Towards Reasoning Pragmatics: State of the Art and Vision for the Semantic Web

Pascal Hitzler

Kno.e.sis Center

Wright State University, Dayton, OH

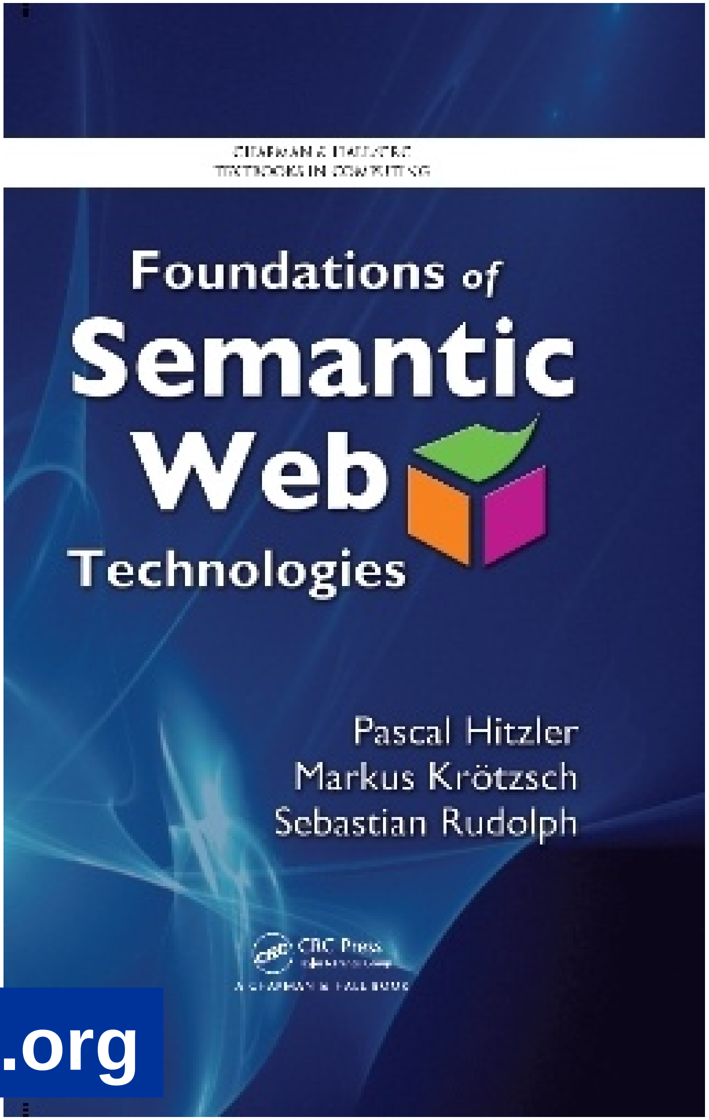
http://www.knoesis.org/pascal/

Book Announcement

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies
Chapman & Hall/CRC, 2009

Grab a flyer!



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

The Past

Remember?

Tim Berners-Lee, James Hendler and Ora Lassila

The Semantic Web

Scientific American, May 17, 2001

- **Talks explicitly about knowledge representation and logic as required ingredient.**

W3C Semantic Web Page, January 2001

"The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people. For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently. The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications."

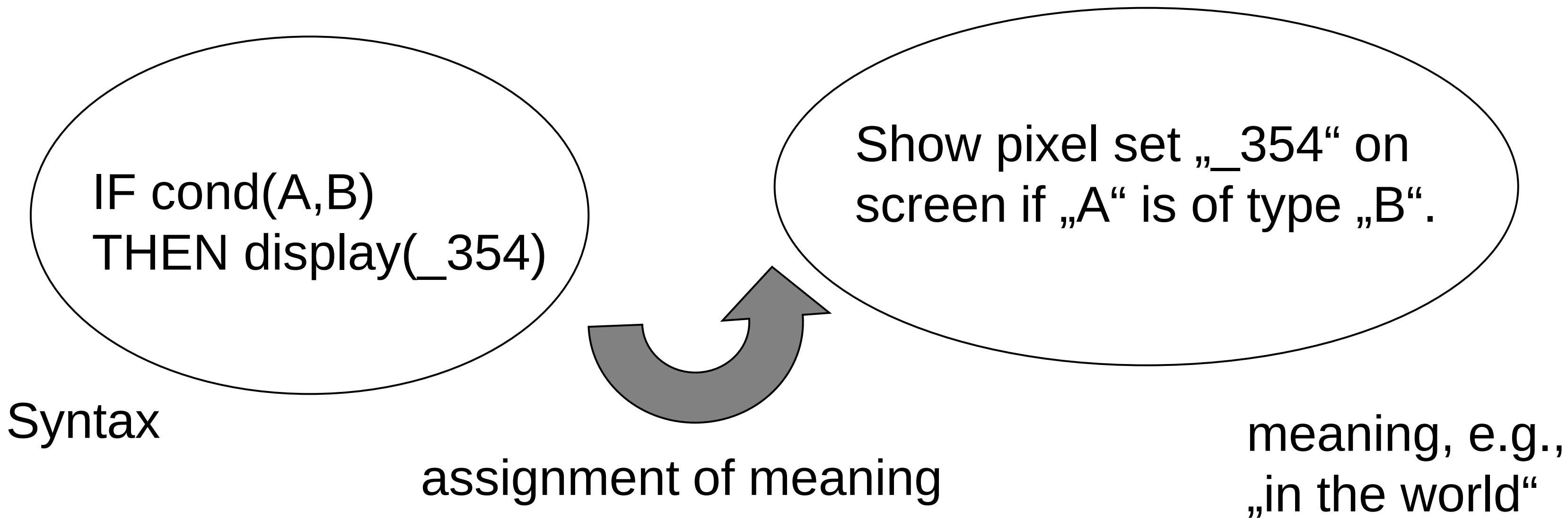
Formal Semantics

- RDF as of 2001 had no formal semantics.
- What actually is semantics? What is formal semantics?

Syntax and Semantics

Syntax: character strings without meaning

Semantics: meaning of the character strings



Semantics of Programming Languages

computing factorial

Intended Semantics

Syntax

```
FUNCTION f(n:natural):natural;
           BEGIN
           IF n=0 THEN f:=1
           ELSE f:=n*f(n-1);
           END;
```

$f : n \mapsto n!$

Formal Semantics

What happens at program execution

Procedural Semantics

So …

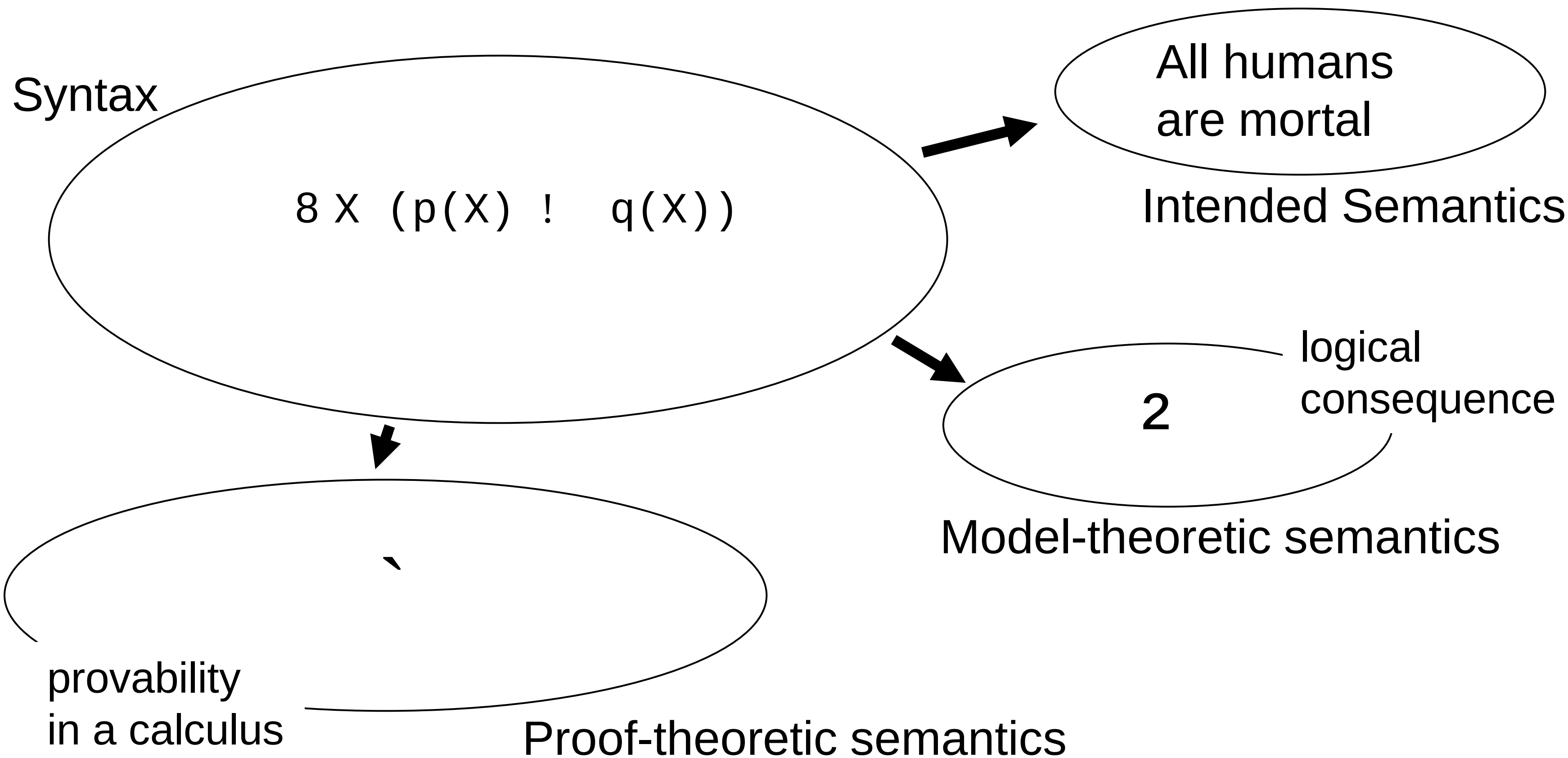
- **Semantics tells us more about something than meets the eye.**
- **Semantics gives access to implicit knowledge.**
- **Semantics helps to focus on the implicit knowledge, and abstracts from concrete representations.**
  - **[there's always more than one way to code something]**

Semantics for the Semantic Web

- **Need semantics for data (not for programs).**
- **How to define semantics? How to encode data?**

- **Idea: Draw on the long history of formal logic and symbolic AI / Knowledge Representation and Reasoning.**
  - **Logic-based formalisms**
  - **Model-theoretic semantics**

Semantics of Logic



So what happened?

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

- The hype started a bit earlier, actually.

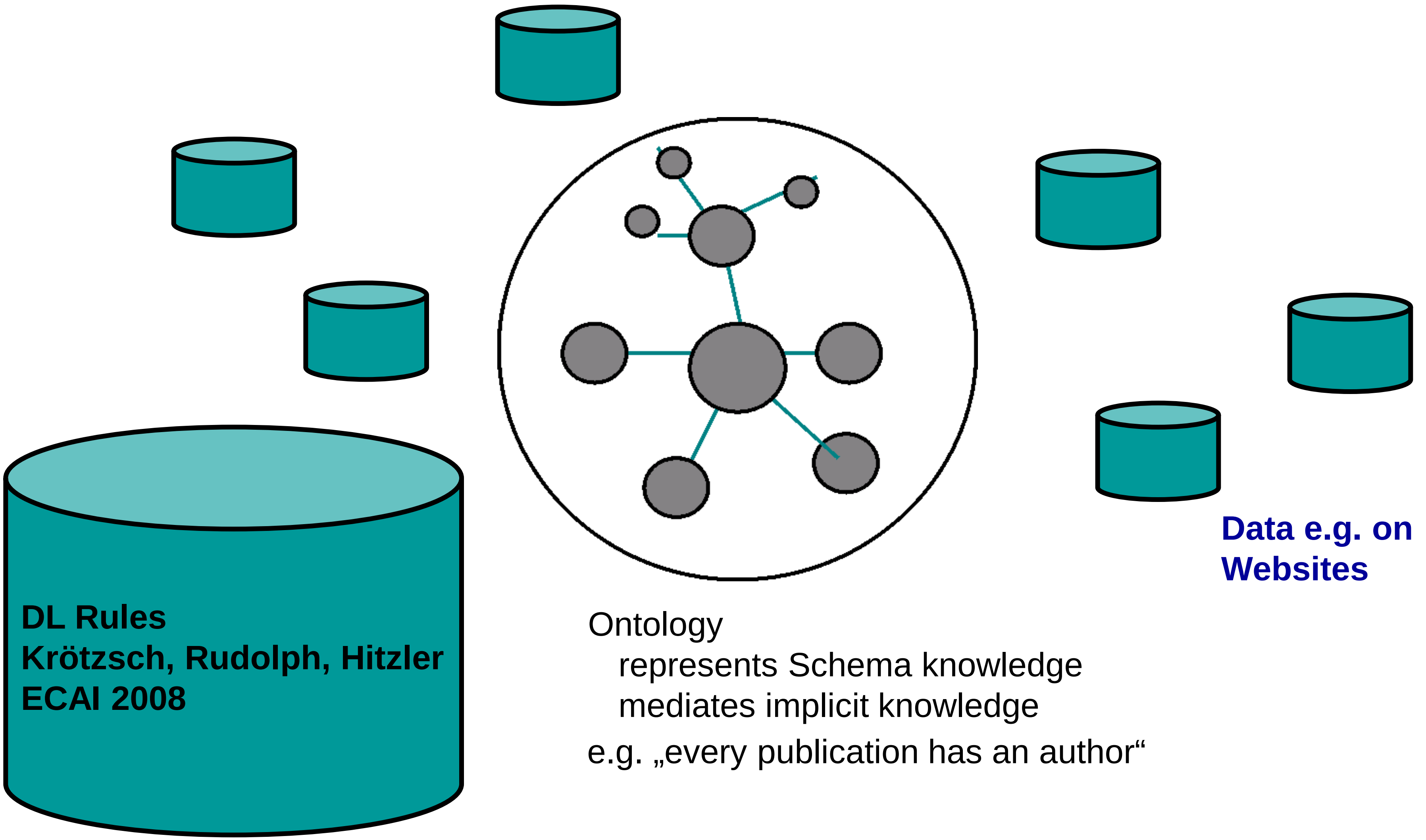
Very brief history of the Semantic Web

- idea as old as 1989.
- 1990s: W3C metadata activity (lead to RDF(S))
- **W3C semantic web activity: chartered 2001.**
- **SciAm article: 2001**

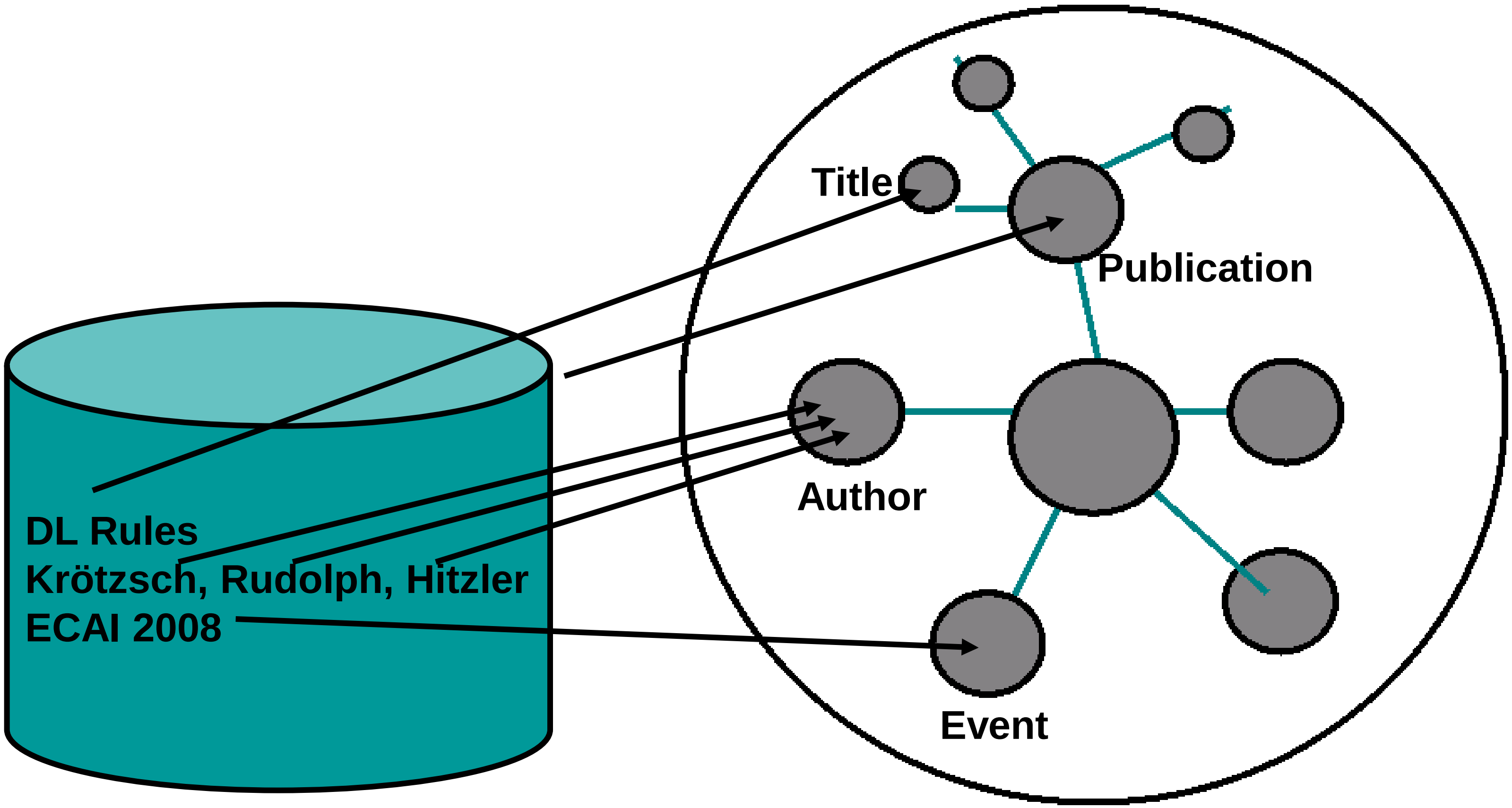
- **USA: DAML-Programme 2000-2005 approx. 70M€.**
- **Many large scale EU projects since 2002 and ongoing.**
  → FP6

- **Now funding mostly application oriented (EU FP7, US NIH)**

Basic Idea of the Semantic Web

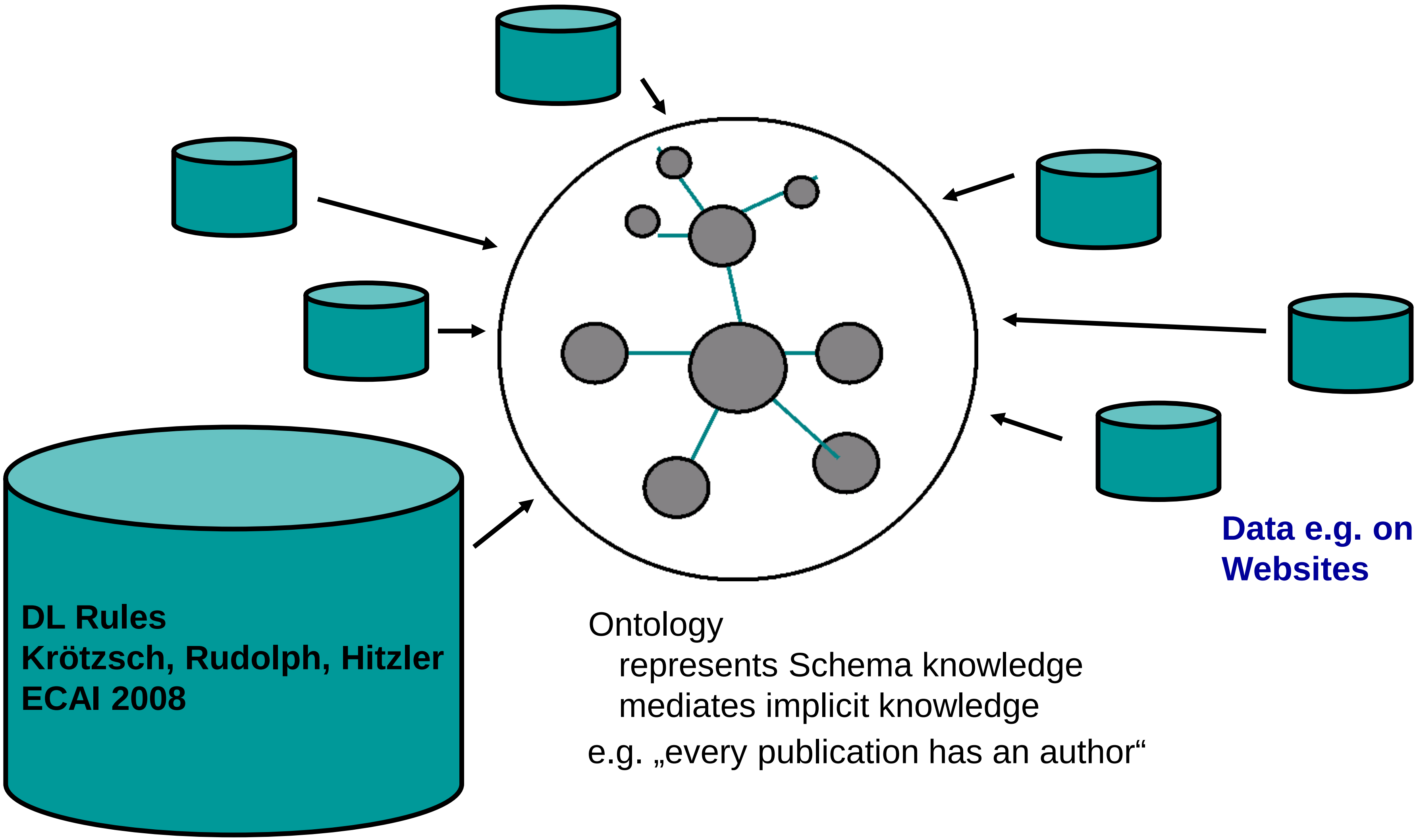


Basic Idea of the Semantic Web



e.g. „every publication has an author“

Basic Idea of the Semantic Web



W3C Semantic Web Page, August 2008

"The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects."

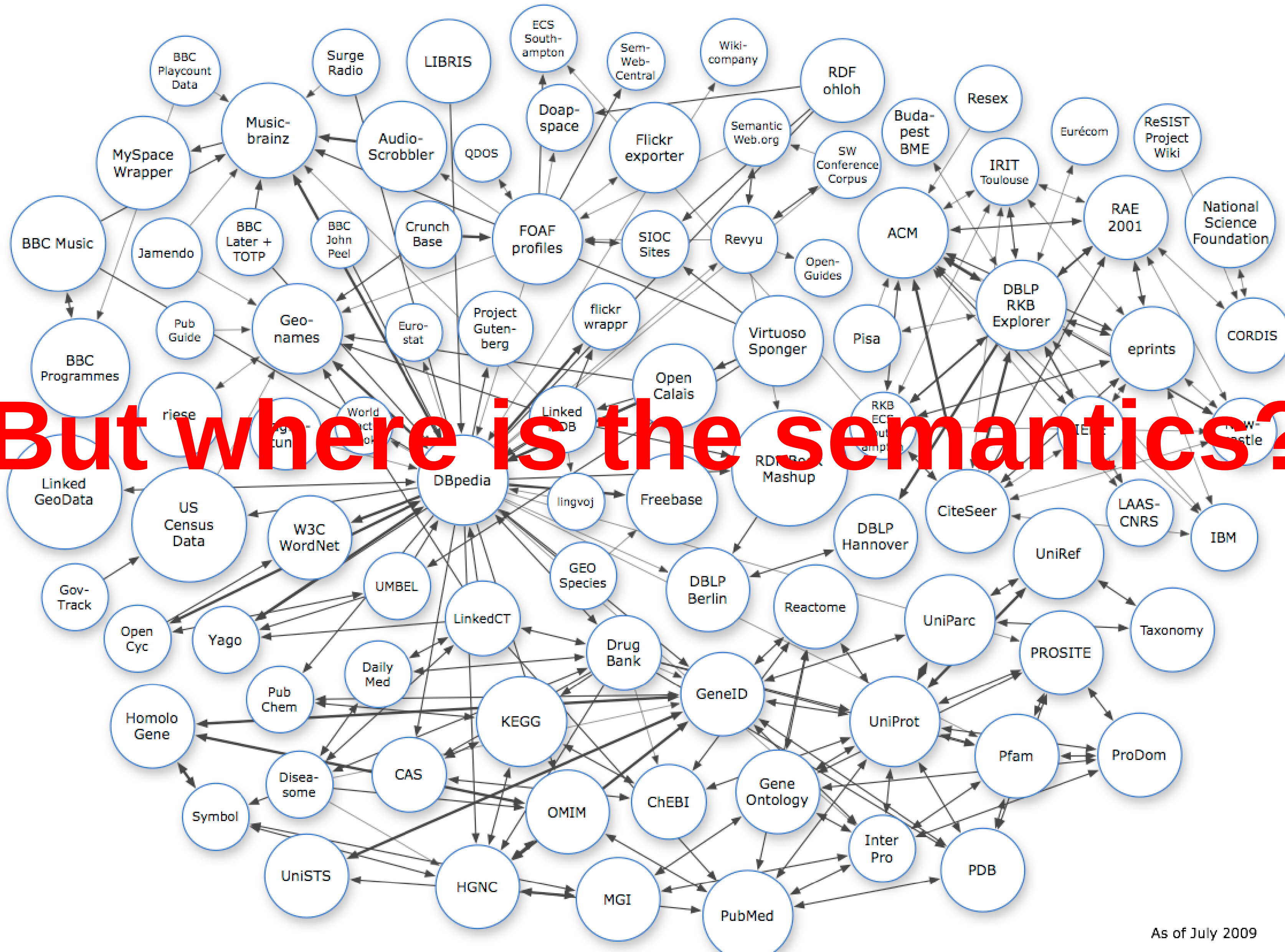
New W3C Semantic Web Page (current)

The new buzzword: Linked Data

Five Aspects mentioned:

- **Linked Data:**
  “The Semantic Web is a Web of data”
- **Vocabularies:**
  OWL, SKOS – “enrich data with additional meaning”
- **Query:**
  “If the Semantic Web is viewed as a global database ...”
- **Inference:**
  “discovering new relationships”
- **Vertical Applications:**
  “innovation adoption through Semantic Web technology”

Linked Open Data



But where is the semantics?

As of July 2009

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-o administrative d	
695	P.PPLA2	seat of a second 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?

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

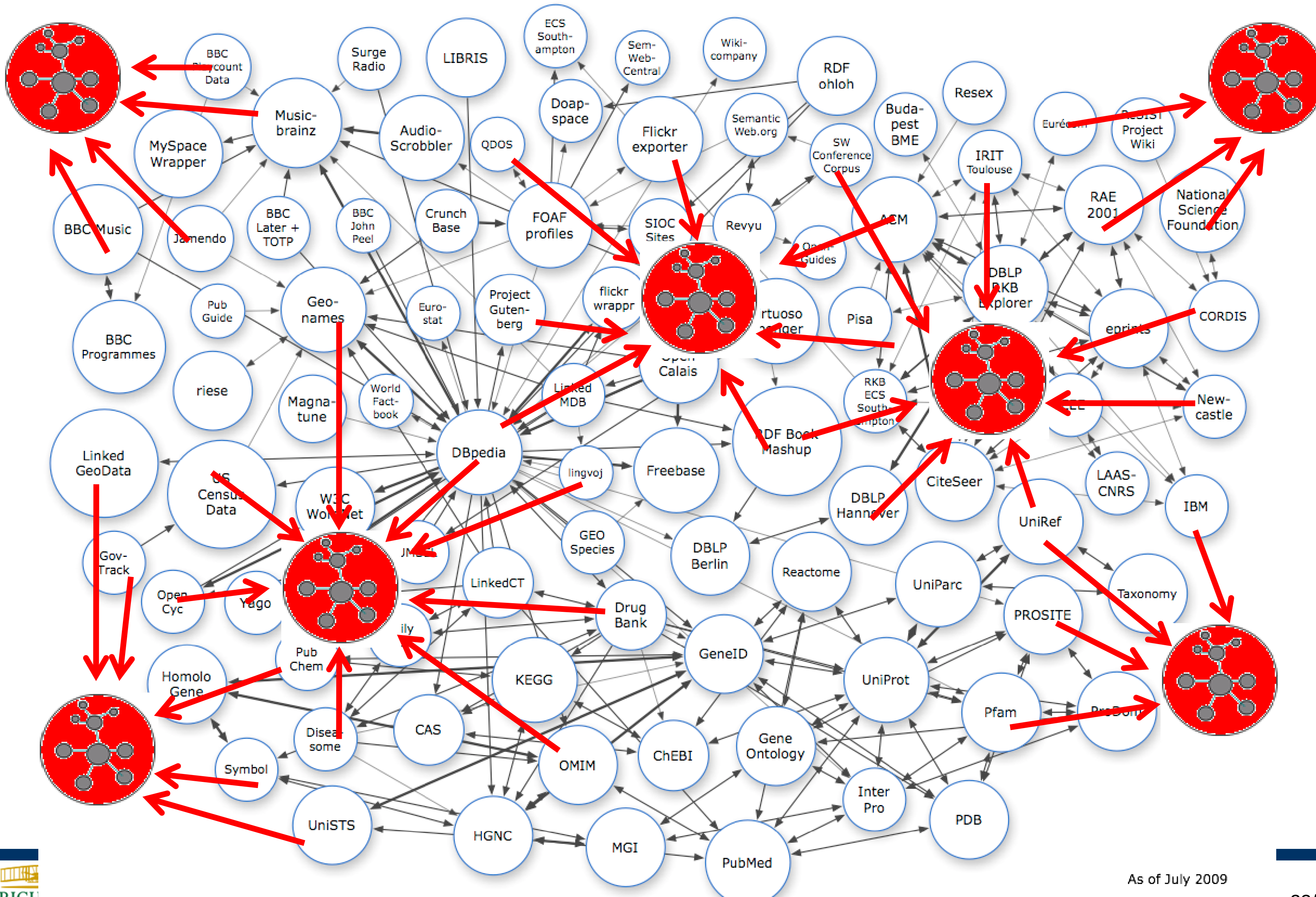
Example querying LoD

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

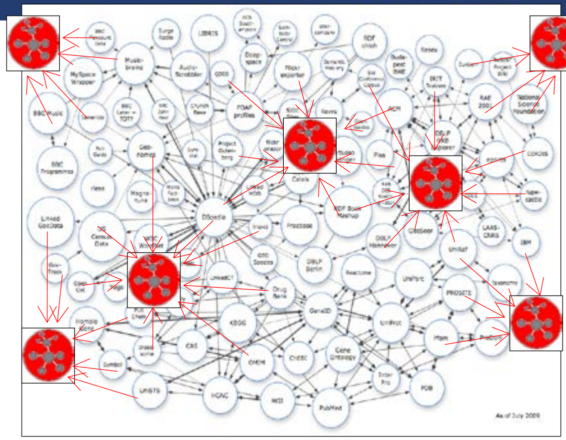
Some missing puzzle pieces:

- **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)

How to get there



How to get there



- Schema ontologies
  - made for specific purposes (e.g., querying)
  - spanning several LoD datasets
  - incorporating schema knowledge hidden in the LoD datasets
  - including additional background knowledge needed for design purpose
- Added reasoning capabilities extending OWL as needed.
  - rules
  - extended datatypes
  - spatial and temporal reasoning etc.
- Making use of ontology lifecycle state-of-the-art tools
  - ontology evaluation
  - ontology learning from texts
  - ontology evolution etc.

From no semantics to low semantics

1. **Take a no-semantics or low-semantics solution.**
   E.g., naive LoD querying using SPARQL.

2. **Identify where added value could be obtained by formal semantics.**
   E.g., by using schema knowledge as query entry points; by using schema knowledge to get better answers.

3. **Identify (or develop!) ontology language which has the required features (→ really interesting research!).**
   E.g., spatial reasoning.

4. **Realize application and publish (additional) data as LoD data.**

Important: **Keep it simple, stupid!**
A little semantics can go a long way.

Take-Home Message

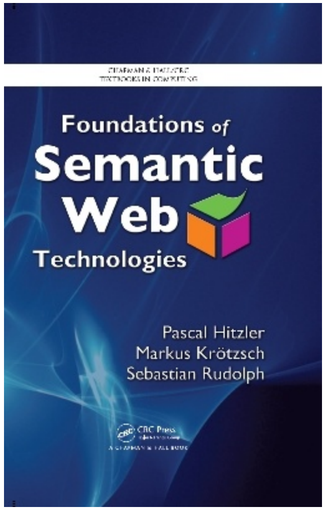
Metadata without formal semantics is simply more data.

Get added value
from using formal semantics / access to implicit knowledge.

Lift your applications *carefully* to the use of deeper semantics.

... and thanks to Prateek Jain for the LoD querying example ...

Thanks!


Foundations of
Semantic
Web
Technologies
Pascal Hitzler
Markus Krötzsch
Sebastian Rudolph

References

- **Krzysztof Janowicz, Pascal Hitzler, *The Digital Earth as Knowledge Engine*. Semantic Web 3 (3), 213-221, 2012.**
- **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.**
- **Pascal Hitzler, Frank van Harmelen, *A reasonable Semantic Web*. Semantic Web 1(1-2), 39-44, 2010.**
- **Pascal Hitzler, Krzysztof Janowicz, *What's Wrong with Linked Data?* http://blog.semantic-web.at/2012/08/09/whats-wrong-with-linked-data/ , August 2012.**
- **Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press, 2009.**

References

- Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, Sebastian Rudolph, OWL 2 Web Ontology Language: Primer. W3C Recommendation, 27 October 2009.
- 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.
- Prateek Jain, Pascal Hitzler, Kunal Verma, Peter Yeh, Amit Sheth, Moving beyond sameAs with PLATO: Partonomy detection for Linked Data. In: Ethan V. Munson, Markus Strohmaier (Eds.): 23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012. ACM, 2012, pp. 33-42.

References

- Amit Krishna Joshi, Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, Mariana Damova, Alignment-based Querying of Linked Open Data. In: Meersman, R.; Panetto, H.; Dillon, T.; Rinderle-Ma, S.; Dadam, P.; Zhou, X.; Pearson, S.; Ferscha, A.; Bergamaschi, S.; Cruz, I.F. (eds.), On the Move to Meaningful Internet Systems: OTM 2012, Confederated International Conferences: CoopIS, DOA-SVI, and ODBASE 2012, Rome, Italy, September 10-14, 2012, Proceedings, Part II. Lecture Notes in Computer Science Vol. 7566, Springer, Heidelberg, 2012, pp. 807-824.
- Prateek Jain, Pascal Hitzler, Kunal Verma, Peter Yeh, Amit Sheth, Moving beyond sameAs with PLATO: Partonomy detection for Linked Data. In: Ethan V. Munson, Markus Strohmaier (Eds.): 23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012. ACM, 2012, pp. 33-42.

References

- Kunal Sengupta, Adila Krisnadhi, Pascal Hitzler, Local Closed World Reasoning: Grounded Circumscription for OWL. In: L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. F. Noy, E. Blomqvist (Eds.): The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I. Lecture Notes in Computer Science Vol. 7031, Springer, Heidelberg, 2011, pp. 617-632.
- Prateek Jain, Peter Z. Yeh, Kunal Verma, Reymonrod G. Vasquez, Mariana Damova, Pascal Hitzler, Amit P. Sheth, Contextual Ontology Alignment of LOD with an Upper Ontology: A Case Study with Proton. In: Grigoris Antoniou, Marko Grobelnik, Elena Paslaru Bontas Simperl, Bijan Parsia, Dimitris Plexousakis, Pieter De Leenheer, Jeff Pan (Eds.): The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29-June 2, 2011, Proceedings, Part I. Lecture Notes in Computer Science 6643, Springer, 2011, pp. 80-92.