lageplan = Hesse DA(city).svg Höhe = 144 Landkreis = Fläche = 122.24 PLZ = 64283-64297 PLZ-alt = 6100 Kfz = DA Gemeindeschlüssel = 06411000 LOCODE = DE DAR Gliederung = 9 [[Stadtteil]]e Bürgermeister = [[Jochen Partsch]] Bürgermeistertitel = Oberbürgermeister Partei = Grüne }}</pre>	
Bundesland:	
Regierungsbezirk:	
Höhe:	
Fläche:	
Einwohner:	
Bevölkerungsdichte:	
Postleitzahlen:	
Vorwahlen:	
Kfz-Kennzeichen:	
Gemeindeschlüssel:	
Stadtgliederung:	
Adresse der Stadtverwaltung:	
Webpräsenz:	www.darmstadt.de
Oberbürgermeister:	Jochen Partsch (Grüne)
Lage von Darmstadt in Hessen	
	



```
--<rdf:RDF>
-<rdf:Description rdf:about="http://dbpedia.org/resource/Karl_Wolff">
  <dbpedia-owl:birthPlace rdf:resource="http://dbpedia.org/resource/Darmstadt"/>
</rdf:Description>
<rdf:Description rdf:about="http://dbpedia.org/resource/Darmstadt">
  <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#float">8.649999618530273</geo:long>
</rdf:Description>
<rdf:Description rdf:about="http://dbpedia.org/resource/Johanna_of_Hesse_and_by_Rhine">
  <dbpedia-owl:deathPlace rdf:resource="http://dbpedia.org/resource/Darmstadt"/>
</rdf:Description>
<rdf:Description rdf:about="http://dbpedia.org/resource/Darmstadt">
  <dbpedia-owl:abstract xml:lang="zh">
    达姆施塔特 (Darmstadt) 是位于德国黑森州南部的中型城市，在德国号称“科技城”。城市属于莱茵河和美因河交汇地区，正在成为黑森州新的中心。达姆施塔特是位于法兰克福、威斯巴登和卡塞尔后黑森州第四大城市，地理上最靠近的大城市是位于北部30公里的法兰克福以及南部45公里的曼海姆。作为城市标志的“科学城”称号是1997年由黑森州内政部授予的，作为对达姆施塔特以1827年成立的达姆施塔特工业大学为首的，包括其他一所
```

DBPedia: Funktionsweise

Geography

[\[edit\]](#)

Climate

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Record high °C (°F)	15 (59)	17 (63)	24 (75)	29 (84)	31 (88)	35 (95)	35 (95)	38 (100)	30 (86)	26 (79)	18 (64)	15 (59)	38 (100)
Average high °C (°F)	5 (41)	7 (45)	11 (52)	15 (59)	20 (68)	23 (73)	25 (77)	25 (77)	20 (68)	14 (57)	8 (48)	5 (41)	15
Average low °C (°F)	0 (32)	1 (34)	3 (37)	6 (43)	11 (52)	13 (55)	15 (59)	15 (59)	11 (52)	7 (45)	4 (39)	1 (34)	7
Record low °C (°F)	-14 (7)	-12 (10)	-10 (14)	-4 (25)	1 (34)	3 (37)	6 (43)	7 (45)	2 (36)	-4 (25)	-8 (18)	-11 (12)	-14 (7)
Precipitation mm (inches)	28 (1.1)	27 (1.06)	27 (1.06)	29 (1.14)	29 (1.14)	44 (1.73)	43 (1.69)	41 (1.61)	27 (1.06)	41 (1.61)	25 (0.98)	36 (1.42)	33 (1.3)

Source: Weather2^[14]

Transport

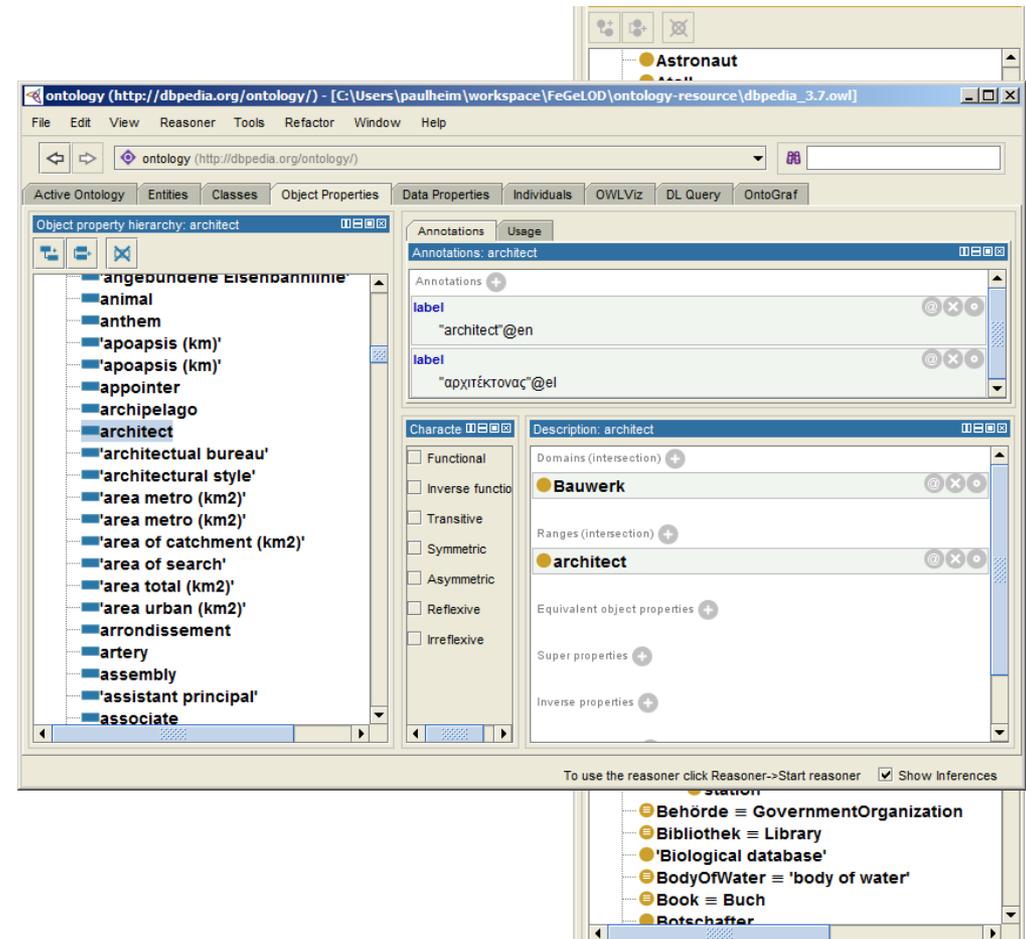
Darmstadt is relatively typical for German cities in terms of its transport system, with the car being the main, but not overwhelmingly dominant mode of transport.

DBPedia: Inhalte

- Daten aus verschiedenen Infoboxen (engl. Wikipedia)
- Informationen über Redirects und Disambiguations
- Externe Weblinks
- Abstracts in verschiedenen Sprachen
- Typinformationen:
 - DBPedia-Ontologie
 - YAGO
 - ...und weitere

Die DBPedia-Ontologie

- **Klassenhierarchie:**
 - ~400 Klassen
 - Teilweise hierarchisch
 - Dagegen YAGO:
~150.000 Klassen
- **Properties:**
 - ~850 Relationen
 - tw. mit Domain/Range
 - ~900 Datenproperties
 - keine Hierarchien



- "Yet Another Great Ontology"
- ~10 Millionen Instanzen, ~80 Millionen Aussagen
- Automatisch aus verschiedenen Quellen erzeugt
 - Wikipedia, WordNet, GeoNames
- Korrektheit: ~95%

Firefox - YAGO 2 spotix Text Browser

mpg.de https://d5gate.ag5.mpi-sb.mpg.de/web/yagospotix/Browser?entity=Darmstadt_University_of_Technology

Wikipedia (en)

Twiki Aigaion xkcd IEEE Xplore - Home SW Forum Lesezeichen

Browse YAGO2

Entity: case insensitive

Darmstadt_University_of_Technology

Show transitive facts

← Bert Rürup ← Hubert Markl ← Kurt H. Debus	worksAt	wasCreatedOnDate	1877-##-## →
← Darmstadt university of technology ← Darmstadt University of Technology ← Technical University of Darmstadt ← Technische Hochschule Darmstadt ← Technische universitaet darmstadt ← Technische Universitaet Darmstadt ← Technische universitaet darmstadt ← Technische Universität Darmstadt ← Technische universität darmstadt ← Technische Universität Darmstadt ← TUD ← Tu darmstadt ← TU Darmstadt	means	isLocatedIn	Earth → Europe → Germany →
← Darmstadt University of Technology	hasPreferredMeaning	hasPreferredName	Darmstadt Univer
← Carl Adam Petri ← Helmut Hölzer ← Peter Grünberg ← Walter Haeussermann	graduatedFrom	hasWikipediaCategory	Darmstadt → Darmstadt Univer Educational institu Technical universi Universities and c Universities and c
		isCalled	Darmstadtin tekni Darmstadts tekni Darmstadt Teknik Darmštato teknik Technische Unive Tegniese Univer Tækniháskóli Dar Πολυτεχνείο του Ν Дармштадтский ঢ়াৰ্মশ্ৰীট উলিভাৰ্শিট 达姆施塔特工业大 Universidade Téc Universidad Técn Universitas littera Universitat Tècnic Université de tech
		hasWikipediaArticleLength	25177 →
			Educational institu Technical universi Universities and c Universities and c

Weitere interessante Datensets aus dem Bereich "Allgemeines"



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Freebase
 - inhaltlich ähnlich zu DBpedia
 - enthält Daten verschiedener Quellen
 - Direkter Community-Ansatz
- OpenCyc
 - Sammlung von allgemeinen Definitionen
 - Mehrere 100.000 Begriffe mit Millionen von Axiomen
 - Open-Source-Variante von Cyc
- LinkedGeoData
 - Dinge mit einer Position
 - Daten aus OpenStreetMaps



Vergleich Freebase und DBPedia



TECHNISCHE
UNIVERSITÄT
DARMSTADT

▪ Beispiel: TU Darmstadt

▪ DBPedia

- enthält mehr Informationen
- z.B. Motto

"mens agitat molem"
(Der Geist bewegt die Materie)

▪ Freebase

- ist exakter

▪ Beispiel: Einwohnerzahl von Darmstadt

- DBpedia: 141.471
- Freebase: 142.761
- tatsächlich: 143.276

▪ Klassischer Trade-Off zwischen Precision und Recall

OpenCyc

- von "encyclopedia"
- mehrere 100.000 Begriffe
- mehrere Millionen Aussagen
- nutzt intern eine eigene Sprache (CycL)



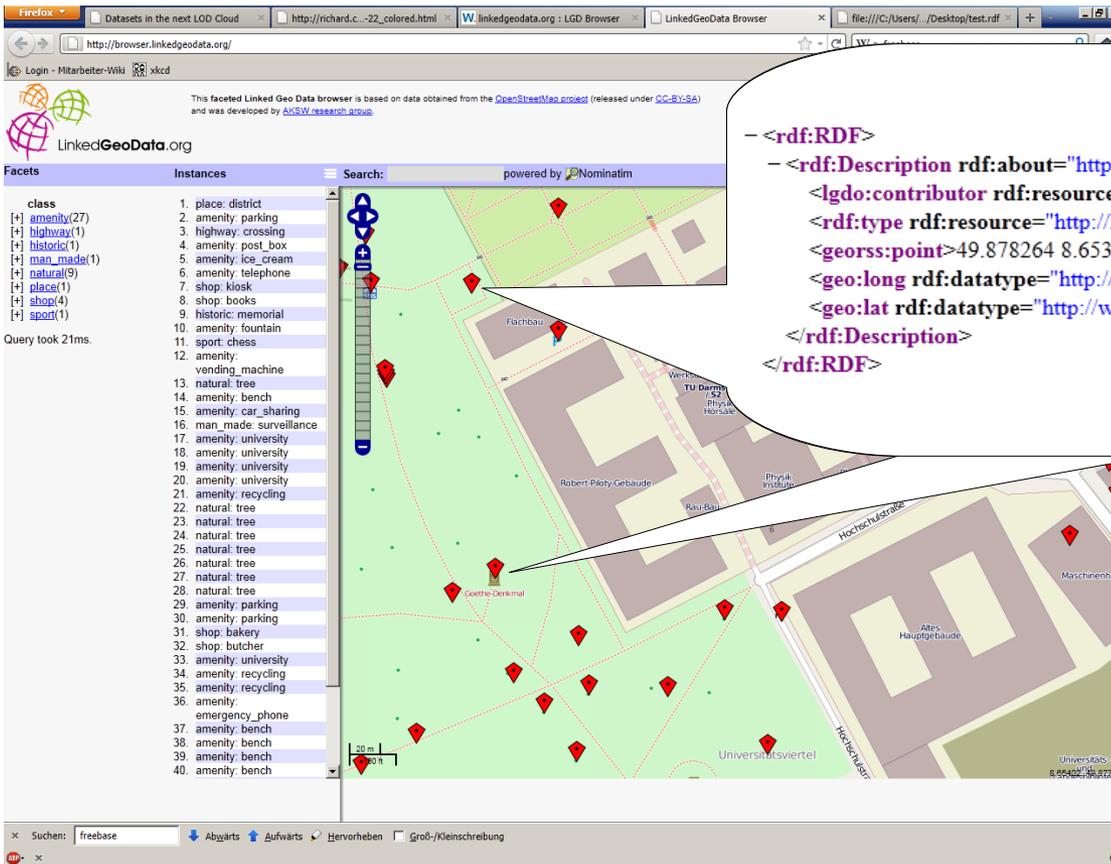
The screenshot shows a Firefox browser window with the following content:

- Address bar: `sw.opencyc.org/concept/Mx4rwPuepJwpEbGdrcN5Y29ycA`
- Page title: **OpenCyc Collection: electric guitar**
- Unique ID: [Mx4rwPuepJwpEbGdrcN5Y29ycA]
- English ID: [Guitar-Electric]
- English Aliases: ["electric guitars"]
- Description: Instances are instances of **Guitar** that produce there sound by turning the vibration of their strings into an electronic signal, which is in turn electronically amplified, then transmitted as sound via an electronic speaker.
- Type of: electric instrument, guitar
- Instance of: kind of artifact not distinguished by brand or model, type of object, type of object whose instances do not physically overlap
- Subtypes: electric bass guitar
- Instances:
- Same as: http://umbel.org/umbel/sc/Guitar_Electric
- Copyright © 2001-2009 Cycorp, Inc.
- Creative Commons License: CC BY

Linked Geo Data

- Basiert auf Open Street Maps
- Links auf DBPedia
- Besonderheit: alles hat eine Position
 - Latitude/Longitude
- Man kann auch Dinge nach Position abfragen

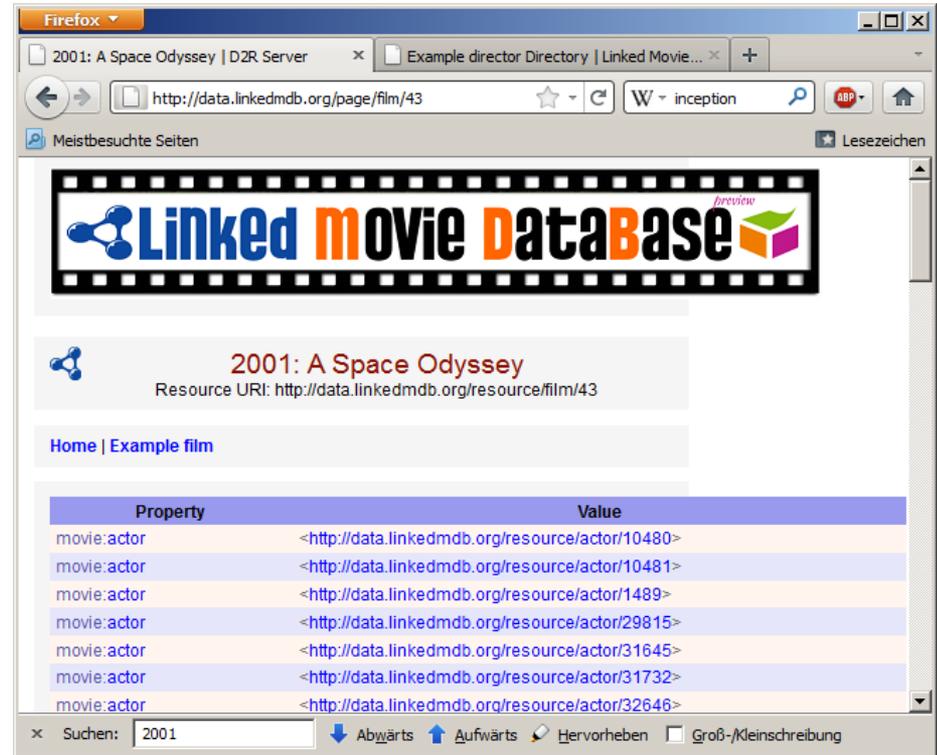
- hat auch ein eigenes User Interface:



```
- <rdf:RDF>
- <rdf:Description rdf:about="http://linkedgedata.org/triplify/node371158882">
  <lgdo:contributor rdf:resource="http://linkedgedata.org/triplify/user61927"/>
  <rdf:type rdf:resource="http://linkedgedata.org/ontology/Chess"/>
  <georss:point>49.878264 8.6537455</georss:point>
  <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">8.6537455</geo:long>
  <geo:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">49.878264</geo:lat>
</rdf:Description>
</rdf:RDF>
```

Beispiele für Datensets

- **Linked Movie Database**
 - Daten über Filme, Schauspieler, Regisseure...
 - 40.000 Filme
- **MusicBrainz**
 - Musiker, Alben, ...
 - 22.000 Alben, 40.000 Musiker
- **Open Library**
 - 2 Mio. Bücher
 - Autoren, Verlage, ...
- **DBLP**
 - Fachpublikationen Informatik
 - 1,6 Mio. Artikel, inkl. Autoren etc.

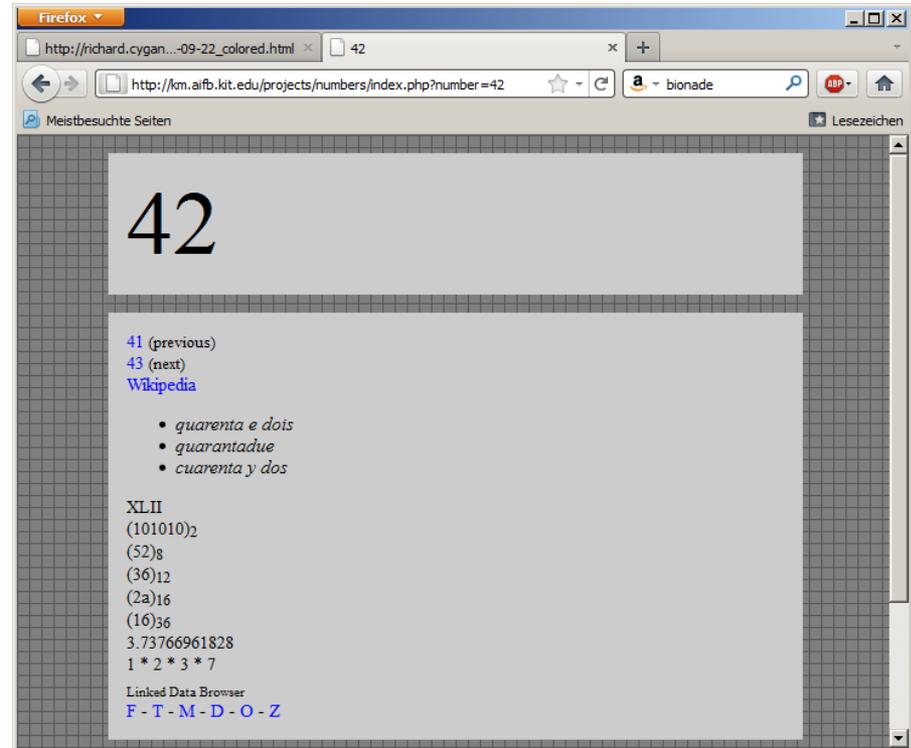


The screenshot shows a Firefox browser window with the URL `http://data.linkedmdb.org/page/film/43`. The page title is "2001: A Space Odyssey" and the resource URI is `http://data.linkedmdb.org/resource/film/43`. The page features a table of properties and values for the movie.

Property	Value
movie:actor	< http://data.linkedmdb.org/resource/actor/10480 >
movie:actor	< http://data.linkedmdb.org/resource/actor/10481 >
movie:actor	< http://data.linkedmdb.org/resource/actor/1489 >
movie:actor	< http://data.linkedmdb.org/resource/actor/29815 >
movie:actor	< http://data.linkedmdb.org/resource/actor/31645 >
movie:actor	< http://data.linkedmdb.org/resource/actor/31732 >
movie:actor	< http://data.linkedmdb.org/resource/actor/32646 >

Beispiele für Datensets

- ProductDB
 - Produkte und Hersteller
 - Mit Crawlern von verschiedenen Webseiten extrahiert
- DrugBank
 - 4700 Medikamentwirkstoffe
 - Produkte, Nebenwirkungen, Wechselwirkungen...
- NASA Data Incubator
 - Daten über NASA-Missionen
- Linked Open Numbers
 - Zahlen, Namen in verschiedenen Sprachen
 - Römische und arabische Ziffern, Binär-, Hex- und andere Systeme, Primfaktoren



Beispiele für Datensets: DrugBank



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- ca. 4500 Medikamente, Beschreibungen, Indikationen und Wechselwirkungen

The screenshot shows a Firefox browser window displaying the DrugBank profile for Ticlopidine. The page title is "Ticlopidine" and the resource URI is "http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00208". The page includes a navigation bar with "Home" and "All drugs". The main content is a table with two columns: "Property" and "Value".

Property	Value
drugbank:absorption	Absorption is greater than 80%. Food increases absorption.
drugbank:affectedOrganism	Humans and other mammals
drugbank:ahfsCode	20:12.18
drugbank:atcCode	B01AC05
drugbank:biotransformation	Metabolized extensively by the liver; only trace amounts of intact drug are detected in the urine. At least 20 metabolites have been identified. It has been proposed that 1 or more active metabolites may account for ticlopidine's activity, because ticlopidine itself is an extremely weak platelet aggregation inhibitor in vitro at the concentrations achieved in vivo. However, no active metabolite has been identified.
drugbank:brandName	Ticlid
drugbank:casRegistryNumber	<http://bio2rdf.org/cas:55142-85-3>
drugbank:chemicalFormula	C14H14ClNS
drugbank:chemicalLupacName	5-[(2-chlorophenyl)methyl]-6,7-dihydro-4H-thieno[3,2-c]pyridine
drugbank:contraindicationInsert	<http://129.128.185.122/drugbank2/drugs/DB00208/inserts/1535/full>
drugbank:creationDate	2005-06-13 13:24:05 UTC
drugbank:description	Ticlopidine is an effective inhibitor of platelet aggregation. The drug has been found to significantly reduce infarction size in acute myocardial infarcts and is an effective antithrombotic agent in arteriovenous fistulas, aorto-coronary bypass grafts, ischemic heart disease, venous thrombosis, and arteriosclerosis. [PubChem]
drugbank:dosageForm	<http://www4.wiwiss.fu-berlin.de/drugbank/resource/dosageforms/tabletOral>
drugbank:dpdDrugIdNumber	02239744
drugbank:drugCategory	<http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugcategory/fibrinolyticAgents>
drugbank:drugCategory	<http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugcategory/plateletAggregationInhibitors>



Ausgewählte Schemata



- Es gibt einige Schemata, die sich für Linked Open Data etabliert haben
- Das ermöglicht
 - Arbeitersparnis
 - intelligentere Agenten
 - Fusion von Informationen aus unterschiedlichen Quellen
 - ...

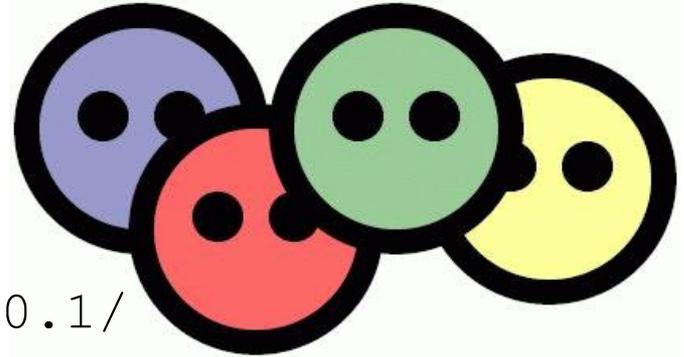
Dublin Core

- Haben wir schon kennengelernt
- Metadaten zu Ressourcen / Dokumenten
- Namespace <http://purl.org/dc/elements/1.1/>
- Präfix meist `dc:`
- definiert Properties, z.B.
 - creator
 - subject
 - date
- Typen von Ressourcen: DCMI Type Vocabulary
 - Text
 - Image
 - Software
 - ...

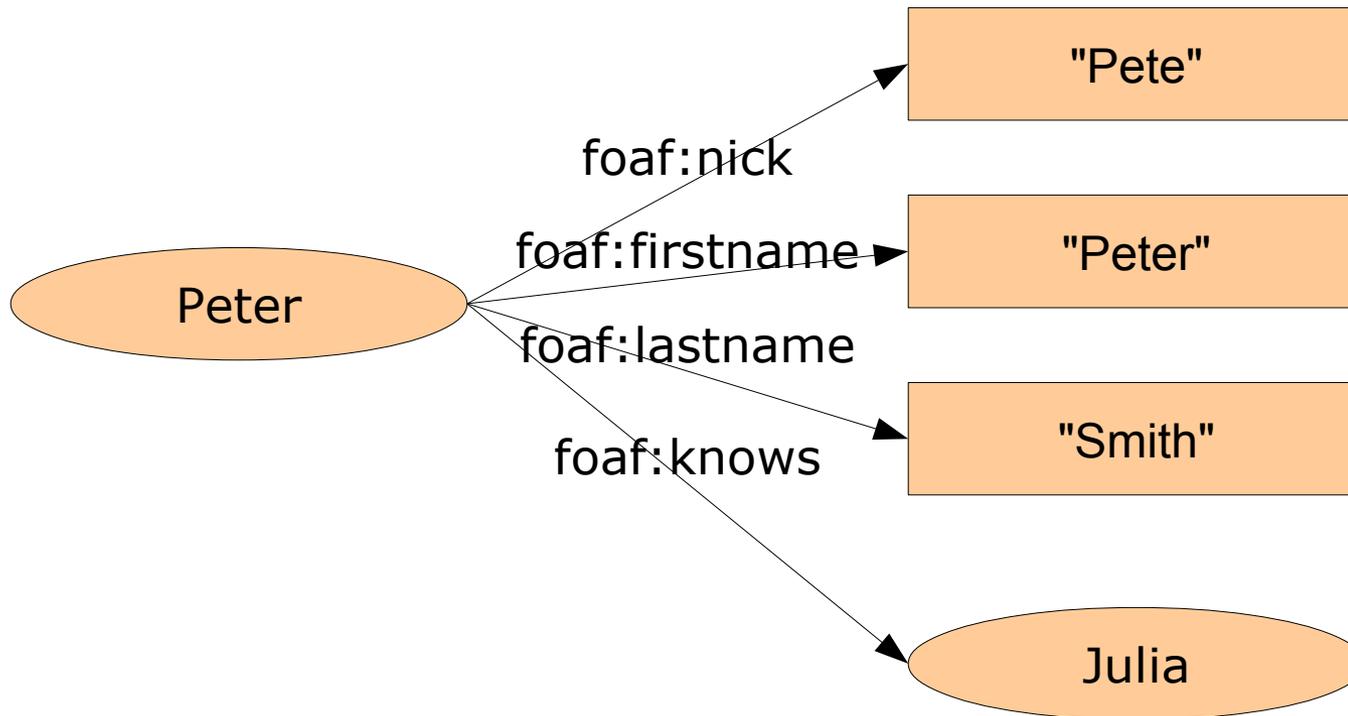


FOAF (Friend of a Friend)

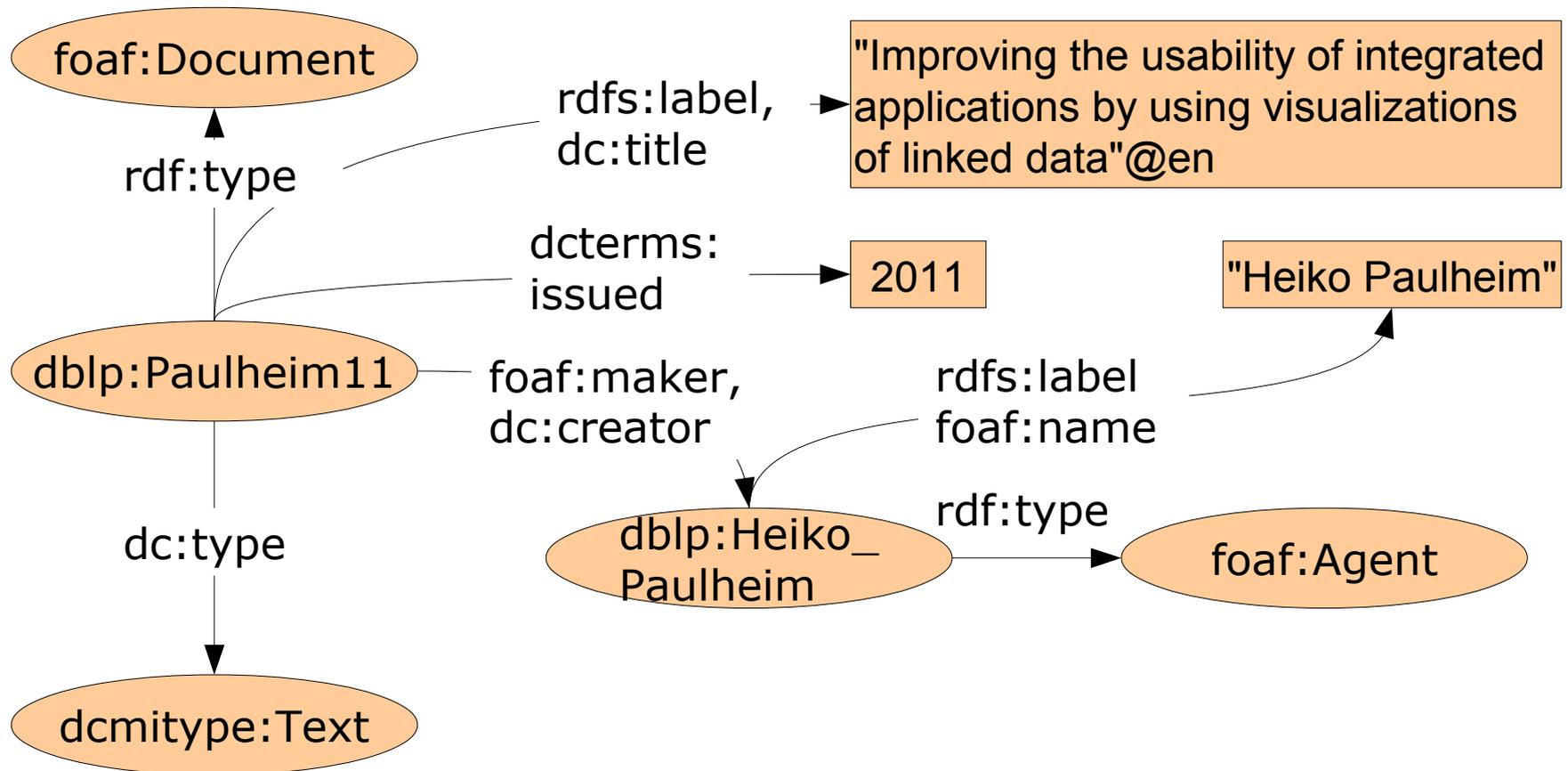
- Beschreibung von Personen und ihrer Relationen
- Für Metadaten privater Homepages (aber nicht nur)
- Namespace `http://xmlns.com/foaf/0.1/`
- Präfix meist `foaf:`
- Wichtige Klassen:
 - Person
 - Group
 - Organization
 - Project
 - ...
- Wichtige Properties:
 - `name`, `firstName`, `lastName`
 - `phone`, `mbox`, `homepage`
 - `knows`
 - `currentProject`, `pastProject`
 - ...



FOAF (Friend of a Friend)



Beispiel für FOAF + DC: DBLP

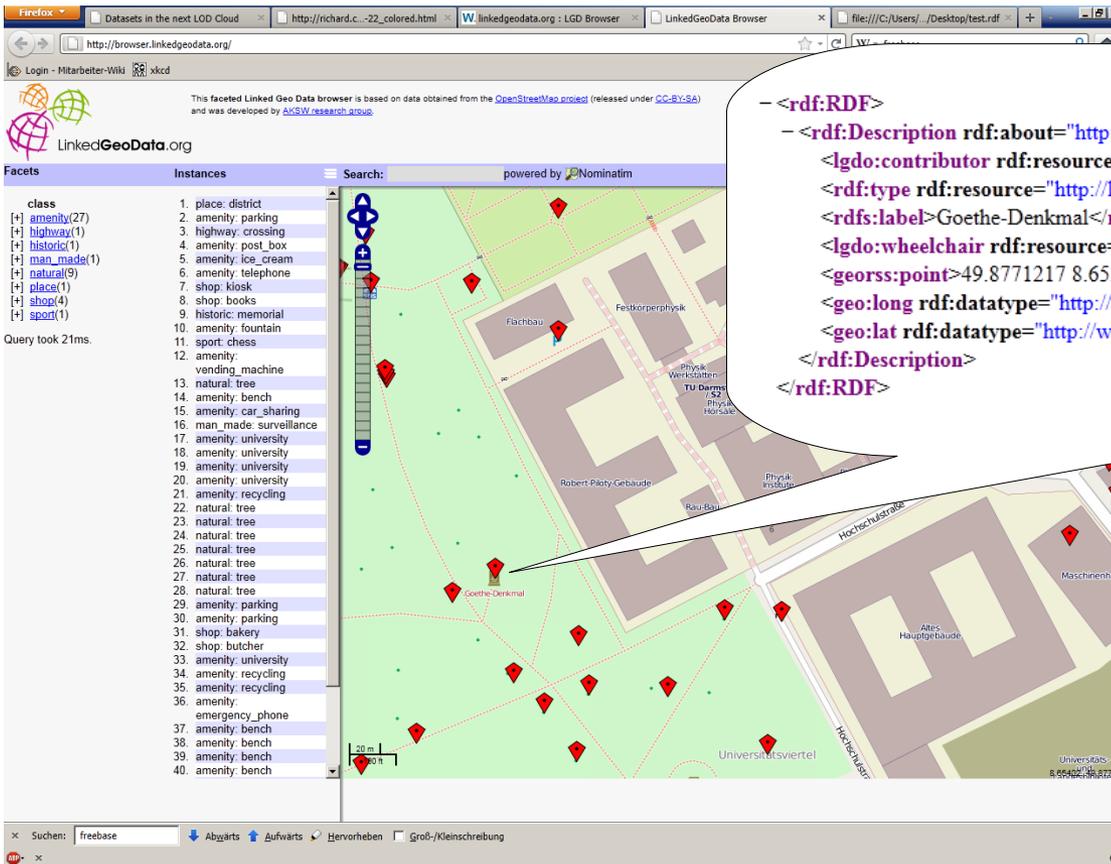


WGS 84

- Codierung von geographischen Daten
 - World Geodetic System 1984
 - 3D-Referenzmodell
 - Namespace http://www.w3.org/2003/01/geo/wgs84_pos#
 - Präfix oft `geo:`
-
- Klassen:
 - SpatialThing
 - Point
 - Properties:
 - latitude
 - longitude
 - altitude
 - location



■ Beispiel: Linked Geo Data



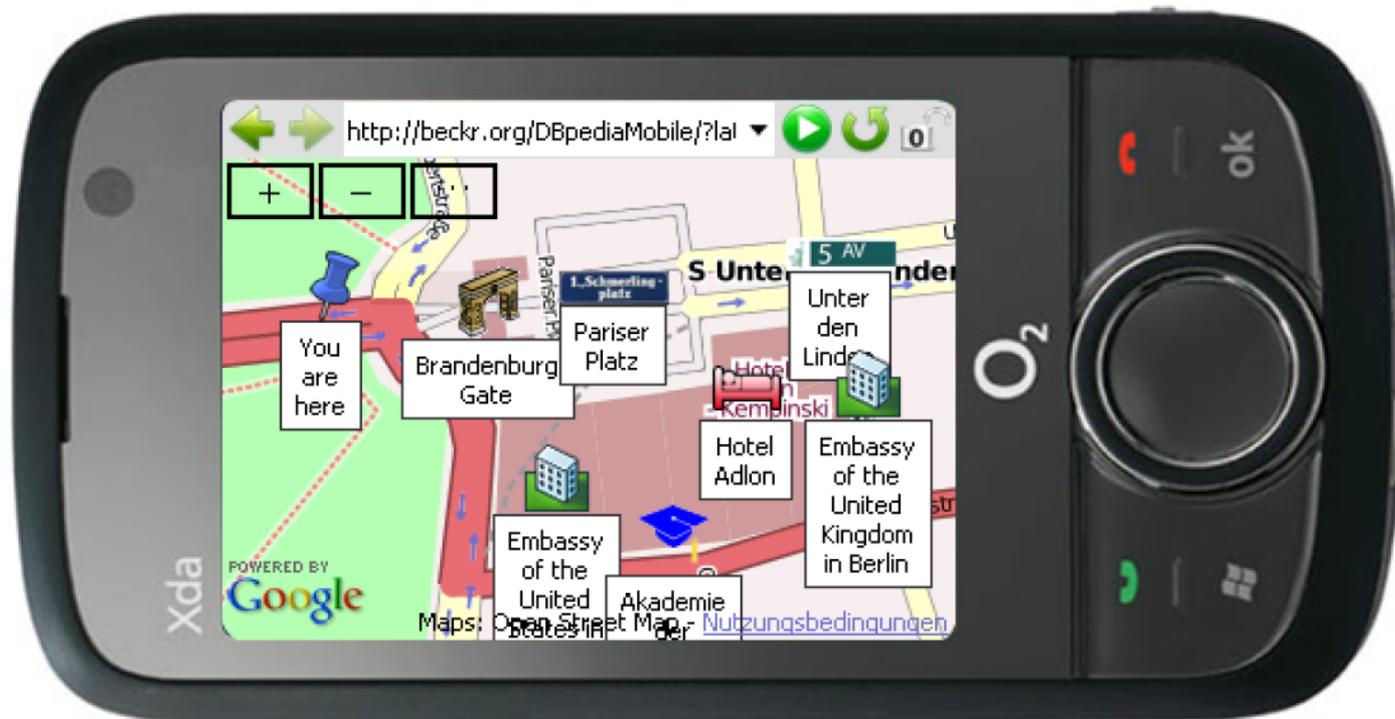
The screenshot shows the LinkedGeoData.org interface. On the left, a 'Facets' panel lists various classes such as 'amenity', 'highway', 'historic', 'natural', 'place', 'shop', and 'sport'. The main area displays a map of Darmstadt with red diamond markers representing geo-entities. A callout bubble contains the following RDF snippet:

```

<rdf:RDF>
  <rdf:Description rdf:about="http://linkedgeodata.org/triplify/node255462188">
    <lgdo:contributor rdf:resource="http://linkedgeodata.org/triplify/user290680"/>
    <rdf:type rdf:resource="http://linkedgeodata.org/ontology/Memorial"/>
    <rdfs:label>Goethe-Denkmal</rdfs:label>
    <lgdo:wheelchair rdf:resource="http://linkedgeodata.org/ontology/yes_%28WheelChair%29"/>
    <georss:point>49.8771217 8.6538865</georss:point>
    <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">8.6538865</geo:long>
    <geo:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">49.8771217</geo:lat>
  </rdf:Description>
</rdf:RDF>
  
```

WGS 84

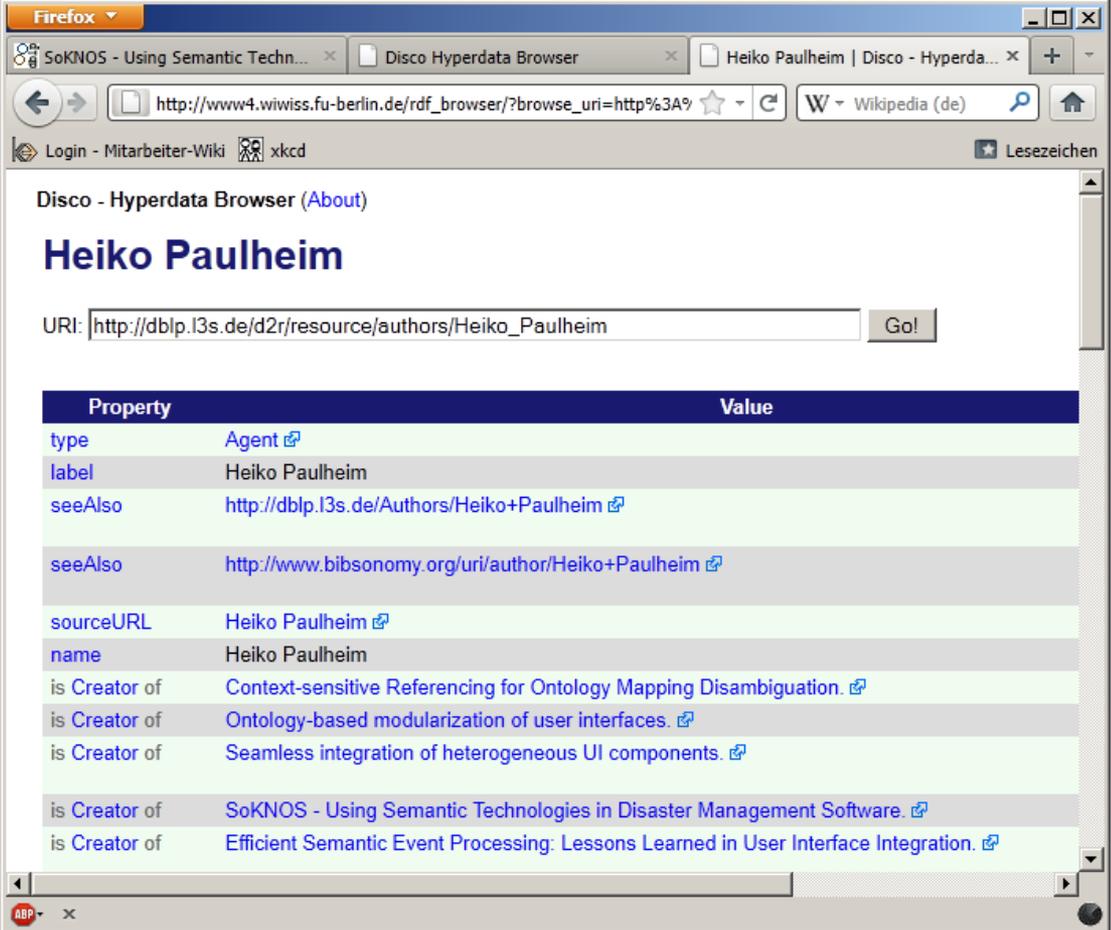
- ermöglicht auch mobile Anwendungen
- auch DBPedia enthält Geo-Daten → DBPedia Mobile



Becker & Bizer (2008): DBpedia Mobile: A LocationEnabled .Linked Data Browser

Browser für Linked Open Data

- Disco (FU Berlin)
- tabellarische Ansicht aller Tripel
- über Links kann man browsen

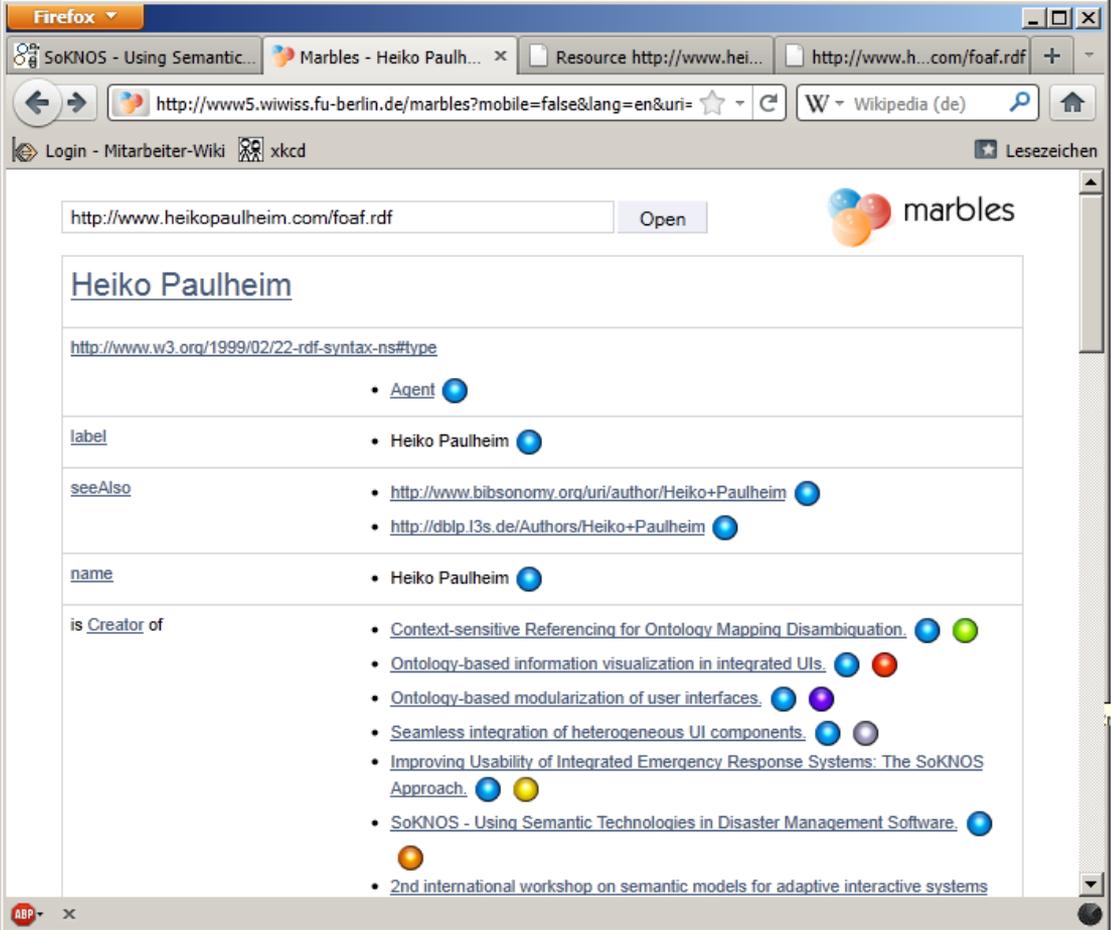


The screenshot shows a Firefox browser window displaying the Disco Hyperdata Browser interface. The browser's address bar shows the URL: `http://www4.wiwiss.fu-berlin.de/rdf_browser/?browse_uri=http%3A%2F%2Fdblp.l3s.de/d2r/resource/authors/Heiko_Paulheim`. The page title is "Disco - Hyperdata Browser (About)" and the main heading is "Heiko Paulheim". Below the heading is a search bar with the same URI and a "Go!" button. The main content is a table with two columns: "Property" and "Value".

Property	Value
type	Agent ↗
label	Heiko Paulheim
seeAlso	http://dblp.l3s.de/Authors/Heiko+Paulheim ↗
seeAlso	http://www.bibsonomy.org/uri/author/Heiko+Paulheim ↗
sourceURL	Heiko Paulheim ↗
name	Heiko Paulheim
is Creator of	Context-sensitive Referencing for Ontology Mapping Disambiguation. ↗
is Creator of	Ontology-based modularization of user interfaces. ↗
is Creator of	Seamless integration of heterogeneous UI components. ↗
is Creator of	SoKNOS - Using Semantic Technologies in Disaster Management Software. ↗
is Creator of	Efficient Semantic Event Processing: Lessons Learned in User Interface Integration. ↗

Browser für Linked Open Data

- Marbles (FU Berlin)
- ähnlich wie Disco
- führt verschiedene Quellen zusammen

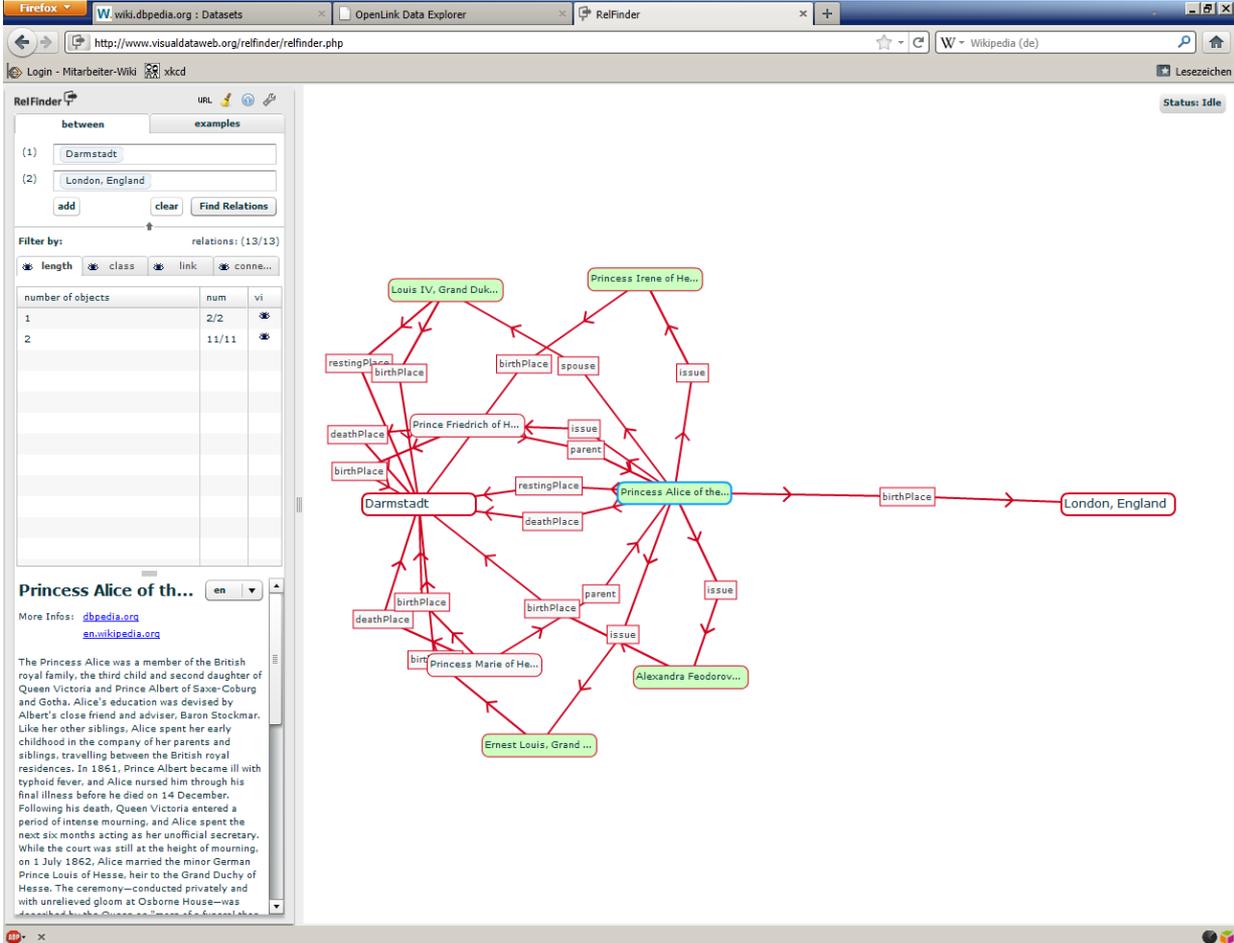


The screenshot shows the Marbles browser interface in a Firefox window. The address bar displays the URL <http://www5.wiwiss.fu-berlin.de/marbles?mobile=false&lang=en&uri=http://www.heikopaulheim.com/foaf.rdf>. The main content area displays the profile for Heiko Paulheim, including his name, a label, and a list of works he is the creator of. The interface is clean and modern, with a search bar and navigation buttons at the top.

Property	Value
Agent	Heiko Paulheim
label	Heiko Paulheim
seeAlso	<ul style="list-style-type: none">http://www.bibsonomy.org/uri/author/Heiko+Paulheimhttp://dblp.13s.de/Authors/Heiko+Paulheim
name	Heiko Paulheim
is Creator of	<ul style="list-style-type: none">Context-sensitive Referencing for Ontology Mapping Disambiguation.Ontology-based information visualization in integrated UIs.Ontology-based modularization of user interfaces.Seamless integration of heterogeneous UI components.Improving Usability of Integrated Emergency Response Systems: The SoKNOS Approach.SoKNOS - Using Semantic Technologies in Disaster Management Software.2nd international workshop on semantic models for adaptive interactive systems

Browser für Linked Open Data

- RelFinder (Visual Data Web)
- Findet Relationen zwischen zwei Ressourcen



The screenshot shows the RelFinder interface in a Firefox browser. The address bar displays `http://www.visualdataweb.org/relfinder/relfinder.php`. The interface includes a search area with two input fields containing "Darmstadt" and "London, England", and a "Find Relations" button. Below this is a "Filter by:" section with a table showing 13 relations. The main area displays a network graph with nodes representing entities and edges representing relationships. The graph shows connections between "Darmstadt" and "London, England" via "birthPlace" relationships, and other entities like "Princess Alice of the..." and "Louis IV, Grand Duk...".

number of objects	num	vi
1	2/2	
2	11/11	

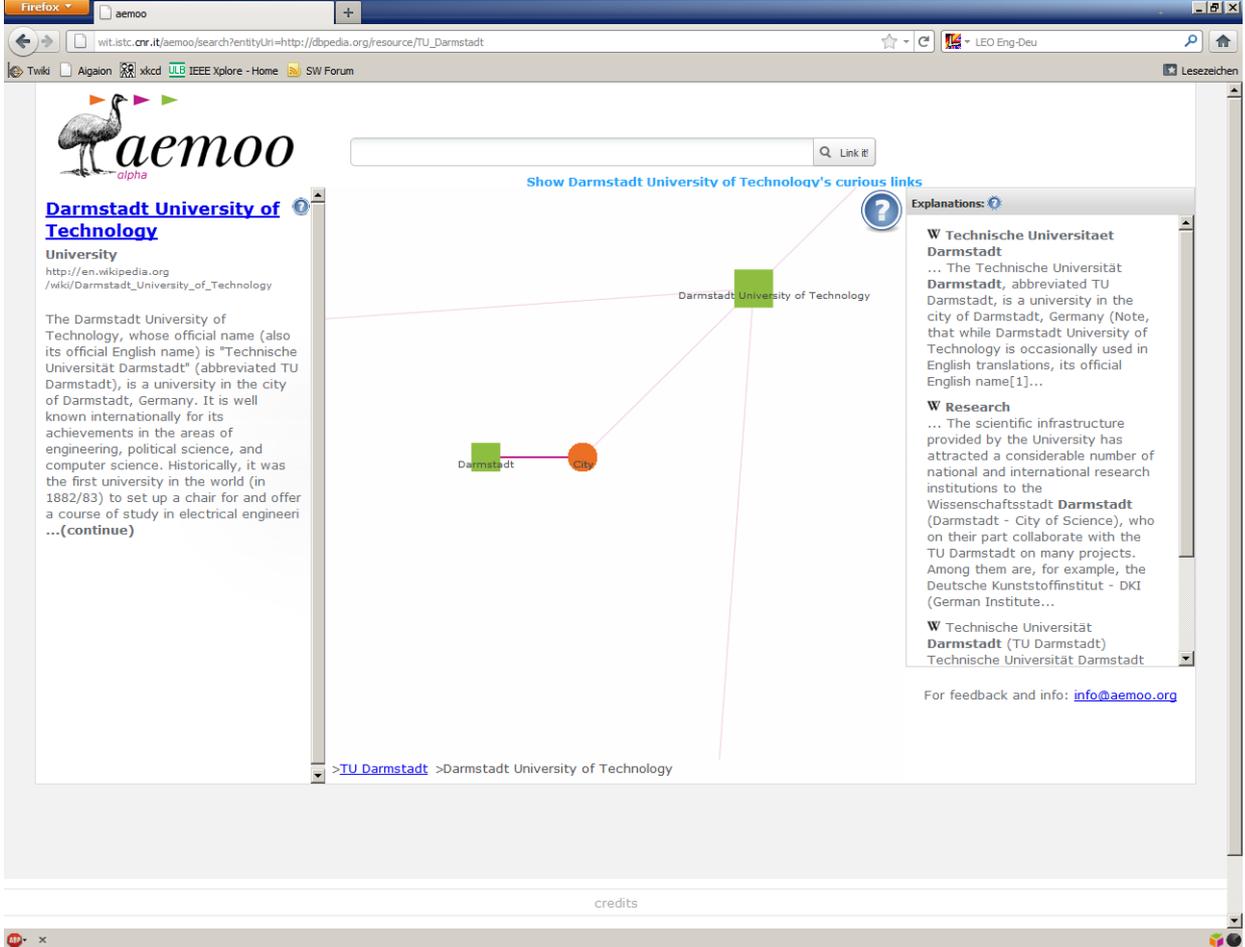
Princess Alice of th... en

More Info: [dbpedia.org](#)
[en.wikipedia.org](#)

The Princess Alice was a member of the British royal family, the third child and second daughter of Queen Victoria and Prince Albert of Saxe-Coburg and Gotha. Alice's education was devised by Albert's close friend and adviser, Baron Stockmar. Like her other siblings, Alice spent her early childhood in the company of her parents and siblings, travelling between the British royal residences. In 1861, Prince Albert became ill with typhoid fever, and Alice nursed him through his final illness before he died on 14 December. Following his death, Queen Victoria entered a period of intense mourning, and Alice spent the next six months acting as her unofficial secretary. While the court was still at the height of mourning, on 1 July 1862, Alice married the minor German Prince Louis of Hesse, heir to the Grand Duchy of Hesse. The ceremony—conducted privately and with unrelieved gloom at Osborne House—was

Browser für Linked Open Data

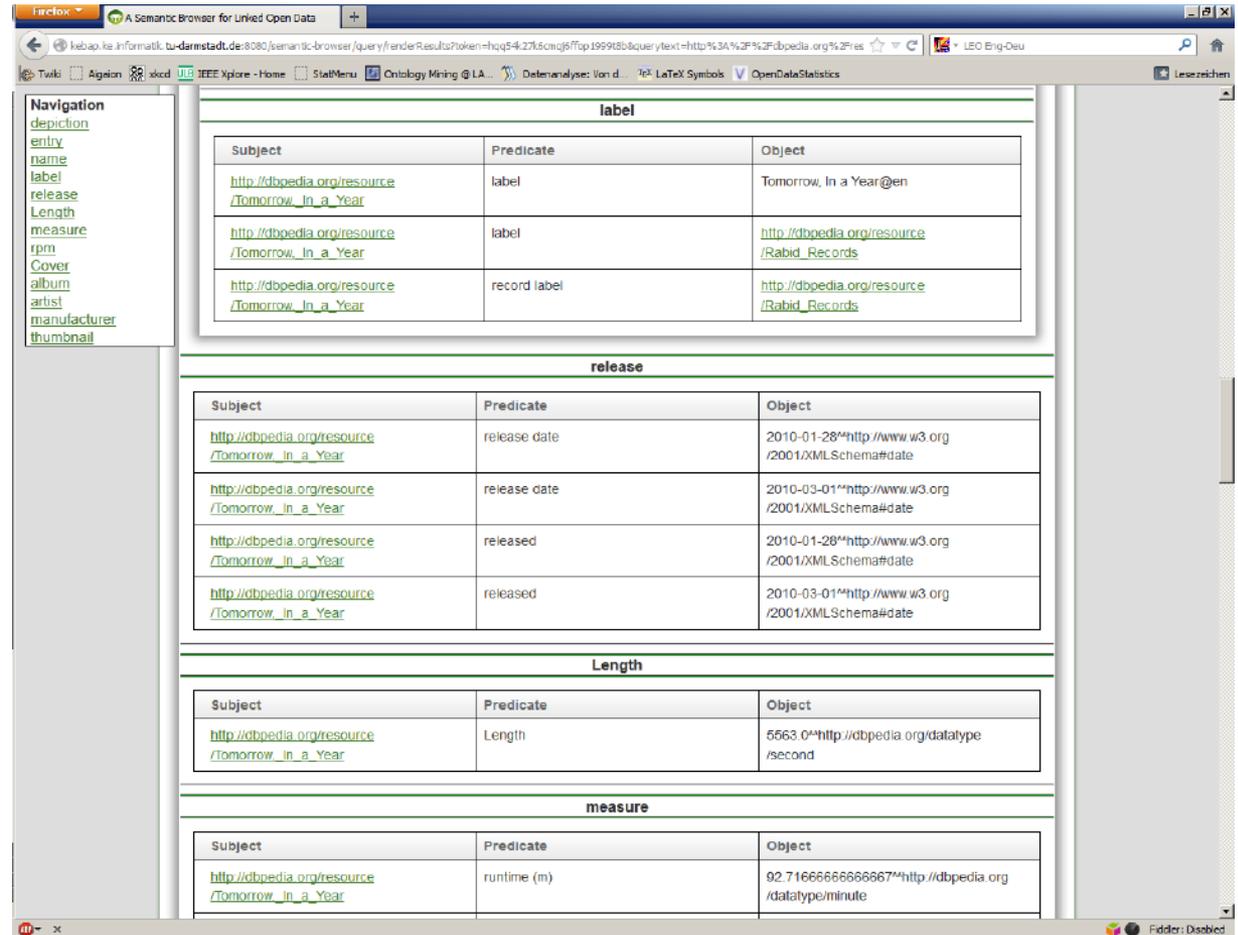
- Aemoo
- kombiniert Graph und Text
- verwendet weitere Texte aus Wikipedia



The screenshot shows the Aemoo browser interface. On the left, there is a sidebar with the Aemoo logo (an ostrich) and the text "Darmstadt University of Technology" with a link to its Wikipedia page. The main content area displays a graph with nodes for "Darmstadt", "City", and "Darmstadt University of Technology". A search bar at the top right of the graph area contains "Link #". Below the graph, there is a section titled "Explanations:" with two entries: "W Technische Universitaet Darmstadt" and "W Research". The "W Technische Universitaet Darmstadt" entry provides a definition of the university. The "W Research" entry describes the university's research infrastructure. At the bottom of the page, there is a "credits" section.

Browser für Linked Open Data

- MoB4LOD
- Entwickelt an der TU Darmstadt
- Modulares Framework
- Beliebige eigene Ansichten konfigurierbar



The screenshot shows a web browser window displaying a query result for the resource `http://dbpedia.org/resource/Tomorrow_in_a_Year`. The interface includes a navigation sidebar on the left and a main content area with several tables.

Navigation sidebar:

- depiction
- entry
- name
- label
- release
- Length
- measure
- rpm
- Cover
- album
- artist
- manufacturer
- thumbnail

label table:

Subject	Predicate	Object
http://dbpedia.org/resource/Tomorrow_in_a_Year	label	Tomorrow, In a Year@en
http://dbpedia.org/resource/Tomorrow_in_a_Year	label	http://dbpedia.org/resource/Rabid_Records
http://dbpedia.org/resource/Tomorrow_in_a_Year	record label	http://dbpedia.org/resource/Rabid_Records

release table:

Subject	Predicate	Object
http://dbpedia.org/resource/Tomorrow_in_a_Year	release date	2010-01-28* http://www.w3.org/2001/XMLSchema#date
http://dbpedia.org/resource/Tomorrow_in_a_Year	release date	2010-03-01* http://www.w3.org/2001/XMLSchema#date
http://dbpedia.org/resource/Tomorrow_in_a_Year	released	2010-01-28* http://www.w3.org/2001/XMLSchema#date
http://dbpedia.org/resource/Tomorrow_in_a_Year	released	2010-03-01* http://www.w3.org/2001/XMLSchema#date

Length table:

Subject	Predicate	Object
http://dbpedia.org/resource/Tomorrow_in_a_Year	Length	5563.0* http://dbpedia.org/datatype/second

measure table:

Subject	Predicate	Object
http://dbpedia.org/resource/Tomorrow_in_a_Year	runtime (m)	92.7166666666667* http://dbpedia.org/datatype/minute

Erstellen von Linked Open Data

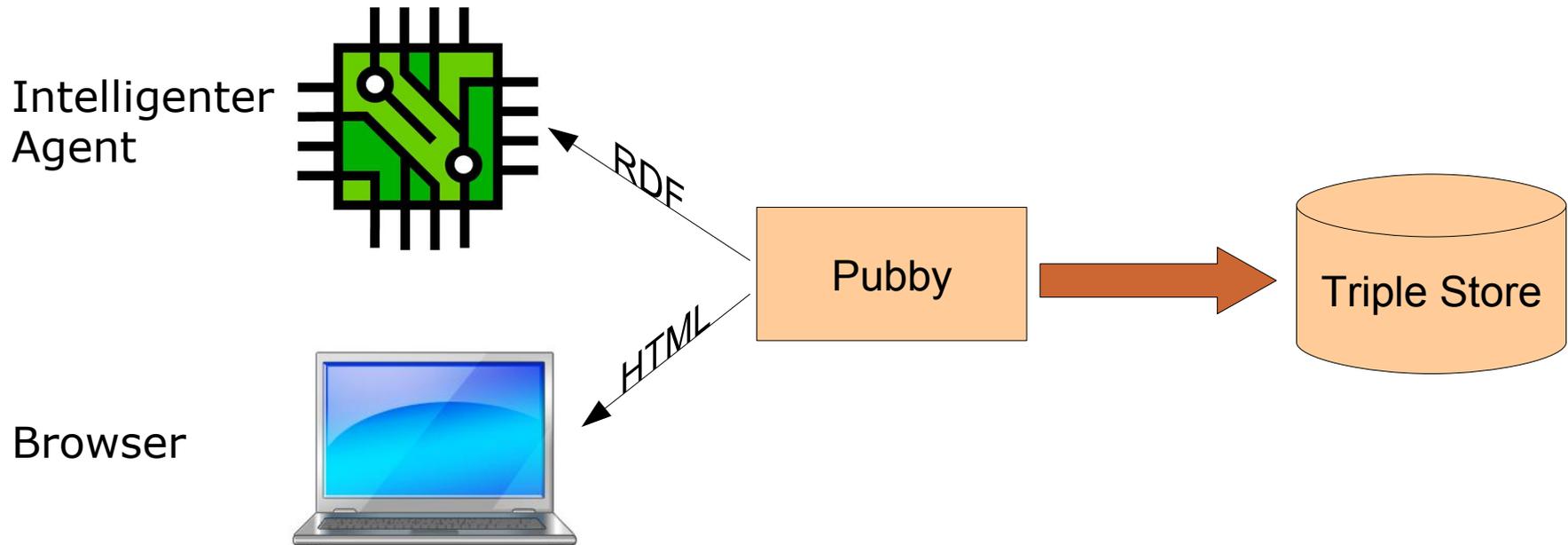


TECHNISCHE
UNIVERSITÄT
DARMSTADT

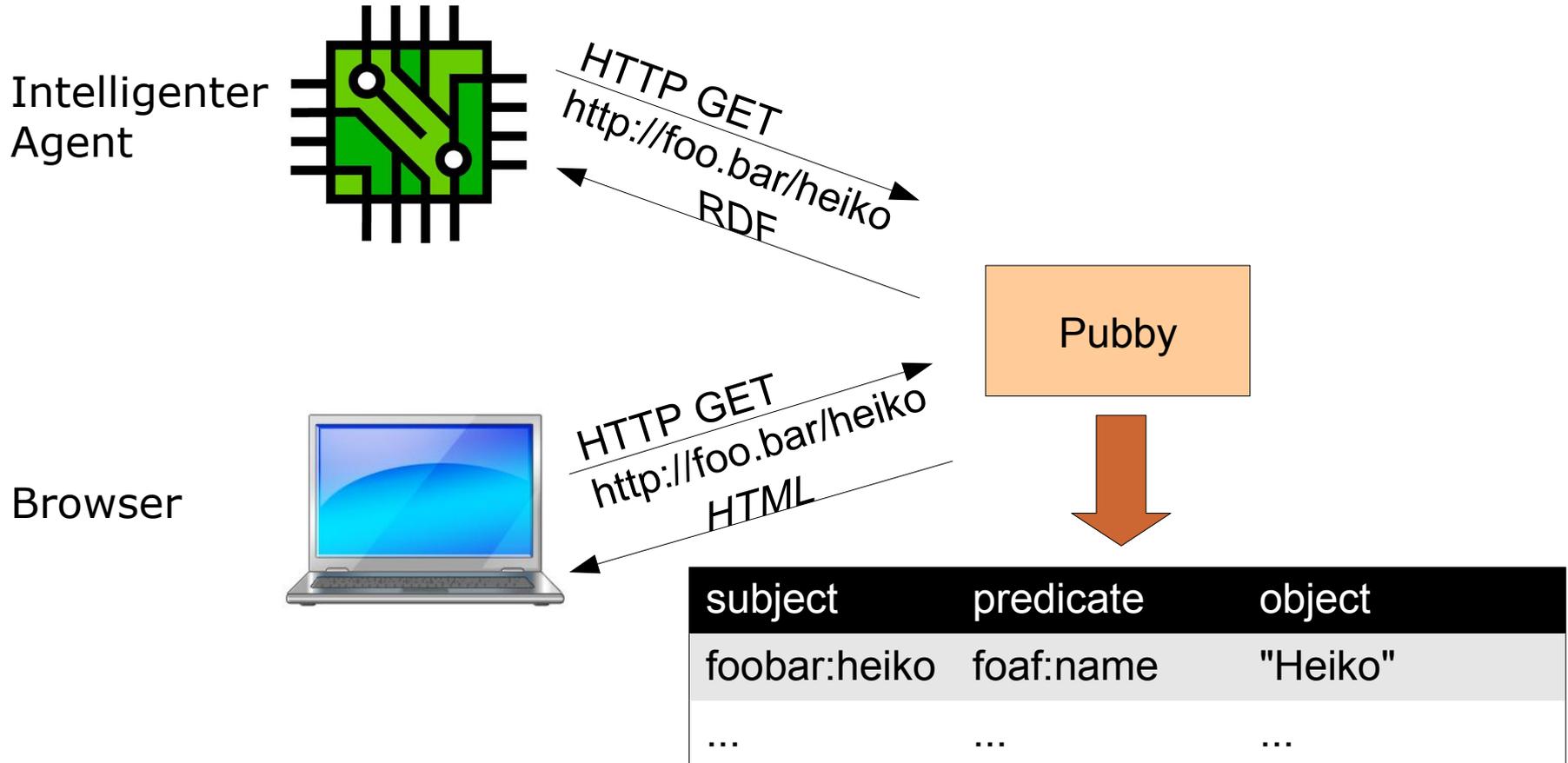
- Momentan verbreitete Varianten
 - handcodiert
 - aus strukturierten Quellen (z.B. Datenbanken)
 - aus (semi-)strukturierten Web-Quellen

Linked Open Data aus Triple Stores

- Pubby ist eine Erweiterung für Triple Stores
- Unterstützt Content Negotiation etc.

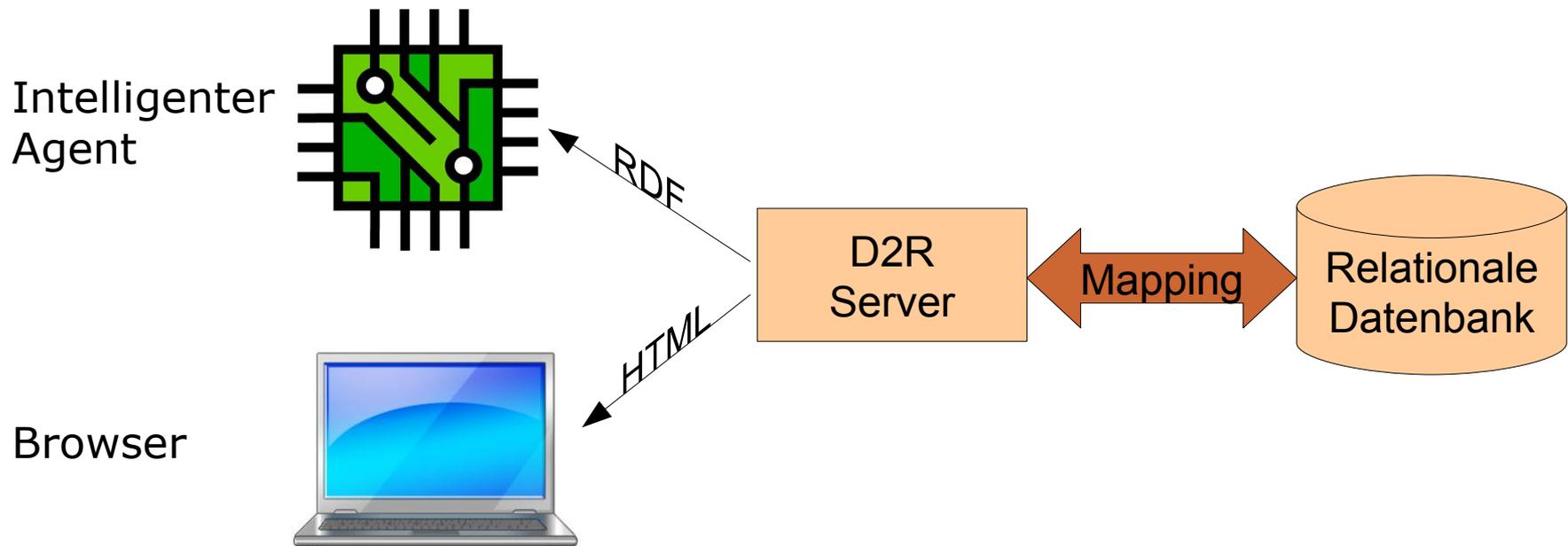


Linked Open Data aus Triple Stores



Linked Open Data aus relationalen Datenbanken

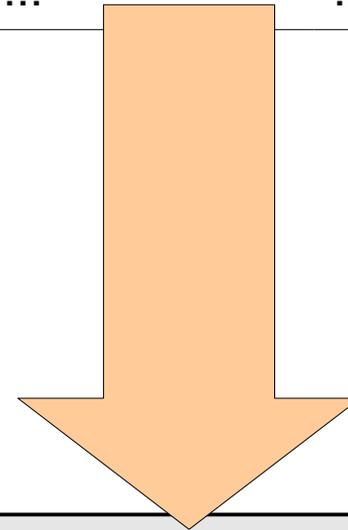
- D2R: Linked-Open-Data-Interface auf relationale Datenbanken
- Unterstützt Content Negotiation etc.



Linked Open Data aus relationalen Datenbanken

```
map:Person a d2rq:ClassMap;
  d2rq:dataStorage map:Database1.
  d2rq:class foaf:Person;
  d2rq:uriPattern "http://foo.bar/p@@Person.ID@";
  .
map:personName a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:Person;
  d2rq:property foaf:name;
  d2rq:column "Person.name";
  d2rq:datatype xsd:string;
  .
map:location a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:Person;
  d2rq:property foaf:basedNear;
  d2rq:column "Person.location";
  d2rq:datatype xsd:string;
  .
```

ID (int)	name (text)	location (int)
1327890123	"Heiko"	"Darmstadt"
...



```
<http://foo.bar/p1327890123> a foaf:Person .
<http://foo.bar/p1327890123> foaf:name "Heiko" .
<http://foo.bar/p1327890123> foaf:basedNear "Darmstadt" .
```

Linked Open Data: Anwendungen

- TrialX: Finden von medizinischen Studien
 - Applied Informatics, USA
- basierend auf
 - Personal Health Records
 - U.S. National Institute of Health
- Semantisches Matchmaking
 - eigene Krankengeschichte
 - geplante Studie

Linked Open Data: Anwendungen



TECHNISCHE
UNIVERSITÄT
DARMSTADT

The screenshot shows a Firefox browser window displaying the ClinicalTrials.gov website. The search results for 'asthma clinical trials' are shown, with 10 results found. The search criteria are 'asthma' and 'NY'. The results list several trials, with 'Evaluation of the Lungs of Individuals with Lung Disease' highlighted by a blue circle. The trial description states: 'The purpose of this protocol is to obtain biologic materials from the blood and lungs from patients with lung disease in order to develop an understanding of the etiology and ...'.



Linked Open Data: Anwendungen

- Explain-A-LOD (TU Darmstadt)
- Finden von Erklärungen für Statistiken

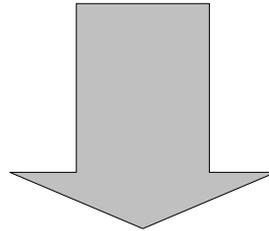
- Beispiele:
 - Mercer Quality of Living Survey
 - Transparency International CPI
 - ...

Linked Open Data: Anwendungen



- Beispiel Mercer:

City	Index
Vancouver	106
Zurich	106
...	...



City	Link	Index
Vancouver	http://dbpedia.org/resource/Vancouver	106
Zurich	http://dbpedia.org/resource/Zurich	106
...

Linked Open Data: Anwendungen

- Verwendete Daten:
 - Typinformation (DBPedia, YAGO, ...)
 - Datenattribute (z.B. Einwohnerzahl)
 - Unqualifizierte Relationen
 - Qualifizierte Relationen

City	Link	Index
Vancouver	http://dbpedia.org/resource/Vancouver	106
Zurich	http://dbpedia.org/resource/Zurich	106
...

Linked Open Data: Anwendungen

- Analysen:
 - Korrelation einzelner Attribute
 - Regellerner für komplexe Zusammenhänge

City	Link	Index
Vancouver	http://dbpedia.org/resource/Vancouver	106
Zurich	http://dbpedia.org/resource/Zurich	106
...

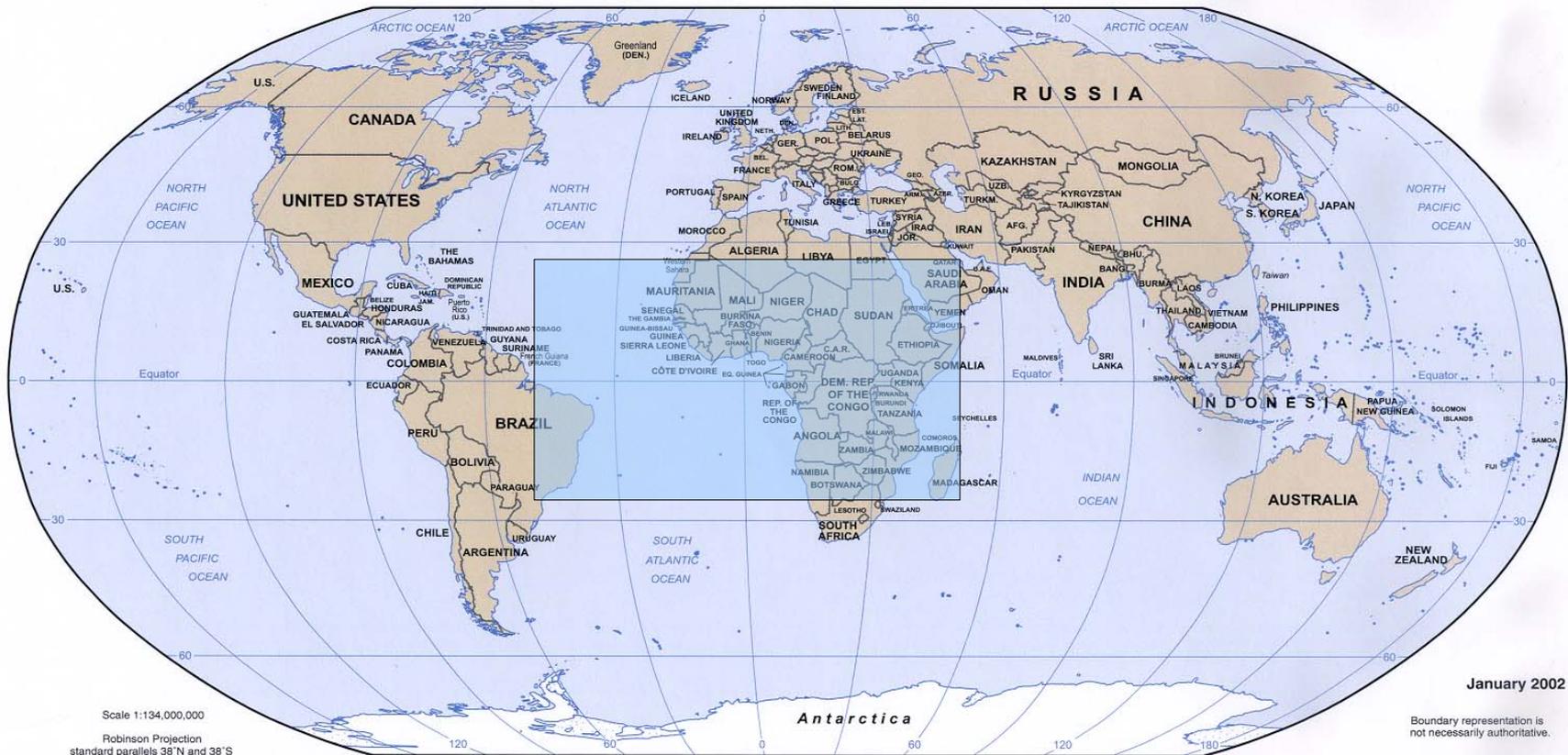
Linked Open Data: Anwendungen

- Beispiele für Städte mit niedriger Qualität
 - große heiße Städte ($\text{janHighC} \geq 27$ and $\text{areaTotalKm} \geq 334$)
 - kalte Städte, in denen keine Schallplatten gemacht werden ($\text{recordedIn_in} = \text{false}$ and $\text{janHighC} \leq 16$)
 - $\text{latd} \leq 24$ and $\text{longd} \leq 47$
 - Breiten- und Längengrad ohne weitere Angabe (N/S/W/O)
 - ?!?

Linked Open Data: Anwendungen



TECHNISCHE
UNIVERSITÄT
DARMSTADT

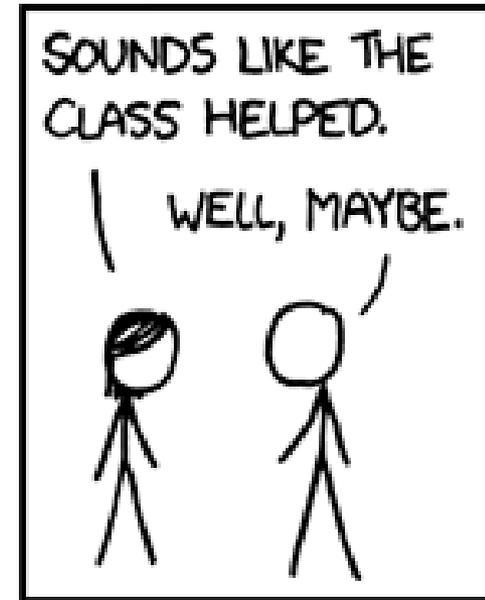
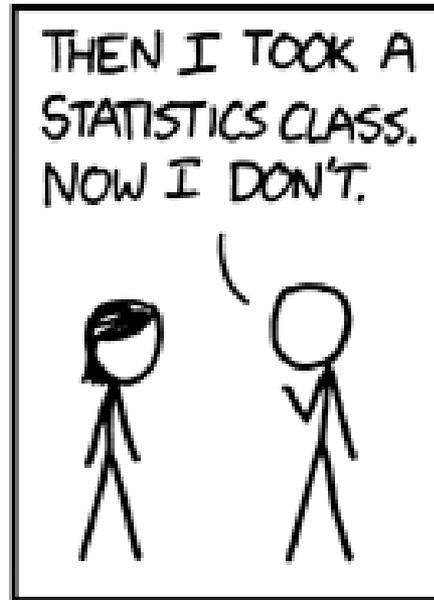
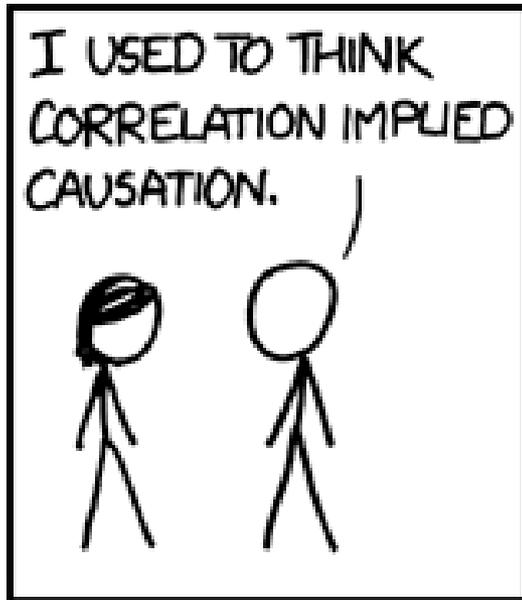


Linked Open Data: Anwendungen

- Beispiele für Länder mit niedriger Korruption
 - HDI > 78%
 - Human Development Index, berechnet aus Lebenserwartung, Bildung, Wirtschaftskraft
 - OECD-Mitgliedsstaaten (YAGO-Typ)
 - Gründungsort von mehr als neun Organisationen
 - Mehr als zehn Berge
 - Mehr als zehn Firmen mit Hauptquartier in diesem Staat, aber weniger als zwei Frachtfluglinien

Linked Open Data: Anwendungen

- Vorsicht:
 - Datenqualität
 - Korrelation vs. Kausalität
 - Open World Assumption



<http://xkcd.com/552/>

Linked Open Data: Zusammenfassung



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Bereitstellung von RDF-Daten
 - Verlinkt zu anderen Daten
 - Auflösbare URIs
- Beispiele
- gebräuchliche Vokabulare
- Techniken
- einige Anwendungen

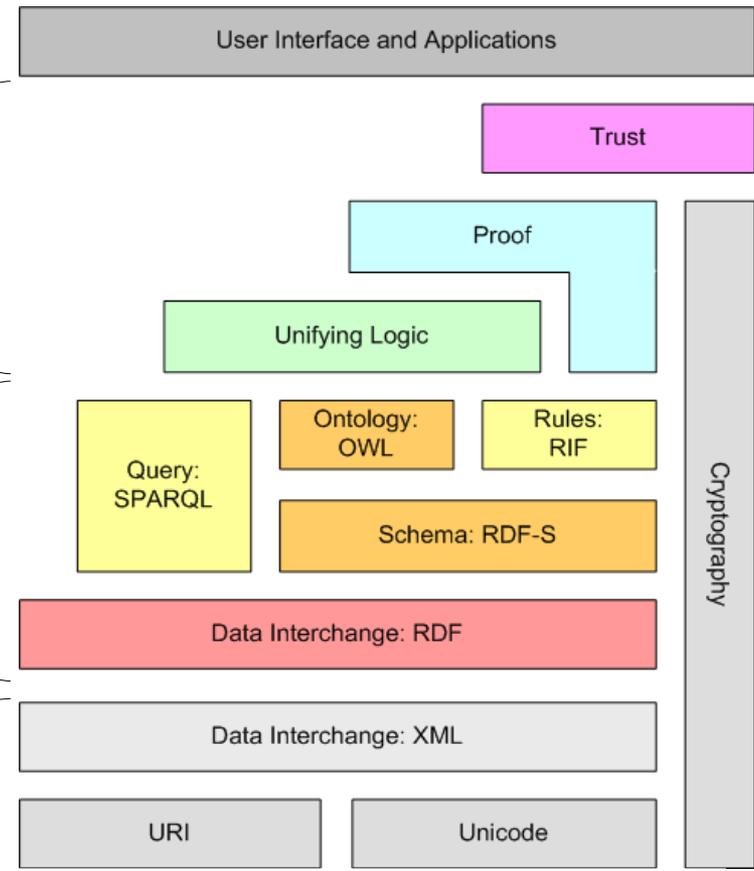
Semantic Web – Aufbau



here be dragons...

Semantic-Web-
Technologie
(Fokus der Vorlesung)

Technische
Grundlagen



Berners-Lee (2009): *Semantic Web and Linked Data*
<http://www.w3.org/2009/Talks/0120-campus-party-tbl/>

Vorlesung Semantic Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vorlesung im Wintersemester 2011/2012

Dr. Heiko Paulheim

Fachgebiet Knowledge Engineering