Vorlesung Semantic Web

TECHNISCHE UNIVERSITÄT DARMSTADT

Vorlesung im Wintersemester 2011/2012

Dr. Heiko Paulheim

Fachgebiet Knowledge Engineering

Linked Open Data

TECHNISCHE UNIVERSITÄT DARMSTADT

- 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"?

TECHNISCHE UNIVERSITÄT DARMSTADT

```
: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?

TECHNISCHE UNIVERSITÄT DARMSTADT

- 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"?

TECHNISCHE UNIVERSITÄT DARMSTADT

```
: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"?

TECHNISCHE UNIVERSITÄT DARMSTADT

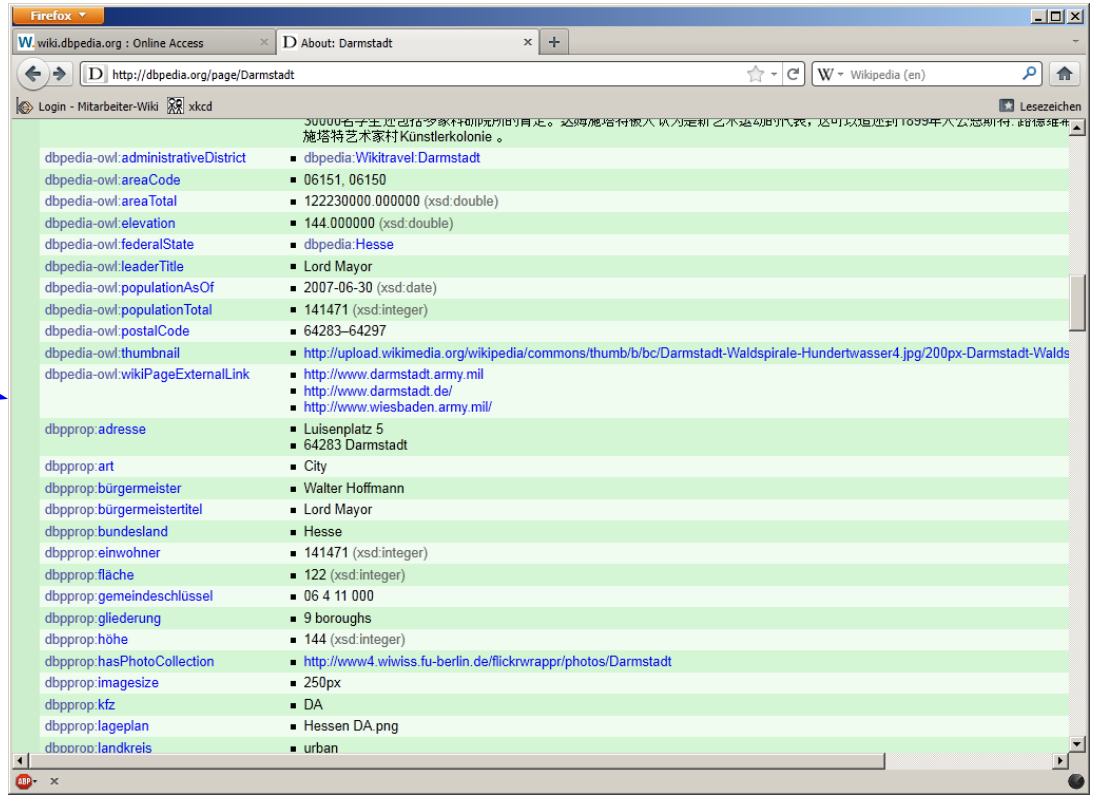
- Linked Open Data sind RDF-Daten
  - verteilt vorgehalten

- 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 Negotation oder RDFa einsetzen

Linked Open Data – was heißt hier "Linked"?

TECHNISCHE UNIVERSITÄT DARMSTADT

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



Linked Open Data – was heißt hier "Linked"?

TECHNISCHE UNIVERSITÄT DARMSTADT

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

Links in HTML und in Linked Open Data

TECHNISCHE UNIVERSITÄT DARMSTADT

- Vergleiche

```
Heiko arbeitet in <a href="http://www.darmstadt.de">Darmstadt</a>.
```

und

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

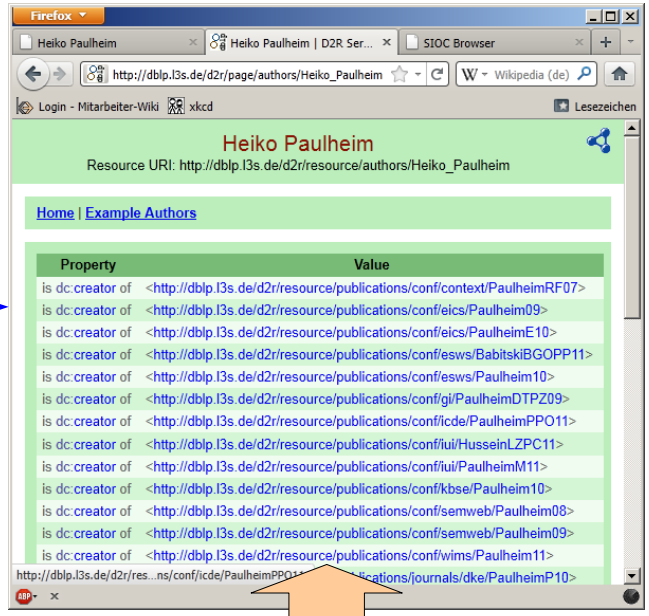
- Merke:
  - In Linked Open Data sind Links immer *explizit* typisiert
  - Der Typ des Links ist damit dem Computer unmittelbar verständlich

Linked Open Data: Mehrere Datensets zu einer Ressource

TECHNISCHE UNIVERSITÄT DARMSTADT

- Wichtiger Spezialfall: owl:sameAs*

```
:Heiko
  owl:sameAs
  <http://dblp.l3s.de/d2r/page/
   authors/Heiko_Paulheim>
```



* 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`

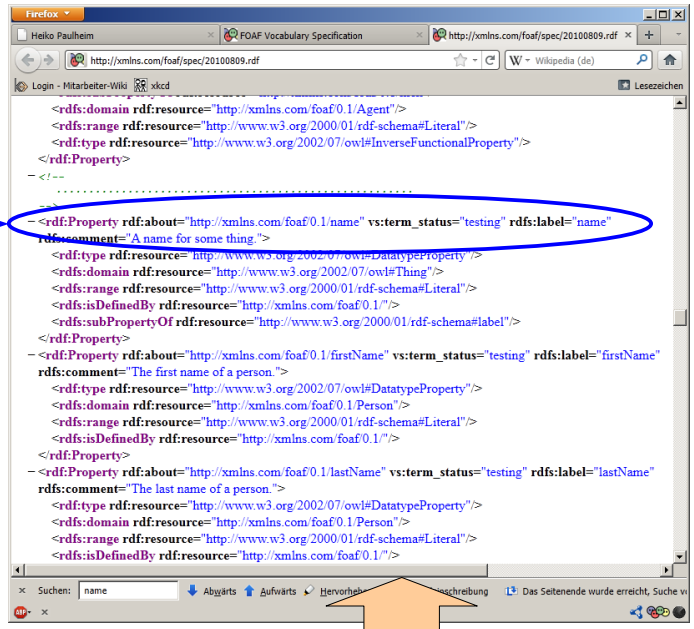
* OWL bekommen wir erst noch – das muss uns jetzt aber nicht stören...

Linked Open Data: Links auf das Schema

TECHNISCHE UNIVERSITÄT DARMSTADT

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

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



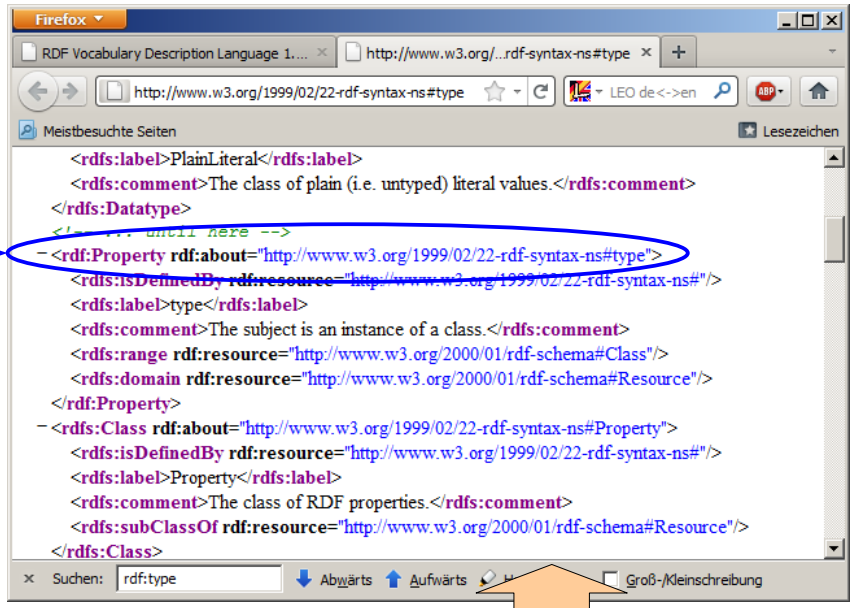
Linked Open Data: Links auf das Schema

TECHNISCHE UNIVERSITÄT DARMSTADT

- Das funktioniert auch mit den "eingebauten" Schemata

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

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



Linked Open Data: Die vier Grundregeln

TECHNISCHE UNIVERSITÄT DARMSTADT

- 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?

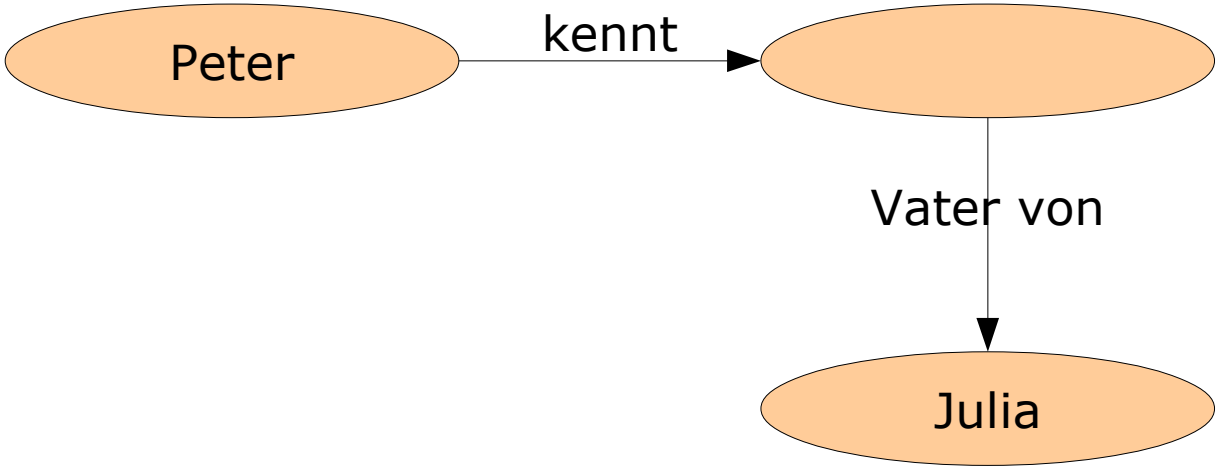
TECHNISCHE UNIVERSITÄT DARMSTADT

- 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

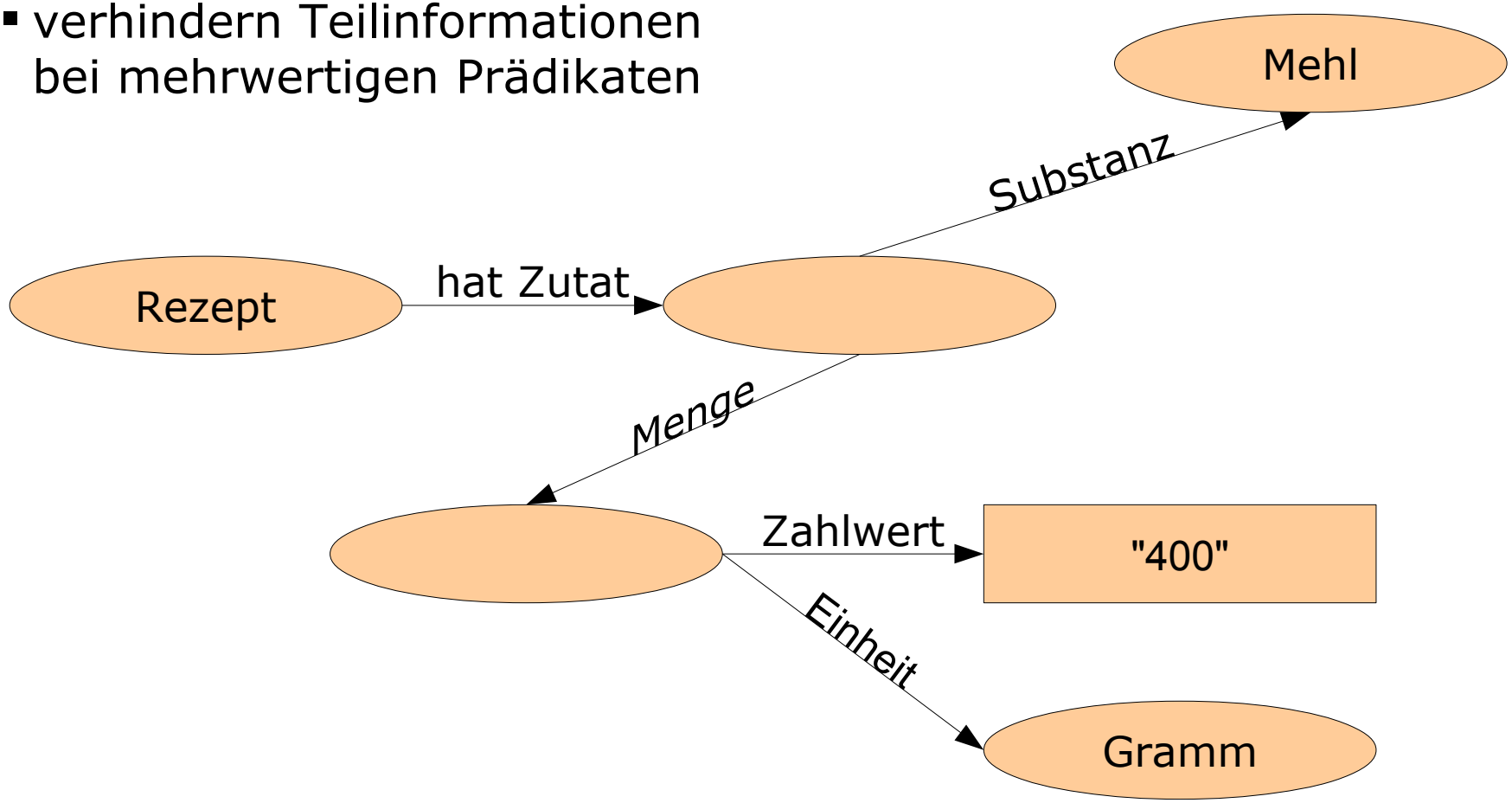
RDF-Moleküle

- Verhindern "Sackgassen" beim Browsen



RDF-Moleküle

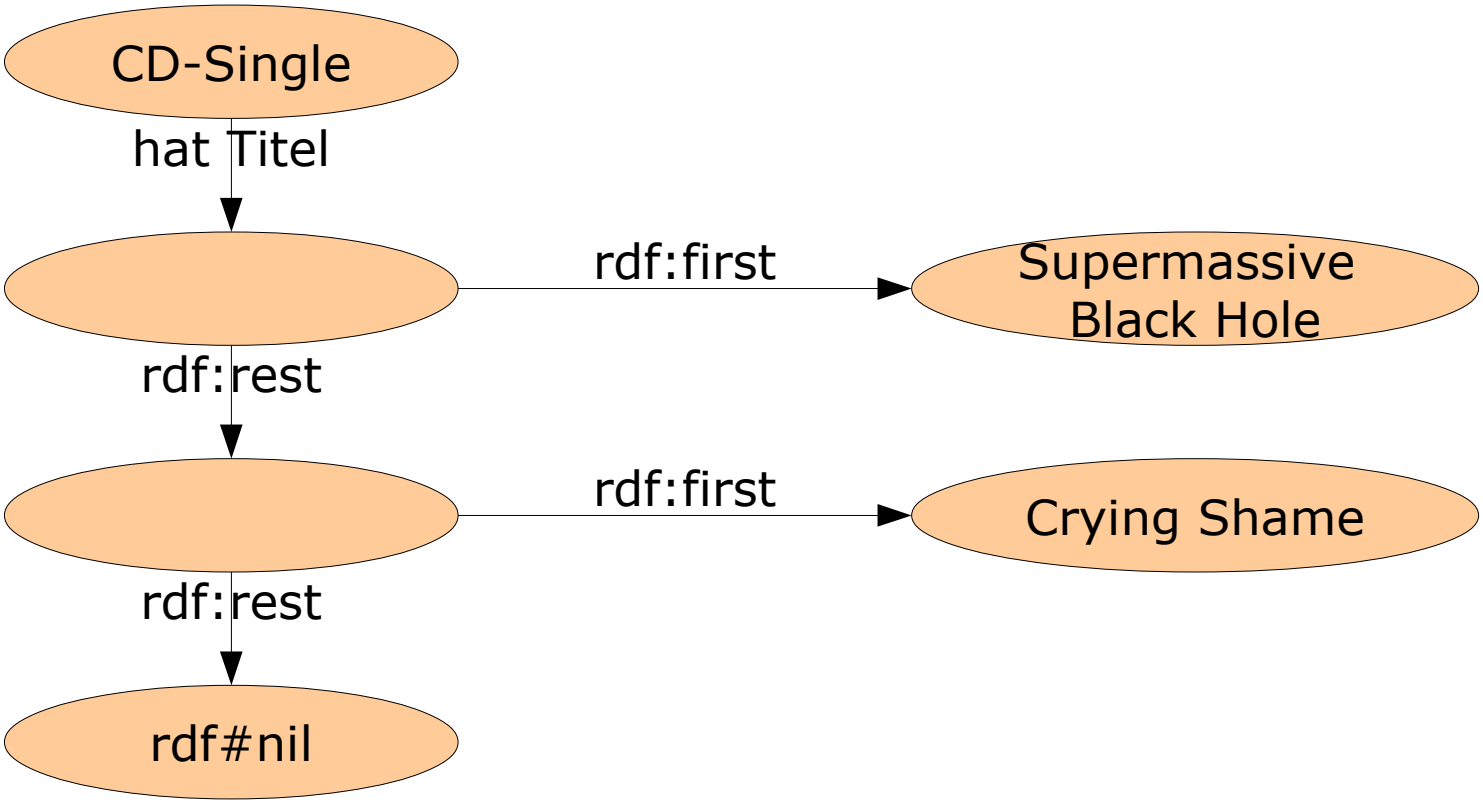
- verhindern Teilinformationen bei mehrwertigen Prädikaten



RDF-Moleküle

TECHNISCHE UNIVERSITÄT DARMSTADT

- verhindern unvollständige Listen



RDF-Moleküle: Theorie und Praxis

TECHNISCHE UNIVERSITÄT DARMSTADT

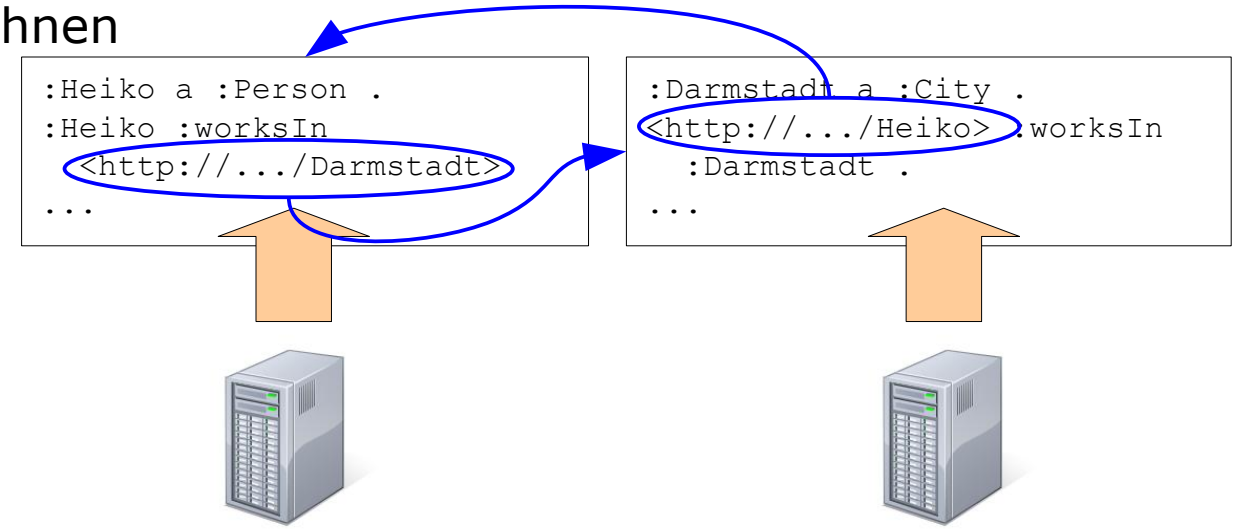
- Was ist ein RDF-Molekül?
  - Alle Tripel, die die Ressource als Subjekt oder Objekt enthalten
  - Alle Blank Nodes haben mindestens zwei ein- oder ausgehende Kanten

- Folgen:
  - Tripel werden u.U. dupliziert (Redundanzen!)
  - Moleküle können sehr groß werden

RDF-Moleküle: Theorie und Praxis

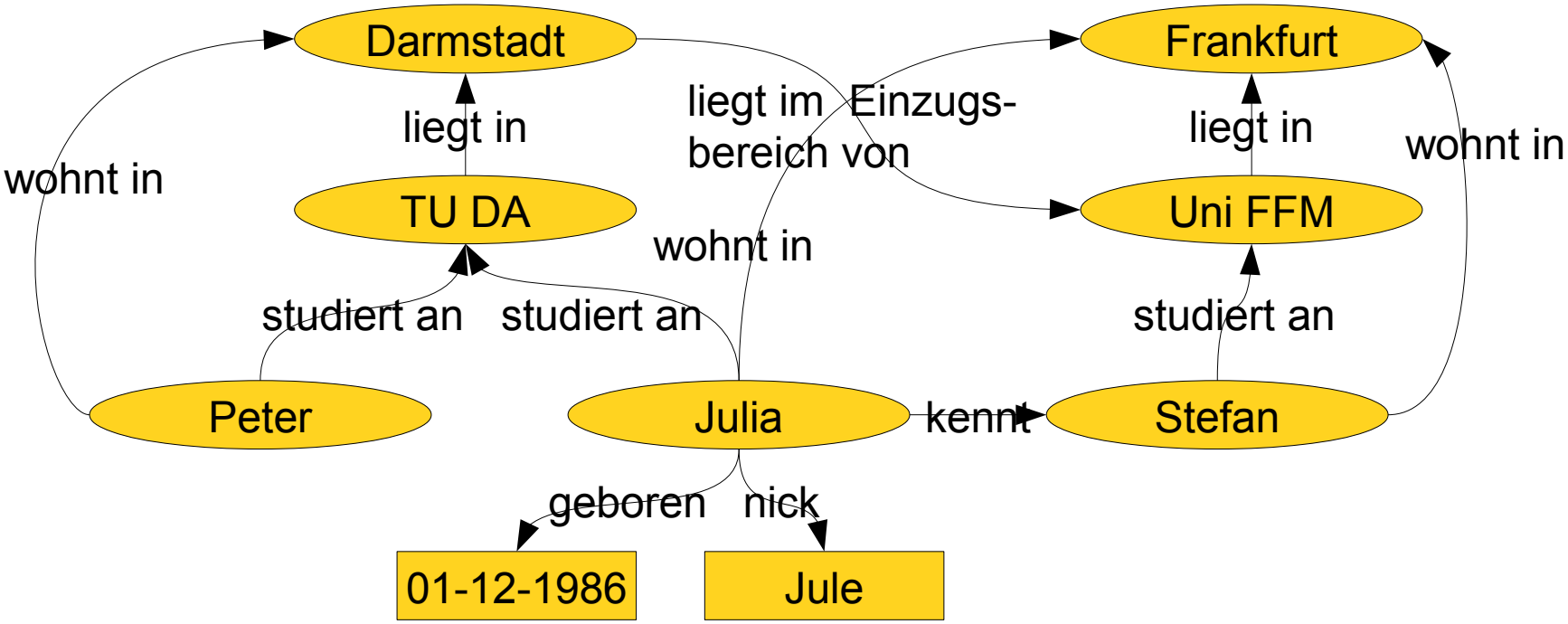
TECHNISCHE UNIVERSITÄT DARMSTADT

- 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

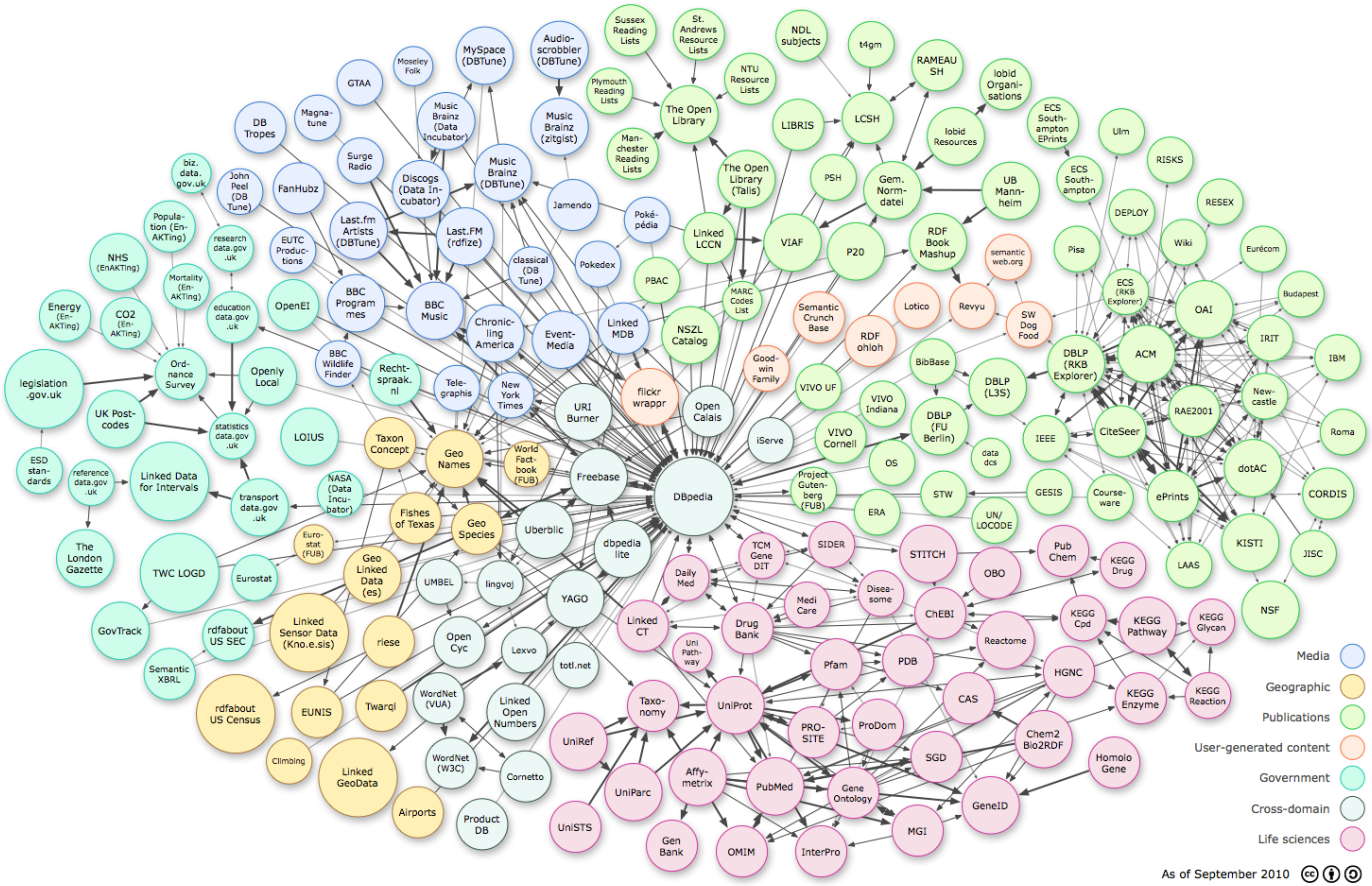


RDF-Moleküle: Pragmatischer Ansatz

TECHNISCHE UNIVERSITÄT DARMSTADT

- Sinnvolles Molekül zu einer Person:
  - komplettes Molekül
- Sinnvolles Molekül zu einer Universität:
  - alle ausgehende Kanten "liegt in"
  - alle eingehenden Kanten "liegt im Einzugsbereich von"
- ...

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

TECHNISCHE UNIVERSITÄT DARMSTADT

- In Zahlen:
  - 200 Datensets
  - ~25 Mrd. Tripel
  - ~830 Mio. Links zwischen Datensets

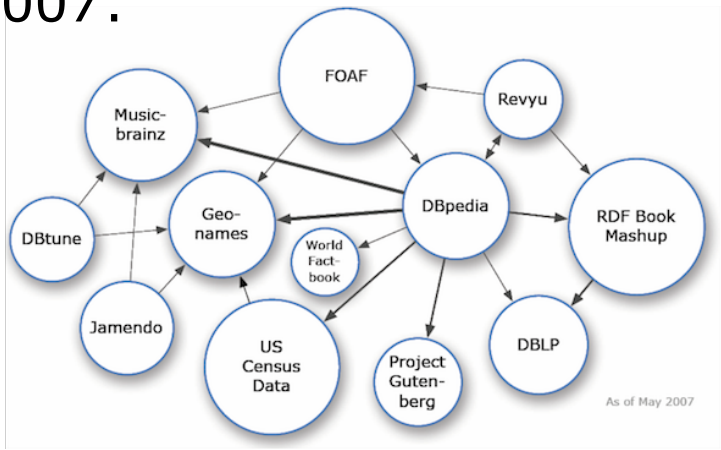
- In Themen:
  - Medien
  - Geodaten
  - Wissenschaftliche Publikationen
  - Regierungsdaten
  - Biologie
  - User Generated Content
  - übegreifende Themen

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

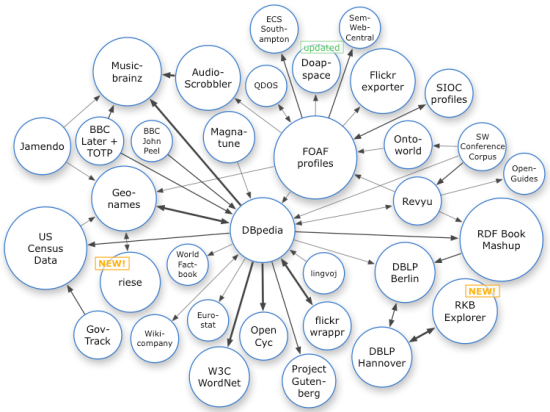
Die Linked Open Data Cloud: Geschichte

TECHNISCHE UNIVERSITÄT DARMSTADT

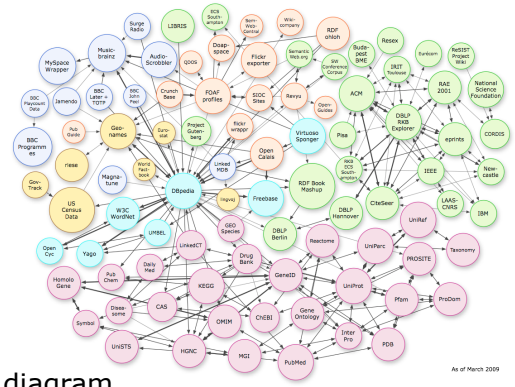
- Mai 2007:



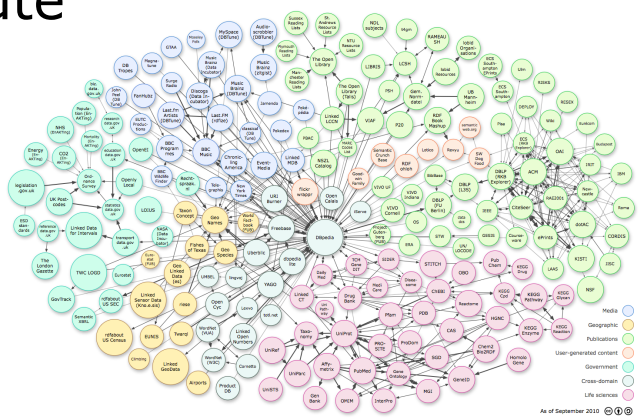
- März 2008:



- März 2009



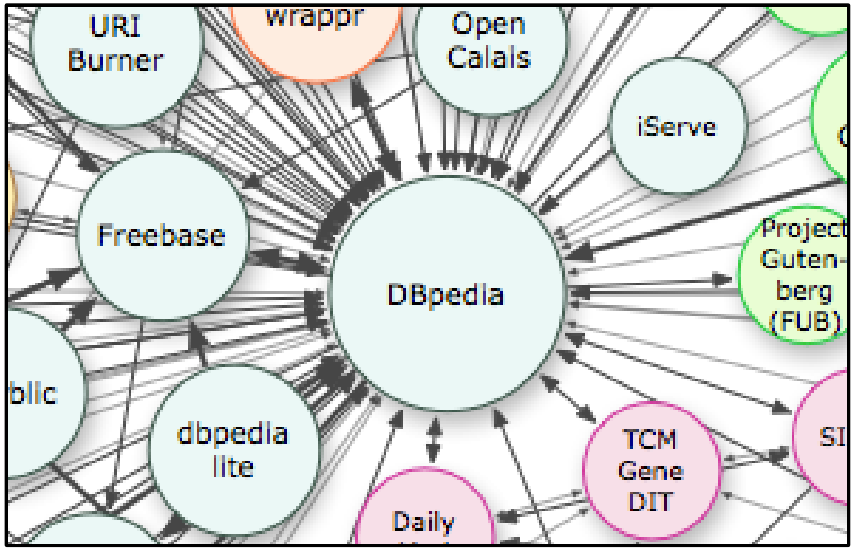
- heute



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

Das "Herz" der LOD Cloud: DBPedia

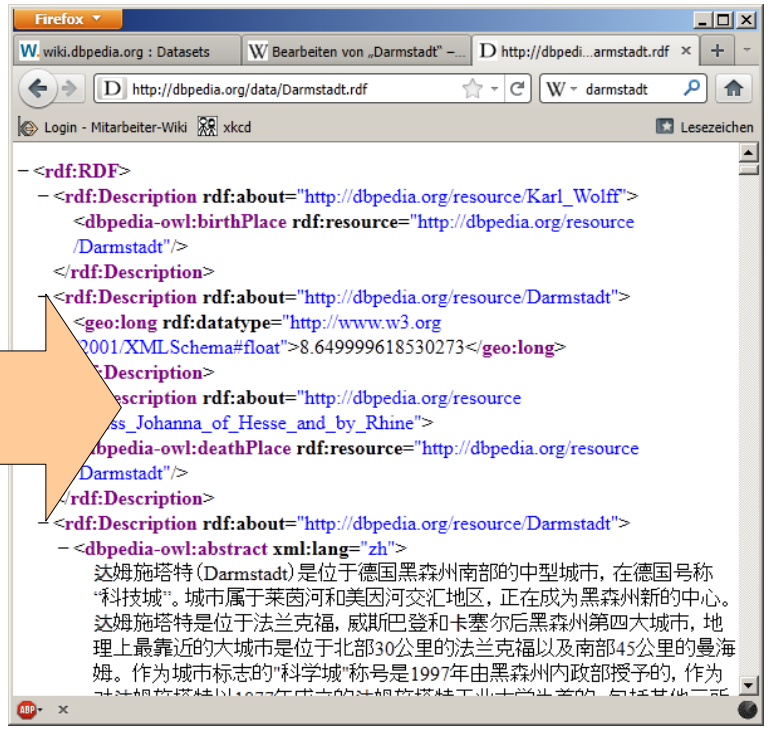
- Allgemeinwissen über 3,5 Millionen Entitäten
- 500 Millionen Tripel
- Mit 100 weiteren Datensets verknüpft
  - Das am stärksten vernetzte Datenset



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

DBPedia: Funktionsweise

```
{{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
...
}}
```



DBPedia: Funktionsweise

Geography

[edit]

Climate

Climate data for Darmstadt [hide]

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 (46)	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

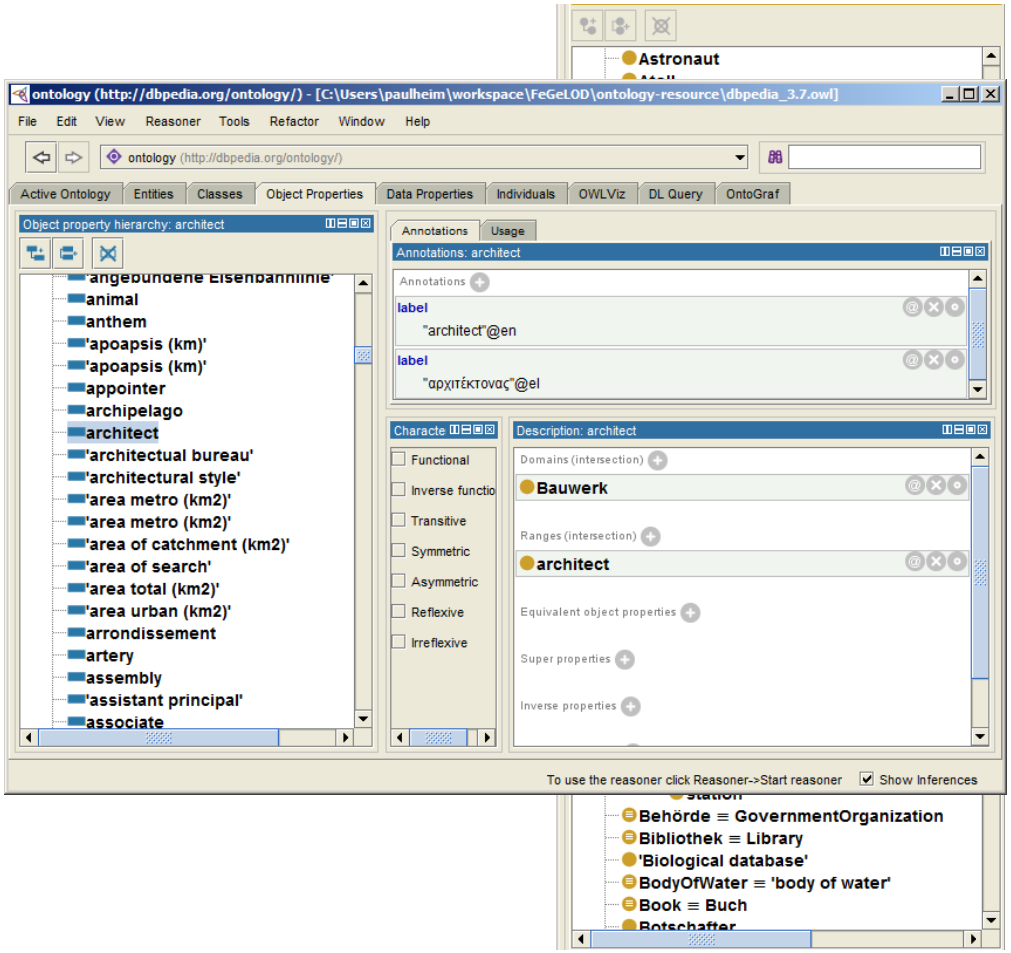
Darmstadt is relatively typical for German cities in terms of its transport system, with the car being the main, but not overwhelmingly dominant mode of transport.

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



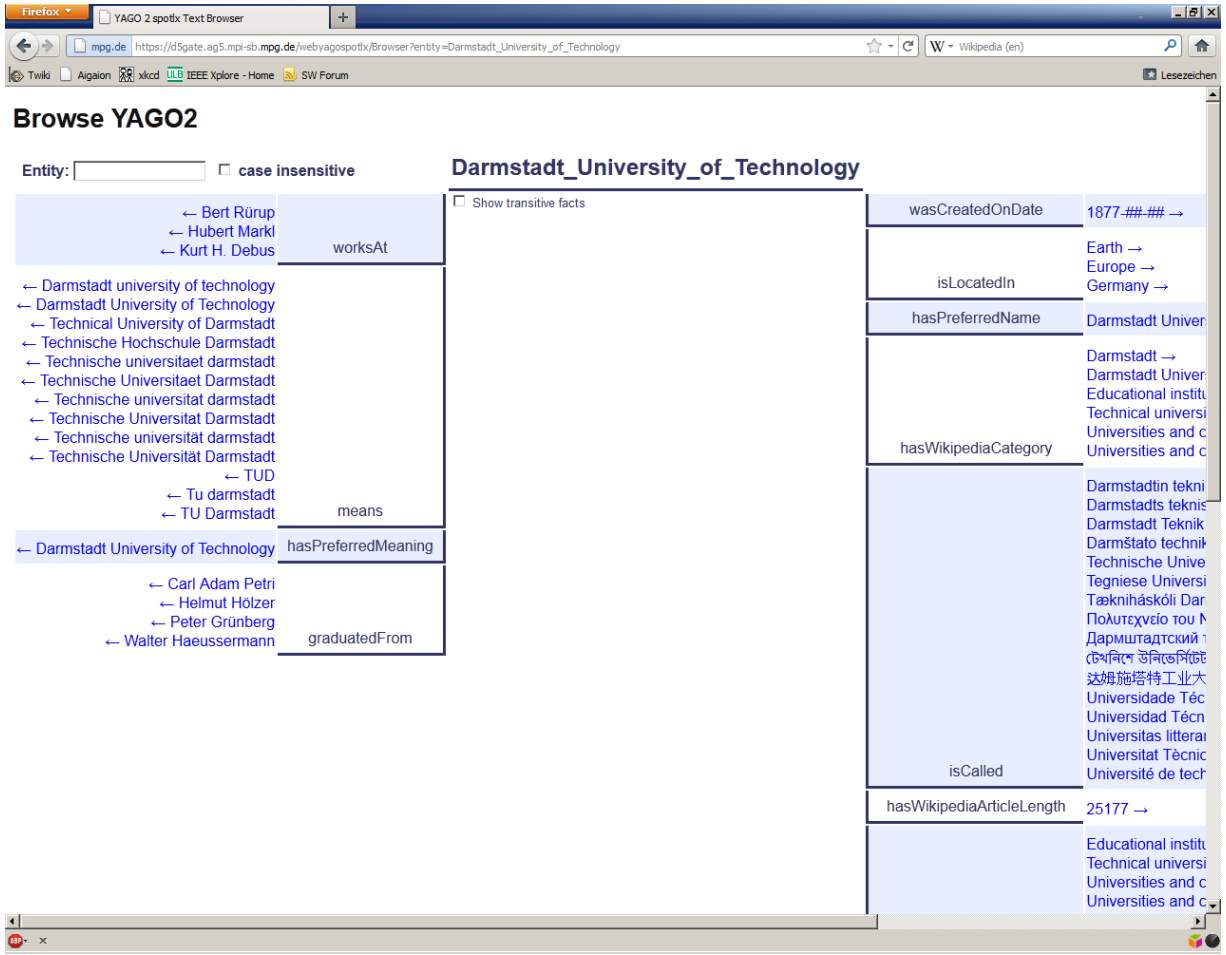
YAGO

- "Yet Another Great Ontology"

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

YAGO

TECHNISCHE UNIVERSITÄT DARMSTADT



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
  - Freebase
    - ist exakter

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

- Beispiel: Anzahl der Studenten
  - laut DBPedia: 23 (vgl. h_da: 11000?!)
  - laut Freebase: 23.100

- Klassischer Trade-Off zwischen Precision und Recall

OpenCyc

- von "en**cyc**lopedia"
- mehrere 100.000 Begriffe
- mehrere Millionen Aussagen
- nutzt intern eine eigene Sprache (CycL)

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

LinkedGeoData

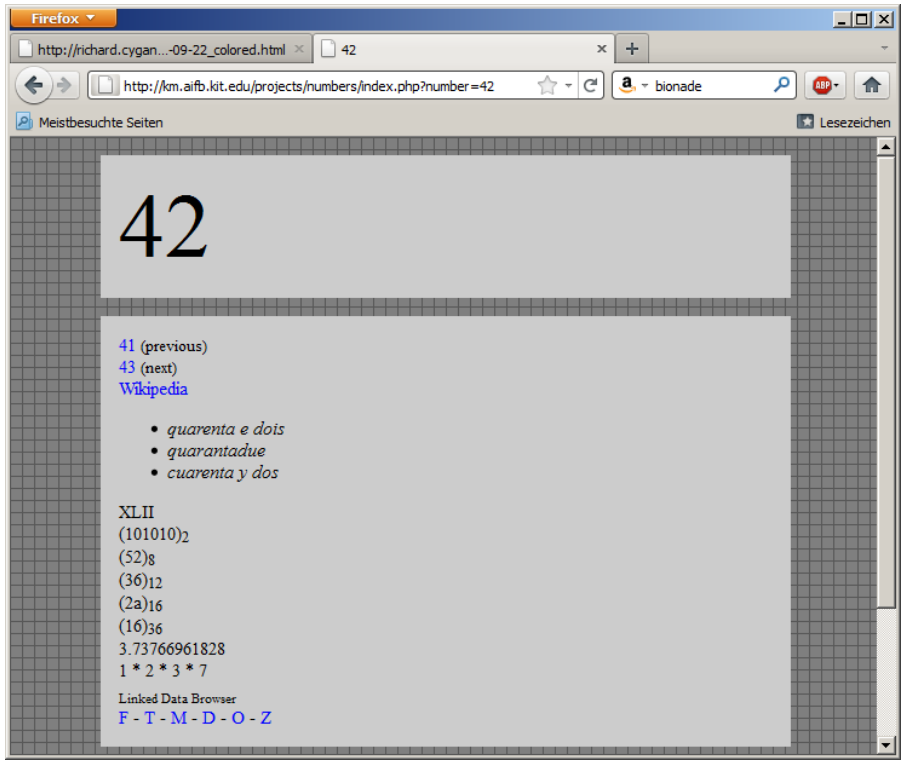
- hat auch ein eigenes User Interface:

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.

Beispiele für Datensets

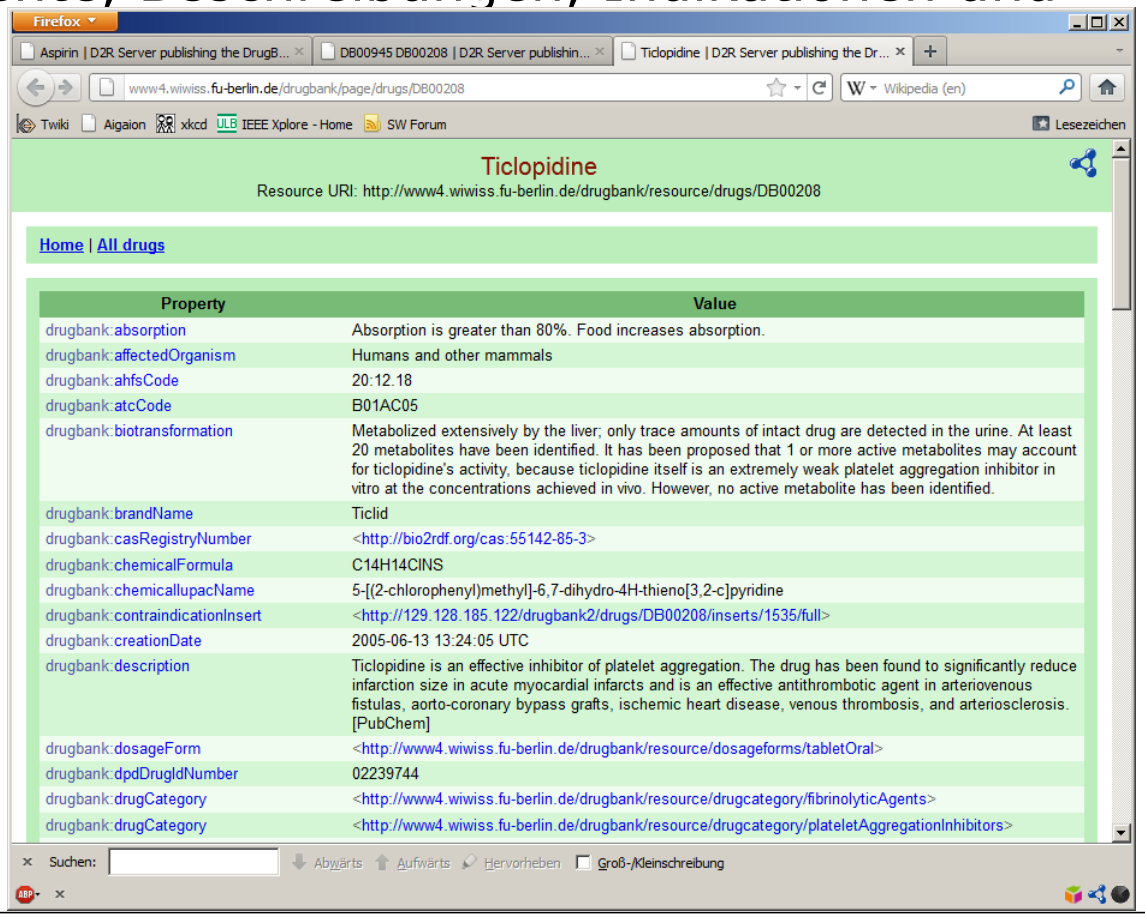
- ProductDB
  - Produkte und Hersteller
  - Mit Crawlern von verschiedenen Webseiten extrahiert
- DrugBank
  - 4700 Medikamentwirkstoffe
  - Produkte, Nebenwirkungen, Wechselwirkungen...
- NASA Data Incubator
  - Daten über NASA-Missionen
- Linked Open Numbers
  - Zahlen, Namen in verschiedenen Sprachen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



Beispiele für Datensets: DrugBank

TECHNISCHE UNIVERSITÄT DARMSTADT

- ca. 4500 Medikamente, Beschreibungen, Indikationen und Wechselwirkungen



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

TECHNISCHE UNIVERSITÄT DARMSTADT

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

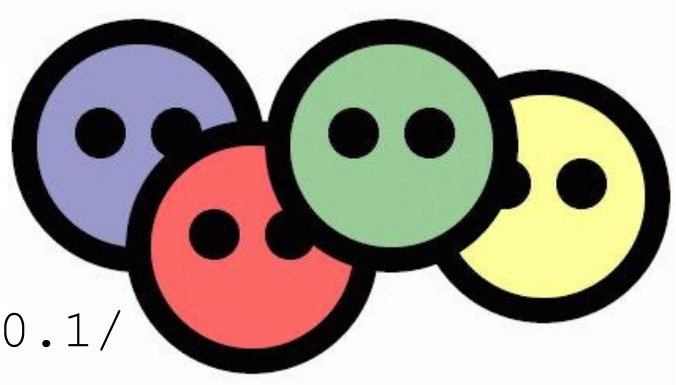
Dublin Core

- 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)

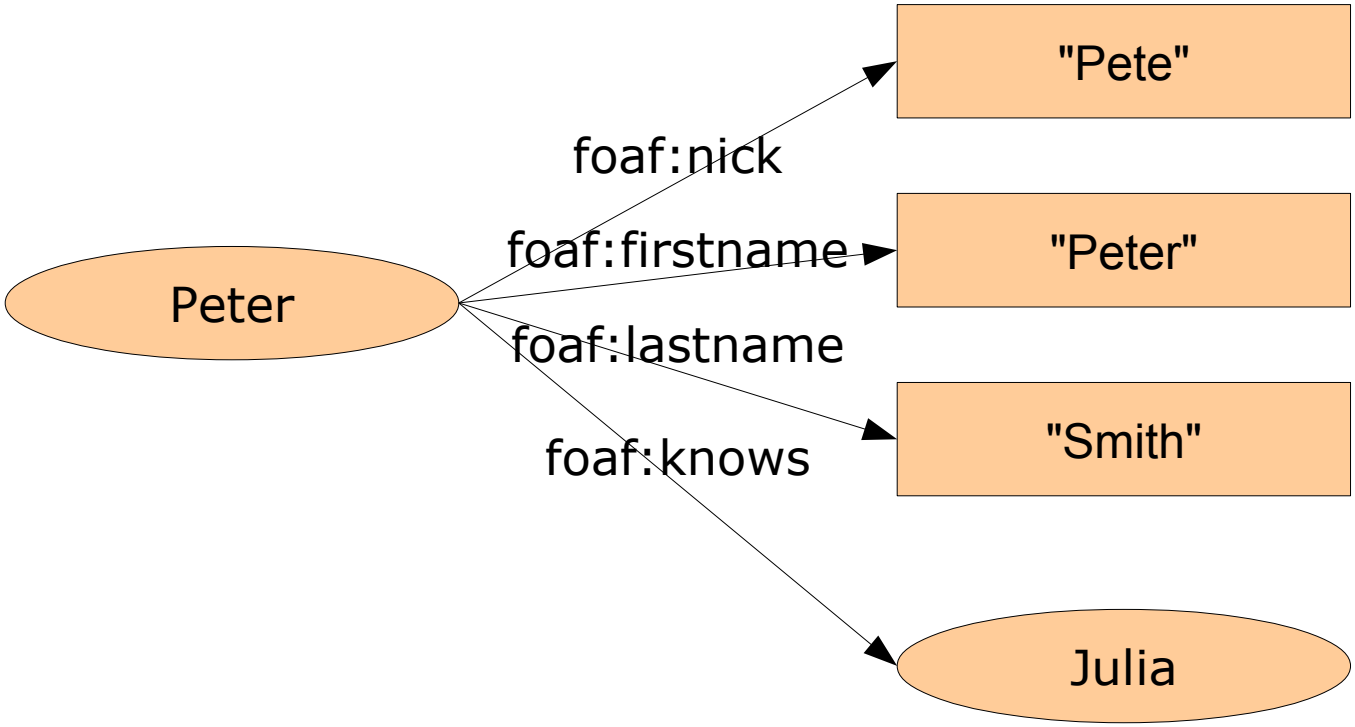
TECHNISCHE UNIVERSITÄT DARMSTADT

- Beschreibung von Personen und ihrer Relationen
- Für Metadaten privater Homepages (aber nicht nur)
- Namespace `http://xmlns.com/foaf/0.1/`
- Präfix meist `foaf:`

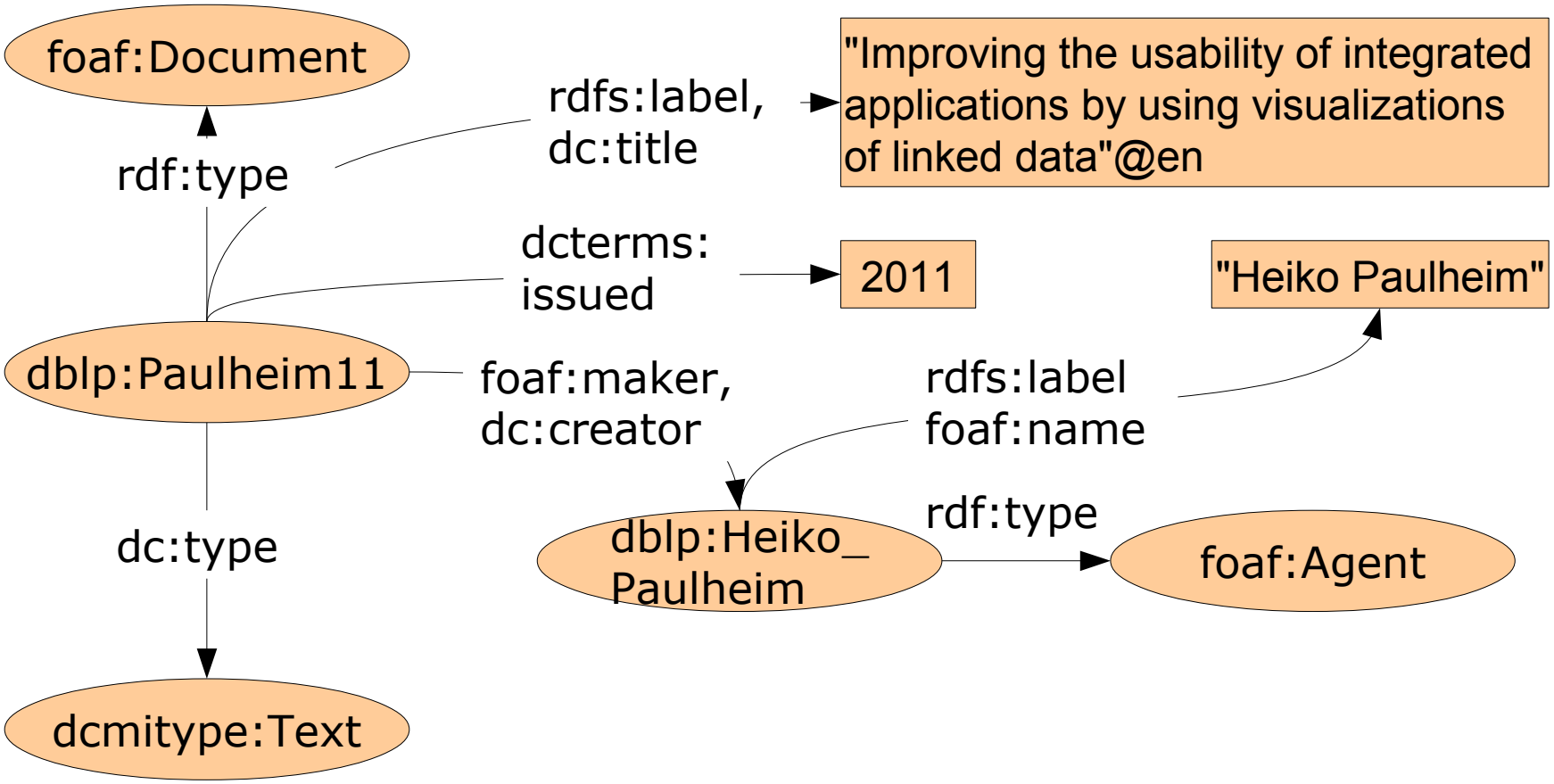


- Wichtige Klassen:
  - Person
  - Group
  - Organization
  - Project
  - ...
- Wichtige Properties:
  - name, firstName, lastName
  - phone, mbox, homepage
  - knows
  - currentProject, pastProject
  - ...

FOAF
(Friend of a Friend)



Beispiel für FOAF + DC: DBLP



WGS 84

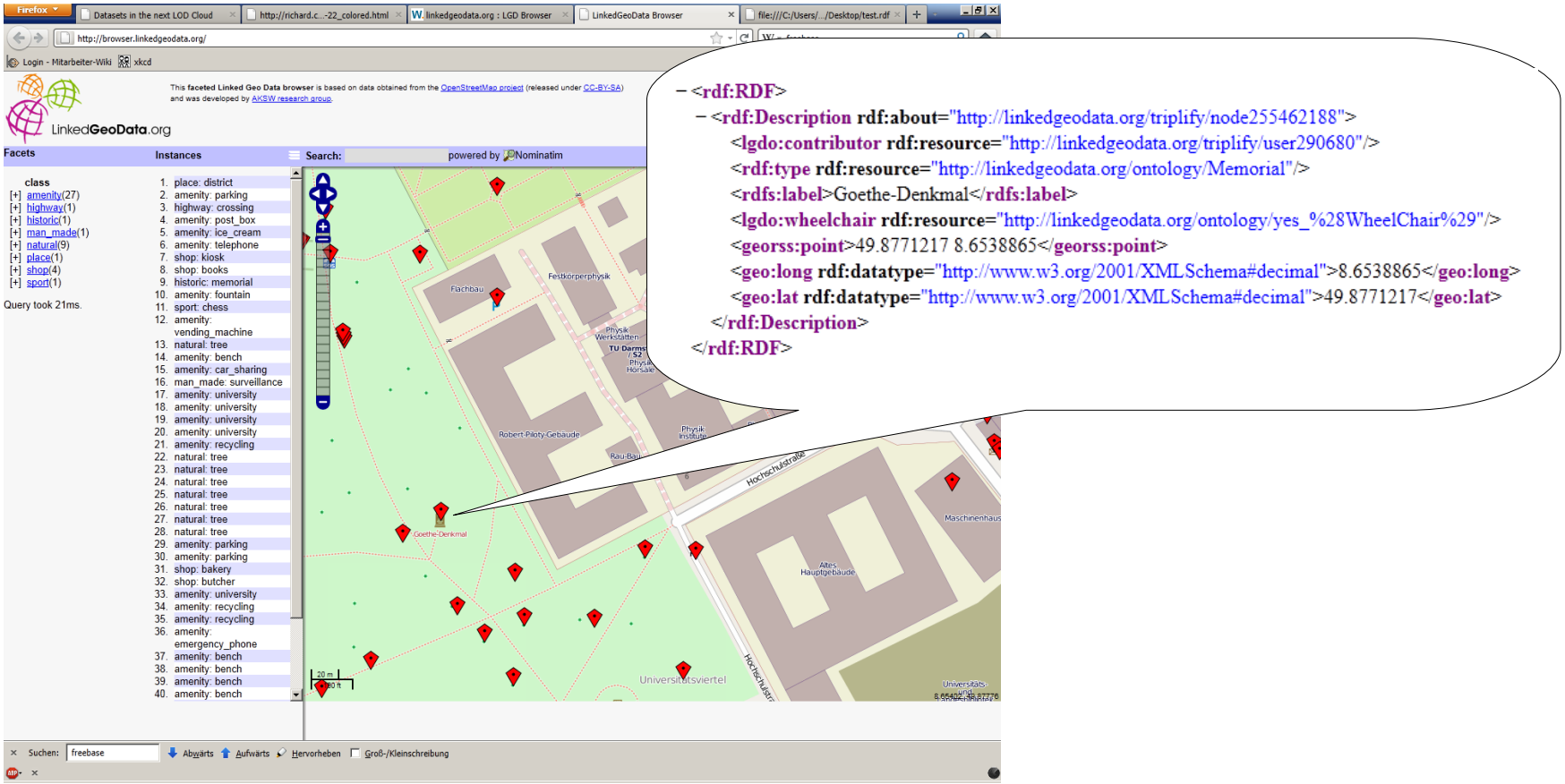
- Codierung von geographischen Daten
- World Geodetic System 1984
- 3D-Referenzmodell
- Namespace `http://www.w3.org/2003/01/geo/wgs84_pos#`
- Präfix oft `geo:`

- Klassen:
  - SpatialThing
  - Point

- Properties:
  - latitude
  - longitude
  - altitude
  - location

WGS 84

- Beispiel: Linked Geo Data



WGS 84

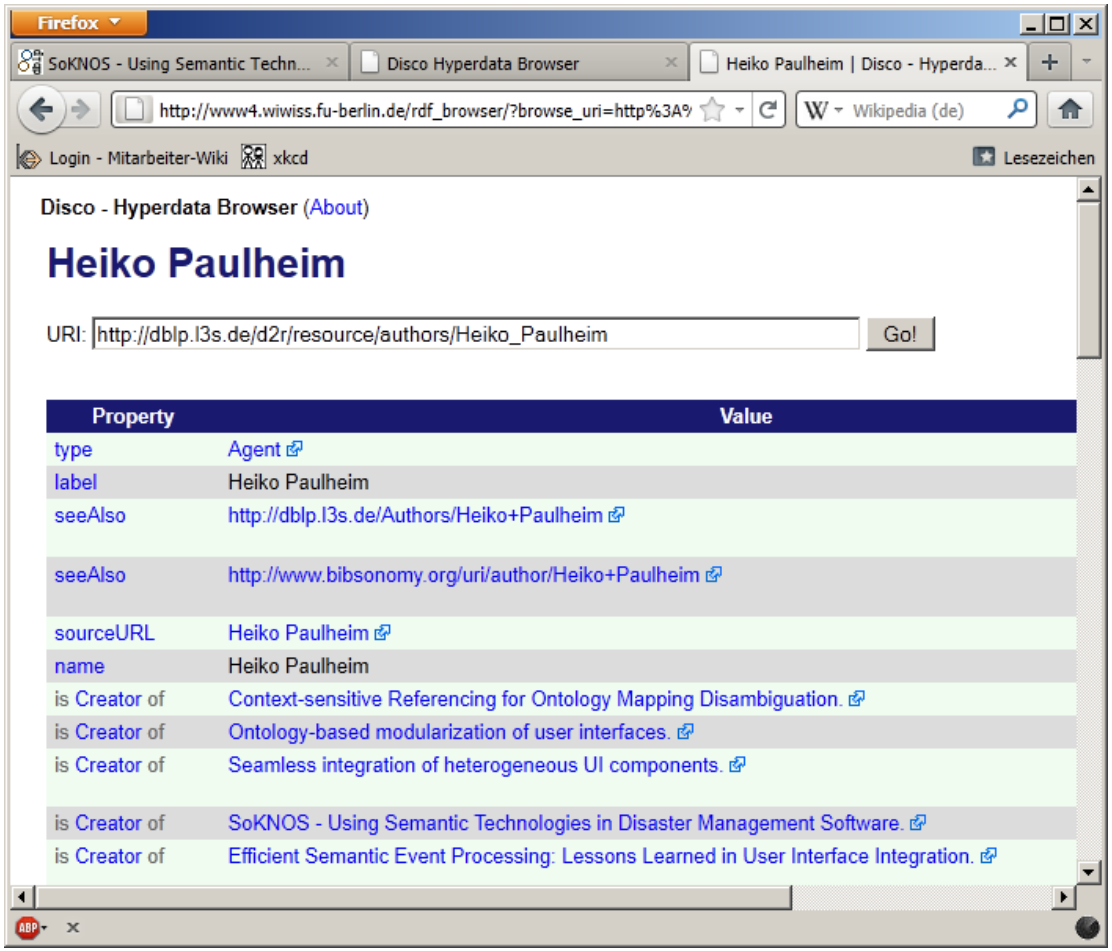
- ermöglicht auch mobile Anwendungen
- auch DBPedia enthält Geo-Daten → DBPedia Mobile

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

Browser für Linked Open Data

TECHNISCHE UNIVERSITÄT DARMSTADT

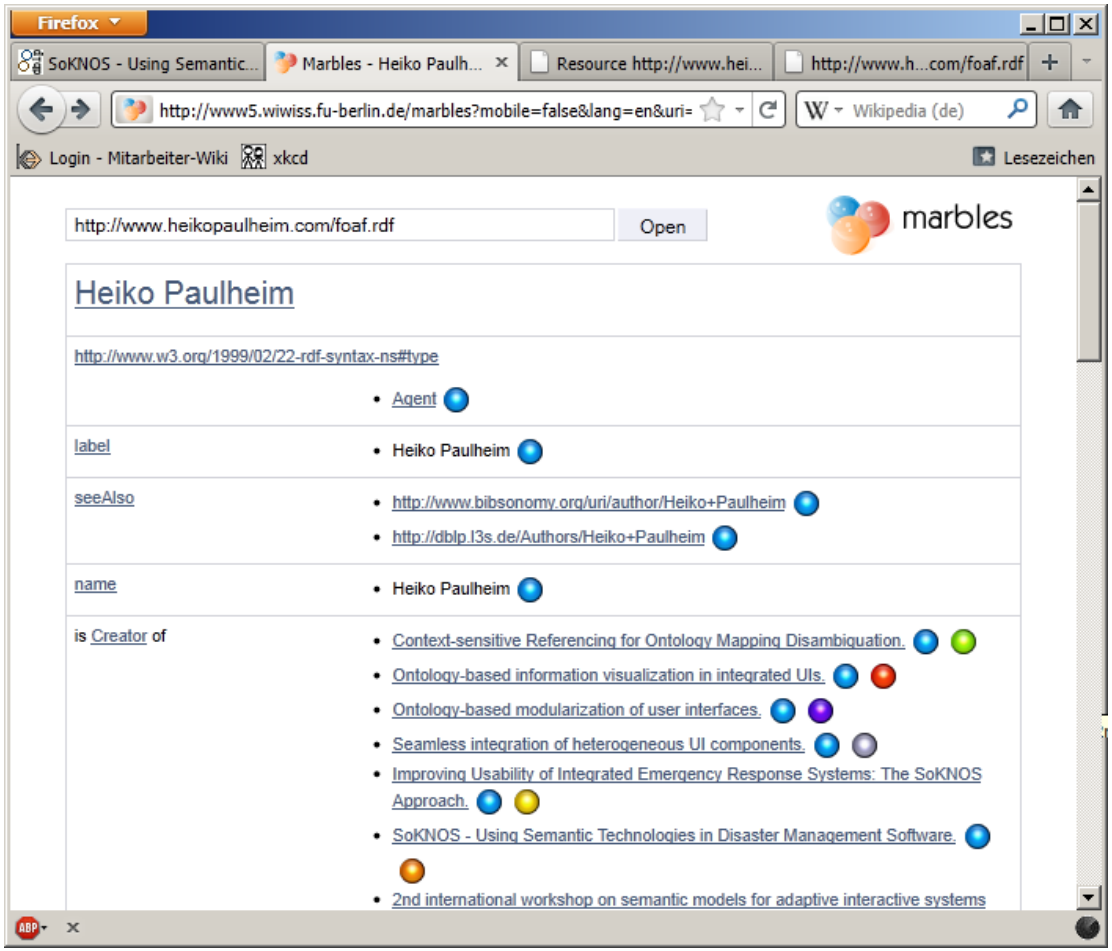
- Disco (FU Berlin)
- tabellarische Ansicht aller Tripel
- über Links kann man browsen



Browser für Linked Open Data

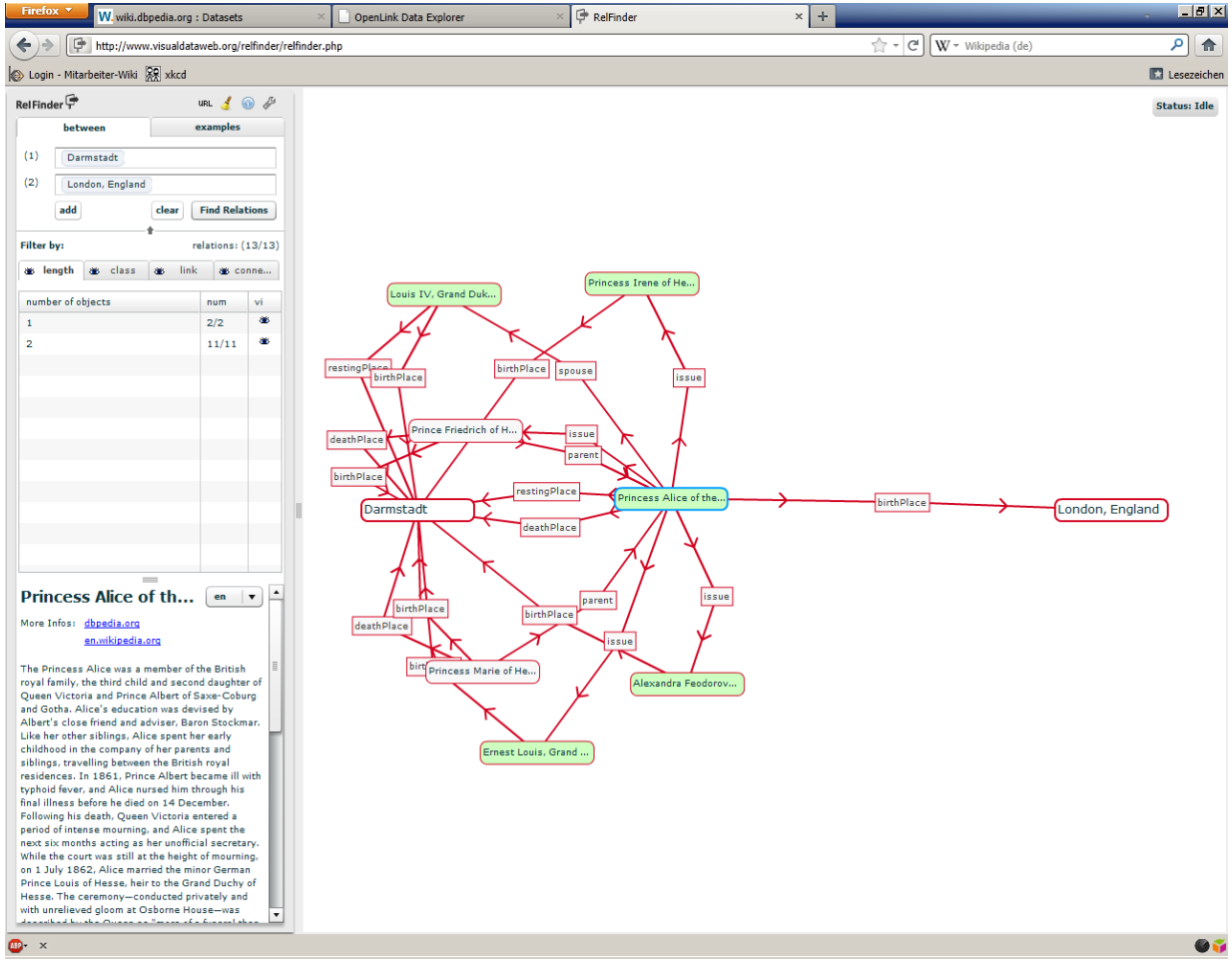
TECHNISCHE UNIVERSITÄT DARMSTADT

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



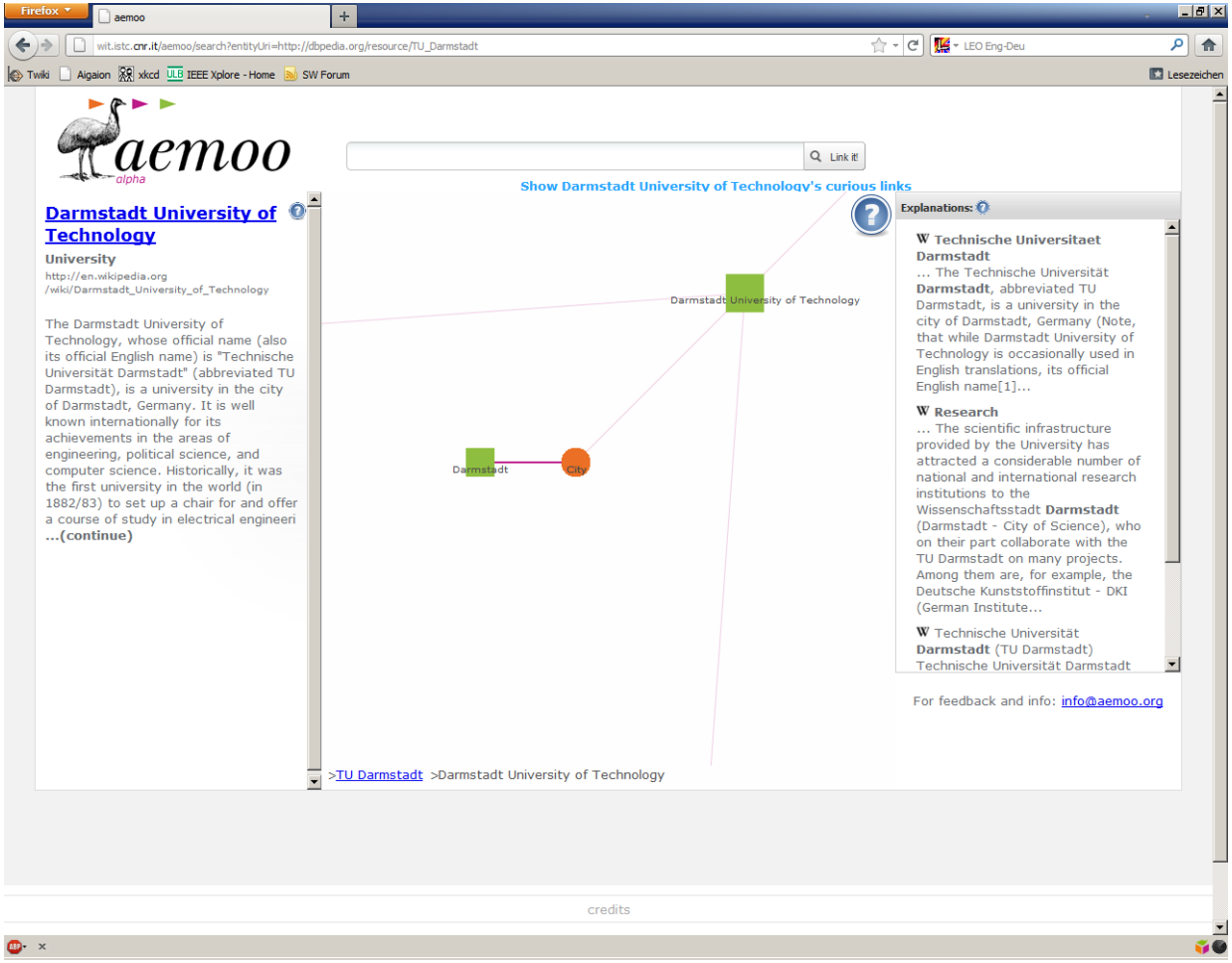
Browser für Linked Open Data

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



Browser für Linked Open Data

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



Erstellen von Linked Open Data

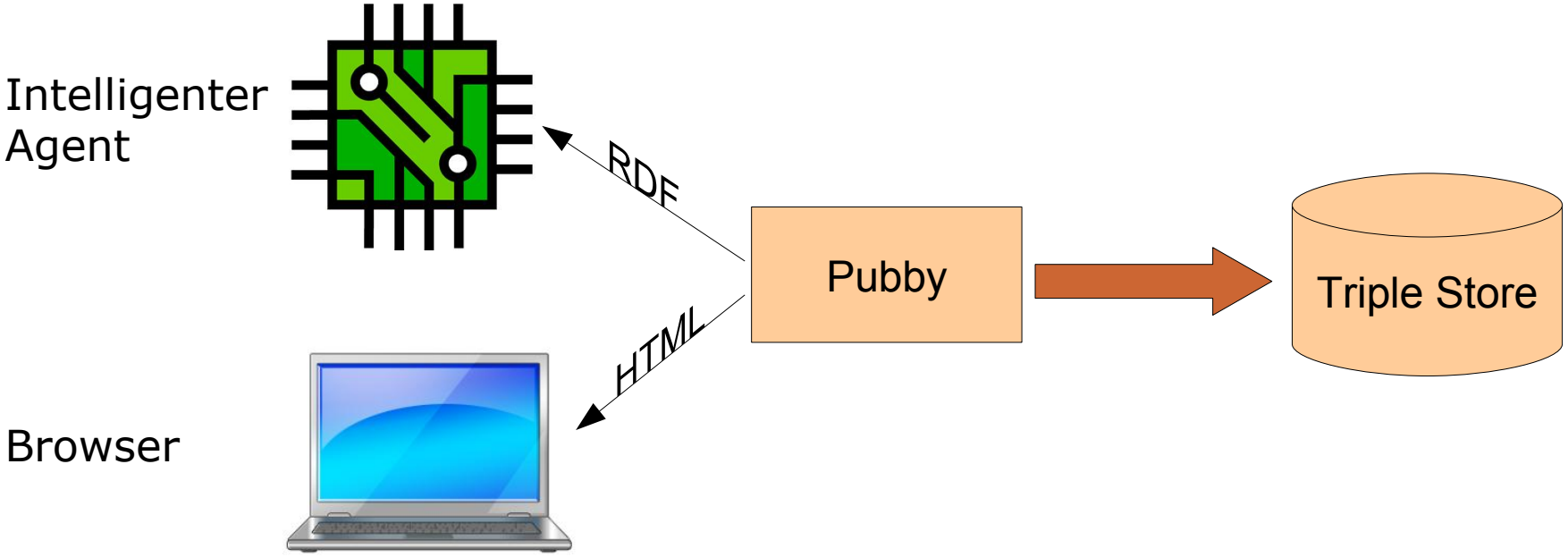
TECHNISCHE UNIVERSITÄT DARMSTADT

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

Linked Open Data aus Triple Stores

TECHNISCHE UNIVERSITÄT DARMSTADT

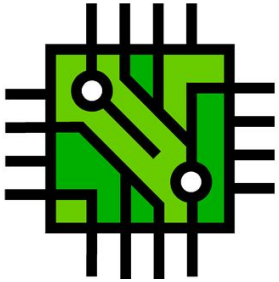
- Pubby ist eine Erweiterung für Triple Stores
- Unterstützt Content Negotiation etc.



Linked Open Data aus Triple Stores

TECHNISCHE UNIVERSITÄT DARMSTADT

Intelligenter Agent



HTTP GET
http://foo.bar/heiko
RDF

Pubby

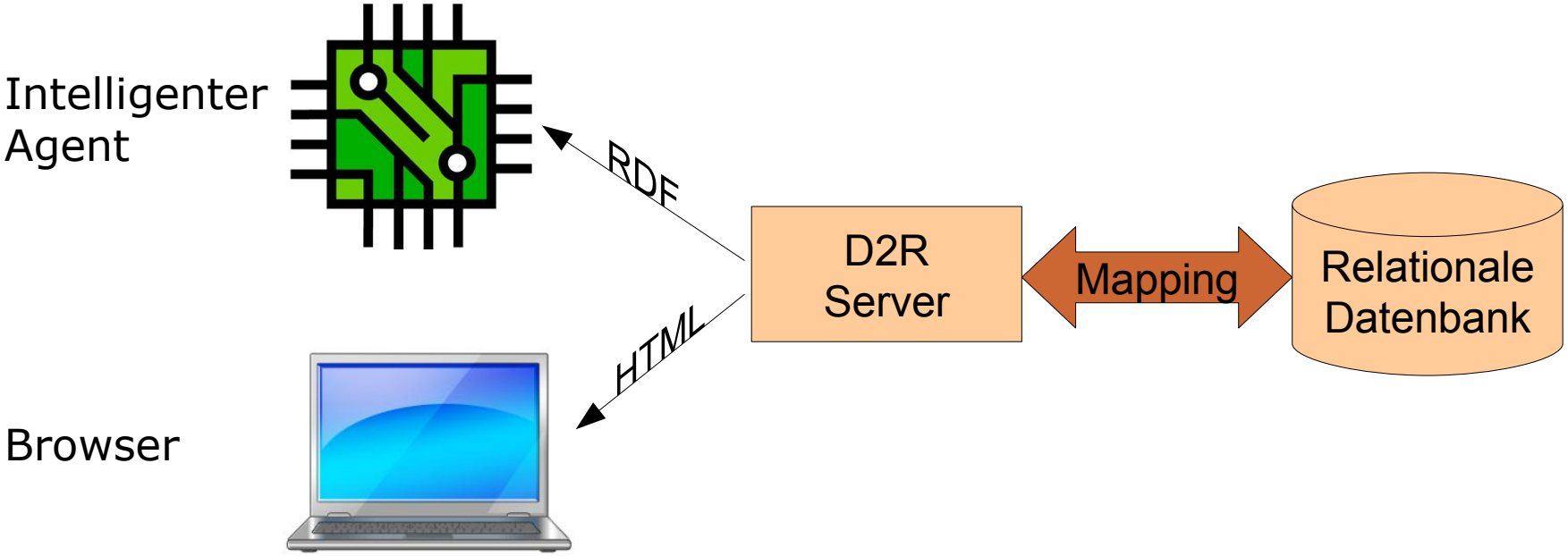
Browser

HTTP GET
http://foo.bar/heiko
HTML

subject	predicate	object
foobar:heiko	foaf:name	"Heiko"
...	...	...

Linked Open Data aus relationalen Datenbanken

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



Linked Open Data aus relationalen Datenbanken

TECHNISCHE UNIVERSITÄT DARMSTADT

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

```
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;
    .
```

```
<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

TECHNISCHE UNIVERSITÄT DARMSTADT

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

Linked Open Data: Anwendungen

TECHNISCHE UNIVERSITÄT DARMSTADT

Firefox

challenge.semanticweb.org | Asthma Clinical Trials, New Treatments a...

trialx.com/match2trials/?keyword=asthma&clocation=NY&#keyword=asthma&clocation=NY&&tb=tx&start=1

Twiki | Aigaion | xkcd | IEEE Xplore - Home | SW Forum | Lesezeichen

TrialX

HOW IT WORKS | PATIENTS | INVESTIGATORS | ENTERPRISES | CURE TALK | LOGIN | JOIN AS PATIENT / INVESTIGATOR

« Start Over

Asthma Clinical Trials (10)

You Searched For...

Disease or Symptom
asthma

Location
NY

Refine Results

Eligible Gender
Male (10)
Female (10)

Trial Phase
Phase 2 (1)
Phase 4 (2)

Study Type
Observational (3)
Interventional (7)

Keep refining »

10 asthma clinical trials found — Showing Results 1 to 10

Study of Asthma and Nasal Steroids
The trial will study the effectiveness of nasal steroids for the treatment of chronic rhinitis and/or sinusitis on **asthma** control.
Phase 4 | Interventional | Research Sites:19 | NA

Allergy Immunotherapy for the Reduction of Asthma
In this clinical study we aim to determine the effect of allergy immunotherapy in decreasing **asthma** and allergy related disease in children who had multiple episodes of wheezing and who ...
Phase NA | Interventional | Research Sites:1 | NA

Evaluation of the Lungs of Individuals With Lung Disease
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 ...
Phase NA | Observational | Research Sites:1 | NA

Disease Modifying Genes in Severe Asthma
Gene and protein bank to study disease modifying and susceptibility genes and proteins in urban patients with **asthma**.
Phase NA | Observational | Research Sites:1 | Accepts Healthy Volunteers

A Community Health Worker Home Intervention to Improve Pediatric Asthma Outcomes
The investigators propose to test the hypothesis that a home-based **asthma** intervention, the Wee Wheezers program, delivered by the Community Health Workers and tailored to the needs of the investigators

Suchen: trial | Abwärts | Aufwärts | Hervorheben | Groß-/Kleinschreibung

Linked Open Data: Anwendungen

TECHNISCHE UNIVERSITÄT DARMSTADT

- 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/Vancouver	106
...	...	...

Linked Open Data: Anwendungen

TECHNISCHE UNIVERSITÄT DARMSTADT

- 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/Vancouver	106
...	...	...

Linked Open Data: Anwendungen

TECHNISCHE UNIVERSITÄT DARMSTADT

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

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

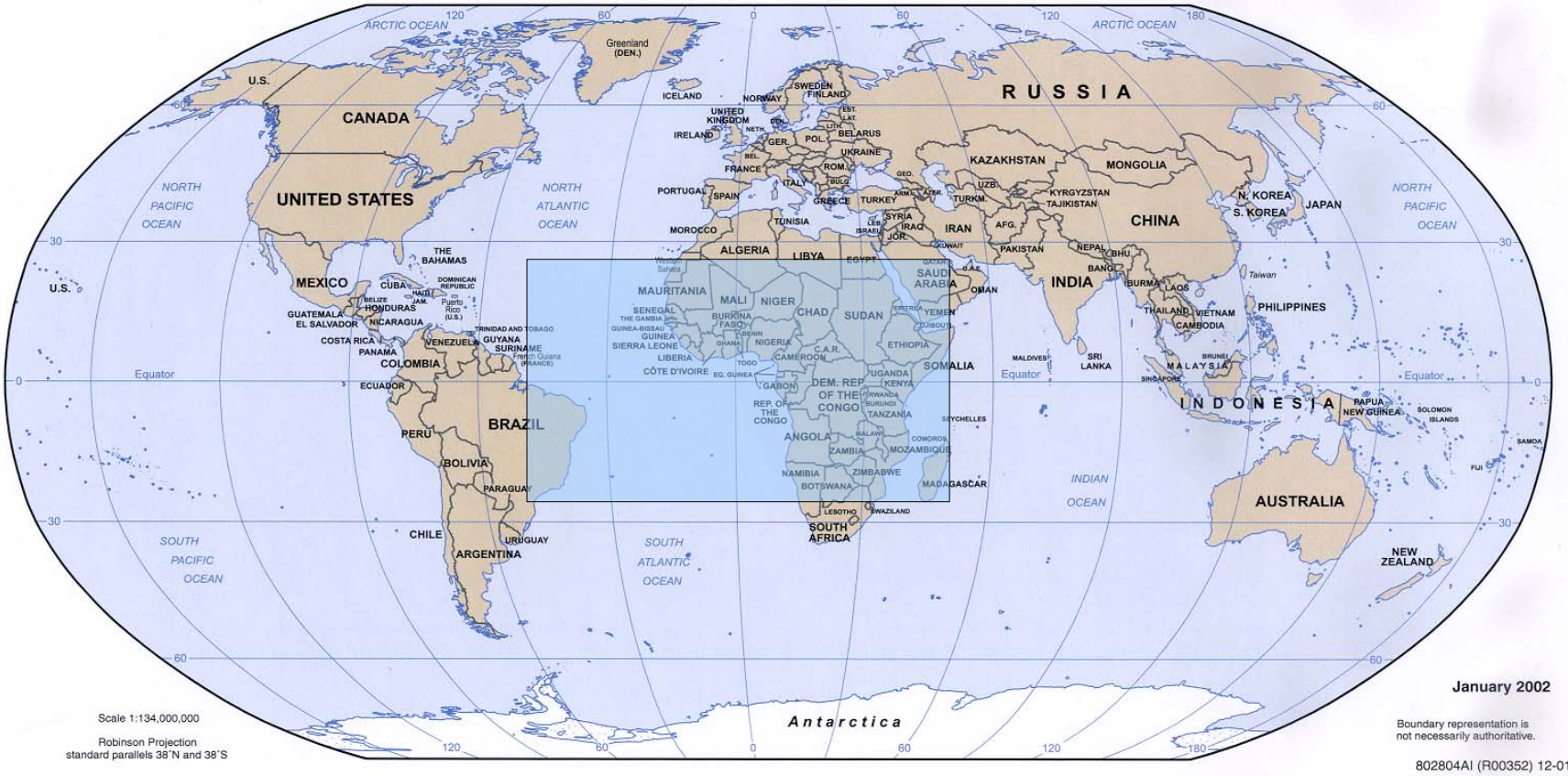
Linked Open Data: Anwendungen

TECHNISCHE UNIVERSITÄT DARMSTADT

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

Linked Open Data: Anwendungen

TECHNISCHE UNIVERSITÄT DARMSTADT



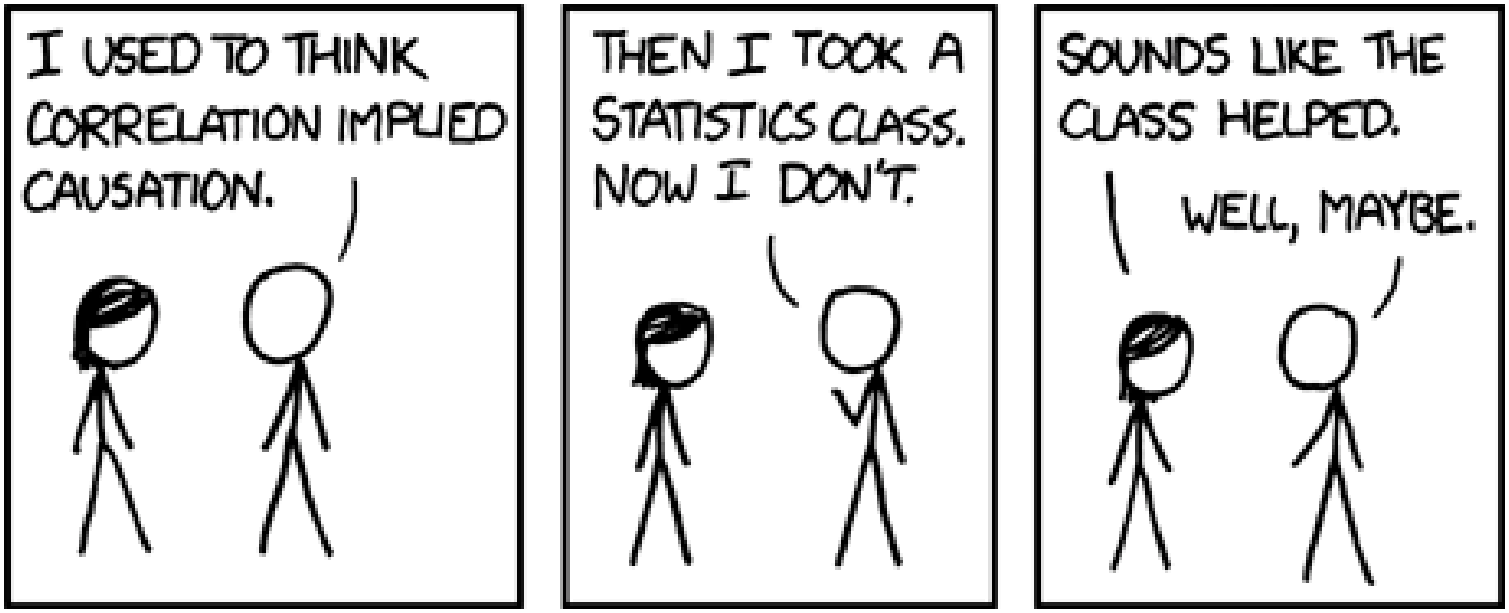
Linked Open Data: Anwendungen

TECHNISCHE UNIVERSITÄT DARMSTADT

- 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 Berger
  - 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

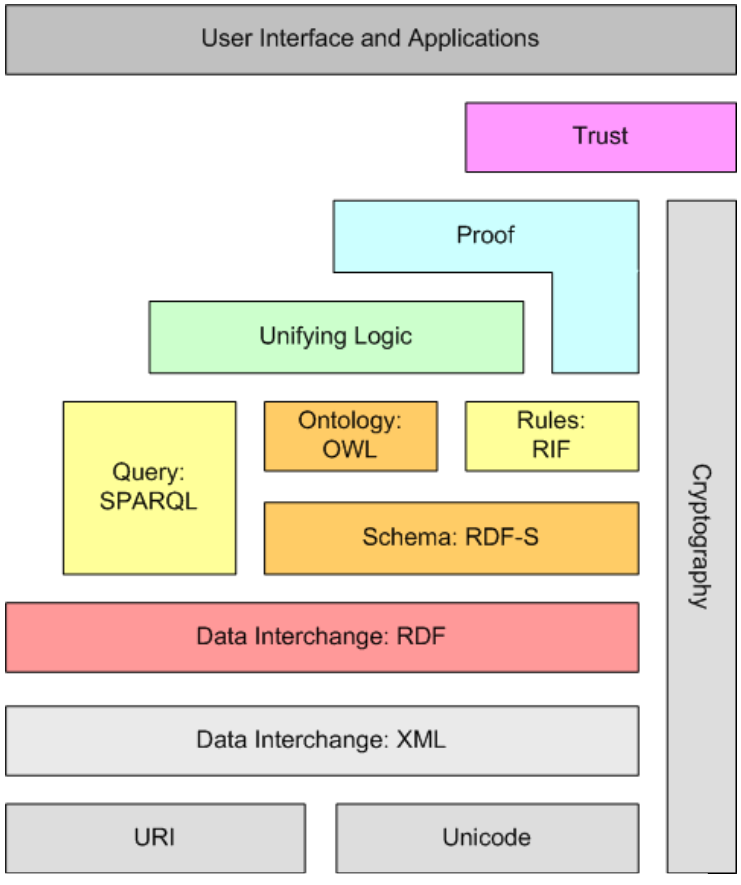
- Bereitstellung von RDF-Daten
  - Verlinkt zu anderen Daten
  - Auflösbare URIs
- Beispiele
- gebräuchliche Vokabulare
- Techniken
- einige Anwendungen

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