

Linked Open Data und strukturiertes Markup im Web



Agenda

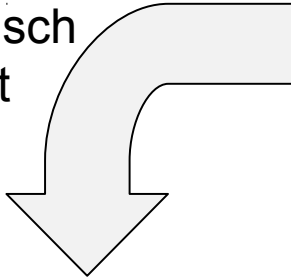
- Motivation
- Linked Open Data
 - Idee
 - Stand der Technik
- Strukturiertes Markup
 - RDFa, Microformats, Microdata
 - schema.org
- Analyse: Evolution und Adoption von schema.org
- Vergleich LOD vs. Microdata+schema.org
- Anwendungen

Disclaimer:

I speak neither for the schema.org nor the LOD community. All views expressed in this talk are my own.

Motivation

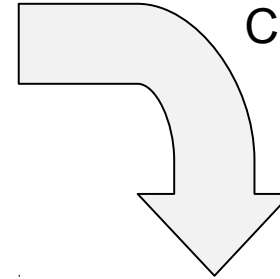
Was ein
Mensch
sieht



```
<html>
...
<b>Dr. Mark Smith</b>
<i>Physician</i>
Main St. 14
Smalltown
Mon-Fri 9-11 am
Wed 3-6 pm
...
</html>
```

Dr. Mark Smith
Physician
Main St. 14
Smalltown
Mon-Fri 9-11 am
Wed 3-6 pm

Was ein
Computer
sieht



Print in bold: „hmf298hmmhuds“
Print in italics: „mj2i9ji0“
Print normal: „fdsah
02hfadsh0um2m0adsmf0ihm
asdfjkköfdsa298ndsfmij32mio
lk2mjpoimjiofdpmsajiomjm“

Suche nach Informationen im Web

In der Regel Schlüsselwortbasiert (z.B. Google):

- „Mark Smith“
- „Physician in Smalltown“
- „Doctor in Smalltown“
- „Physician in Smalltown with opening hours on Wednesday afternoon“
- „Somebody in Smalltown who can fix a broken leg“

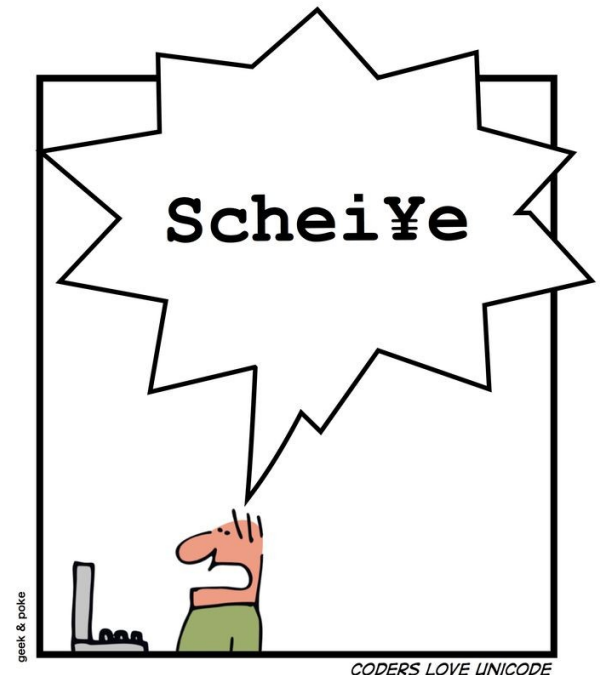
```
<html>
...
<b>Dr. Mark Smith</b>
<i>Physician</i>
Main St. 14
Smalltown
Mon-Fri 9-11 am
Wed 3-6 pm
...
</html>
```

→ je nützlicher die Anfrage, desto schlechter ist sie zu beantworten

→ intelligente Agenten können das Web nur schlecht nutzen

Probleme des “klassischen” Web

- Informationen finden
 - Stichwortbasierte Suche statt echter Fragen
 - Verschiedene Sprachen
 - Synonyme, Hyponyme, Polyseme, ...
 - Uneindeutigkeit natürlicher Sprache
- Informationen verarbeiten
 - Formate und Zeichenkodierungen
- Informationen nutzen
 - Verteilt über verschiedene Seiten
 - z.B.: Autor eines Buches auf Verlagsseite, Adresse des Autors auf seiner privaten Seite



<http://geekandpoke.typepad.com/geekandpoke/2011/08/coders-love-unicode.html>

Klassisches Web vs. Semantic Web

- Artikel von Tim Berners-Lee, Jim Hendler, and Ora Lassila (2001):

„The Web is the killer app of the Internet.
The Semantic Web is another killer app
of that magnitude.“



Berners-Lee et al. (2001): *The Semantic Web*. In: Scientific American, Mai 2001.

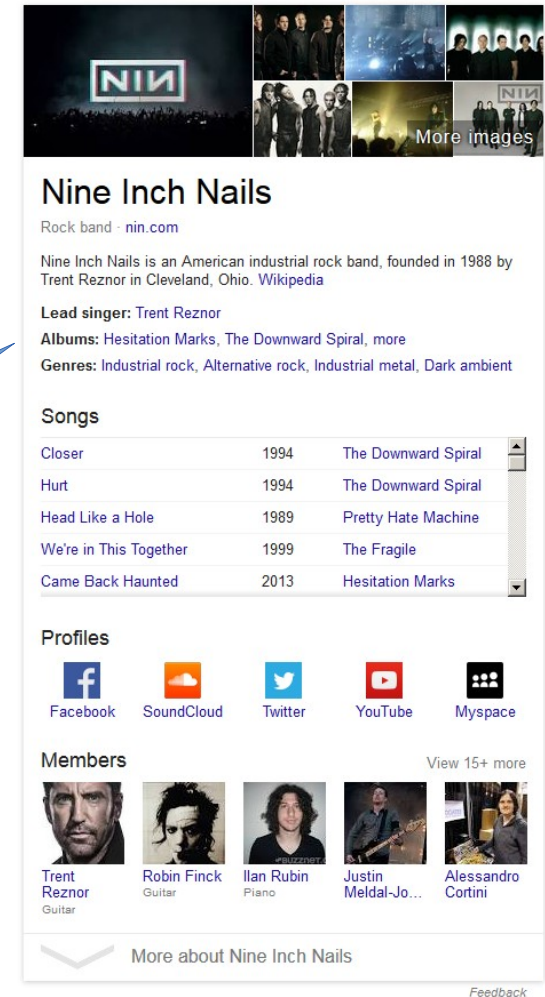
Semantic Web – Implementierungen

- Grundidee
 - Daten im Web für Maschinen verständlich machen
 - Semantisch ausgezeichnet mithilfe interpretierbaren Schemas
- Implementierungen z.B.
 - Google Knowledge Graph
 - Linked Open Data Cloud
 - RDFa
 - Microformats
 - Microdata+schema.org

Google Knowledge Graph

- “Things, not strings” (Google, 2012)
- Strukturierte Daten für Entitäten
- Uniforme Darstellung für Entitätstypen
- Entitäten sind über typisierte Links verbunden

Links haben eine Semantik!



Nine Inch Nails
Rock band · nin.com

Nine Inch Nails is an American industrial rock band, founded in 1988 by Trent Reznor in Cleveland, Ohio. [Wikipedia](#)

Lead singer: Trent Reznor
Albums: [Hesitation Marks](#), [The Downward Spiral](#), [more](#)
Genres: [Industrial rock](#), [Alternative rock](#), [Industrial metal](#), [Dark ambient](#)






Songs

Closer	1994	The Downward Spiral
Hurt	1994	The Downward Spiral
Head Like a Hole	1989	Pretty Hate Machine
We're in This Together	1999	The Fragile
Came Back Haunted	2013	Hesitation Marks

Profiles

[Facebook](#) [SoundCloud](#) [Twitter](#) [YouTube](#) [Myspace](#)

Members [View 15+ more](#)

 Trent Reznor Guitar	 Robin Finck Guitar	 Ilan Rubin Piano	 Justin Meldal-Jo...	 Alessandro Cortini
--	---	---	---	--

[More about Nine Inch Nails](#)

Feedback

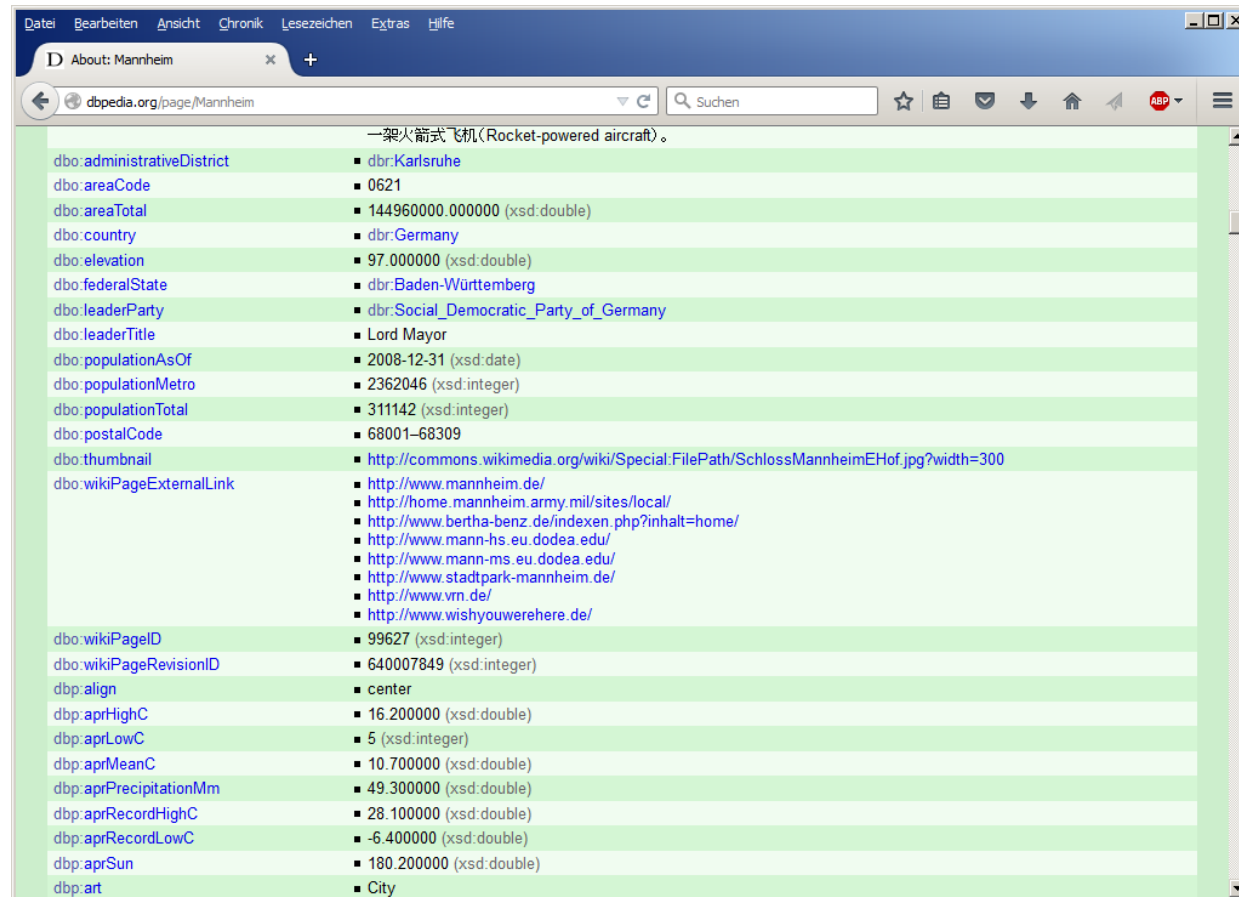
Linked Open Data

- Vier Grundprinzipien (Sir Tim Berners-Lee, 2006)
 - 1) URIs verwenden, um Dinge zu identifizieren
 - 2) URIs sollen dereferenzierbar sein
 - 3) Hinter dereferenzierbaren URIs nützliche Informationen hinterlegen; Standards nutzen
 - 4) Verweise zu anderen Datensets setzen



Linked Open Data – Beispiel

- <http://dbpedia.org/resource/Mannheim>



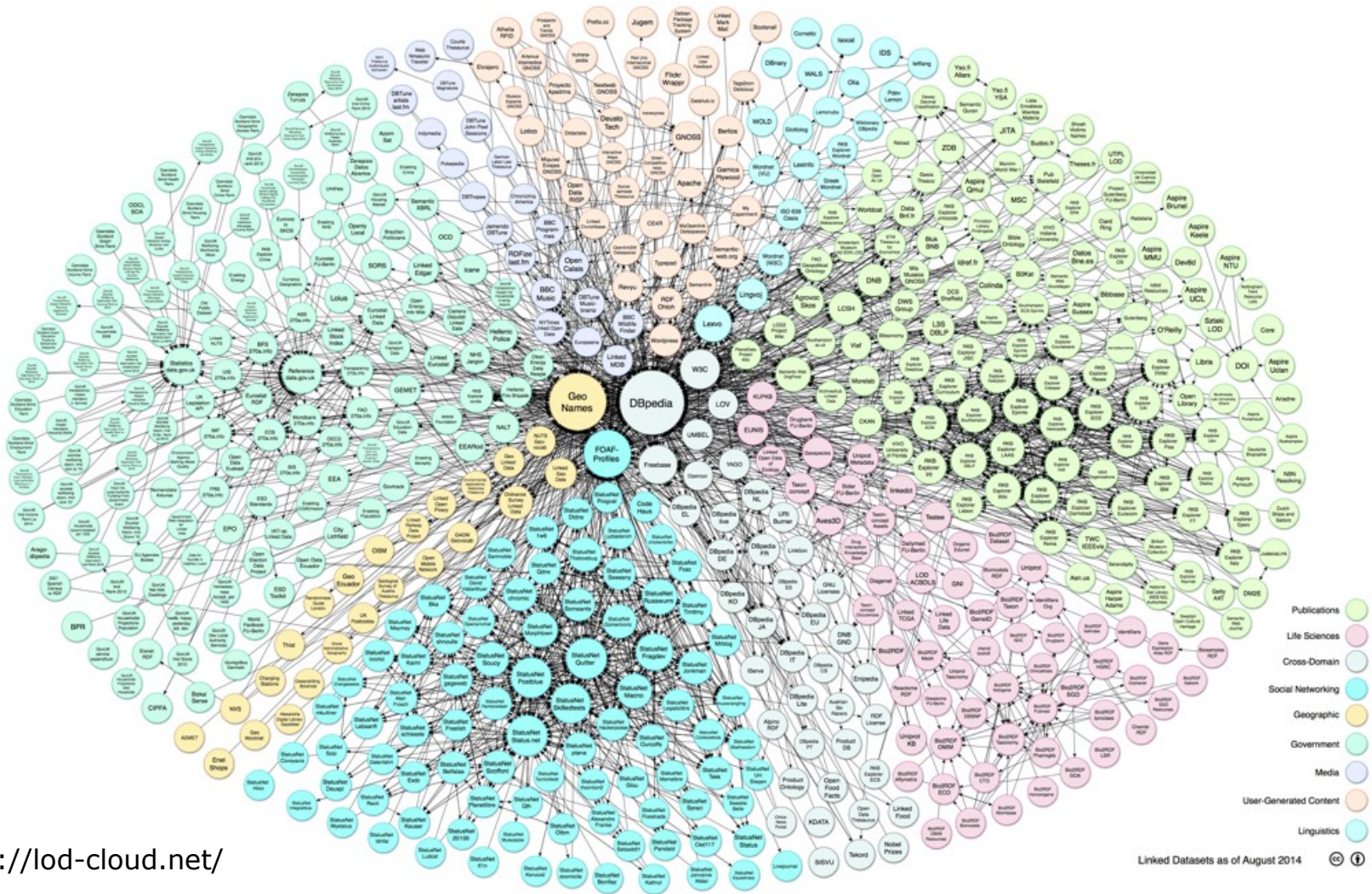
一架火箭式飞机(Rocket-powered aircraft)。	
dbo:administrativeDistrict	■ dbr:Karlsruhe
dbo:areaCode	■ 0621
dbo:areaTotal	■ 144960000.000000 (xsd:double)
dbo:country	■ dbr:Germany
dbo:elevation	■ 97.000000 (xsd:double)
dbo:federalState	■ dbr:Baden-Württemberg
dbo:leaderParty	■ dbr:Social_Democratic_Party_of_Germany
dbo:leaderTitle	■ Lord Mayor
dbo:populationAsOf	■ 2008-12-31 (xsd:date)
dbo:populationMetro	■ 2362046 (xsd:integer)
dbo:populationTotal	■ 311142 (xsd:integer)
dbo:postalCode	■ 68001–68309
dbo:thumbnail	■ http://commons.wikimedia.org/wiki/Special:FilePath/SchlossMannheimEHof.jpg?width=300
dbo:wikiPageExternalLink	■ http://www.mannheim.de/ ■ http://home.mannheim.army.mil/sites/local/ ■ http://www.bertha-benz.de/indexen.php?inhalt=home/ ■ http://www.mann-hs.eu.dodea.edu/ ■ http://www.mann-ms.eu.dodea.edu/ ■ http://www.stadtpark-mannheim.de/ ■ http://www.vn.de/ ■ http://www.wishyouwerehere.de/
dbo:wikiPageID	■ 99627 (xsd:integer)
dbo:wikiPageRevisionID	■ 640007849 (xsd:integer)
dbp:align	■ center
dbp:aprHighC	■ 16.200000 (xsd:double)
dbp:aprLowC	■ 5 (xsd:integer)
dbp:aprMeanC	■ 10.700000 (xsd:double)
dbp:aprPrecipitationMm	■ 49.300000 (xsd:double)
dbp:aprRecordHighC	■ 28.100000 (xsd:double)
dbp:aprRecordLowC	■ -6.400000 (xsd:double)
dbp:aprSun	■ 180.200000 (xsd:double)
dbp:art	■ City

Linked Open Data

- Fünf-Sterne-Schema (Sir Tim Berners-Lee, 2010)
 - * Im Netz verfügbar mit offener Lizenz
 - ** Maschinenlesbare, strukturierte Daten
 - *** wie **, nutzt ein nicht-proprietäres Format
 - **** wie ***, nutzt einen Standard des W3C
 - ***** wie ****, enthält zusätzlich Verweise zu anderen Datensets

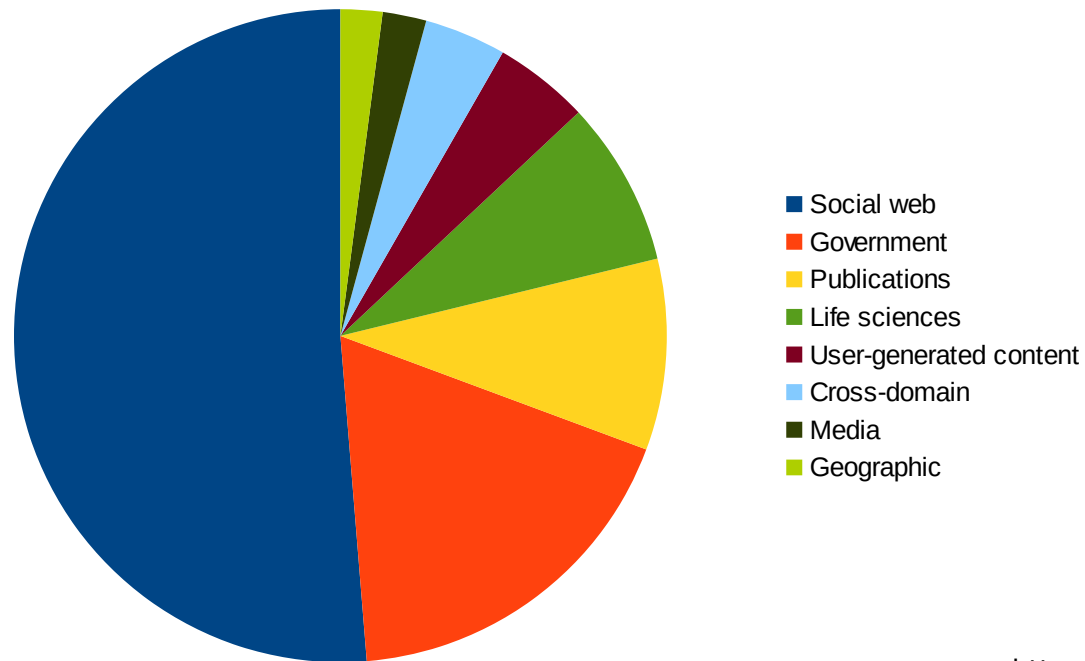


Die Linked Open Data Cloud



Die Linked Open Data Cloud

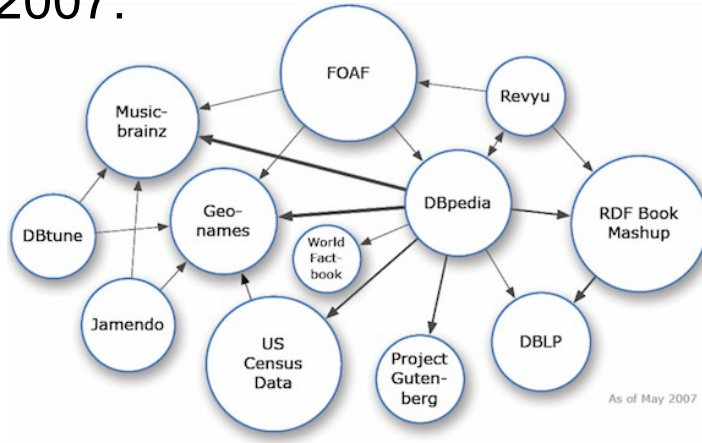
- In Zahlen:
 - 570 Datensets
 - Mehrere Milliarden Fakten
 - Mehrere Millionen Verknüpfungen zwischen Datensets



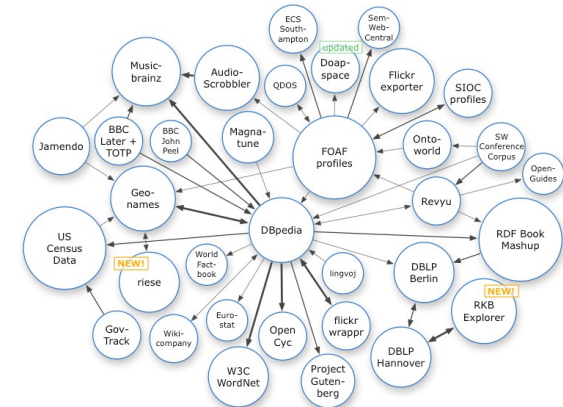
<http://lod-cloud.net/>

Wachstum der Linked Open Data Cloud

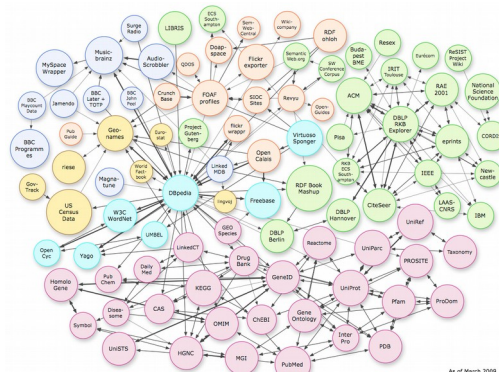
- Mai 2007:



- März 2008:

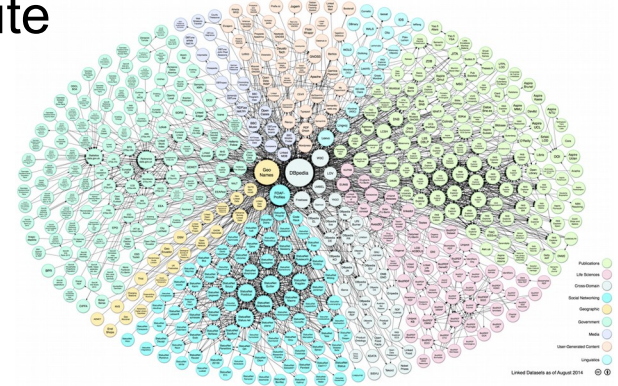


- März 2009



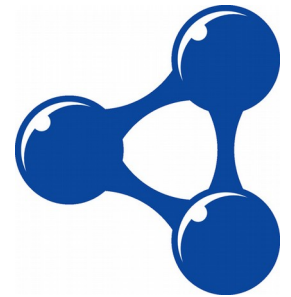
Linking Open Data cloud diagram,
by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

- heute



Technisches Fundament: RDF

- „Resource Description Framework“
- W3C-Standard seit 2004
- Beschreibung von beliebigen Dingen
- Sicht 1: Sätze in der Form <Subjekt, Prädikat, Objekt>
„Heiko arbeitet bei der Universität Mannheim.“
- Sicht 2: Gerichteter Graph mit benannten Kanten



Relevante Vokabulare

- FOAF
 - Verwendet von >69% aller LOD datasets
 - Verwendet von >69% aller LOD datasets getaggt mit “publication”
- Dublin Core
 - Verwendet von >56% aller LOD datasets
 - Verwendet von >81% aller LOD datasets getaggt mit “publication”
- Bibo (Bibliographic Ontology)
 - Verwendet von >6% aller LOD datasets
 - Verwendet von >41% aller LOD datasets getaggt mit “publication”

RDF in Attributes (RDFa)

- Problem mit Linked Open Data
 - Daten werden meist nach LOD konvertiert
 - Create once, then forget...
 - Dieselben Daten liegen mehrfach vor
 - Für Menschen in HTML
 - Für Maschinen als LOD
- Doppelte Buchführung
- Potentielle Quelle für Inkonsistenzen

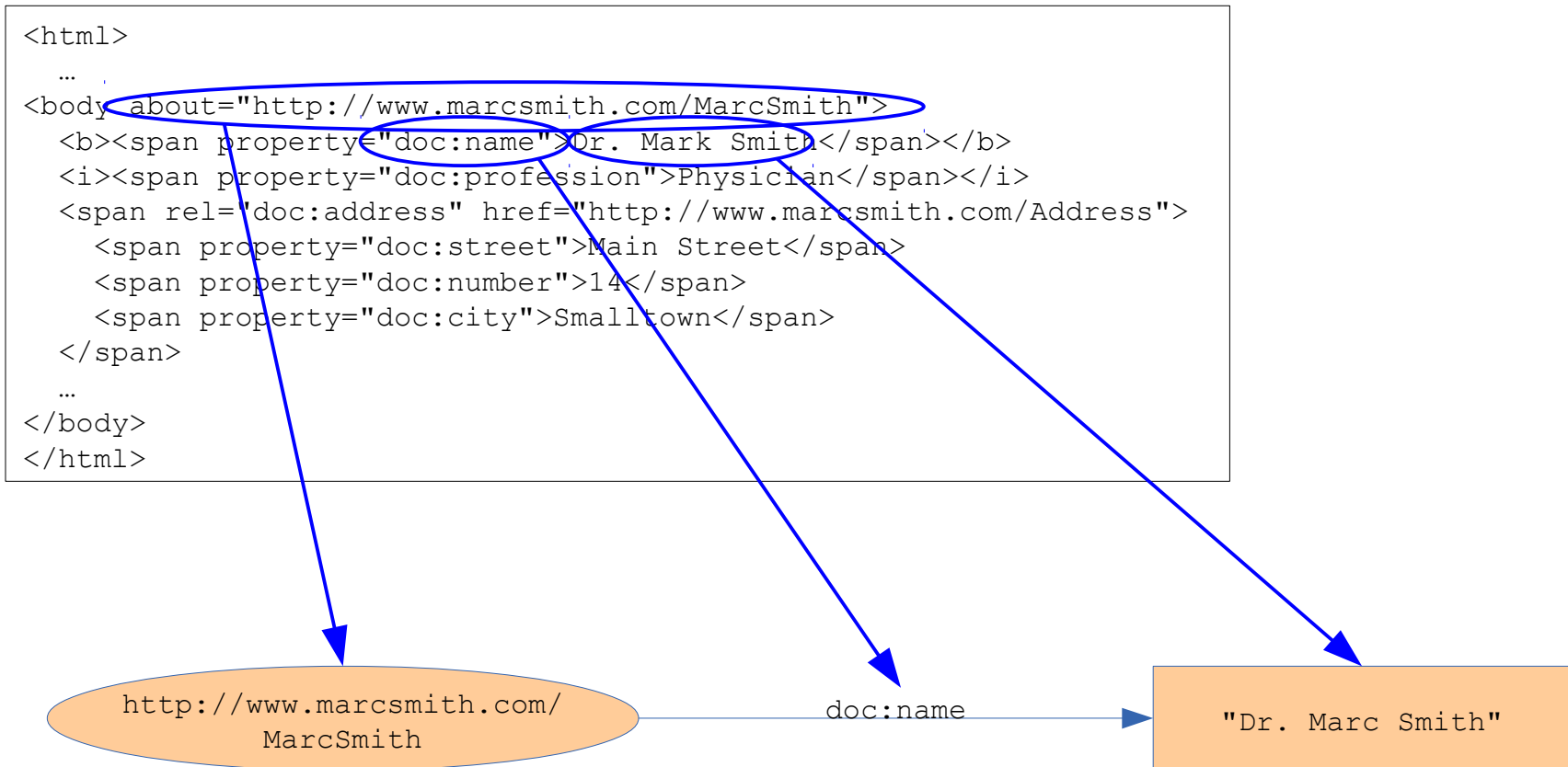
RDF in Attributes (RDFa)

- Idea von RDFa
 - HTML und RDF in einem Dokument codieren
 - Verhindert doppelte Buchführung
- RDFa kombiniert XHTML mit RDF



```
<html>
...
<body about="http://www.marcsmith.com/MarcSmith">
  <b><span property="doc:name">Dr. Mark Smith</span></b>
  <i><span property="doc:profession">Physician</span></i>
  <span rel="doc:address" href="http://www.marcsmith.com/Address">
    <span property="doc:street">Main Street</span>
    <span property="doc:number">14</span>
    <span property="doc:city">Smalltown</span>
  </span>
  ...
</body>
</html>
```

RDF in Attributes (RDFa)



Microformats

- RDFa erlaubt beliebige Vokabulare
- Microformats sind eine Menge Standard-Vokabulare für häufig im Web dargestellte Daten, z.B.
 - Personen (h-card)
 - Events (h-calendar)
 - Reviews (h-reviews)
 - Rezepte (h-recipe)
 - ...
- Bibliographische Metadaten werden kaum unterstützt

Microdata

- Fügt strukturierte Annotationen in HTML hinzu
 - Einzelne Bereiche werden markiert und annotiert
 - Beliebige Vokabulare sind möglich
- Ähnlich zu RDFa



```
<div itemscope
  itemtype="http://schema.org/PostalAddress">
  <span itemprop="name">Data and Web Science Group</span>
  <span itemprop="addressLocality">Mannheim</span>,
  <span itemprop="postalCode">68131</span>
  <span itemprop="addressCountry">Germany</span>
</div>
```

Microdata

- Markup kann nach RDF extrahiert werden
 - Siehe W3C Interest Group Note: Microdata to RDF [1]

HTML



```
<div itemscope  
  itemtype="http://schema.org/PostalAddress">  
  <span itemprop="name">Data and Web Science Group</span>  
</div>
```

```
_:1 a <http://schema.org/PostalAddress> .  
_:1 <http://schema.org/name> "Data and Web Science Group" .  
_:1 <http://schema.org/addressLocality> "Mannheim" .  
_:1 <http://schema.org/postalCode> "68131" .  
_:1 <http://schema.org/adressCounty> "Germany" .
```

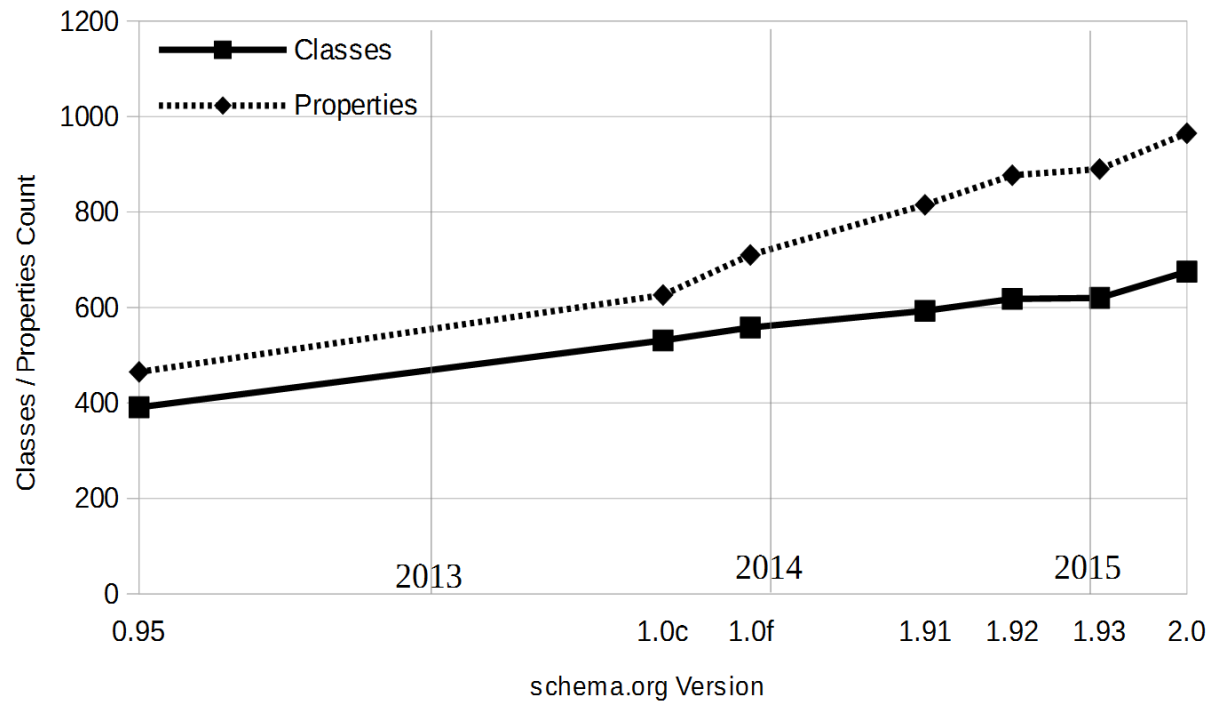
[1] <http://www.w3.org/TR/microdata-rdf/>

schema.org

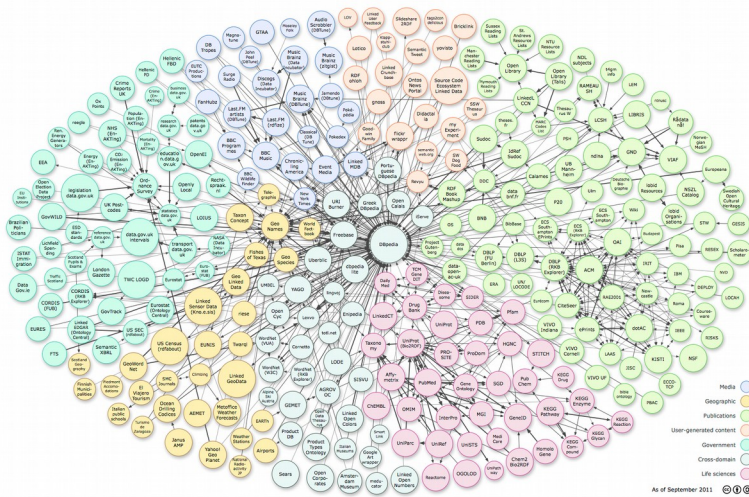
- Microdata kann mit *beliebigen* Schemata benutzt werden
 - Praktisch wird nur schema.org auf breiter Front genutzt
- schema.org
 - Vokabular für Inhalte im Web
 - Vorangetrieben von großen Suchmaschinenbetreibern:
 - Google, Bing, Yahoo! und Yandex
 - Kann auch für LOD und RDFa genutzt werden
 - praktisch relevant ist nur die Kombination Microdata+schema.org

schema.org

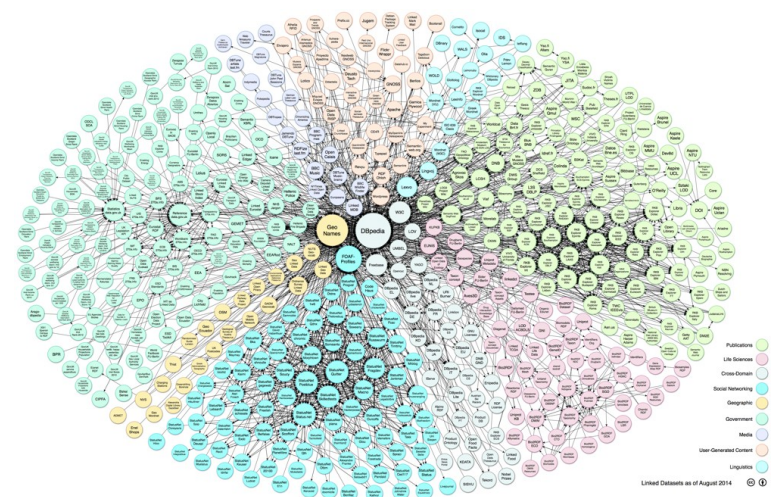
- schema.org (Release 2.0)
 - 675 classes
 - 965 properties



Wachstum LOD, RDFa, Microdata

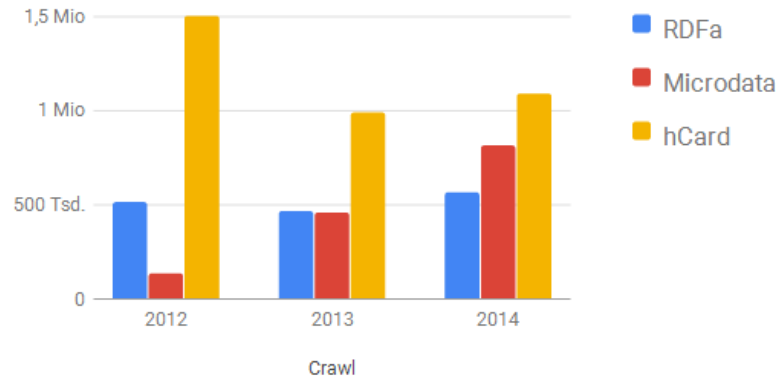


LOD Cloud 2012

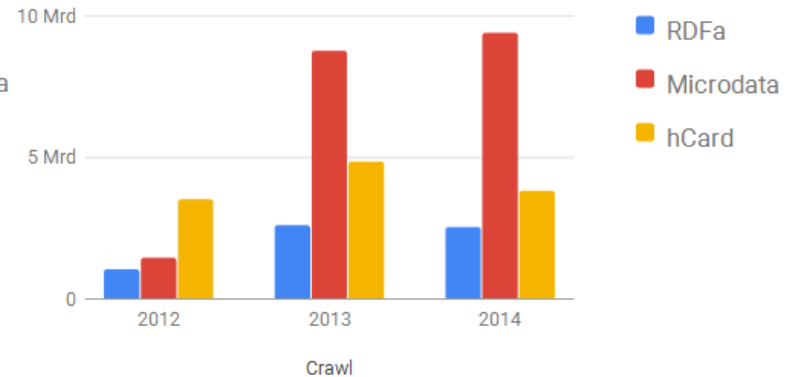


LOD Cloud 2014

Number of PLDs Deploying the three Major Markup Formats



Number of Triples marked up by the three Major Markup Formats



schema.org in a Nutshell

- schema.org hat einige bekannte Vokabulare aufgenommen, u.a.
 - Good Relations (2012)
 - W3C BibExtend (2014)
 - MusicBrainz vocabulary (2015)
 - Automotive Ontology (2015)



THE AUTOMOTIVE
ONTOLOGY
WORKING GROUP



Relevante Konzepte in schema.org

- Klassen
 - Book
 - Article
- Werden mit >80 Attributen/Relationen beschrieben, z.B.
 - author
 - publisher
 - citation

Quantitative Analysen: schema.org Microdata

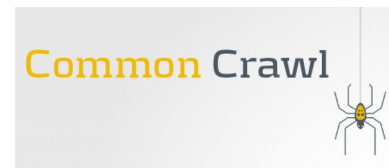
- Das Web Data Commons Projekt (Universität Mannheim)
 - Extrahiert verschiedene Datensets aus dem Common Crawl
 - RDFa, Microdata, Microformats
 - Graphstruktur des Web
 - Relationale Tabellen
- <http://webdatacommons.org/>

Korpus für quantitative
Analysen



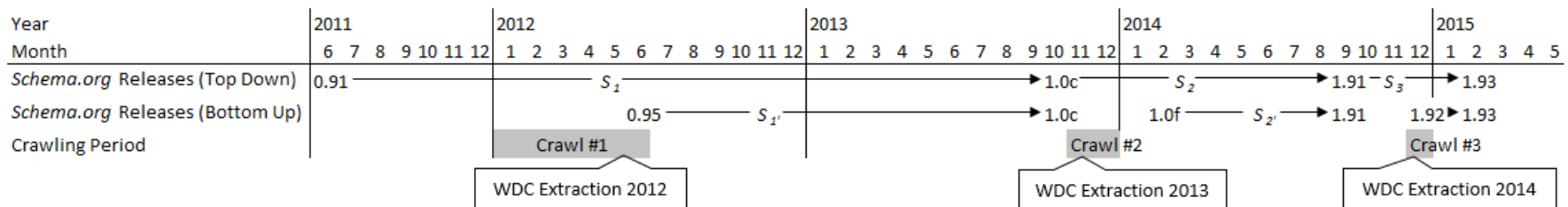
Evolution von schema.org

- Einzigartiges Datenmaterial
 - Mehrere hunderttausend Nutzer
 - Mehrere schema.org-Releases pro Jahr
 - Mehrere Crawls pro Jahr
- Diachronische Analyse möglich
 - Globale Konvergenz oder Divergenz
 - Top-Down-Prozesse: Geschwindigkeit, in der neue Konzepte verwendet werden
 - Bottom-Up-Prozesse: Nutzer-/Nutzengetriebene Veränderungen des Standards



Evolution von schema.org

- Datengrundlage der Studie
- Drei Releases des WebDataCommons Microdata Corpus
 - 2012, 2013 und 2014
- Versionen von schema.org, die jeweils gültig waren
 - Zu Beginn des Crawls
 - Am Ende des Crawls

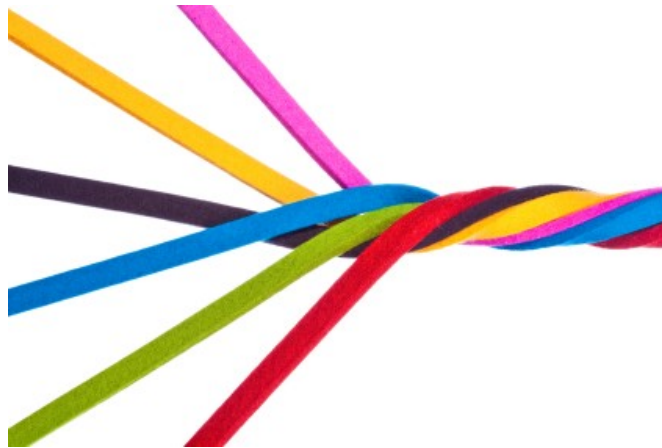


Globale Konvergenz/Divergenz

- Microdata-Dokumente als Menge von Pfaden
- Messung der Entropie als

$$H = \sum_{i=1}^n -p(path_i) \log(p(path_i))$$

- Geringe Entropie = hohe Homogenität
- Normalisiert nach maximaler Entropie und Anzahl der Pfade



Globale Konvergenz/Divergenz

- Beobachtung
 - Globale Entropie fällt → höhere Homogenität

- Klassen mit hoher Konvergenz

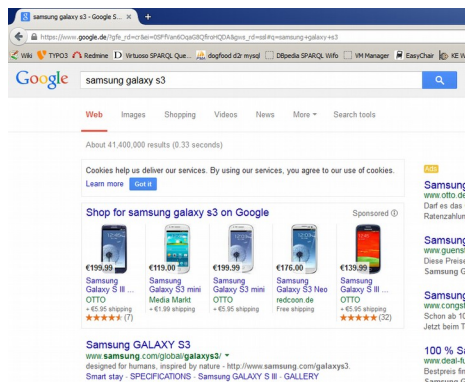
- WebSite, Blog, ...
- Hotel, Restaurant, ...
- Product, Offer, ...
- Rating, Review

schema.org in CMS

Yellow pages

Google Rich Snippets

...alles davon



Top-Down-Prozesse

- Wie schnell werden Änderungen angenommen?
 - Neue Klassen/Properties
 - Deprecations
 - Änderung von domain und range
- Herausforderungen bei der Messung
 - Verschiedene Crawls
 - Globales Wachstum von schema.org Microdata
- Maßzahl: *normalized usage increase (nui)* von i nach j :
 - $nui(s) > 1.05$: signifikante Zunahme der Nutzung
 - $nui(s) < 0.95$: signifikante Abnahme der Nutzung



$$nui_{ij}(s) := \frac{\#PLD_i(s)}{\#PLD_j(s) + 1} / \frac{\#PLD_i}{\#PLD_j} \quad (i > j)$$

Top-Down-Prozesse

- Verwendung neuer Klassen und Properties
 - Fast die Hälfte aller neuen Klassen wird *nie* benutzt!
 - Ähnlich für neue Properties
- Gründe
 - En-bloc-Übernahme ganzer Vokabularer
 - Nicht jedes Konzept wird gleichermaßen verwendet
 - z.B. medizinisches Vokabular
 - Blinder Fleck des Ansatzes
 - Manche Konzepte werden hauptsächlich für *e-mail*-Markup verwendet
 - z.B. Actions



Top-Down-Prozesse

- Wo die Übernahme funktioniert

- Metadaten für Webcontent
(schema.org/Website has the highest NUI)
- Broadcasting (z.B. TV Episodes)
- Questions & Answers
- Anschriften

schema.org in CMS

Kollaboration
mit BBC und EBU

Q&A Pages, z.B.
Stackoverflow

Yellow Pages
Suchmaschinen

Top-Down-Prozesse

- Implementieren von domain/range-Änderungen
 - Eher langsam
- Ausnahmen
 - Produkte (height, width, itemCondition, ...)
 - Broadcasting (episode, season, actor, ...)

Suchmaschinen
Business Incentive

Collaboration
with BBC and EBU



Samsung GALAXY S6 Android Phone 32 GB - WCDMA (UMTS) / GS...

699,00 € bei über 25 Händlern

★★★★★ 1 Produktbewertung Nr. 5 in "Mobiltelefone › Samsung"

Samsung · Smartphone · 4G LTE · Android · 13 cm display · GSM · UMTS · 16 MP Kamera · Quadband

Inspiriert durch die Arbeit von Glasbläsern und Kunstschmieden repräsentiert das **Samsung Galaxy S6** eine edle Symbiose aus Glas und Metall. Mit ...



Samsung GALAXY S6 Android Phone 32 GB - WCDMA (UMTS) / GS...

387,00 € bei über 50 Händlern

Nr. 3 in "Mobiltelefone"

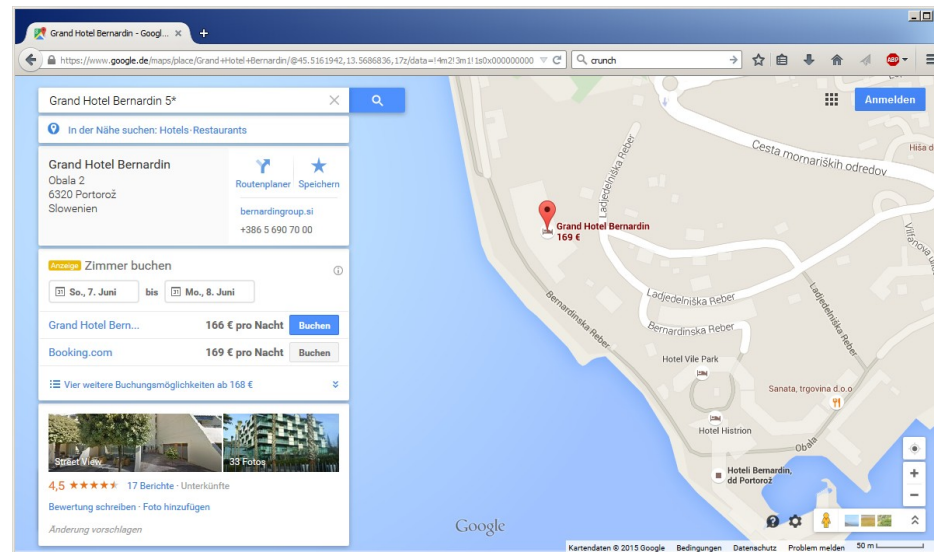
Samsung · Smartphone · 4G LTE · Android · 13 cm display · GSM · UMTS · 16 MP Kamera · Quadband

Inspiriert durch die Arbeit von Glasbläsern und Kunstschmieden repräsentiert das **Samsung Galaxy S6** eine edle Symbiose aus Glas und Metall. Mit ...

Top-Down-Prozesse

- Übernahme von Deprecations
 - Funktioniert gut (29 von 32 haben signifikant niedrigen NUI)
- Ausnahmen
 - s:map (\leftarrow s:hasMap)
 - s:maps (\leftarrow s:hasMap)

Für Google Maps
Viele veraltete Tutorials!



Bottom-Up-Prozesse

- Martin Luther
 - Gründer des Protestantismus
 - A success story, too (like schema.org)
(i.e., 800 million adopters worldwide)

Disclaimer:
I do not speak for the
protestant church.

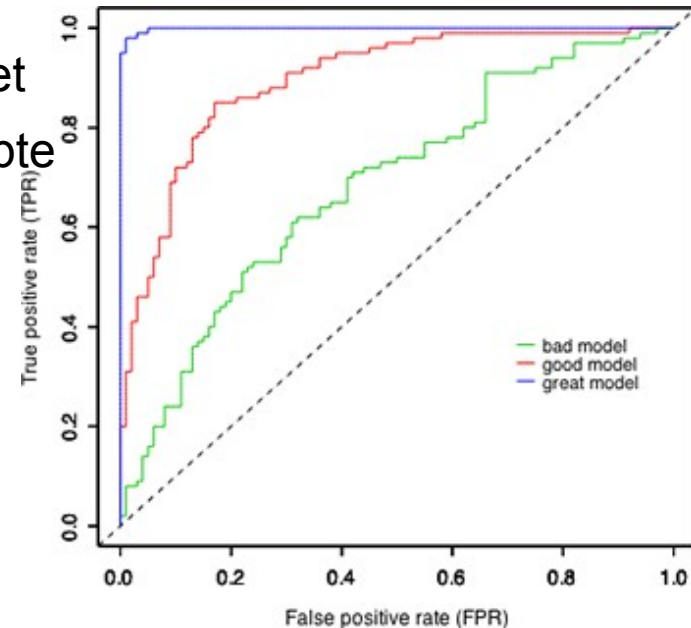


Martin Luther,
1483-1546

- Bekanntes Zitat:
 - “Man muss [...] dem gemeinen Mann aufs Maul schauen”

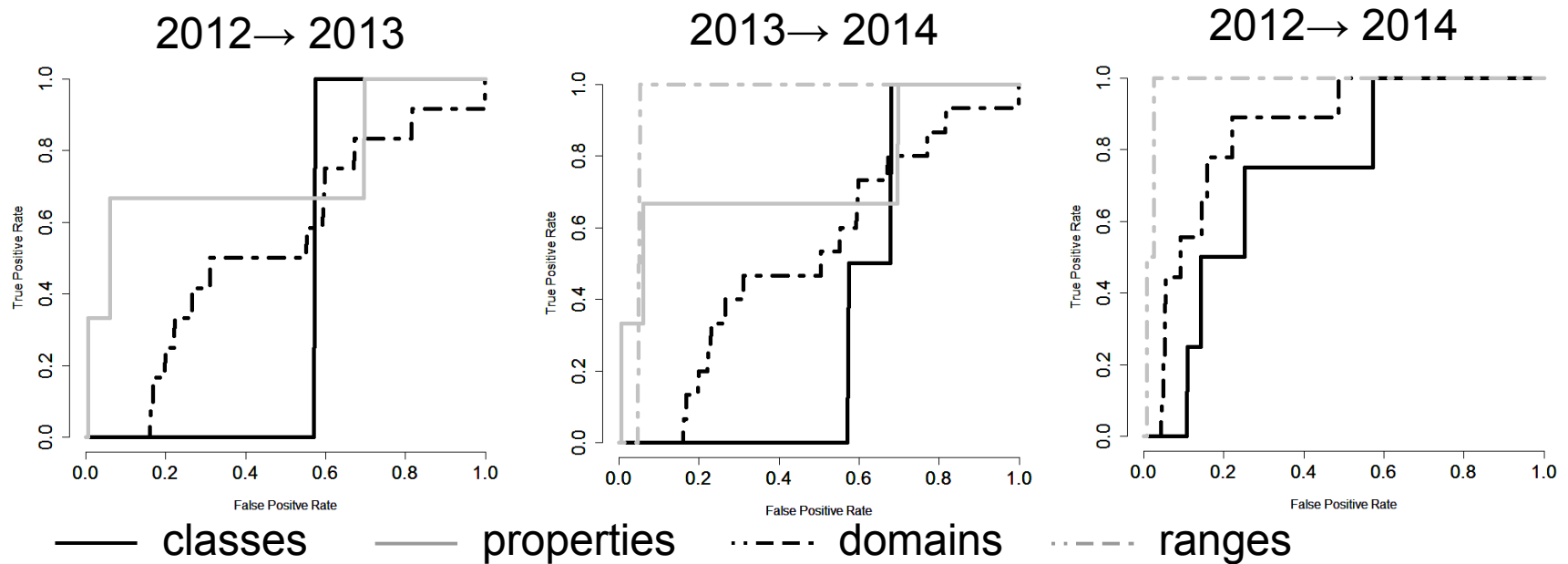
Bottom-Up-Prozesse

- Werden neue Konzepte erst “inoffiziell” verwendet?
 - Neue Klassen/Properties
 - Domain/range-Anpassungen
- Messinstrument: ROC-Kurven
 - True positives gegen false positives geplottet
 - tp: vor Standardisierung verwendete Konzepte
 - fp: nicht vorher verwendete Konzepte
 - Ranking nach #PLDs



Bottom-Up-Prozesse

- Einige Einflüsse erkennbar
 - Stärker für domain/range-Anpassungen
 - Schwächer für neue Klassen/Properties



Erfolgsfaktoren von schema.org

- Suchmaschinenoptimierung
 - Webseitenbetreiber wollen bei Google ganz oben stehen
 - Direkter Business Incentive
- Verwendung in gängigen Systemen
 - Vor allem Content Management Systeme
- Agilität des Standards
 - schema.org: 25 Versionen in den letzten drei Jahren
 - vgl. FOAF: sechs Versionen in den letzten acht Jahren
- Ease of use vs. ease of misuse
 - “Fehler” von Nutzern nicht bestrafen, sondern analysieren



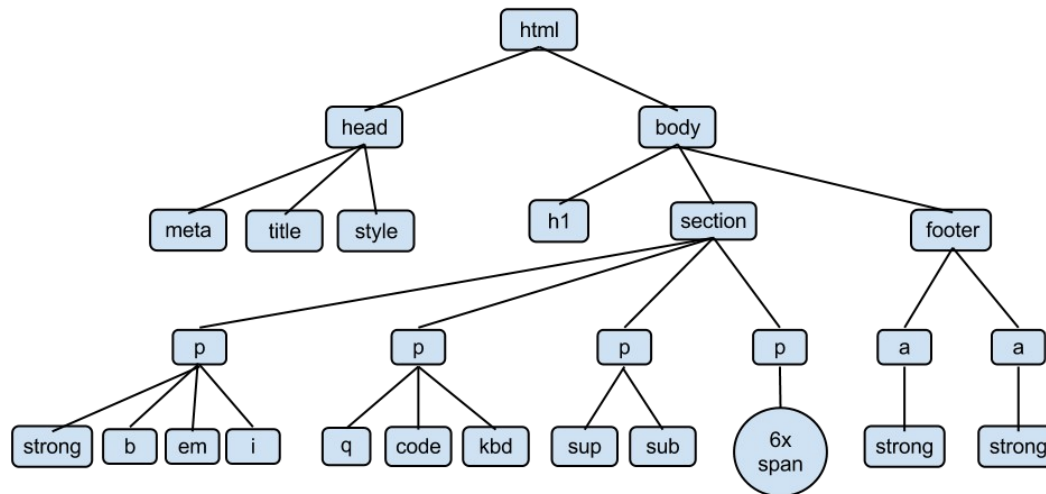
Vergleich Microdata+schema.org und LOD

- Gemeinsamkeiten
 - Maschineninterpretierbares Wissen
 - Standardisierte Vokabulare
 - Beide können RDF-Graphen kodieren



Vergleich Microdata+schema.org und LOD

- Unterschiede
 - Microdata ist in den HTML-Baum eingebettet
 - d.h., der RDF-Graph ist immer eine Menge von Bäumen
 - Graph enthält keine Zyklen
 - Denkanstoß: Zitationsnetzwerke



Vergleich Microdata+schema.org und LOD

- Vier LOD-Grundprinzipien (Sir Tim Berners-Lee)
 - URIs verwenden, um Dinge zu identifizieren MD2RDF erzeugt Blank Nodes
 - URIs sollen dereferenzierbar sein Blank Nodes sind nicht dereferenzierbar
 - Hinter dereferenzierbaren URIs nützliche Informationen hinterlegen, Standards nutzen HTML5+MD ist ein Standard
 - Verweise zu anderen Datensets setzen Möglich mit schema.org/sameas



Microdata/schema.org vs. LOD

- Vier LOD-Grundprinzipien (Sir Tim Berners-Lee)
 - URIs verwenden, um Dinge zu identifizieren
 - URIs sollen dereferenzierbar sein
 - Hinter dereferenzierbaren URIs nützliche Informationen hinterlegen, Standards nutzen
- MD2RDF erzeugt Blank Nodes
Blank Nodes sind nicht dereferenzierbar
HTML5+MD ist ein Standard

```
<div itemscope  
  itemtype="http://schema.org/PostalAddress">  
  <span itemprop="name">Data and Web Science Group</span>
```

```
<http://foo.bar/#1> a <http://schema.org/PostalAddress> .  
<http://foo.bar/#1> <http://schema.org/name> "Data and Web  
Science Group" .  
<http://foo.bar/#1> <http://schema.org/addressLocality>  
"Mannheim" .  
<http://foo.bar/#1> <http://schema.org/postalCode> "68131" .  
<http://foo.bar/#1> <http://schema.org/adressCounty> "Germany" .
```


Microdata/schema.org vs. LOD

- Vier LOD-Grundprinzipien (Sir Tim Berners-Lee, 2006)
 - URIs verwenden, um Dinge zu identifizieren
 - URIs sollen dereferenzierbar sein
 - Hinter dereferenzierbaren URIs nützliche Informationen hinterlegen; Standards nutzen
 - Verweise zu anderen Datensets setzen

Möglich mit
schema.org/sameas



- Verlinkung in schema.org Microdata:
 - schema.org/sameas wird nur von 0.02% aller Provider genutzt

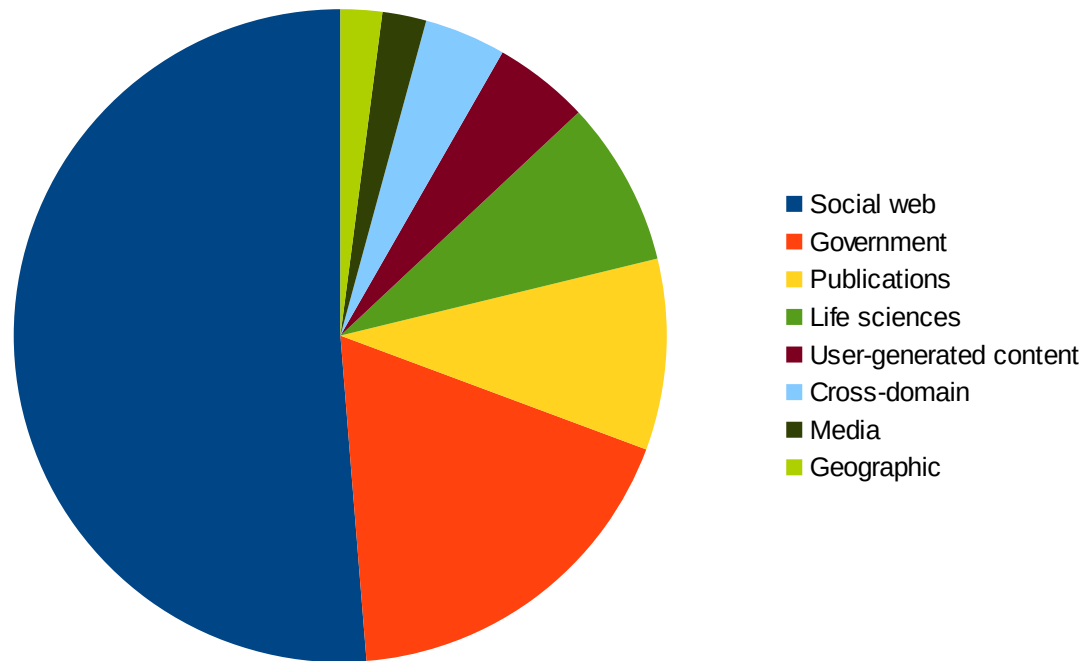
Microdata/schema.org vs. LOD

- Fünf-Sterne-Schema (Sir Tim Berners-Lee, 2010)
 - * Im Netz verfügbar mit offener Lizenz
 - ** Maschinenlesbare, strukturierte Daten
 - *** wie **, nutzt ein nicht-proprietäres Format
 - **** wie ***, nutzt einen Standard des W3C
 - ***** wie ****, enthält zusätzlich Verweise zu anderen Datensets
- Was ist die Lizenz von Daten im Web?



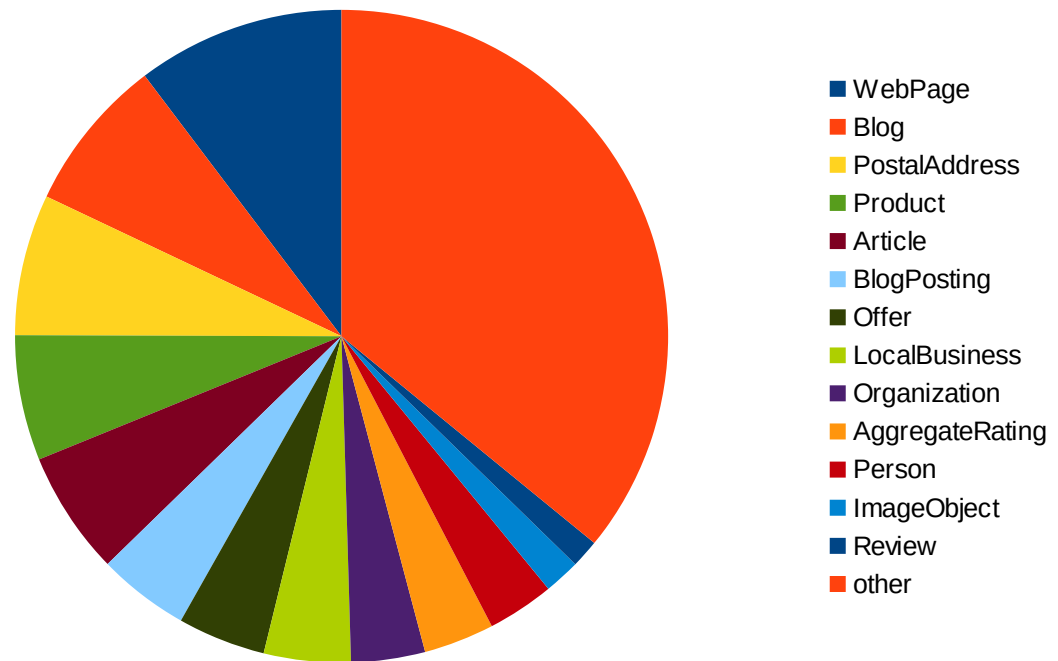
schema.org vs. LOD: Domänen

- Domänen nach Anzahl der Datensets
 - Stand 2014
 - Klassifiziert auf Basis von nutzervergebenen Tags
 - ca. die Hälfte ist Social Web



schema.org vs. LOD: Domänen


- Hauptthemen von schema.org:
 - Metadaten zu Webinhalten (web page, blog...)
 - Businessdaten (products, offers, ...)
 - Kontaktdaten (businesses, persons, ...)
 - (Produkt-) Bewertungen
- ...und ein massiver Long Tail



Beispiele: Nutzer von Strukturierten Daten

- SpringerLink (schema.org + Microdata)


Home • Contact Us

 » Download Book (PDF, 14617 KB)

Chapter
Knowledge Engineering and Management by th
Volume 6317 of the series Lecture Notes in Co

RDFauthor: Employ Knowledge Engineer

Sebastian Tramp, Norman Heino, Söre

 Download Book (PDF, 14617 KB)


Results - [Filter by use case](#) ▾


▾ ScholarlyArticle (1) 2 Errors !

ScholarlyArticle:	http://link.springer.com/chapter/chapter
image:	https://static-content.springer.com/cover/book/978-3-642-16438-5.jpg
image:	https://static-content.springer.com/lookinside/chp%3A10.1007%2F978-3-642-16438-5_7/000.png
genre:	Artificial Intelligence (incl. Robotics) Computer Communication Networks Software Engineering/Programming and Operating Systems Information Systems Applications (incl. Internet) Information Storage and Retrieval Computer Appl. in Administrative Data Processing
genre:	Pharma Materials & Steel Automotive Chemical Manufacturing Biotechnology Electronics IT & Software Telecommunications Consumer Packaged Goods Aerospace Engineering
editor [Person]:	
name:	http://link.springer.com/search?facet-creator=%22Philipp+Cimiano%22
editor [Person]:	

Beispiele: Nutzer von Strukturierten Daten

- O'Reilly: Microformats

Search Inside and Read 



Interactive Data Visualization for the Web
An Introduction to Designing with D3


By [Scott Murray](#)
Publisher: O'Reilly Media
Final Release Date: March 2013
Pages: 272

★★★★★ 4.6

[Read 23 Reviews](#) | [Write a Review](#)

Create and publish your own interactive data visualization. If you have little or no experience with data visualization or web development, this practical, hands-on introduction will get you up to speed with the concepts and methods of D3, a JavaScript library that...

[Larger Cover](#) [Full description](#)


▼ hcard (1) All good 

hcard	
tel [tel]:	
value:	1-800-889-8969
tel [tel]:	
value:	707-827-7019
email [email]:	
value:	orders@oreilly.com

100% guarantee [More Info](#)

[what's this?](#)

[Add to Cart](#)

▼ hreview-aggregate (1) All good 

hreview-aggregate	
item:	Interactive Data Visualization for the Web
count:	23
rating [complex-rating]:	
average:	4.6

[Add to Cart](#)

[Add to Cart](#)

Beispiele: Nutzer von Strukturierten Daten

- ResearchGate: Microdata+schema.org

Article

Global Measures of Data Utility for Microdata Masked for Disclosure Limitation



Mi-Ja Woo



Jerome P Reiter



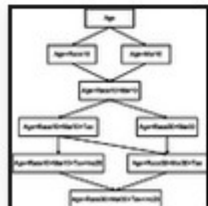
Alan F. Karr

Journal of Privacy and Confidentiality 01/20

ABSTRACT

When releasing microdata to the public, data must be masked to protect the confidentiality of database subjects.

Figures in this publication



Results - [Filter by use case](#)

ScholarlyArticle 2:	https://www.researchgate.net/publication/rgw2_56fba53bd6416
name:	Global Measures of Data Utility for Microdata Masked for Disclosure Limitation
headline:	Global Measures of Data Utility for Microdata Masked for Disclosure Limitation
image:	https://i1.rgstatic.net/publication/254571012_Global_Measures_of_Data_Utility_for_Microdata_Masked_for_Disclosure_Limitation/links/02e7e53c552590ed44000000/smallpreview.png
datePublished:	2009-01
description:	ABSTRACT
author [Person]:	
name:	Mi-Ja Woo
author [Person]:	
name:	Jerome P Reiter
author [Person]:	https://www.researchgate.net/publication/rgw12_56fba53bd6416
sameAs:	https://www.researchgate.net/profile/Anna_Oganian
name:	Anna Oganian
author [Person]:	https://www.researchgate.net/publication/rgw14_56fba53bd6416
sameAs:	https://www.researchgate.net/profile/Alan_Karr

Beispiele: Nutzer von Strukturierten Daten

- DBLP: JSON-LD+schema.org



[+] Heiko Paulheim

> Home > Persons

[+] Person information

■ affiliation: University of Mannheim, Germany

[+] Journal Articles

2016

■ [12] Petar Ristoski, Heiko Paulheim:
Semantic Web in data mining and knowledge discovery: A comprehensive survey. J. Web Sem.

2015

■ [11] Heiko Paulheim, Robert Meusel:
A decomposition of the outlier detection problem into a set of supervised learning problems. J. Web Sem.

■ [10] Roberto García, Heiko Paulheim, Paola Di Maio:
Special Issue on Semantic Web Interfaces. Semantic Web 6(3): 213-214 (2015)

■ [9] Petar Ristoski, Christian Bizer, Heiko Paulheim:
Mining the Web of Linked Data with RapidMiner. J. Web Sem. 35: 142-151 (2015)

■ [8] Oliver Lehmberg, Dominique Ritze, Petar Ristoski, Robert Meusel, Heiko Paulheim, Christian Bizer:
The Mannheim Search Join Engine. J. Web Sem. 35: 159-166 (2015)

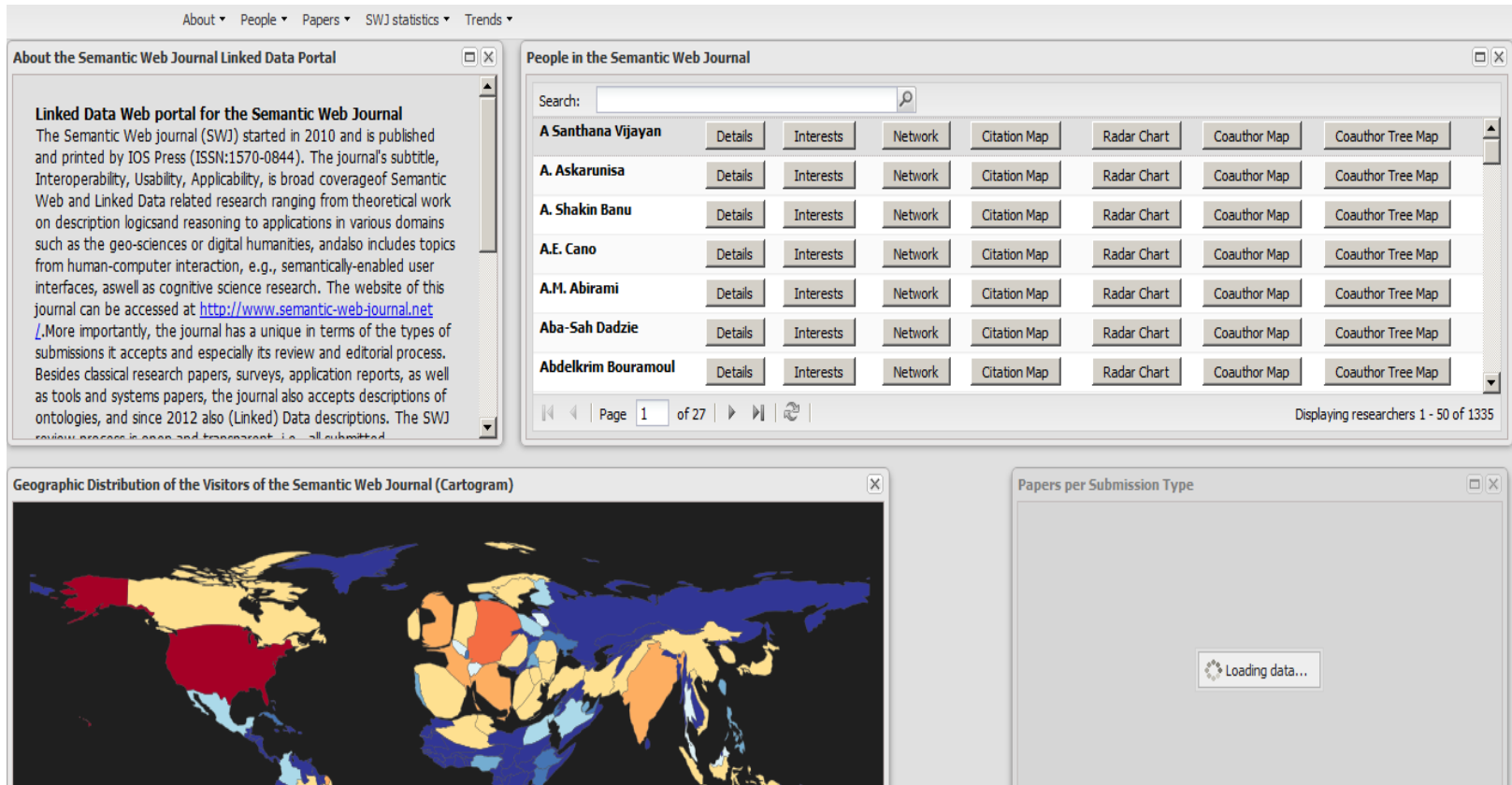
Book (8)

Keine Fehler

Book 1:	http://dblp.uni-trier.de/pers/hd/p/conf/esws/2015sumpre-hswi
additionalType:	http://dblp.uni-trier.de/rdf/schema-2015-01-26#Publication
url:	http://dblp.uni-trier.de/db/conf/esws/sumpre-hswi2015.html
url:	http://ceur-ws.org/Vol-1556
name:	Joint Proceedings of the 1st International Workshop on Summarizing and Presenting Entities and Ontologies and the 3rd International Workshop on Human Semantic Web Interfaces (SumPre 2015, HSWI 2015) co-located with the 12th Extended Semantic Web Conference (ESWC 2015), Portoroz, Slovenia, June 1, 2015.
datePublished:	2016
author [Person]:	
url:	http://dblp.uni-trier.de/pers/hd/c/Cheng:Gong
name:	Gong Cheng
author [Person]:	
url:	http://dblp.uni-trier.de/pers/hd/g/Gunaratna:Kalpa
name:	Kalpa Gunaratna
author [Person]:	
url:	http://dblp.uni-trier.de/pers/hd/t/Thalhammer:Andreas
name:	Andreas Thalhammer
author [Person]:	
name:	Heiko Paulheim
author [Person]:	
url:	http://dblp.uni-trier.de/pers/hd/v/Voigt:Martin
name:	Martin Voigt

Beispiele: Nutzer von Strukturierten Daten

- Semantic Web Journal: Linked Open Data



Beispiele: Nutzer von Strukturierten Daten

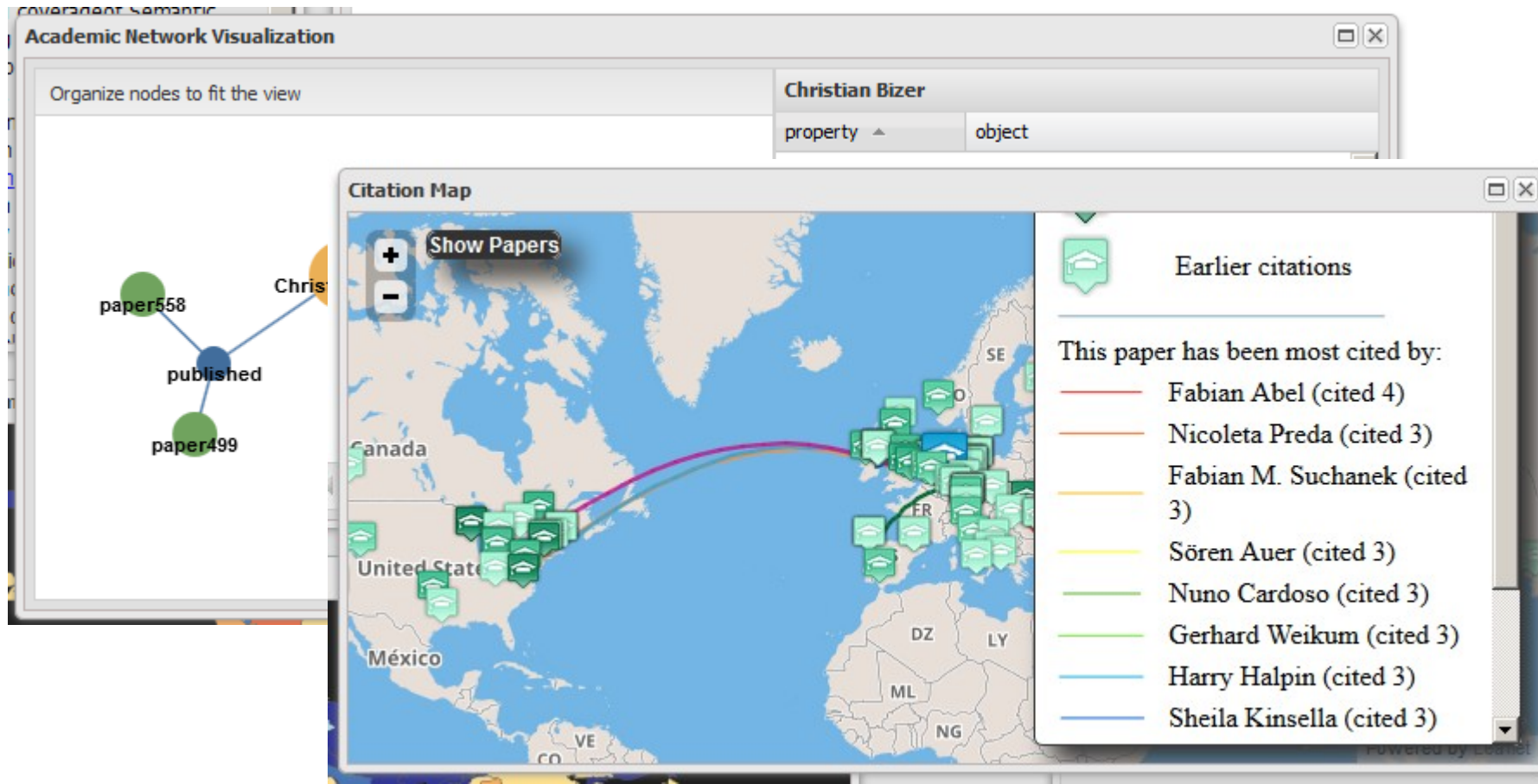
- Semantic Web Journal: Linked Open Data

The screenshot shows a web application titled "People in the Semantic Web Journal". It features a search bar with the text "Bizer" and a magnifying glass icon. Below the search bar, the name "Christian Bizer" is displayed, followed by several tabs: "Details", "Interests", "Network", "Citation Map", and "Radar". The "Details" tab is active, showing a table of properties and their corresponding objects.

property ^	object
affiliation	http://data.semanticweb.org/organization/freie-universitaet-berlin
hasCollaborator	Dimitris Kontokostas
hasCollaborator	Mohamed Morsey
hasCollaborator	Pablo N. Mendes
hasCollaborator	Robert Isele
hasCollaborator	Max Jakob
hasCollaborator	Sören Auer
hasCollaborator	Anja Jentzsch
hasCollaborator	Christian Bizer
hasCollaborator	Sebastian Hellmann
hasCollaborator	Patrick van Kleef
hasCollaborator	Jens Lehmann
holdsRole	http://data.semanticweb.org/conference

Beispiele: Nutzer von Strukturierten Daten

- Semantic Web Journal: Linked Open Data



Weitere mögliche Use Cases

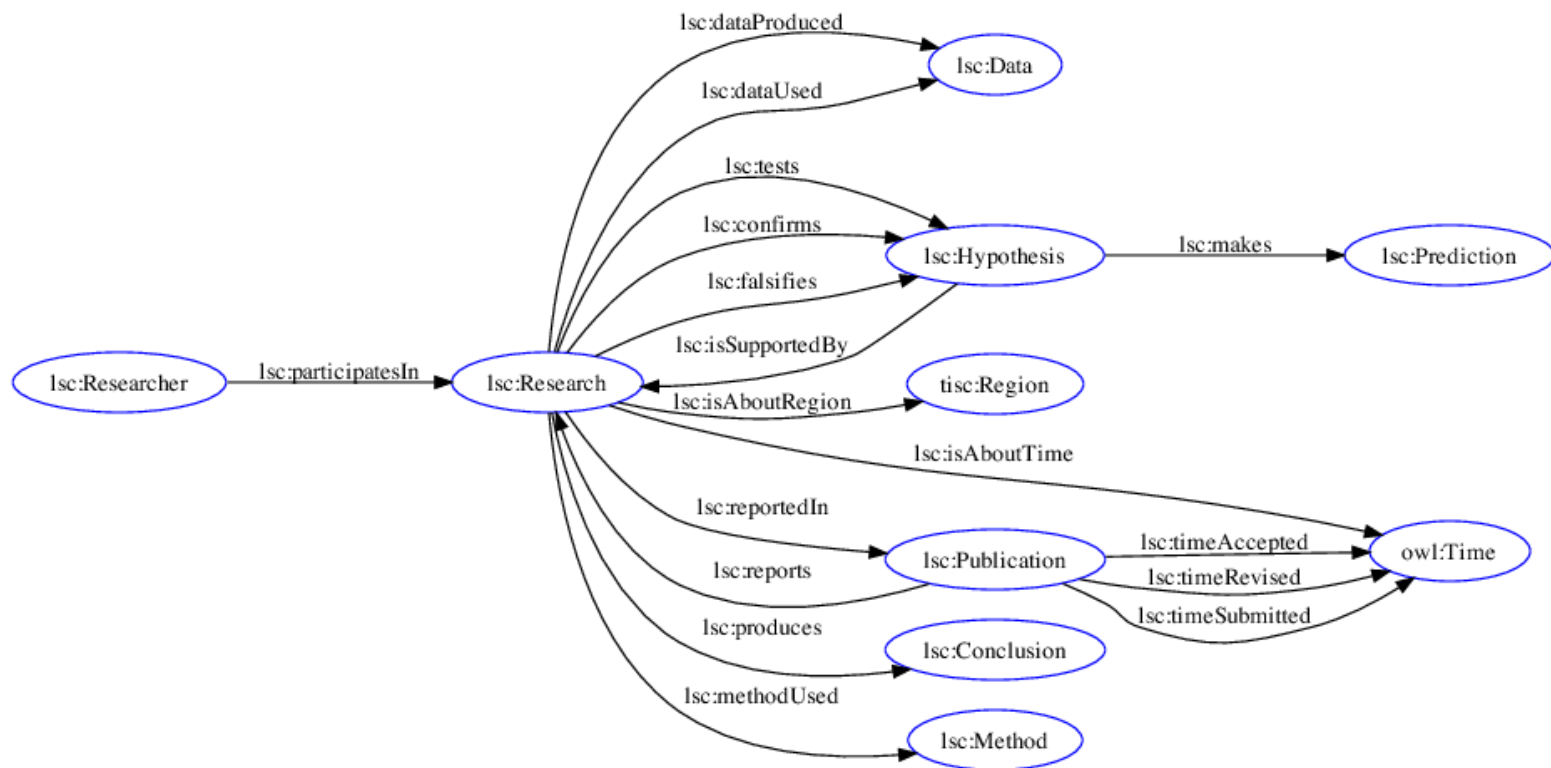
- Für Konferenz-/Journalherausgeber:
 - Zuordnung von Reviewern zu Papern optimieren
- Für Autoren:
 - Bessere Suche nach verwandten Arbeiten
 - Suche nach Kollaborationspartnern
 - Suche nach passenden Konferenzen/CfPs
- Für die Forschung
 - Identifikation von Trendthemen
 - Analyse von Netzwerken
 - Analyse von geographischen/temporalen Verteilungen

Weitere mögliche Use Cases

- Noch einmal Linked Open Data
 - Kernzutat: Links
 - Verknüpfung von bibliographischen Daten mit anderen Datensets
- Ermöglicht Analyse von interessanten Faktoren
 - Korrelation von Publikationen und Wirtschaftsdaten?
 - Korrelation von Forschungsthemen und Wirtschaftszweigen?
- Beispiel heute nachmittag
 - RapidMiner Linked Open Data Extension

Annotationen auf Sub-Publikationsebene

- Linked Science Core Vocabulary: <http://linkedscience.org/lsc/ns/>



Annotationen auf Sub-Publikationsebene

- Reine Metadatenabfrage
 - Finde Arbeiten, die X zitieren
- Mit detaillierteren Annotationen
 - Finde Arbeiten, die X zitieren und Methode M anwenden
 - Finde Arbeiten, die X zitieren und Hypothese H falsifizieren
 - ...
- Herausforderungen
 - Schwerer zu extrahieren als Metadaten
 - Einbindung der Autoren?
 - Identity Resolution

Zusammenfassung

- Überblick über Verschiedene Metadatenformate
 - Linked Open Data
 - RDFa
 - Microdata
- Alle bilden Graphen
- Komplexe Analysen sind möglich

Quellen & Danksagung

- Linked-Open-Data-Statistiken:
 - Max Schmachtenberg, Christian Bizer, Heiko Paulheim:
Adoption of the Linked Data Best Practices in Different Topical Domains, 2014
- WebDataCommons-Statistiken:
 - Robert Meusel, Petar Petrovski, Christian Bizer:
The WebDataCommons Microdata, RDFa and Microformat Dataset Series, 2014
 - Robert Meusel, Christian Bizer and Heiko Paulheim:
A Web-scale Study of the Adoption and Evolution of the schema.org Vocabulary over Time, 2015



Fragen?



Linked Open Data und strukturiertes Markup im Web



Heiko Paulheim, Data and Web Science Group