Semantic Web and the Web of Data

Pascal Hitzler
Kno.e.sis Center
Wright State University, Dayton, OH
http://www.knoesis.org/pascal/

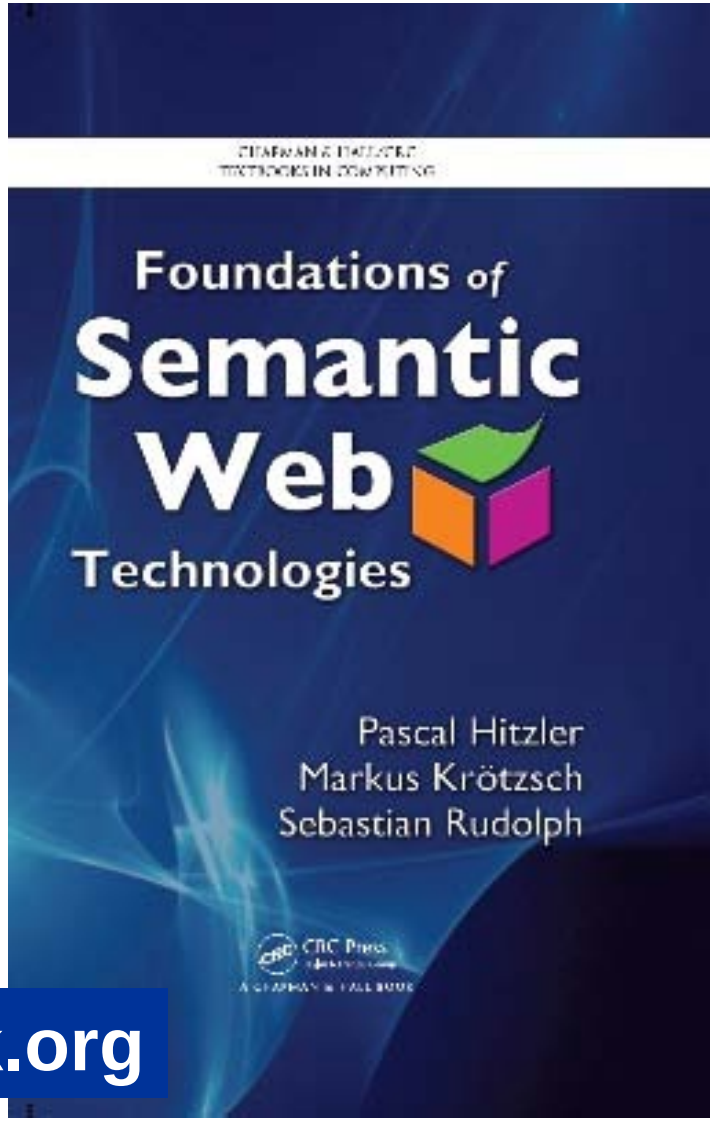
Textbook

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies

Chapman & Hall/CRC, 2010

Choice Magazine Outstanding Academic Title 2010 (one out of seven in Information & Computer Science)



http://www.semantic-web-book.org

Textbook – Chinese translation

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

语义Web技术基础

Tsinghua University Press (清华大学出版社)，2012，to appear

Translators:
Yong Yu, Haofeng Wang, Guilin Qi (俞勇，王昊奋，漆桂林)

http://www.semantic-web-book.org

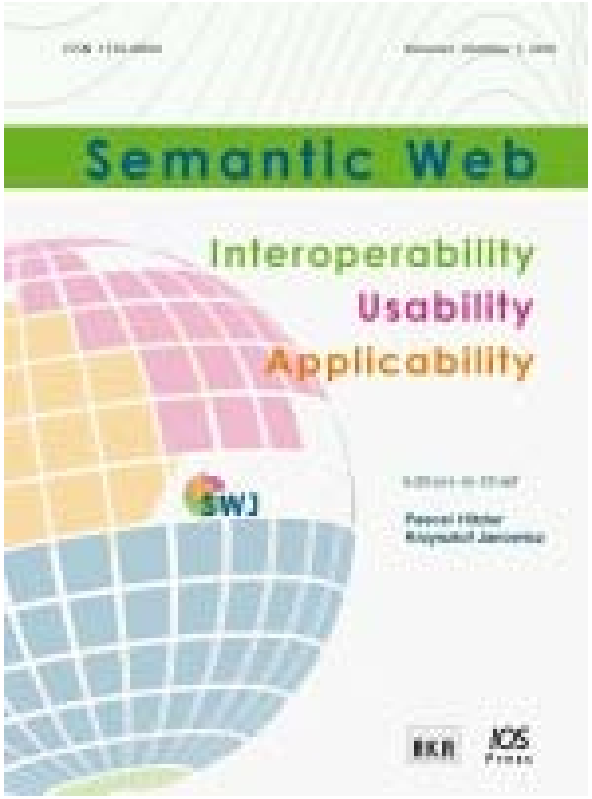
Semantic Web journal

- **EiCs: Pascal Hitzler**
  Krzysztof Janowicz

- **New journal with significant initial uptake.**

- **We very much welcome contributions at the "rim" of traditional Semantic Web research – e.g., work which is strongly inspired by a different field.**

- **Non-standard (open & transparent) review process.**



- **http://www.semantic-web-journal.net/**

Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
- The Web of Data: Linked Data
- Querying Linked Data

The current (World Wide) Web

- Immensely successful.
- Huge amounts of data.
- Syntax standards for transfer of structured data.
- Machine-processable, human-readable documents.

BUT:

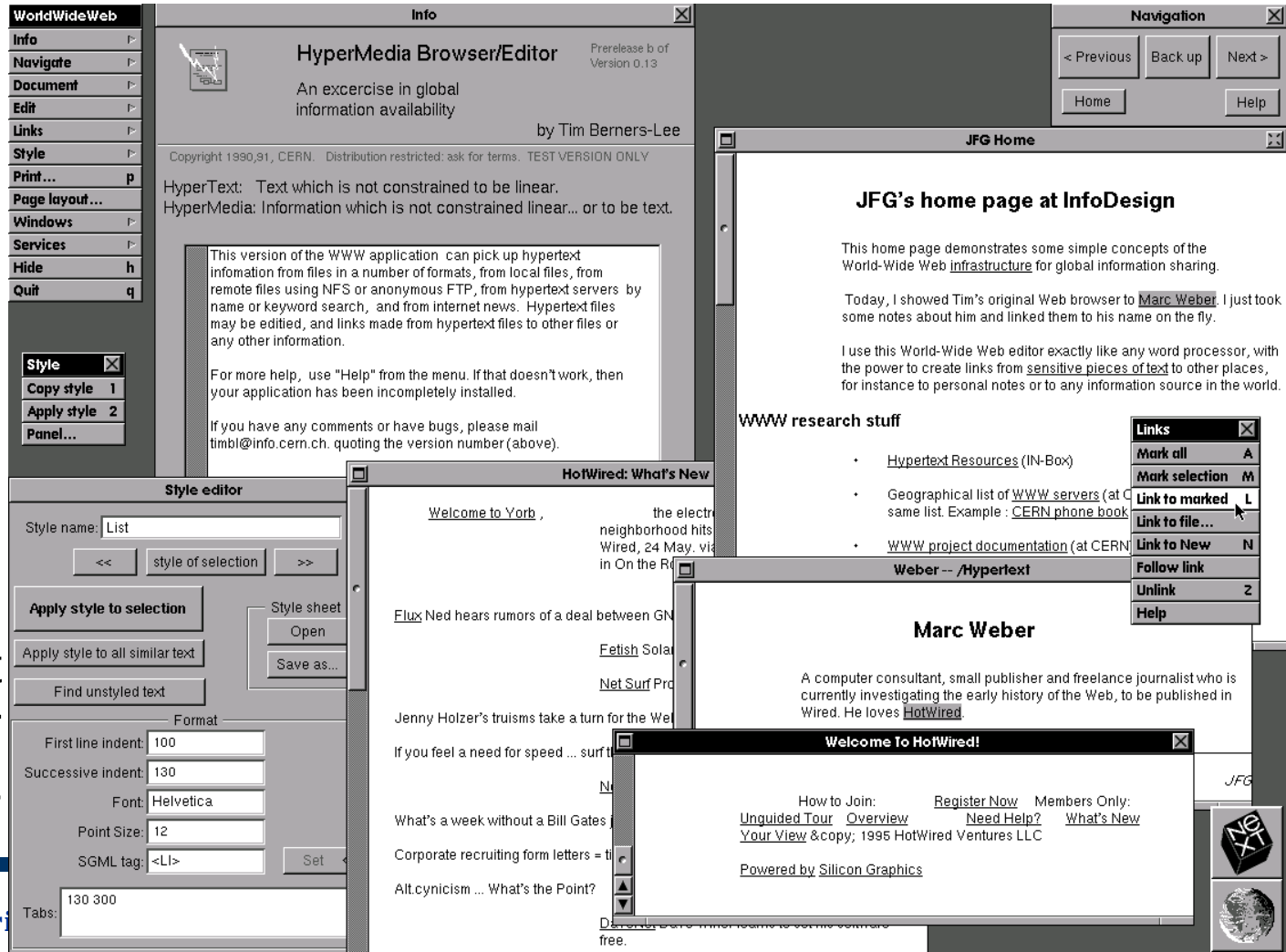
- Content/knowledge cannot be accessed by machines.
  Meaning (semantics) of transferred data is not accessible.

World Wide Web

1992: Internet release (CERN)

Basic ideas for the *Web* were fixed 1989 by Tim Berners-Lee.

Ideas for the (today so-called) *Semantic Web* have already been part of the initial ideas!



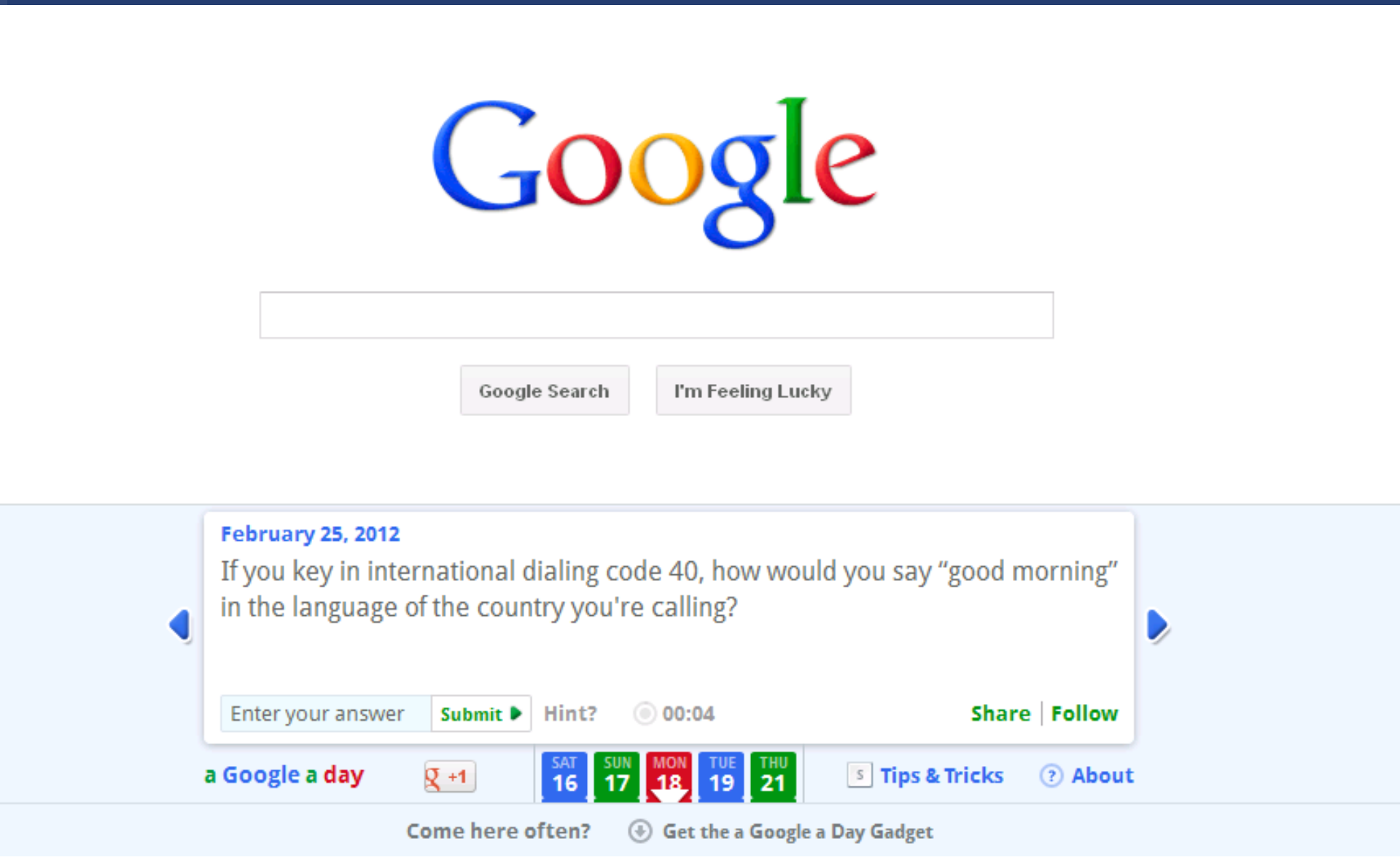
First browser by TBL 1991/92

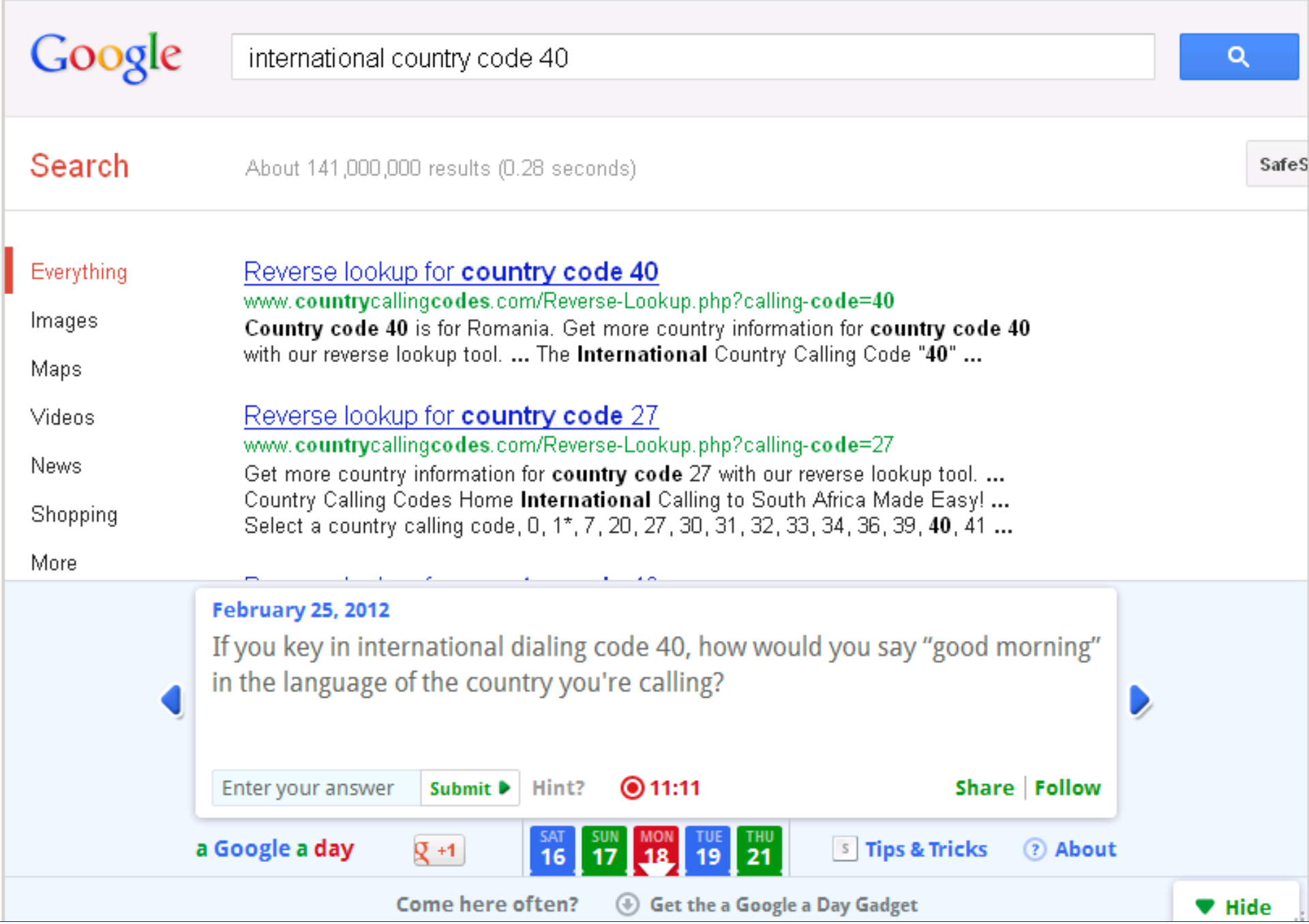
Very brief history of the Semantic Web

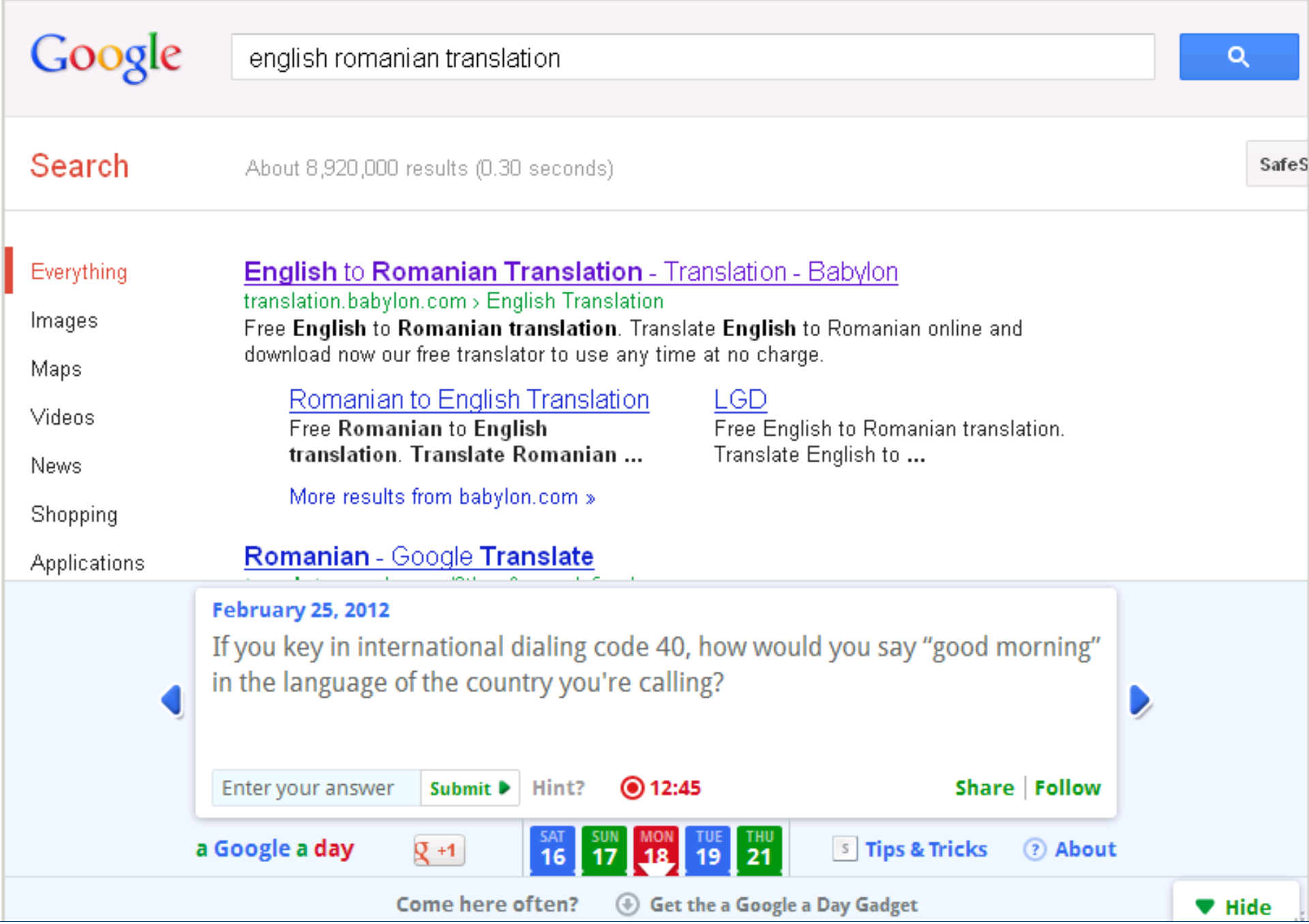
- invented ca. 1989.
- 1990s: W3C metadata activity (lead to RDF(S))
- W3C semantic web activity: chartered 2001.

- USA: DAML-Programme 2000-2005
  approx. $90M.
- Many large scale EU projects since 2002 and ongoing.
  → FP6/FP7
- Major IT companies now
  getting on board.

A Google a Day – agoogleaday.com




Google
international country code 40
Search
About 141,000,000 results (0.28 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Reverse lookup for country code 40
www.countrycallingcodes.com/Reverse-Lookup.php?calling-code=40
Country code 40 is for Romania. Get more country information for country code 40 with our reverse lookup tool. ... The International Country Calling Code "40" ...
Reverse lookup for country code 27
www.countrycallingcodes.com/Reverse-Lookup.php?calling-code=27
Get more country information for country code 27 with our reverse lookup tool. ...
Country Calling Codes Home International Calling to South Africa Made Easy! ...
Select a country calling code, 0, 1*, 7, 20, 27, 30, 31, 32, 33, 34, 36, 39, 40, 41 ...
February 25, 2012
If you key in international dialing code 40, how would you say "good morning" in the language of the country you're calling?
Enter your answer
Submit
Hint?
11:11
Share
Follow
a Google a day
+1
SAT 16
SUN 17
MON 18
TUE 19
THU 21
Tips & Tricks
About
Come here often?
Get the a Google a Day Gadget
Hide


Google
english romanian translation
Search
About 8,920,000 results (0.30 seconds)
SafeS
Everything
Images
Maps
Videos
News
Shopping
Applications
English to Romanian Translation - Translation - Babylon
translation.babylon.com › English Translation
Free English to Romanian translation. Translate English to Romanian online and download now our free translator to use any time at no charge.
Romanian to English Translation
Free Romanian to English translation. Translate Romanian ...
LGD
Free English to Romanian translation. Translate English to ...
More results from babylon.com »
Romanian - Google Translate
February 25, 2012
If you key in international dialing code 40, how would you say "good morning" in the language of the country you're calling?
Enter your answer
Submit
Hint?
12:45
Share
Follow
a Google a day
+1
SAT 16
SUN 17
MON 18
TUE 19
THU 21
Tips & Tricks
About
Come here often?
Get the a Google a Day Gadget
Hide


good morning
English
To:
Romanian
Translate
Human Translation
Buna dimineata
Babylon 9
Free download
Tweet
Try Our E
Grammar
Check
Corrects English te
Free Dow
AT&T U-verse
the best
Get Babylon's Translation Software
Free Download Now!
Get a free translation software for your desktop
100 million downloads and counting...
Free download
February 25, 2012
If you key in international dialing code 40, how would you say "good morning" in the language of the country you're calling?
Enter your answer
Submit
Hint?
02:15
Share
Follow
a Google a day
+1
SAT 16
SUN 17
MON 18
TUE 19
THU 21
Tips & Tricks
About

But computers can't do that

Question:

美国总统是谁

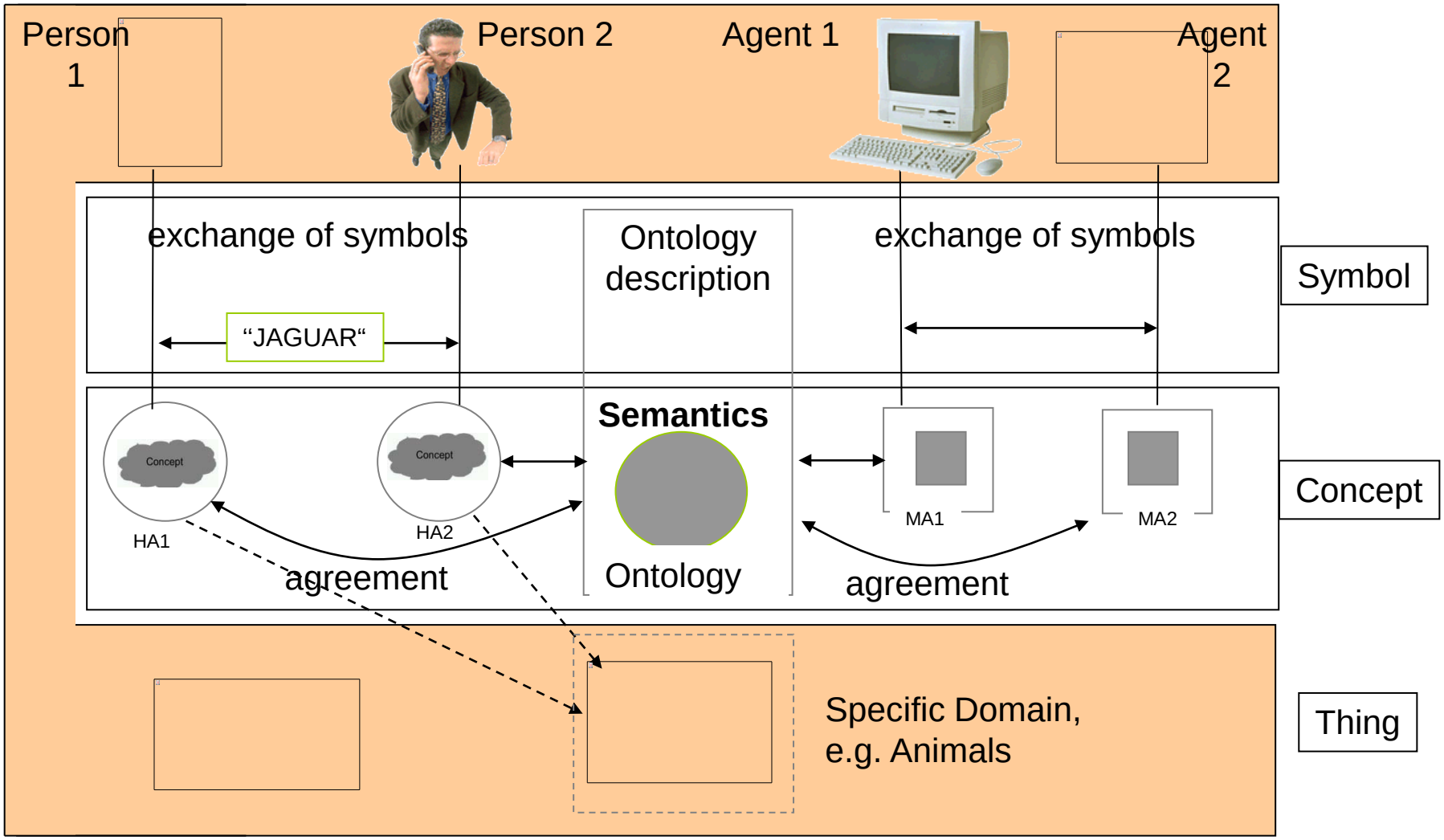
Answer:

巴拉克•奥巴马

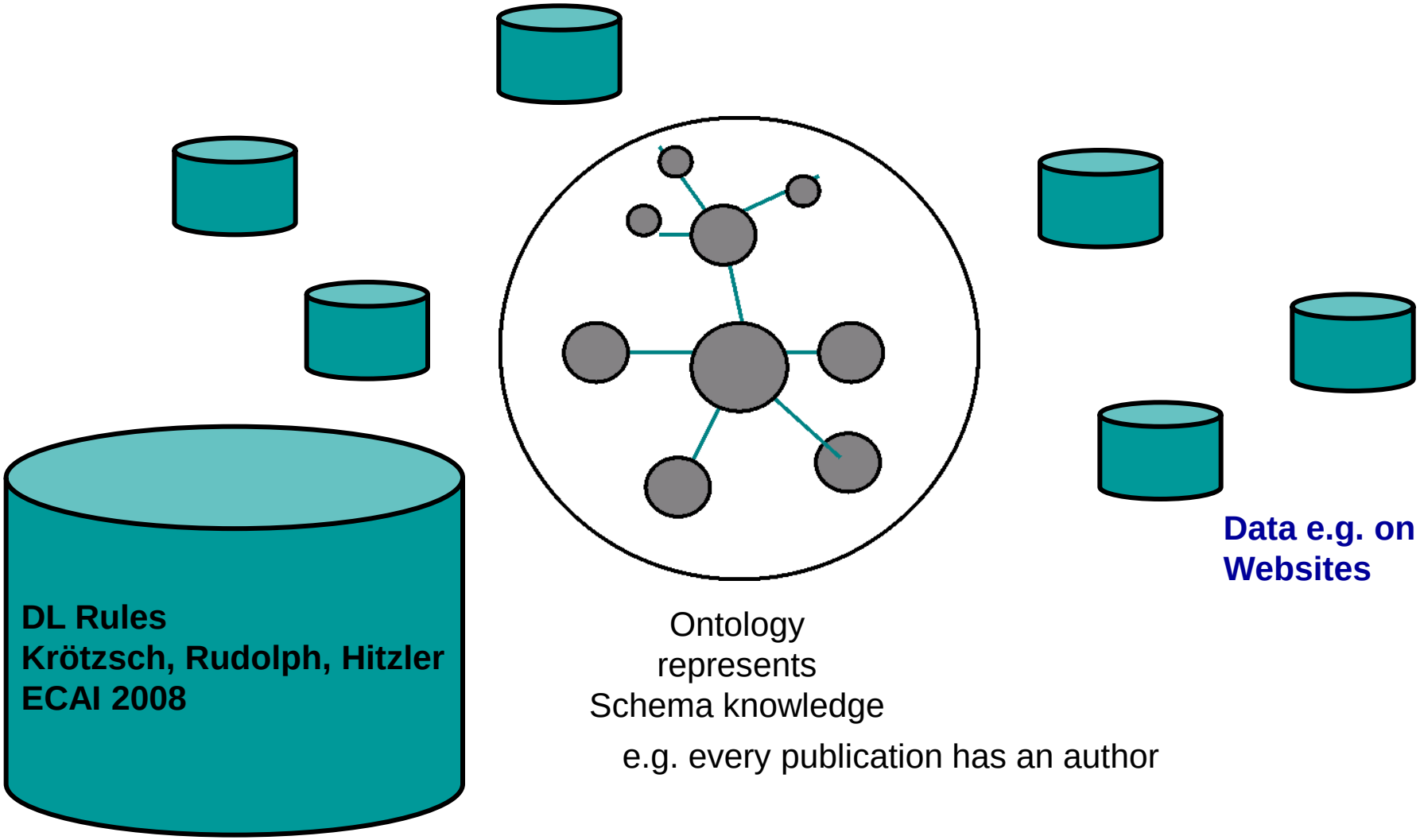
Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
- The Web of Data: Linked Data
- Querying Linked Data

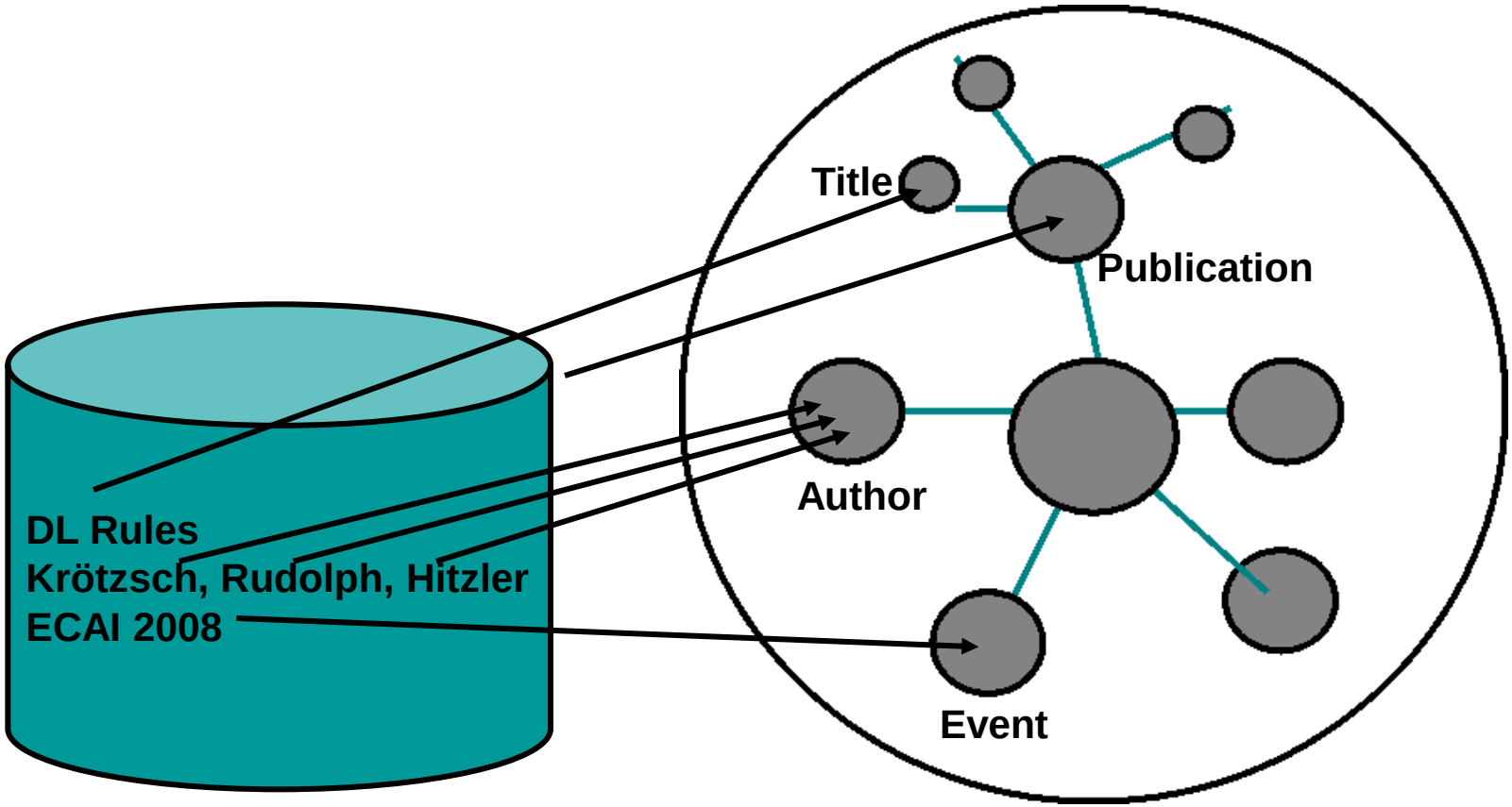
Semantic Web – Main Idea



Basic Idea of the Semantic Web

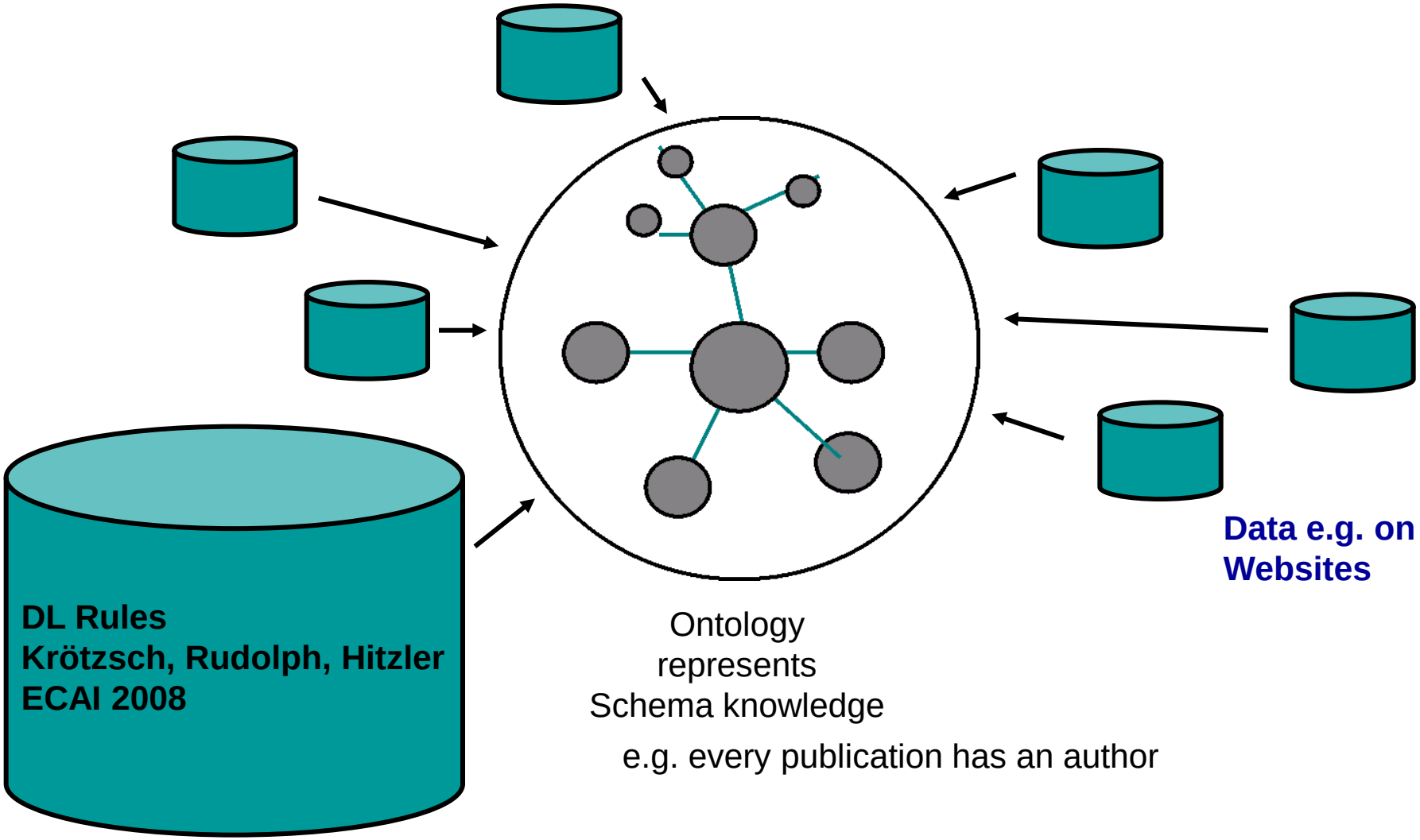


Basic Idea of the Semantic Web



e.g. every publication has an author

Basic Idea of the Semantic Web



What Is Semantic Web Semantics?

- **Opinions Differ. Here's my take.**

- **Semantic Web requires a shareable, declarative and *computable* semantics.**
- **I.e., the semantics must be a formal entity which is clearly defined and automatically computable.**

- **Ontology languages provide this by means of their formal semantics.**
- **Semantic Web Semantics is given by a relation – the *logical consequence* relation.**

- **Note: This is considerably more than saying that the semantics of an ontology is the set of its logical consequences!**

In other words

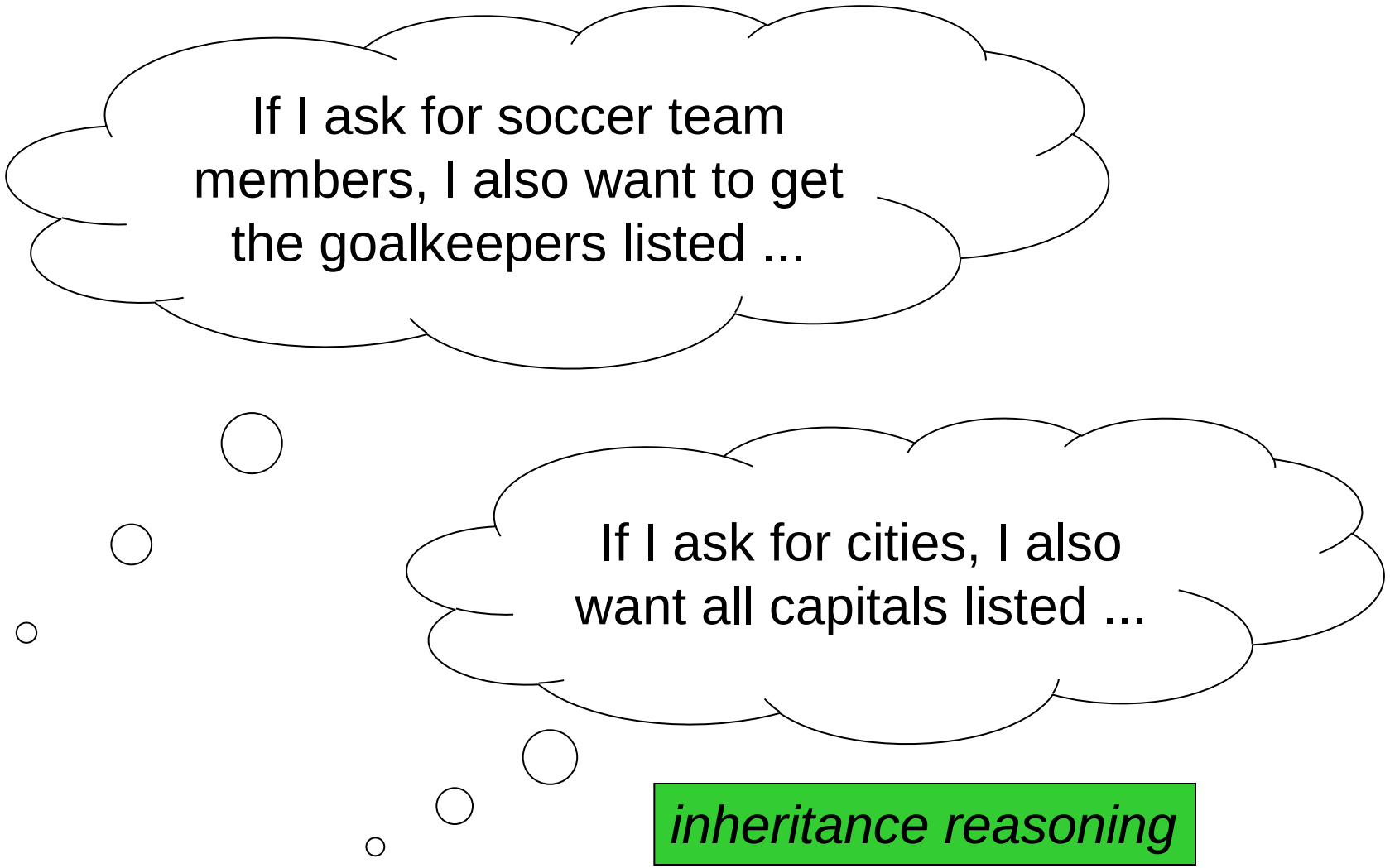
We capture the meaning of information

not by specifying its meaning (which is impossible)
but by specifying

how information interacts with other information.

We describe the meaning indirectly through its effects.

Simple Logical Reasoning



Reasoning Needs

Inspired by presentation by Evan Sandhaus, ISWC2010

x newsFrom rome .
rome locatedIn italy .

we want to conclude:
x newsFrom italy .

Take your news database.
Take location info from somewhere on linked data.
Materialize the new newsFrom triples.

Reasoning Needs

x	newsFrom	rome .	**newsFrom(x,y)**
rome	locatedIn	italy .	**locatedIn(y,z)**

we want to conclude:

x	newsFrom	italy .	**newsFrom(x,z)**

$newsFrom(x,y) \wedge locatedIn(y,z) \rightarrow newsFrom(x,z)$

$newsFrom \circ locatedIn \sqsubseteq newsFrom$

using owl:propertyChainAxiom

Reasoning as first-class citizen

Why would we want to have knowledge/rules such as

$$\text{newsFrom}(x,y) \wedge \text{locatedIn}(y,z) \rightarrow \text{newsFrom}(x,z)$$

if we can also just do this with some software code?

- **It declaratively describes what you do.**
- **It separates knowledge (as knowledge base) from programming.**
- **It makes knowledge shareable.**
- **It makes knowledge easier to maintain.**

From agoogleaday.com

answering requires merging of knowledge from many websites and using background knowledge.

- **What tribe has lived since 1300 AD near the canyon you'd explore from Bright Angel Trail?**
- **The highway that runs through Rachel, Nevada draws enthusiasts who probably enjoy what movie genre?**
- **If you key in international dialing code 40, how would you say "good morning" in the language of the country you're calling?**
- **What word will you use for "taxi" if the airport code of your destination is OSL?**
- **What single state is home to all of the following U.S. cities: Madrid, Toronto, Cincinnati, Denver, Hartford, and Norway?**

So what happened?

- In 2004, two W3C Recommendations were completed:
  - RDF + RDF Schema with formal model-theoretic semantics
  - OWL with formal model-theoretic semantics

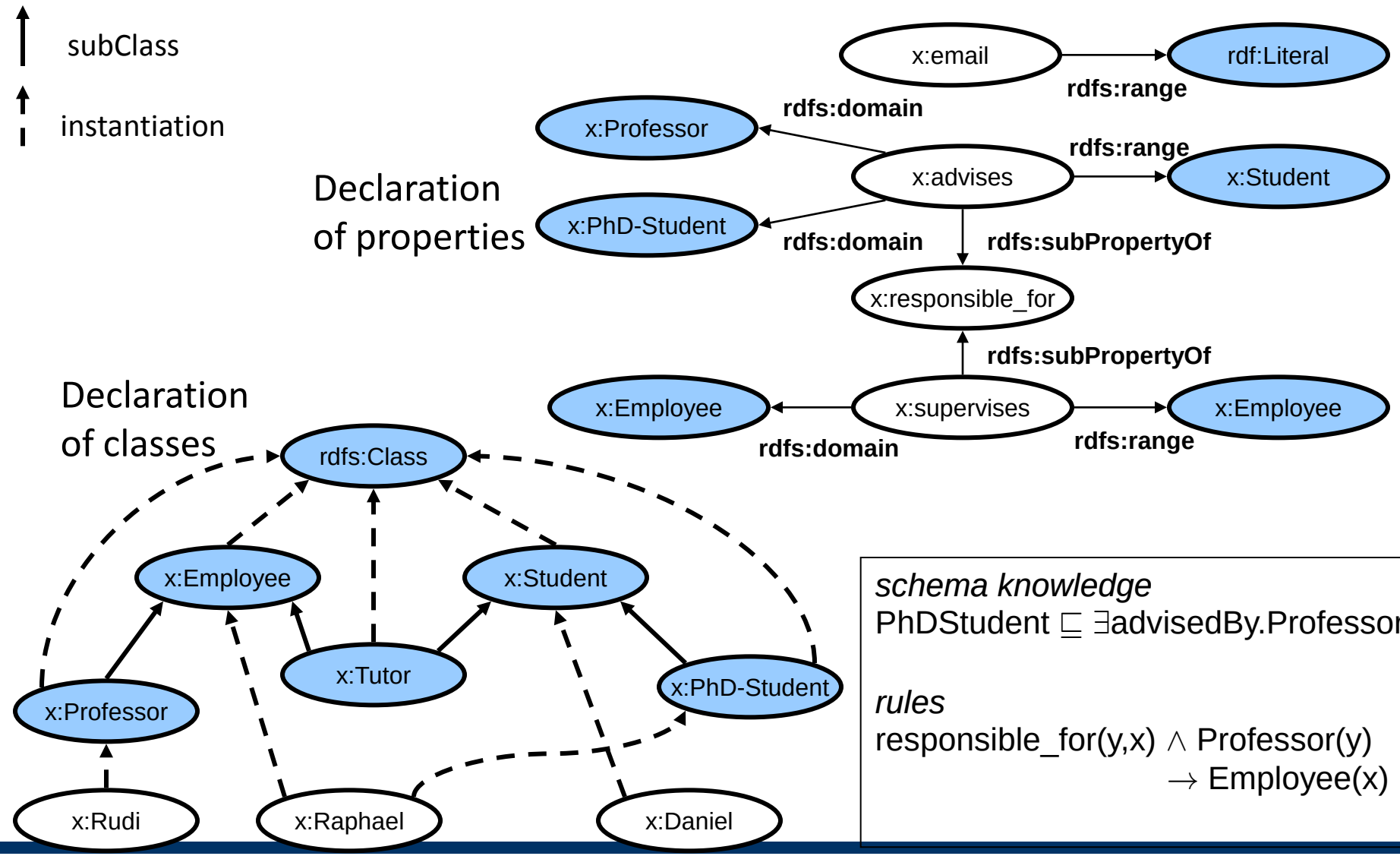
- OWL 2 update emerged 2009.
- RDF update is being discussed right now.

Ontology languages

- **Of central importance for the realisation of Semantic Technologies are suitable representation languages.**
- **Meaning (semantics) provided via logic and deduction algorithms.**
- **Scalability is a challenge.**

Language standards recommended by W3C

Ontology Example



schema knowledge

PhDStudent $\sqsubseteq \exists$advisedBy.Professor

rules

responsible_for(y,x) $\wedge$ Professor(y) $\rightarrow$ Employee(x)

Contents

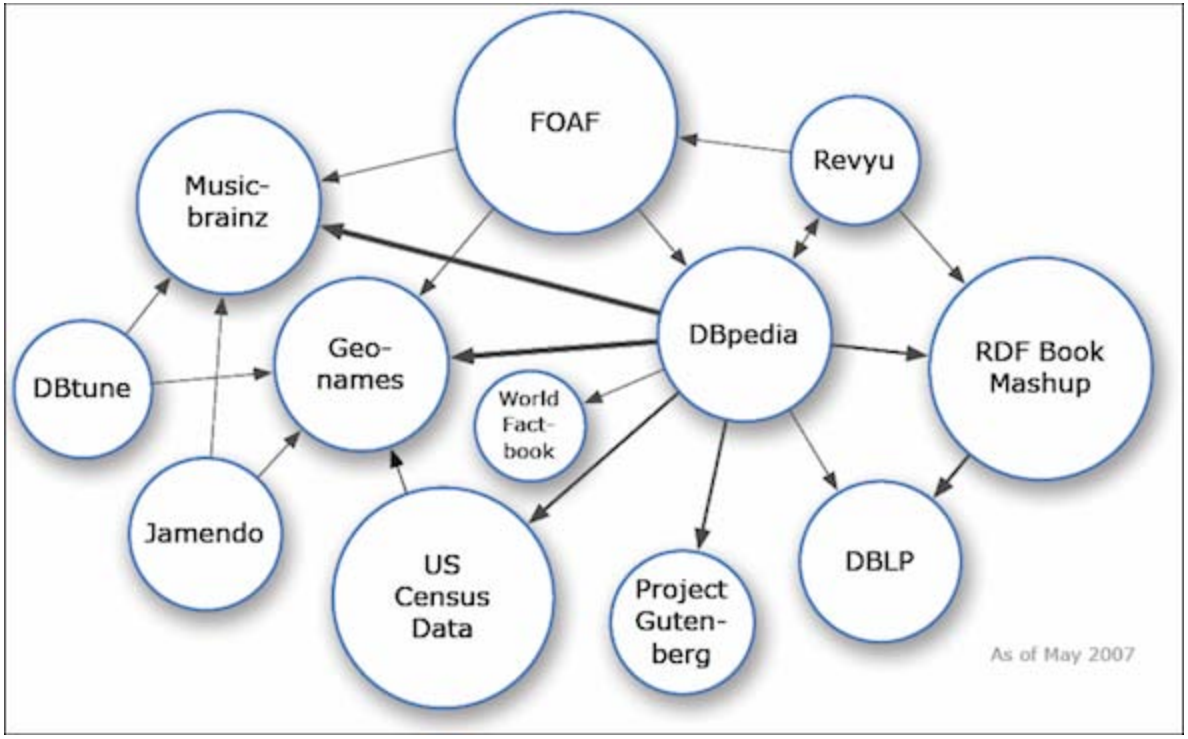
- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
- The Web of Data: Linked Data
- Querying Linked Data

Tim Berners-Lee 2006

- **from http://www.w3.org/DesignIssues/LinkedData.html**

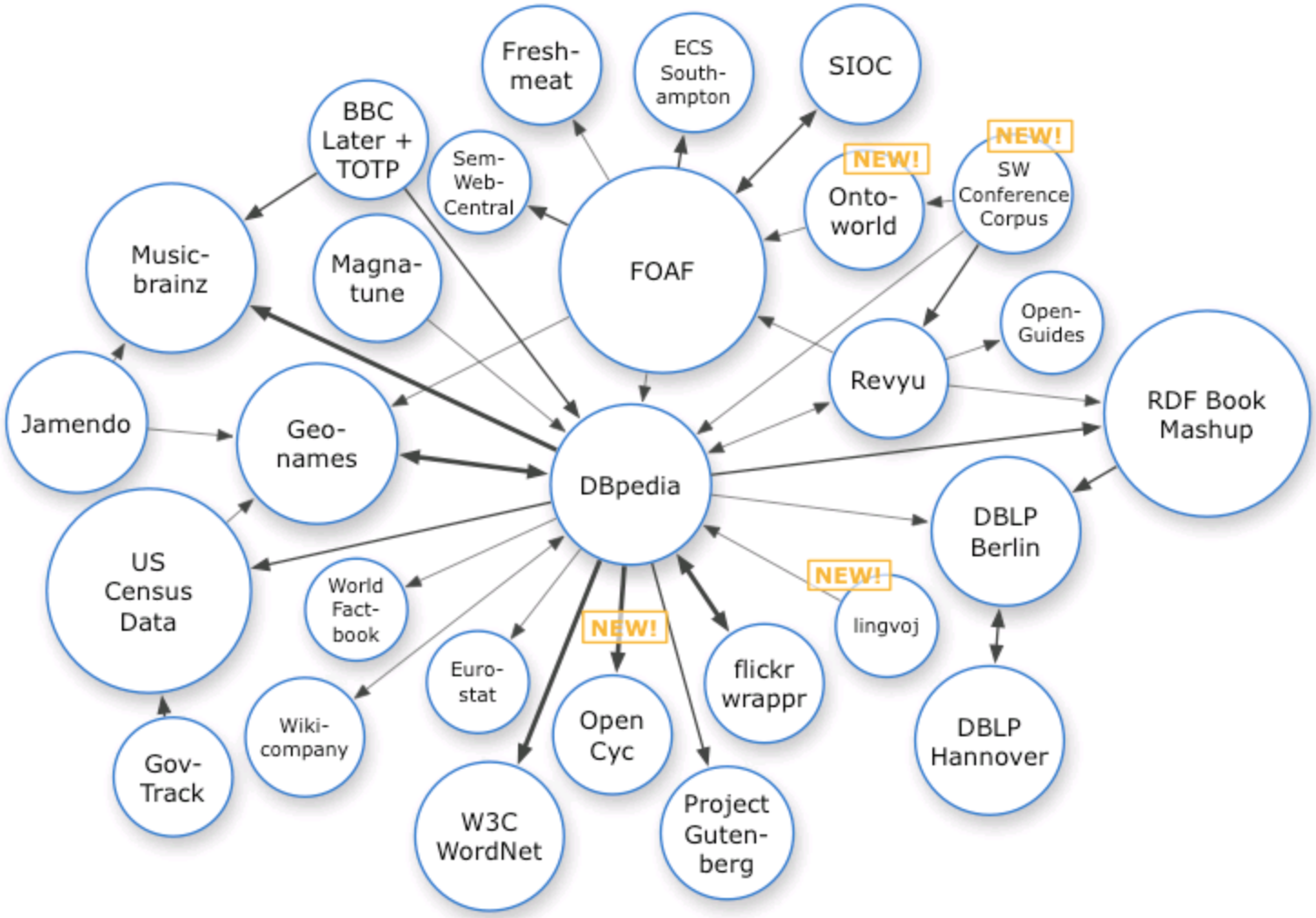
1. **Use URIs as names for things**
2. **Use HTTP URIs so that people can look up those names.**
3. **When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)**
4. **Include links to other URIs. so that they can discover more things.**

Linked Open Data 2007 (May)

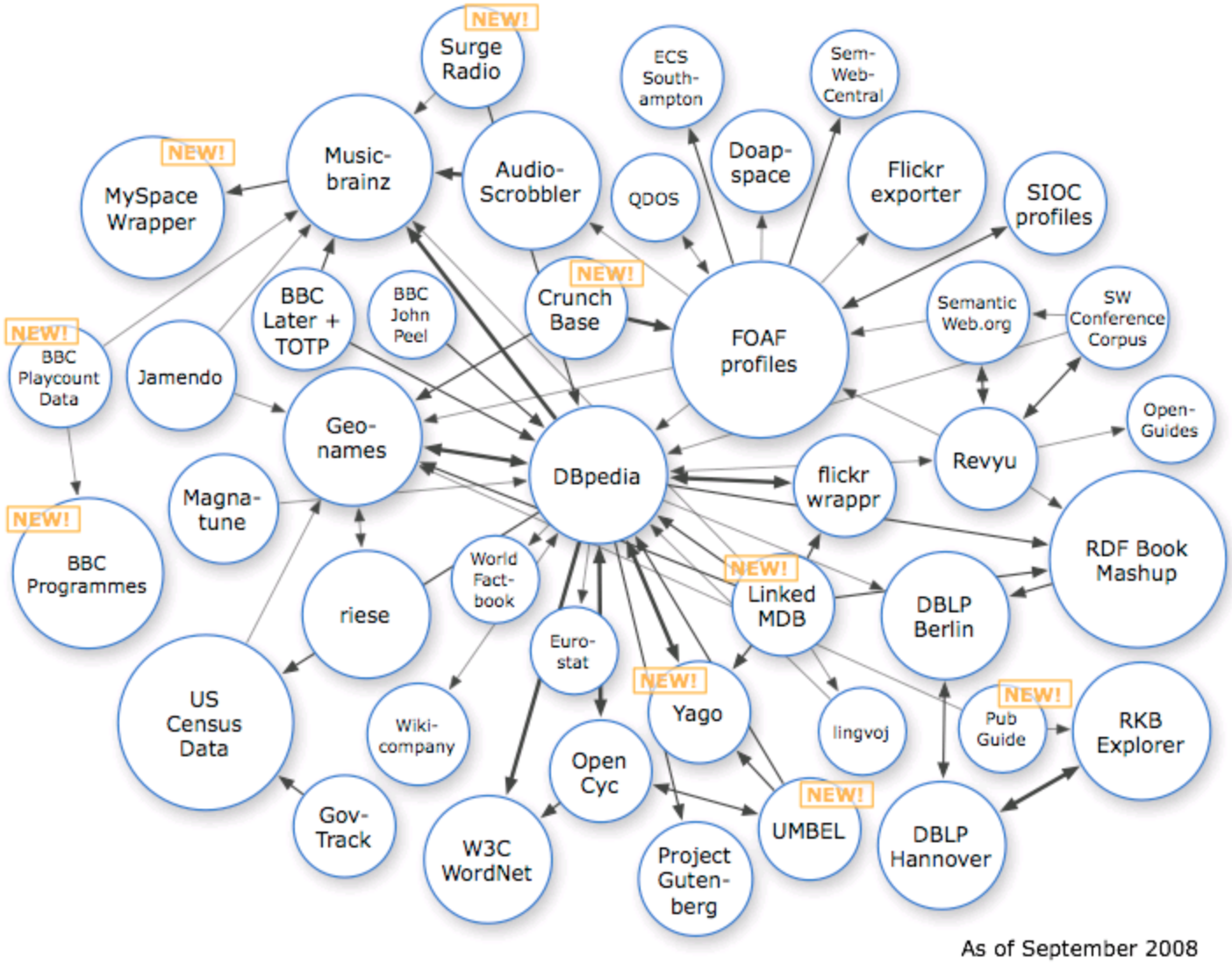


Linking Open Data cloud diagram, this and subsequent pages, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

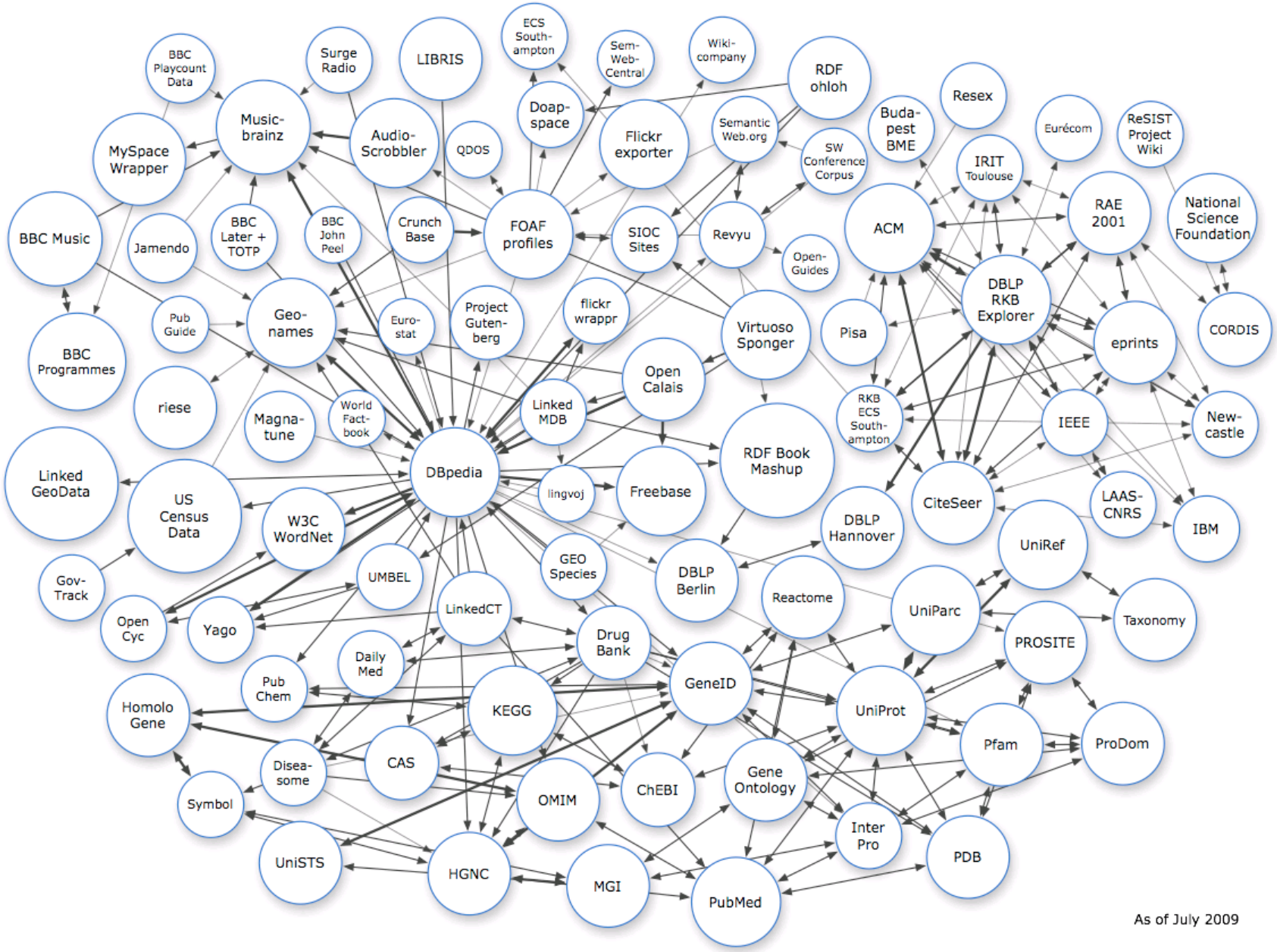
Linked Open Data 2007 (Oct)



Linked Open Data 2008

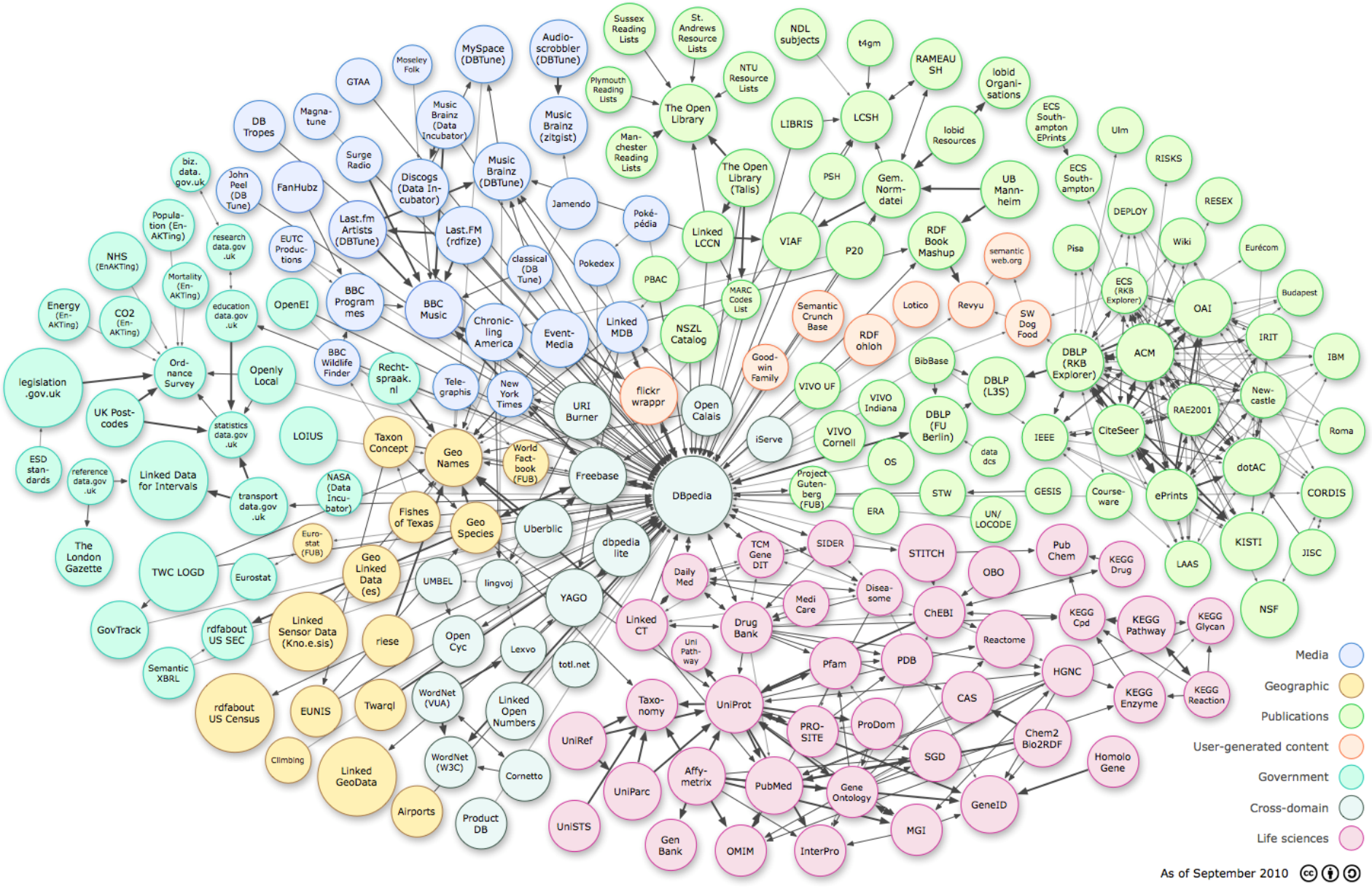


Linked Open Data 2009

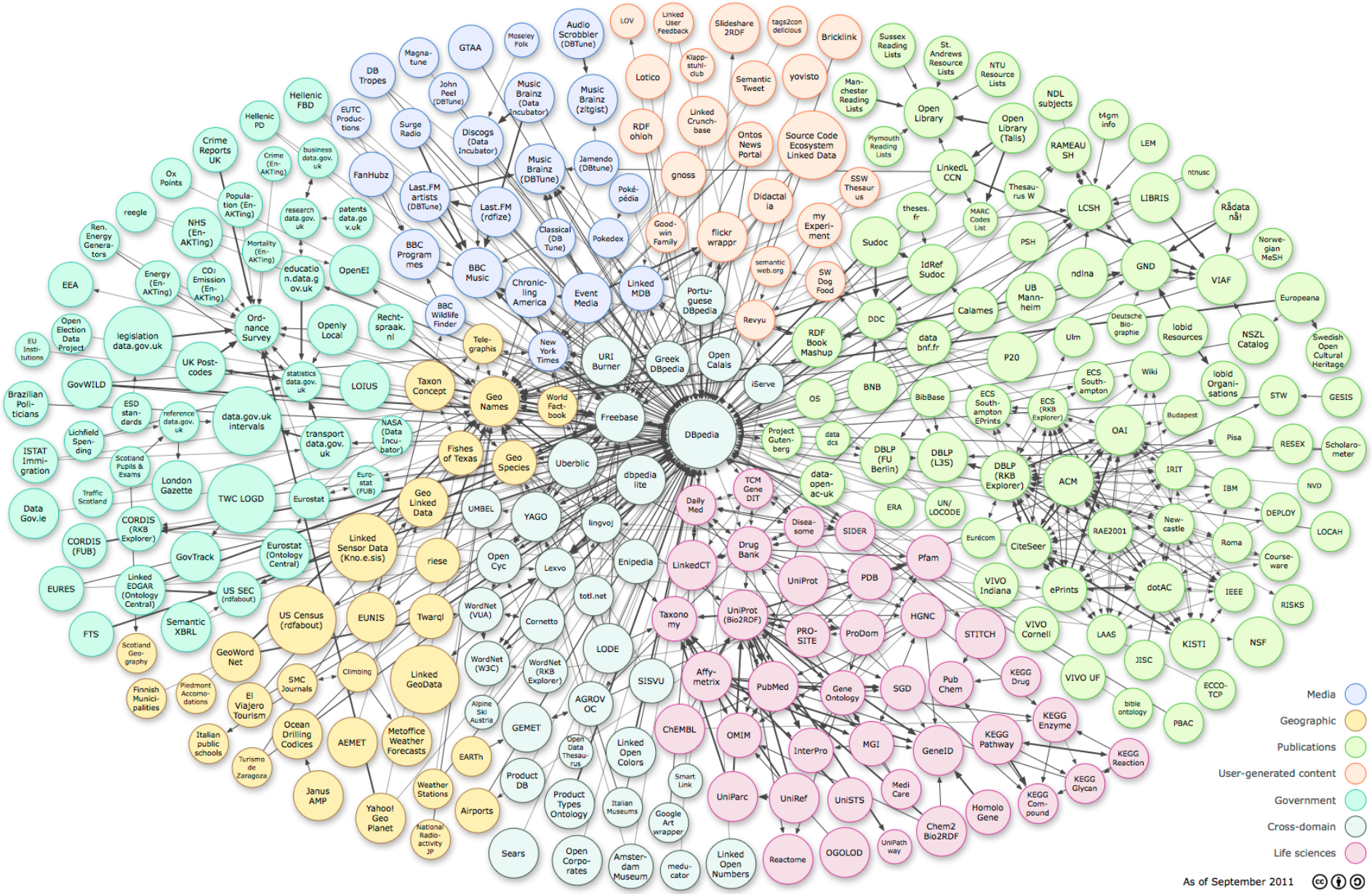


As of July 2009

Linked Open Data 2010



Linked Open Data 2011



Linked Open Data

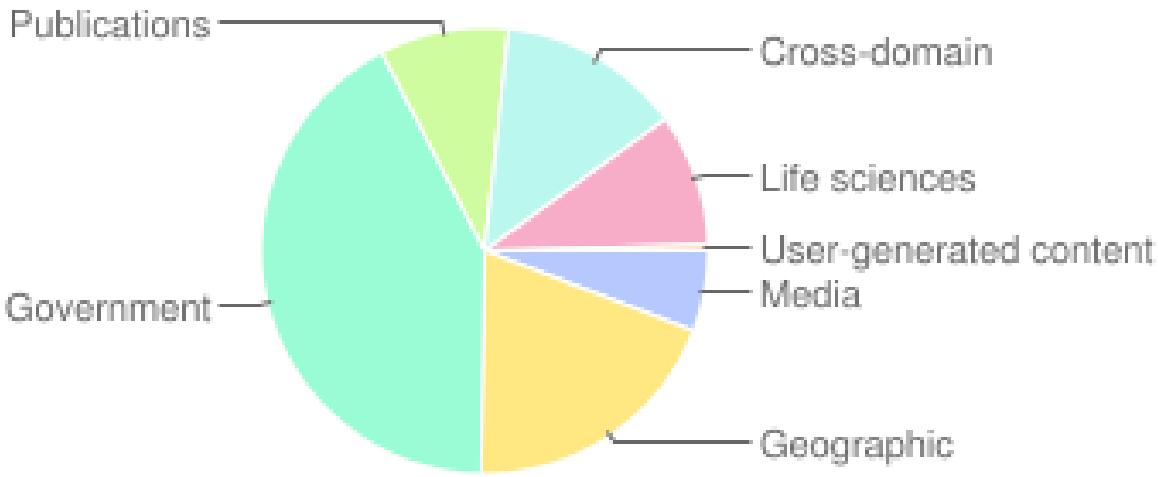
Number of Datasets

2011-09-19	**295**
2010-09-22	**203**
2009-07-14	**95**
2008-09-18	**45**
2007-10-08	**25**
2007-05-01	**12**

Number of triples (Sept 2011)

31,634,213,770

with 503,998,829 out-links



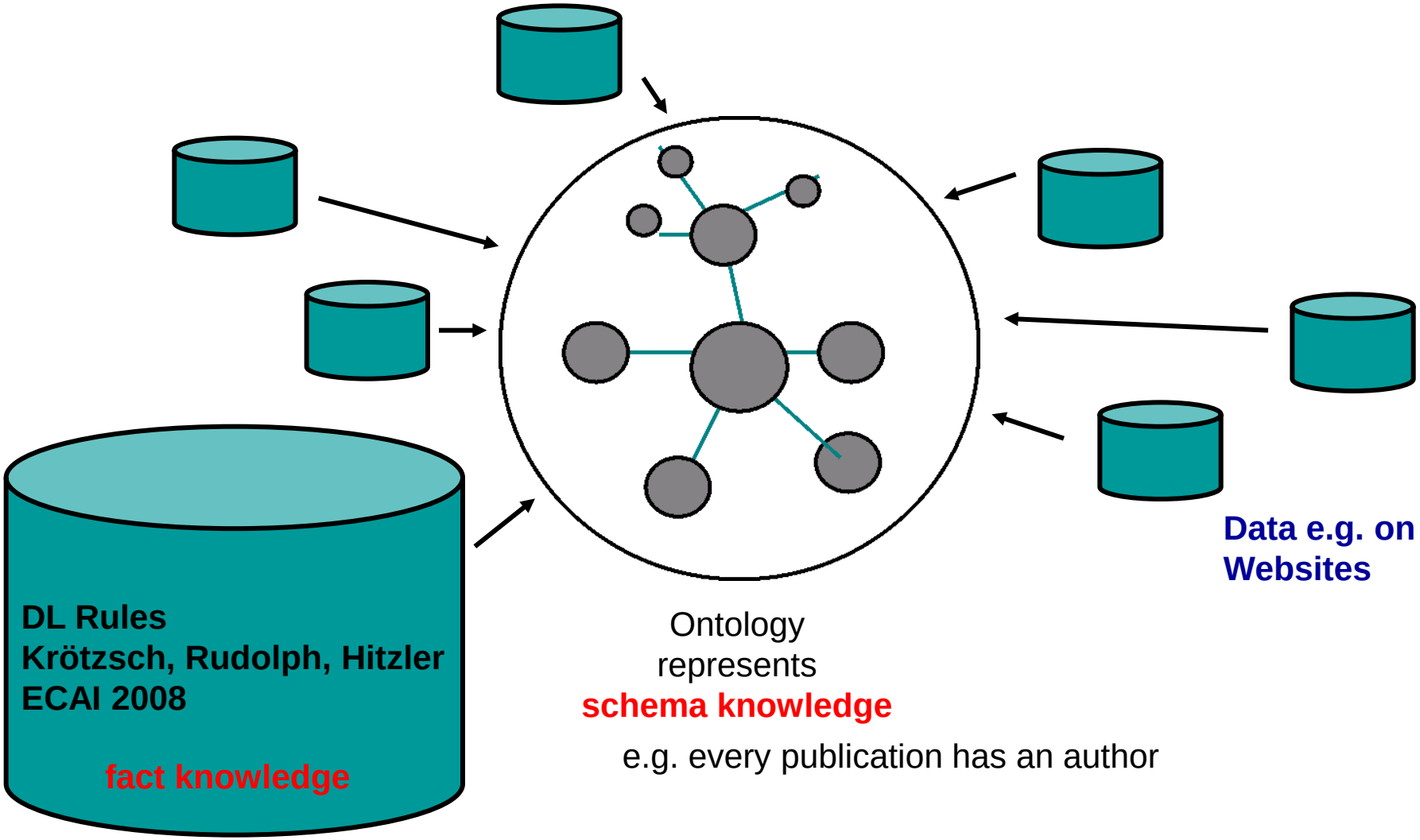
From http://www4.wiwiss.fu-berlin.de/lodcloud/state/

Example: GeoNames

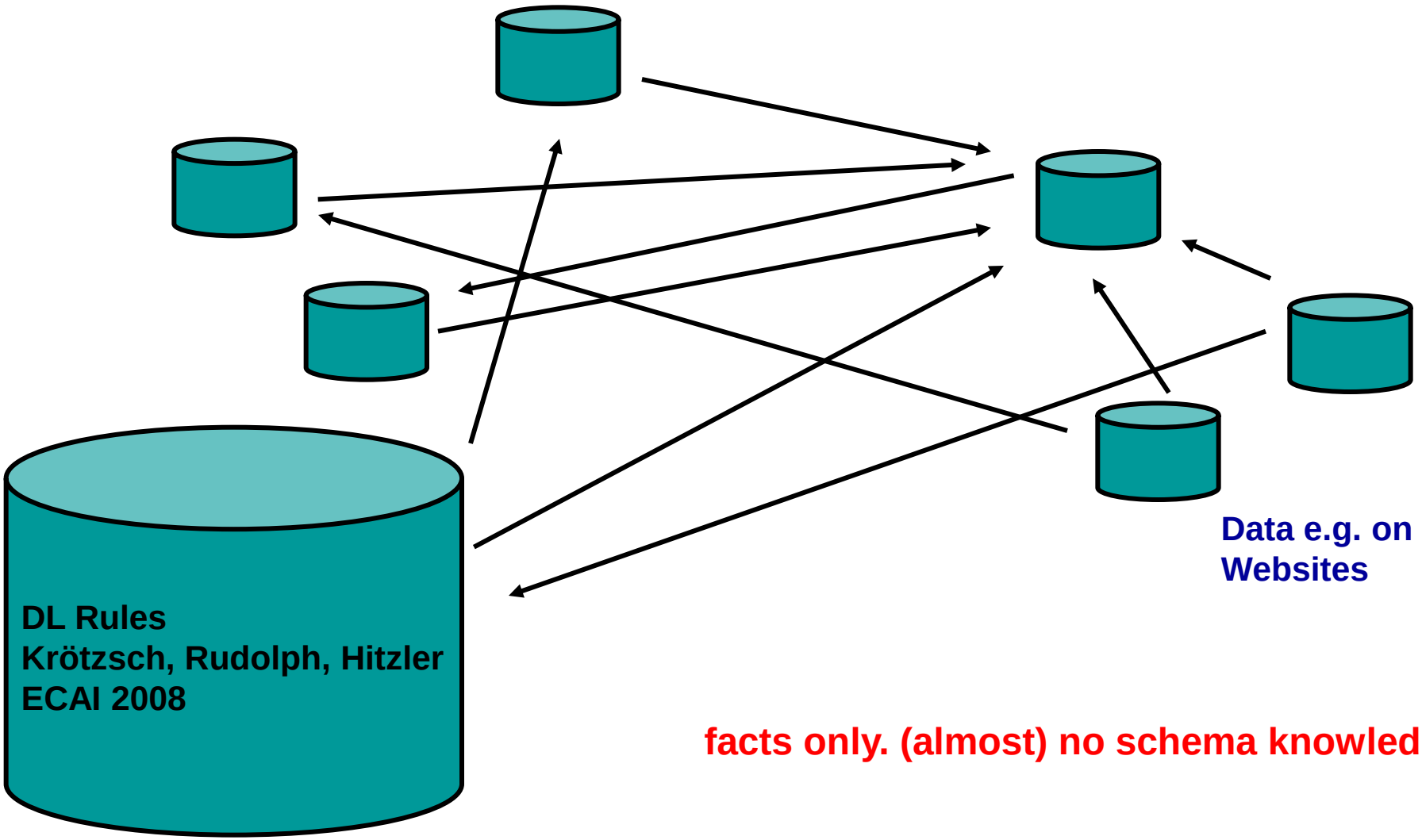
Populated Place Features (city, village,...)			
2,518,403	P.PPL	populated place	a city, town, village, or other agglomeration of buildings where people live and work
48,483	P.PPLX	section of populated place	
39,336	P.PPLL	populated locality	an area similar to a locality but with a small group of dwellings or other buildings
13,306	P.PPLQ	abandoned populated place	
2,684	P.PPLA4	seat of a fourth-order administrative division	
2,028	P.PPLA	seat of a first-order administrative division	seat of a first-order administrative division (PPLC takes precedence over PPLA)
1,847	P.PPLW	destroyed populated place	a village, town or city destroyed by a natural disaster, or by war
1,006	P.PPLF	farm village	a populated place where the population is largely engaged in agricultural activities
930	P.PPLA3	seat of a third-order administrative division	
695	P.PPLA2	seat of a second-order administrative division	
253	P.PPLS	populated places	cities, towns, villages, or other agglomerations of buildings where people live and work
249	P.STLMT	israeli settlement	
235	P.PPLC	capital of a political entity	
57	P.		
29	P.PPLR	religious populated place	a populated place whose population is largely engaged in religious occupations
6	P.PPLG	seat of government of a political entity	
2,629,547	Total for P		

rdfs:subClassOf?

Basic Idea of the Semantic Web



Currently it‘s looking like this



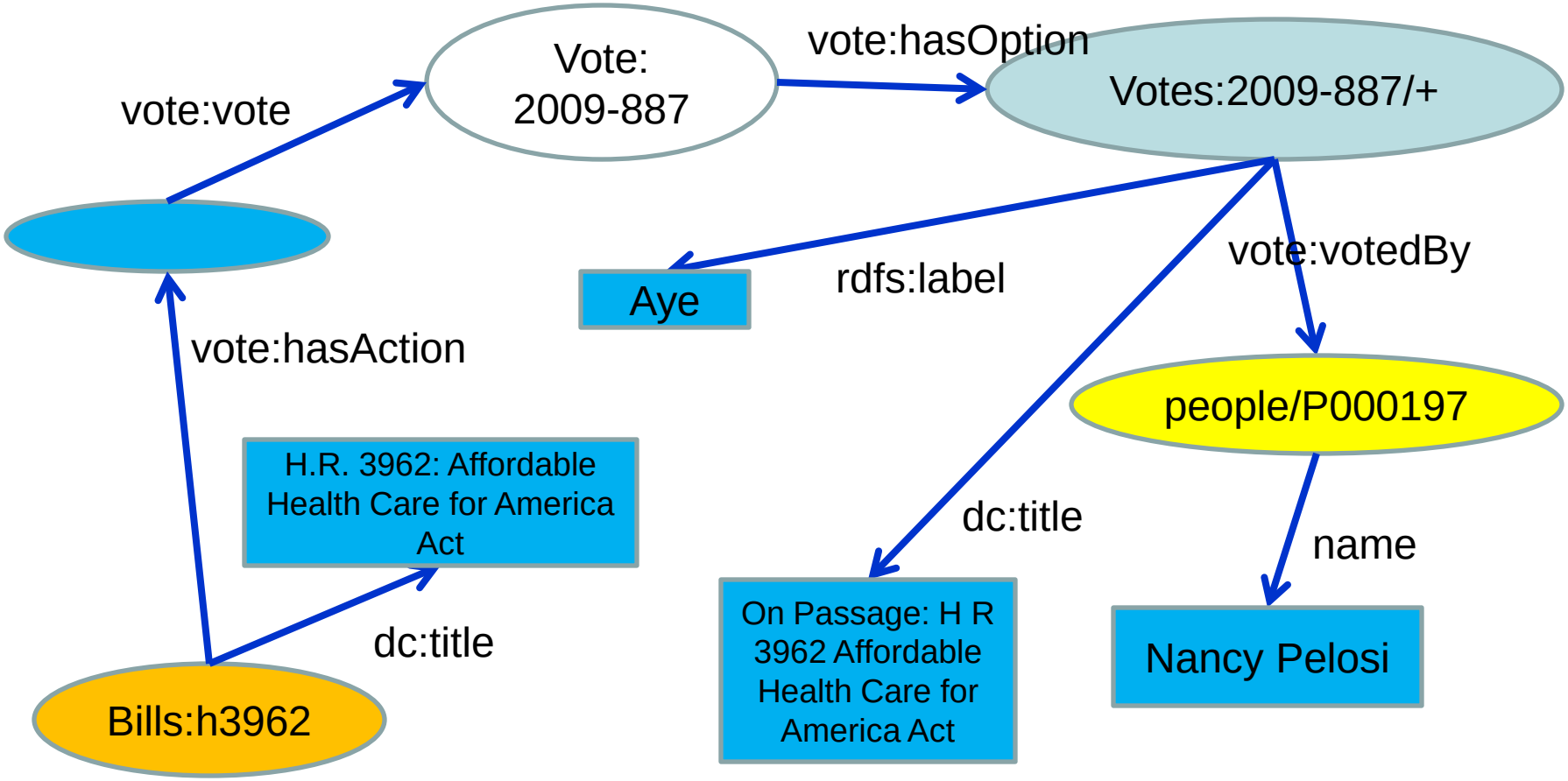
facts only. (almost) no schema knowledge

From agoogleaday.com

- What tribe has lived since 1300 AD near the canyon you'd explore from Bright Angel Trail?
- The highway that runs through Rachel, Nevada draws enthusiasts who probably enjoy what movie genre?
- If you key in international dialing code 40, how would you say "good morning" in the language of the country you're calling?
- What word will you use for "taxi" if the airport code of your destination is OSL?
- What single state is home to all of the following U.S. cities: Madrid, Toronto, Cincinnati, Denver, Hartford, and Norway?

Example: GovTrack

“Nancy Pelosi voted in favor of the Health Care Bill.”



Example querying LoD

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

In principle, all the knowledge is there:

- **GovTrack**
- **GeoNames**
- **DBPedia**
- **US Census**

But even with LoD we cannot answer this query.

Example querying LoD

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

Some missing puzzle pieces:

- Where is the data?
  - GovTrack
    GeoNames
    US Census
    requires intimate knowledge of the LoD data sets

Example querying LoD

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

Some missing puzzle pieces:

- Where is the data?
  (smart federation needed)
- Missing background (schema) knowledge.
  (enhancements of the LoD cloud)
- Crucial info still hidden in texts.
  (ontology learning from texts)
- Added reasoning capabilities (e.g., spatial).
  (new ontology language features)

Don’t get me wrong

Linked Open Data is great, useful, cool, and a very important step.

But we need to make use of the added value of formal semantics in order to advance towards the Semantic Web vision!

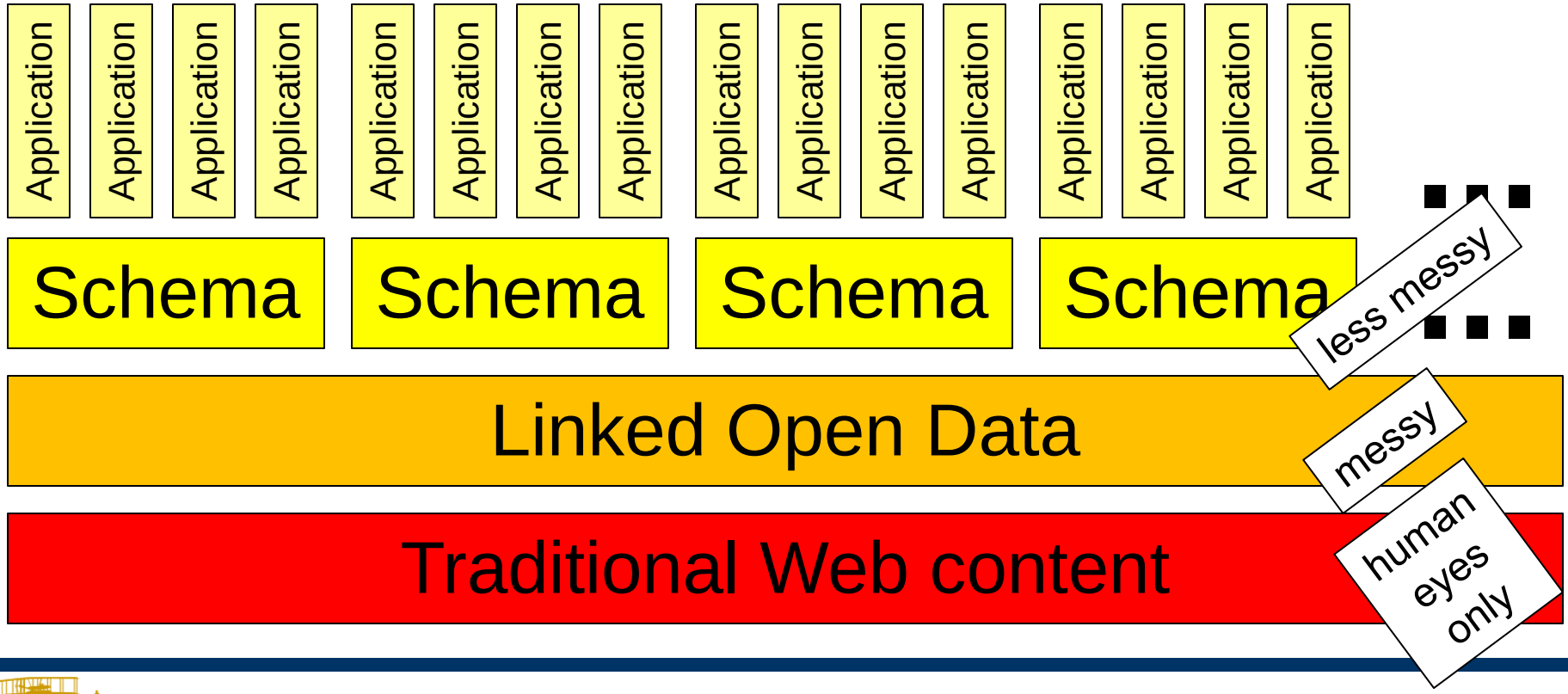
Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
- The Web of Data: Linked Data
- Querying Linked Data

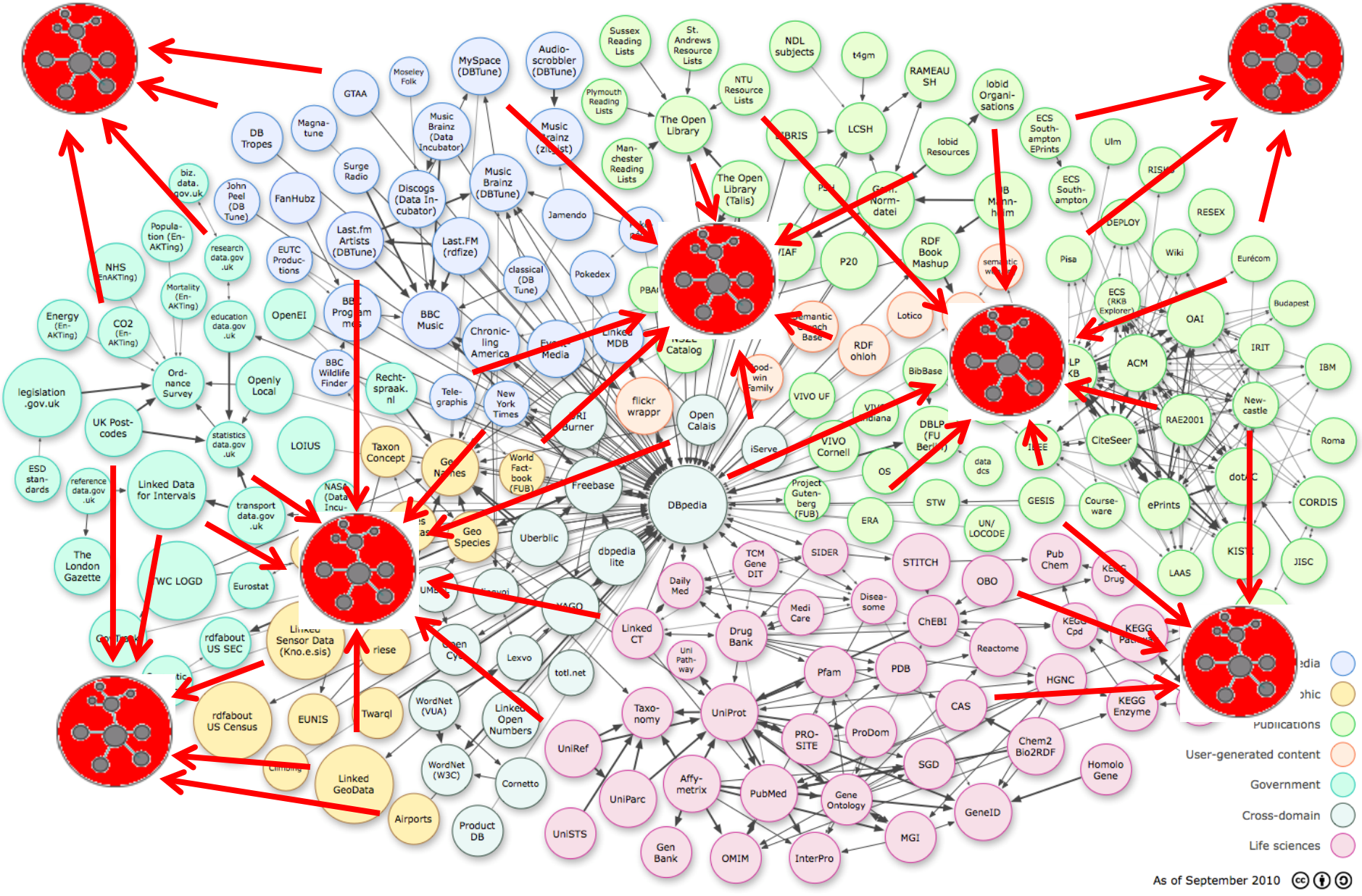
The Semantic Data Web Layer Cake

To leverage LoD, we require schema knowledge

- **application-type driven (reusable for same kind of application)**
- **less messy than LoD (as required by application)**
- **overarching several LoD datasets (as required by application)**



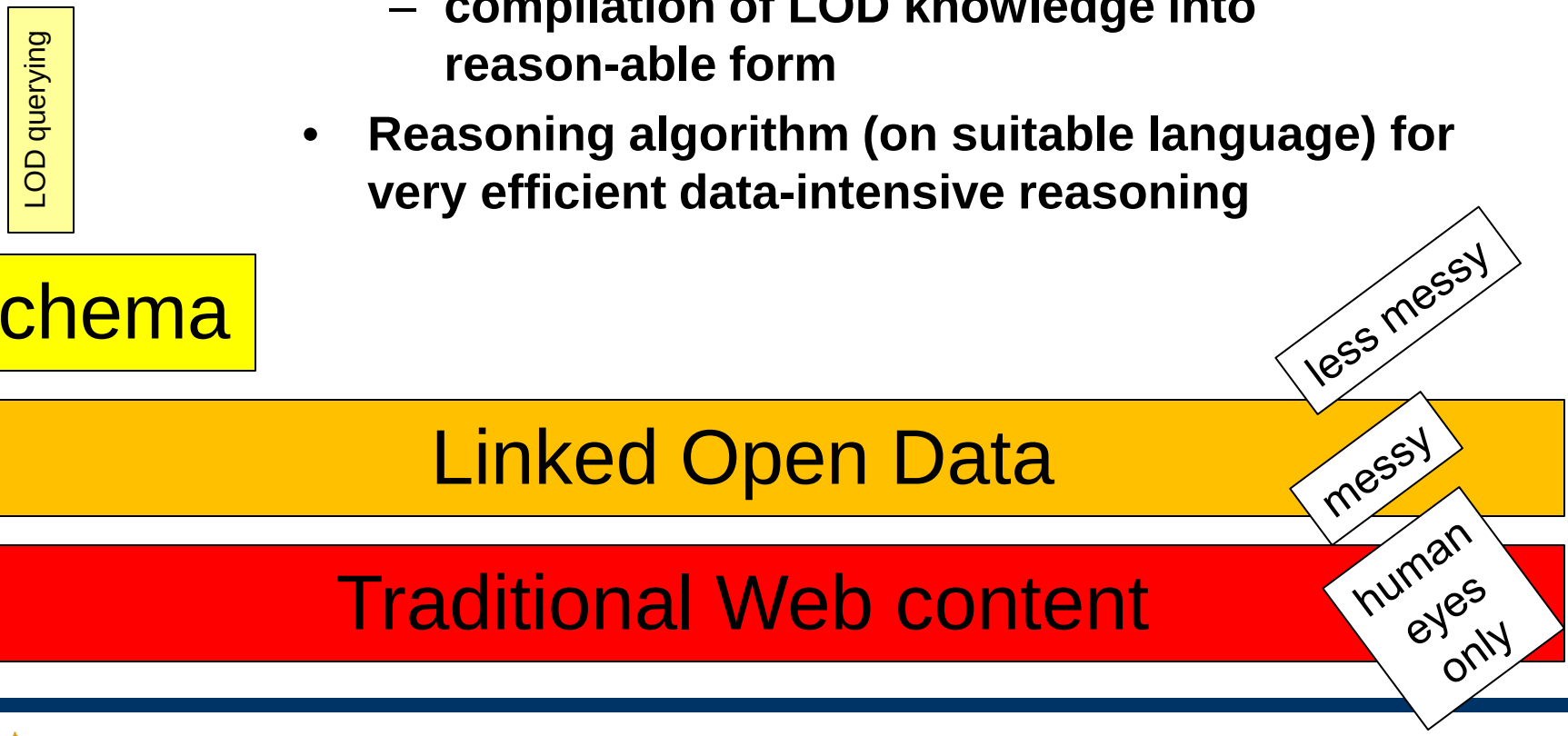
Schema on top of the LoD cloud



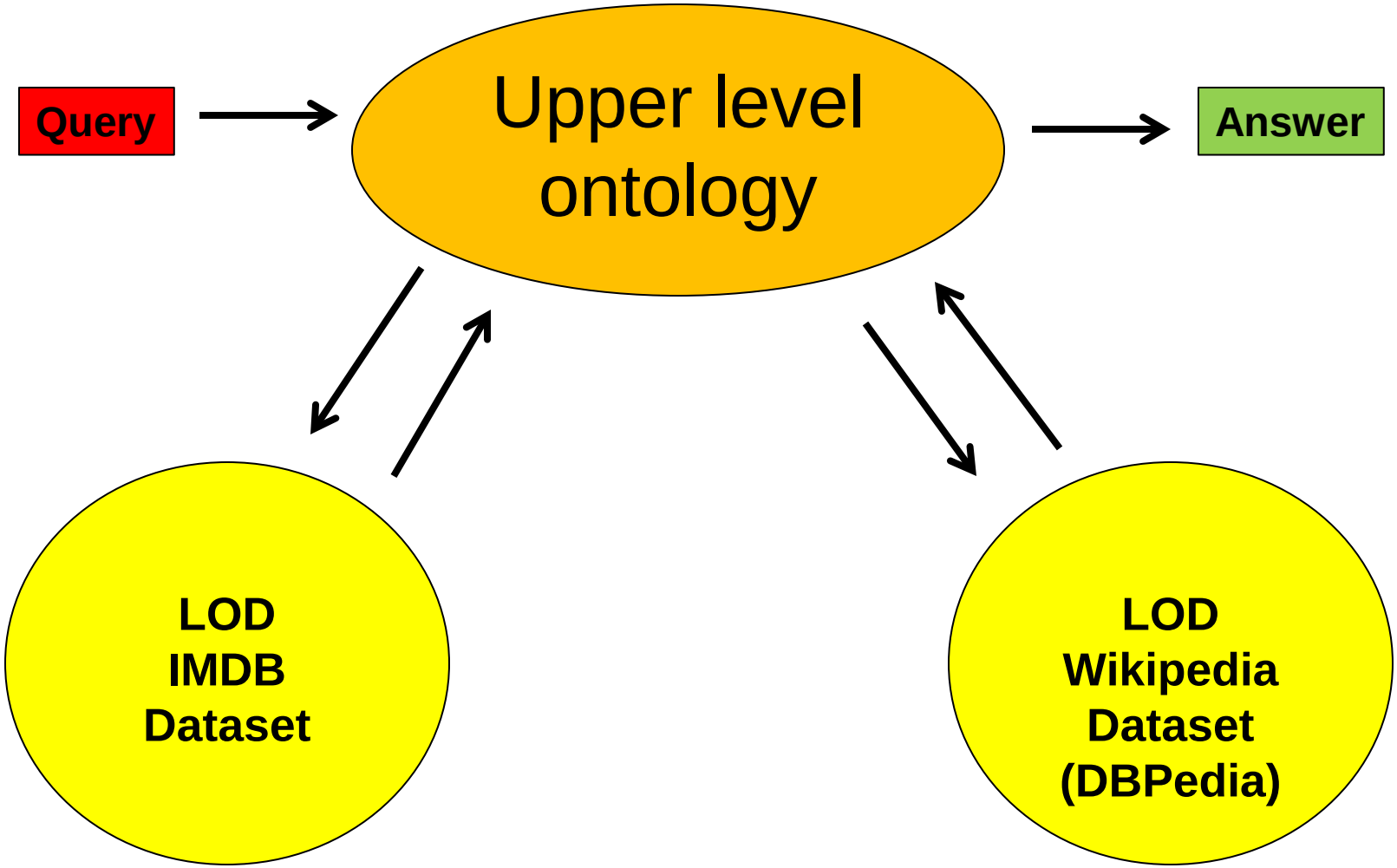
Idea – Querying Linked Open Data

Work in progress.

- **Schema creation for**
  - **query federation**
  - **utilizing background knowledge**
  - **compilation of LOD knowledge into reason-able form**
- **Reasoning algorithm (on suitable language) for very efficient data-intensive reasoning**



Idea


Query
Upper level ontology
Answer
LOD IMDB Dataset
LOD Wikipedia Dataset (DBPedia)

BLOOMS

Table 4. Results of various systems for LOD Schema Alignment. Legends: Prec=Precision, Rec=Recall, M=Music Ontology, B=BBC Program Ontology, F=FOAF Ontology, D=DBpedia Ontology, G=Geonames Ontology, S=SIOC Ontology, W=Semantic Web Conference Ontology, A=AKT Portal Ontology, err=System Error, NA=Not Available

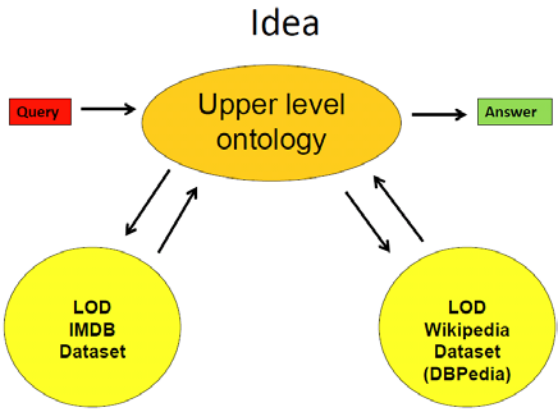
Linked Open Data Schema Ontology Alignment												
	Alignment API		OMViaUO		RiMoM		S-Match		AROMA		BLOOMS	
Test	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec
M,B	0.4	0	1	0	err	err	0.04	0.28	0	0	0.63	0.78
M,D	0	0	0	0	err	err	0.08	0.30	0.45	0.01	0.39	0.62
F,D	0	0	0	0	err	err	0.11	0.40	0.33	0.04	0.67	0.73
G,D	0	0	0	0	err	err	0.23	1	0	0	0	0
S,F	0	0	0	0	0.3	0.2	0.52	0.11	0.30	0.20	0.55	0.64
W,A	0.12	0.05	0.16	0.03	err	err	0.06	0.4	0.38	0.03	0.42	0.59
W,D	0	0	0	0	err	err	0.15	0.50	0.27	0.01	0.70	0.40
Avg.	0.07	0.01	0.17	0	NA	NA	0.17	0.43	0.25	0.04	0.48	0.54

Jain, Hitzler et al, ISWC2010

ALOQUS Illustration

"Identify films, the nations where they were shot and the population of these countries"

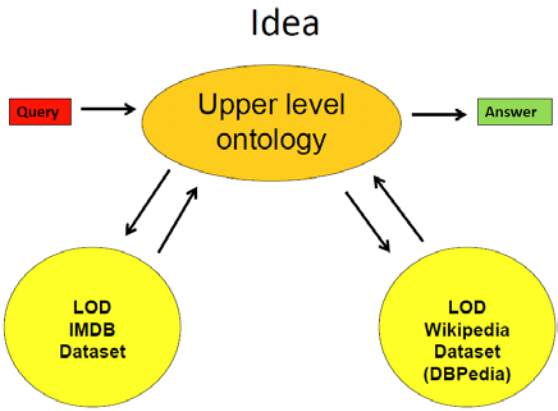
```
SELECT ?film ?nation ?pop

WHERE {
?film     protonu:ofCountry          ?nation.
?film     rdf:type                   protonu:Movie.
?film     rdfs:label                 ?film_name.
?nation protont:populationCount      ?pop.
}
```



ALOQUS Illustration – Blooms Alignment

protonu:ofCountry	maps to	lmdb:country
protonu:Movie	maps to	lmdb:film
protont:populationCount	maps to	dbprop:populationCount

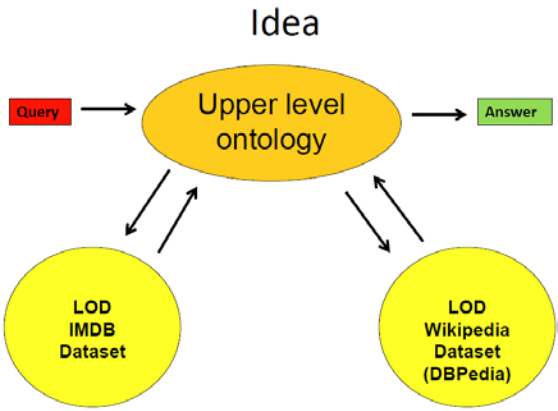
Alignment confidence > 0.9



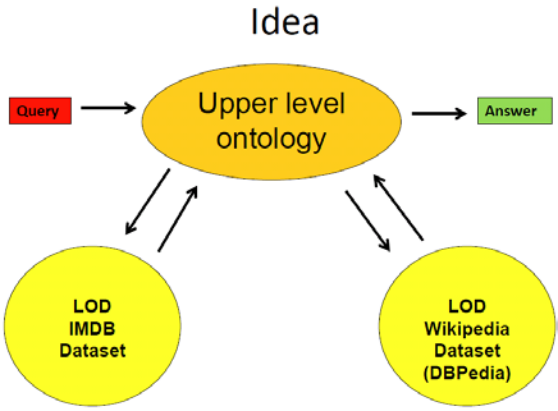
ALOQUS Illustration – Sub Queries

(a) SELECT ?film ?nation ?pop
WHERE {
?film lmdb:country ?nation.
?film rdf:type lmdb:film.
?film rdfs:label ?film_name.
}

(b) SELECT ?nation ?pop
WHERE {
?nation dbprop:populationCensus ?pop.



ALOQUS Illustration – Query Results



(a)

lmdb-film:11446	protonu:ofCountry	lmdb-country:IN.
lmdb-film:11446	rdf:type	protonu:Movie.
lmdb-film:11446	rdfs:label	"Run".
lmdb-film:17091	protonu:ofCountry	lmdb-country:LK.
lmdb-film:17091	rdf:type	protonu:Movie.
lmdb-film:17091	rdfs:label	"Getawarayo".
lmdb-film:16973	protonu:ofCountry	lmdb-country:IN.
lmdb-film:16973	rdf:type	protonu:Movie.
lmdb-film:16973	rdfs:label	"Kabeela".

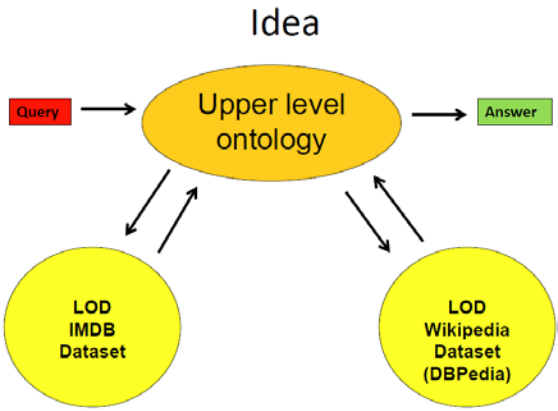
(b)

dbpedia:Sri_Lanka	protont:PopulationCount	21324791.
dbpedia:India	protont:PopulationCount	1210193422.

ALOQUS Illustration – Result I

- **With proxy identifiers**

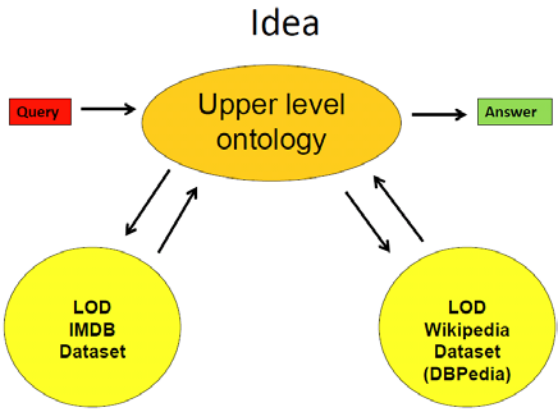
film	name	nation	population
lmdb-film:17091	"Getawarayo"	aloqus:2908ba82	21324791
lmdb-film:16973	"Kabeela"	aloqus:9bc35ca1	1210193422
lmdb-film:11446	"Run"	aloqus:9bc35ca1	1210193422



ALOQUS Illustration – Result II

- **Without proxy identifiers**

film	name	nation	population
lmdb-film:17091	"Getawarayo"	lmdb-country:LK	21324791
lmdb-film:16973	"Kabeela"	lmdb-country:IN	1210193422
lmdb-film:11446	"Run"	lmdb-country:IN	1210193422
lmdb-film:11446	"Run"	nytimes:india_geo	1210193422

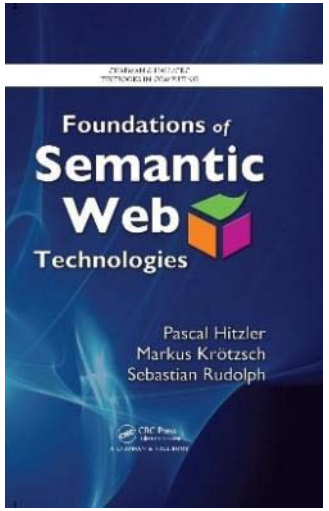


Comparison with other systems

Features	ALOQUS	DARQ	SQUIN
Approach	Uses upper level ontology(PROTON) or any other ontology as primary ontology for query serialization and execution.	Requires formal description of datasets in the form of Service Description.	Requires an initial URI to execute queries.
Query Creation	Creates query corresponding to every mapping for a concept.	Creates queries only corresponding to the concepts mentioned in the query.	Creates queries only corresponding to the concepts mentioned in the query.
Failsafe	Executes all sub-queries for multiple mappings. Hence retrieves at least partial answers if a specific endpoint doesn't work.	X	X
Detect Entity co-references	Crawls and also consumes sameAs.org webservices.	X	X
Result Processing	Query answers, retrieved from different datasets are merged and presented to user.	Retrieves answers from multiple dataset based on service description.	Retrieves answers from multiple dataset through link traversal.
Write queries using ontology not present in LOD	Yes	X	X
Support for open-ended queries like ?s ?p ?o	Yes	X	X
Result Storage for later Retrieval	Yes	X	X
DESCRIBE Query Form	Yes	N/A	Yes

Thanks!

http://www.semantic-web-book.org
http://www.semantic-web-journal.net



References

- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, Foundations of Semantic Web Technologies. CRC Press / Chapman and Hall, 2010.
- Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, Sebastian Rudolph, OWL 2 Web Ontology Language: Primer. W3C Recommendation, 27 October 2009.
- Pascal Hitzler, Frank van Harmelen, A reasonable Semantic Web. Semantic Web 1(1-2), 39-44, 2010.
- Pascal Hitzler, Towards Reasoning Pragmatics. In: Krzysztof Janowicz, Martin Raubal, Sergei Levashkin (Eds.), GeoSpatial Semantics, Third International Conference, GeoS 2009, Mexico City, Mexico, December 3-4, 2009. Proceedings. Lecture Notes in Computer Science Vol. 5892 Springer 2009, pp. 9-25.

References

- http://linkeddata.org
- Tim Berners-Lee: http://www.w3.org/DesignIssues/LinkedData.html
- Christian Bizer, Tom Heath, Tim Berners-Lee: Linked Data – The Story So Far. Int. J. Semantic Web Inf. Syst. 5(3): 1-22 (2009)
- Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, Linked Data is Merely More Data. In: Dan Brickley, Vinay K. Chaudhri, Harry Halpin, Deborah McGuinness: Linked Data Meets Artificial Intelligence. Technical Report SS-10-07, AAAI Press, Menlo Park, California, 2010, pp. 82-86. ISBN 978-1-57735-461-1. Proceedings of LinkedAI at the AAAI Spring Symposium, March 2010.

References

- Prateek Jain, Pascal Hitzler, Amit P. Sheth, Kunal Verma, Peter Z. Yeh, Ontology Alignment for Linked Open Data. In P. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Pan, I. Horrocks, B. Glimm (eds.), The Semantic Web - ISWC 2010. 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I. Lecture Notes in Computer Science Vol. 6496. Springer, Berlin, 2010, pp. 402-417.
- Amit Krishna Joshi, Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, Mariana Damova, Alignment-based Querying of Linked Open Data. Submitted.
- Prateek Jain, Pascal Hitzler, Kunal Verma, Peter Yeh, Amit Sheth, Moving beyond sameAs with PLATO: Partonomy detection for Linked Data. In: Proceedings of Hypertext 2012, the 23rd ACM Conference on Hypertext and Social Media, Milwaukee, WI, USA, June 2012. To appear.