Knowledge Representation for the Semantic Web

Winter Quarter 2012

Slides 1 – 01/03/2012

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

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

About me

- 1998 Diplom (Master) in Mathematics, Tübingen, Germany
- 2001 PhD in Mathematics, Cork, Ireland
- 2001-2004 Postdoc in Artificial Intelligence, Dresden, Germany
- 10-12.2003 Research Associate CWRU Cleveland OH
- 2004-2009 Assistant Prof. in Appl. Comp. Sci., Karlsruhe, Germany
- since September 2009: Assistant Prof., Kno.e.sis Center at WSU

- Active Semantic Web researcher – this course is a basic introduction to my current core research area.

My expectations

- **PhD students beyond the Master do not require teaching through courses: If you're pursuing a PhD you should be able to learn all by yourself those things which you need to learn.**

- **Still, learning through courses can be helpful, and can in particular provide shortcuts to things you need.**

- **And it's rarely wrong to acquire broad knowledge.**

- **I expect:**
  That you take responsibility yourself, and decide for yourself how much work you want to put into this course.
  Regretfully, however, I will have to give grades at the end ...

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

Who are you?

- **Which year?**
- **Master or PhD?**
- **Which specialization area?**
- **Why are you here?**
  - **topic relates to my specialization area;**
  - **topic sounds interesting;**
  - **need another course and it doesn't matter which;**
  - **not sure if I'll stay;**
  - **supervisor told me to come;**
  - **a friend dragged me along;**
  - **isn't this the Algebraic Topology class?**
  - **...**

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

The WWW is penetrating our society

- Social contacts (social networking platforms, blogging, ...)
- Economics (buying, selling, advertising, ...)
- Administration (eGovernment)
- Education (eLearning, Web as information system, ...)
- Work life (information gathering and sharing)
- Recreation (games, role play, creativity, ...)

The current 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.

Limitations of the current Web

- Too much information with too little structure
  and made for human consumption
  - Content search is very simplistic
  - future requires better methods
- Web content is heterogeneous
  in terms of content
  in terms of structure
  in terms of character encoding
  - future requires intelligent information integration
- Humans can derive new (implicit) information from given pieces of information
  but on the current Web we can only deal with syntax
  - requires automated reasoning techniques

Examples

- **Find that landmark article on data integration written by an Indian researcher in the 1990s.**
  [If you manage this without knowing the answer, let me know how you did it.]

- **Are lobsters spiders?**
  [This is getting easier these days, but was impossible a few years ago. It still needs finding and integrating over different websites, as well as some background knowledge.]

- **Which car is called a “duck” in German?**
  [This needs some intelligent integration of content from different websites plus background knowledge.]

Another example

"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 required knowledge is on the Web – most of it even in machine-readable form.

However, without automated processing and reasoning we cannot obtain a useful answer.

Basic ingredients for the Semantic Web

- **Open Standards for describing information on the Web**
- **Methods for obtaining further information from such descriptions**

We'll talk about these matters in this course.

Basic ingredients for the Semantic Web

- **Methods for obtaining further information from such descriptions**

Main approach: Logical deduction (aka automated reasoning)

E.g.,

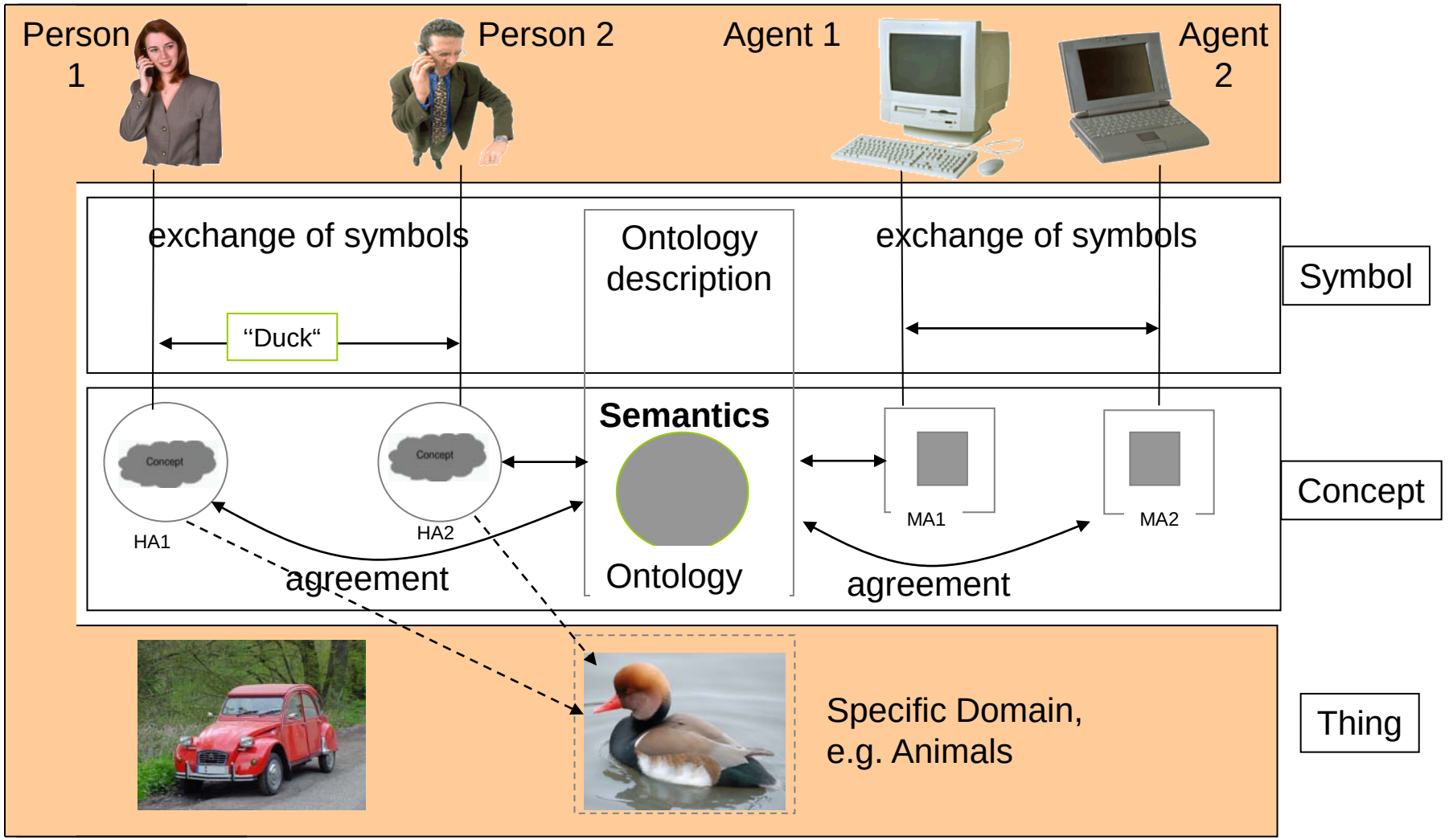
D.C. is a capital

Every capital is a city

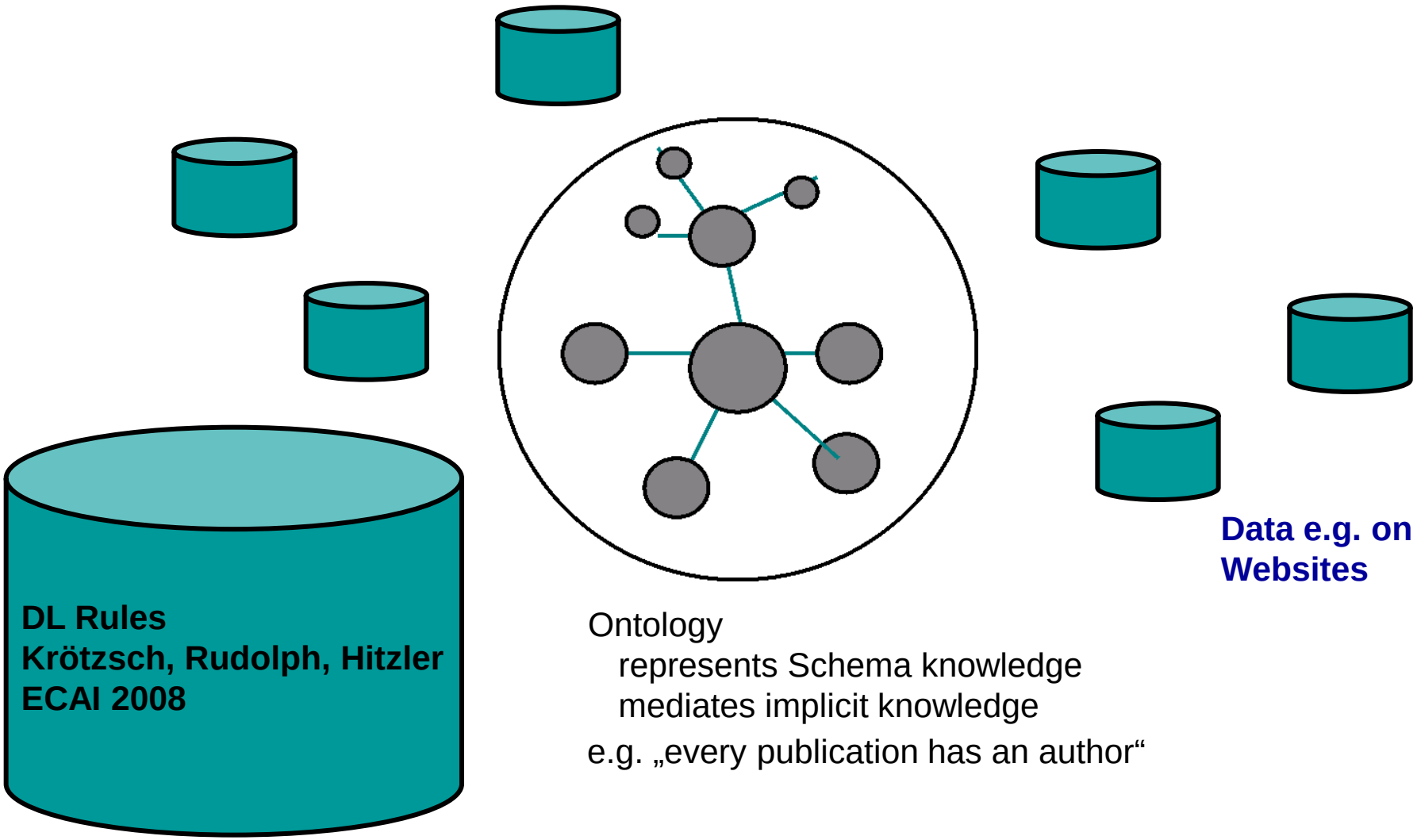
Hence: D.C. is a city

Based on predicate logic. – it needs to be specified which conclusions are valid. Plus, we need algorithms for these.

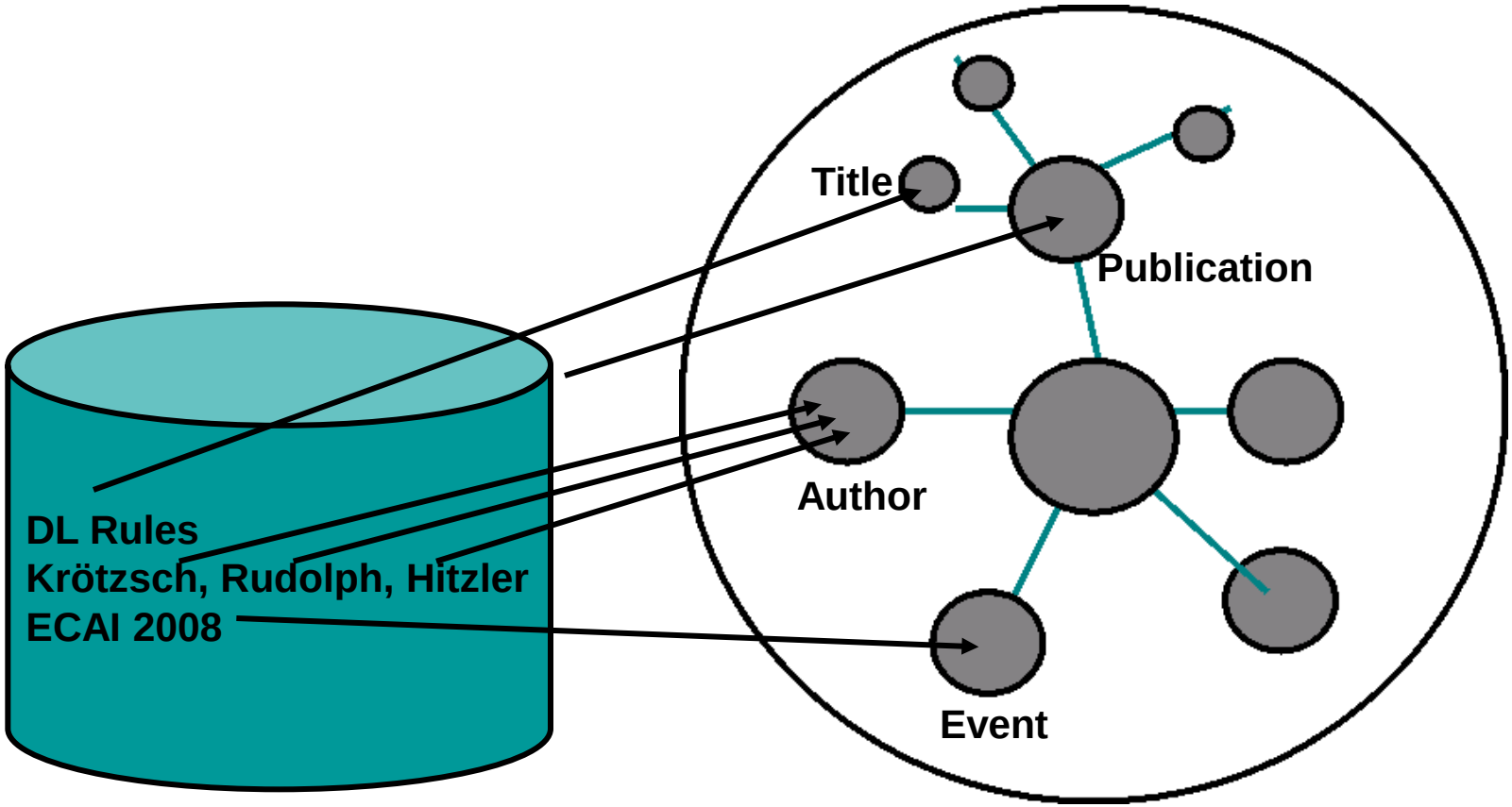
Basic Idea of the Semantic Web



Basic Idea of the Semantic Web

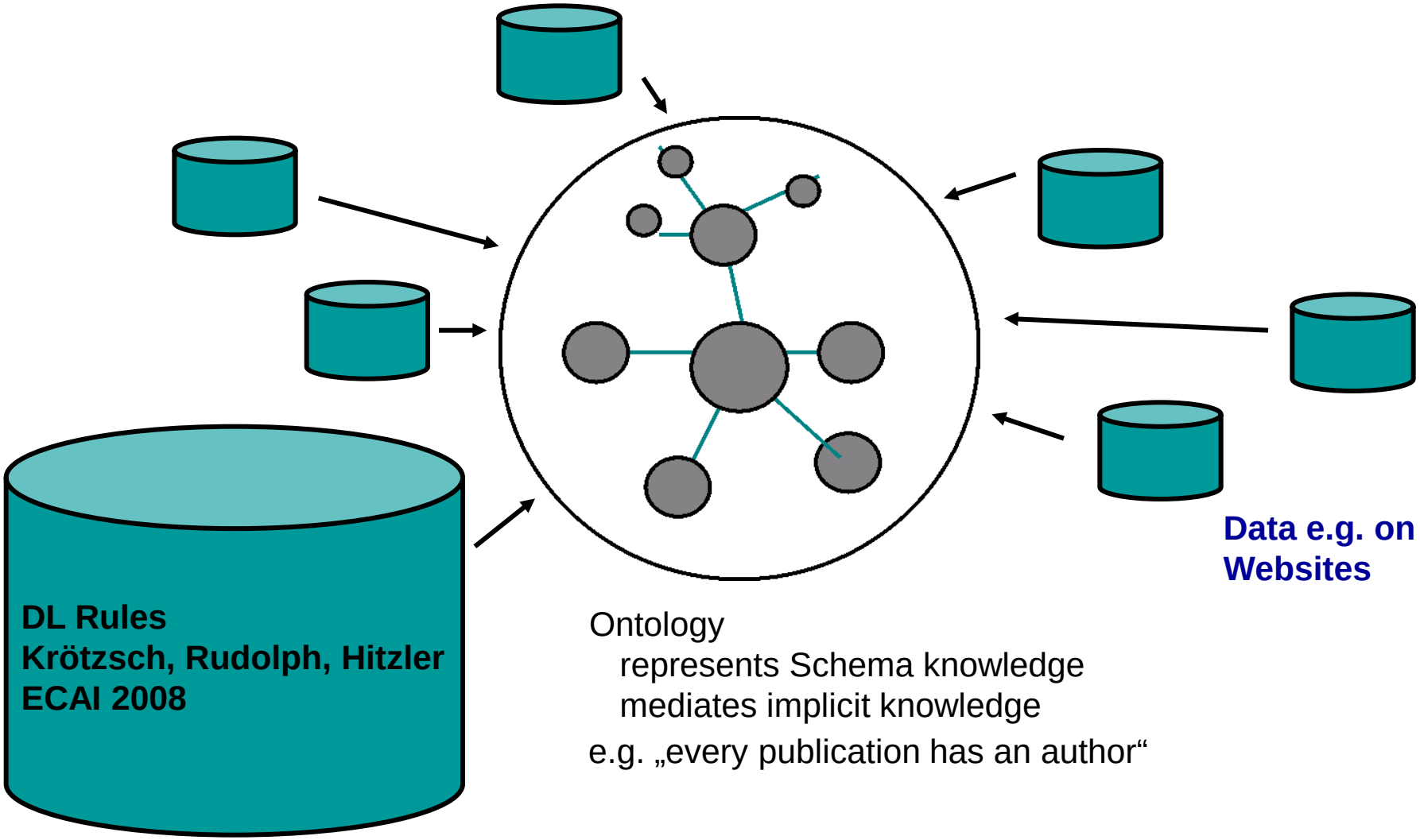


Basic Idea of the Semantic Web



e.g. „every publication has an author“

Basic Idea of the Semantic Web



Ontology
- represents Schema knowledge
- mediates implicit knowledge

e.g. „every publication has an author“

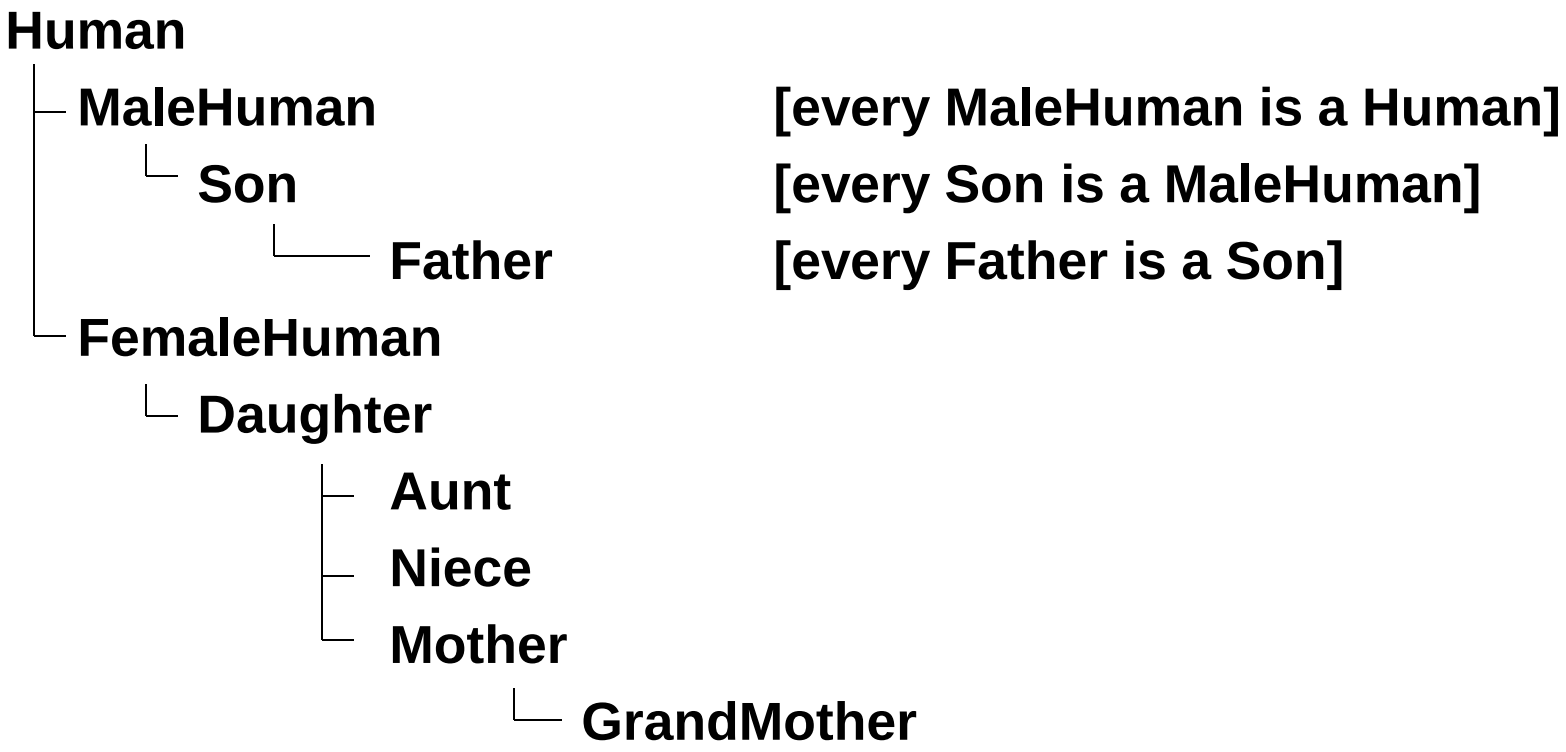
Ontology languages

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

Expressivity Scalability

Ontologies

- The core of an ontology is usually a *taxonomy*:
  - classes of things, arranged in a hierarchy



Ontologies

- **We will talk more (in fact, a lot) about ontologies**

- **For now, let's focus on taxonomies**

Partonomy vs. Taxonomy

- The following is a *partonomy* (and *not* a taxonomy):
  - classes of things, arranged in a hierarchy of "part-of" relationships

America
- LatinAmerica
  - SouthAmerica
    - Brazil [Brazil is *part of* South America]
- NorthAmerica
  - USA
    - Indiana
    - Florida
    - Ohio
      - GreeneCounty

Partonomy vs. Taxonomy

- **Partonomy:**

  A *is part of* B

  hand is part of body
  Germany is part of Europe
  Wing is part of aircraft
  Engine is part of car

- **Taxonomy**

  ***every* A *is a* B**

  every father is a man
  every dog is a mammal
  every bottle is a container
  every arm is a limb

Very brief history of the Semantic Web

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

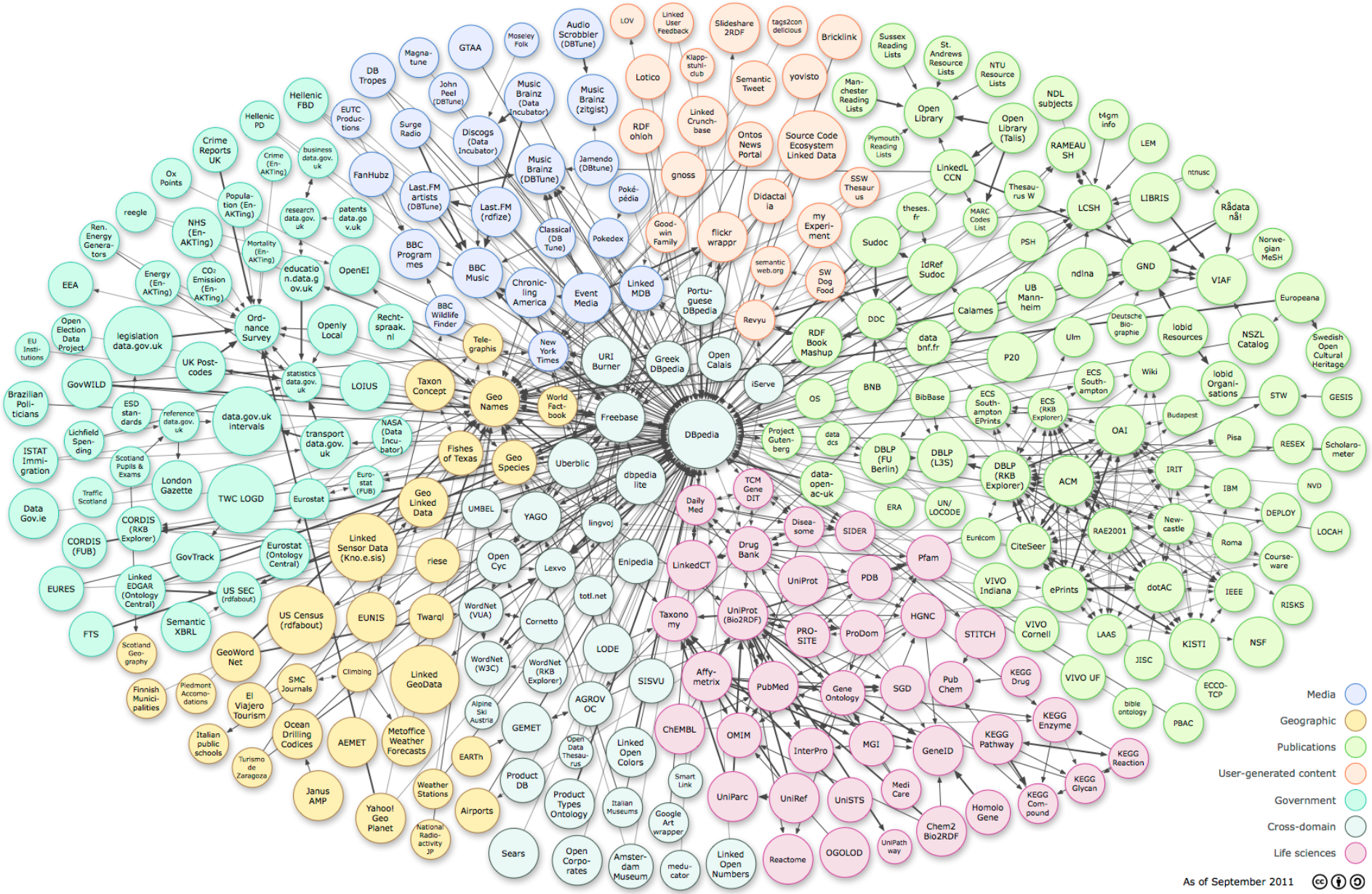
- USA: DAML-Programme 2000-2005
  approx. $90M.
- Many large scale EU projects since 2002 and ongoing.
  → FP6/FP7
- Major IT companies and
  venture capital now investing.

Semantic Technologies in the US

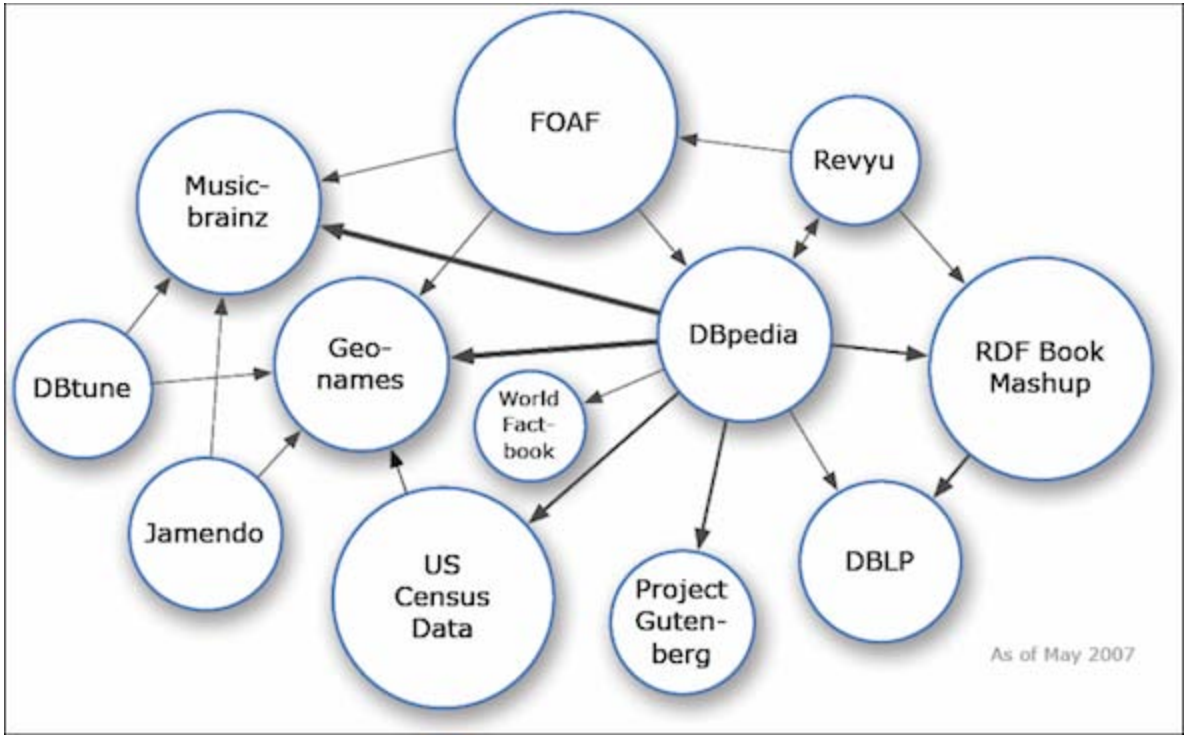
- **Funding available e.g. via**
  - **NIH**
  - **NSF**
  - **DoD, DoE, AFRL**
  - **IARPA, DARPA**
  - **...**

- **Considerable industrial take-up**
  - **Annual Semantic Technology Conference in CA**
    Taylored towards industry
  - **Major IT players (Oracle, IBM, HP, Accenture, Siemens) invest**
  - **Major government contractors (BBN, Lockheed, ...)**
  - **Venture capital (e.g. Vulcan, Inc.).**
  - **Structured data on the Web (BBC, nytimes.com, data.gov,...)**

Linked Open Data

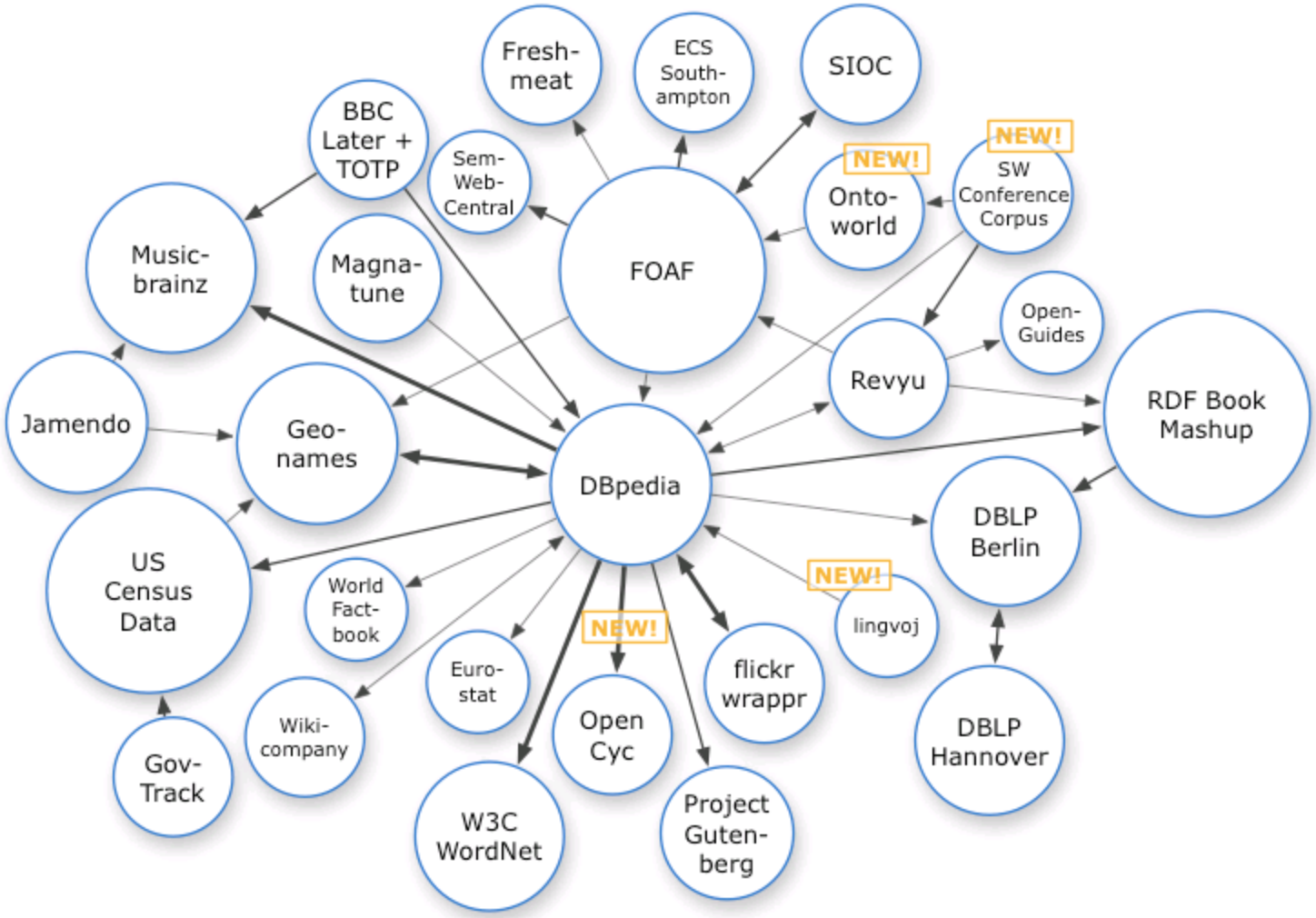


Linked Open Data 2007 (May)

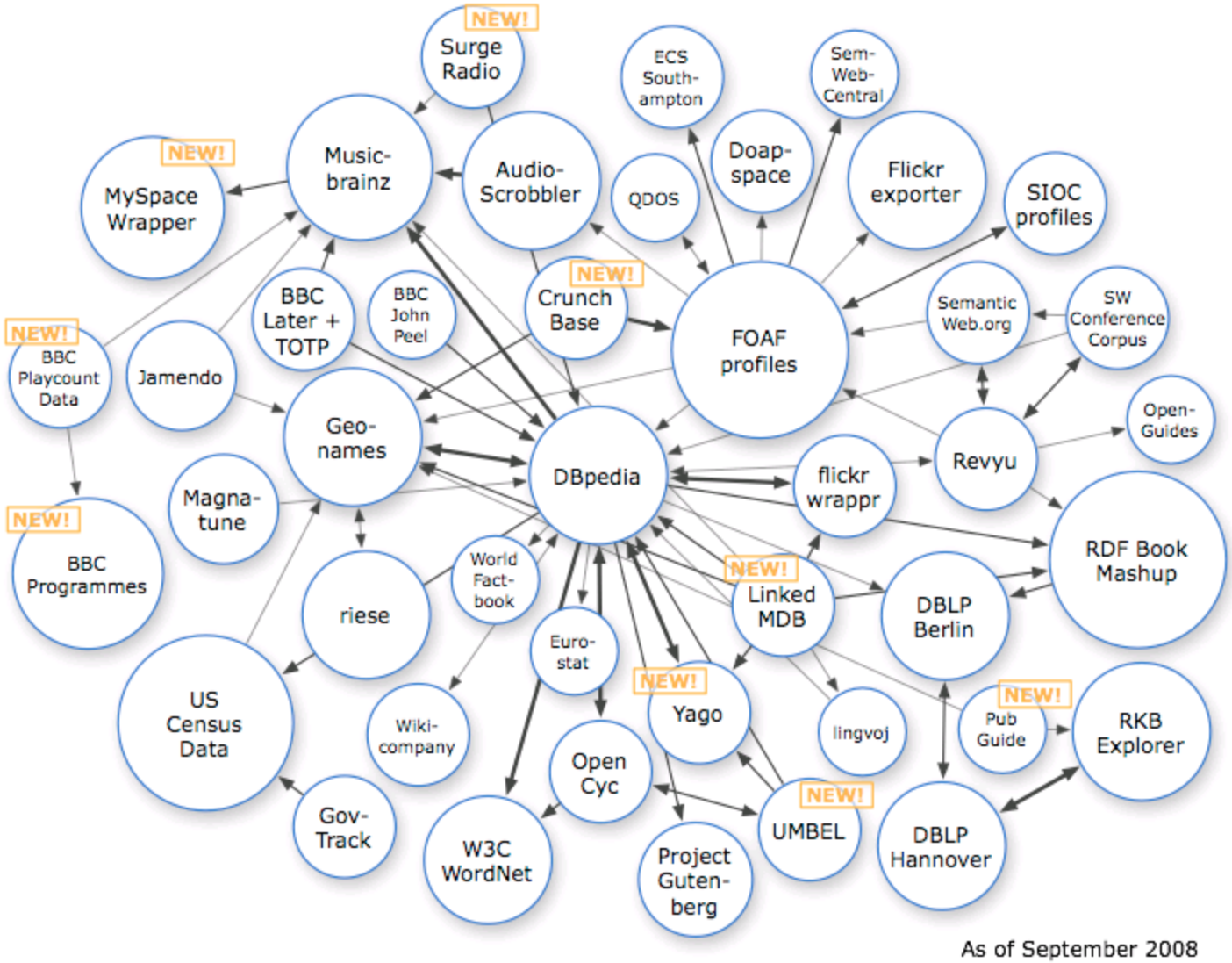


Linking Open Data cloud diagram, this and subsequent pages, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

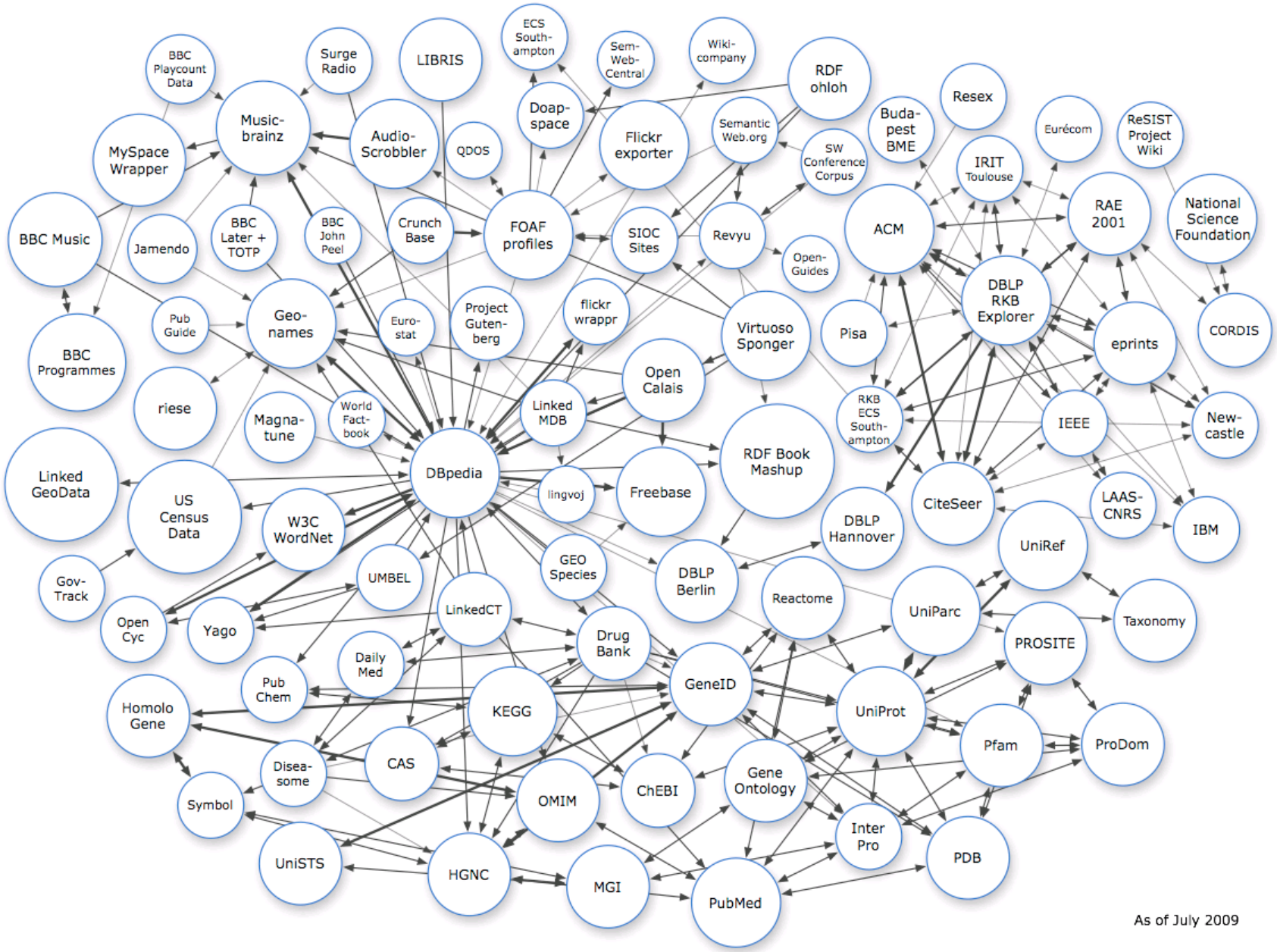
Linked Open Data 2007 (Oct)



Linked Open Data 2008



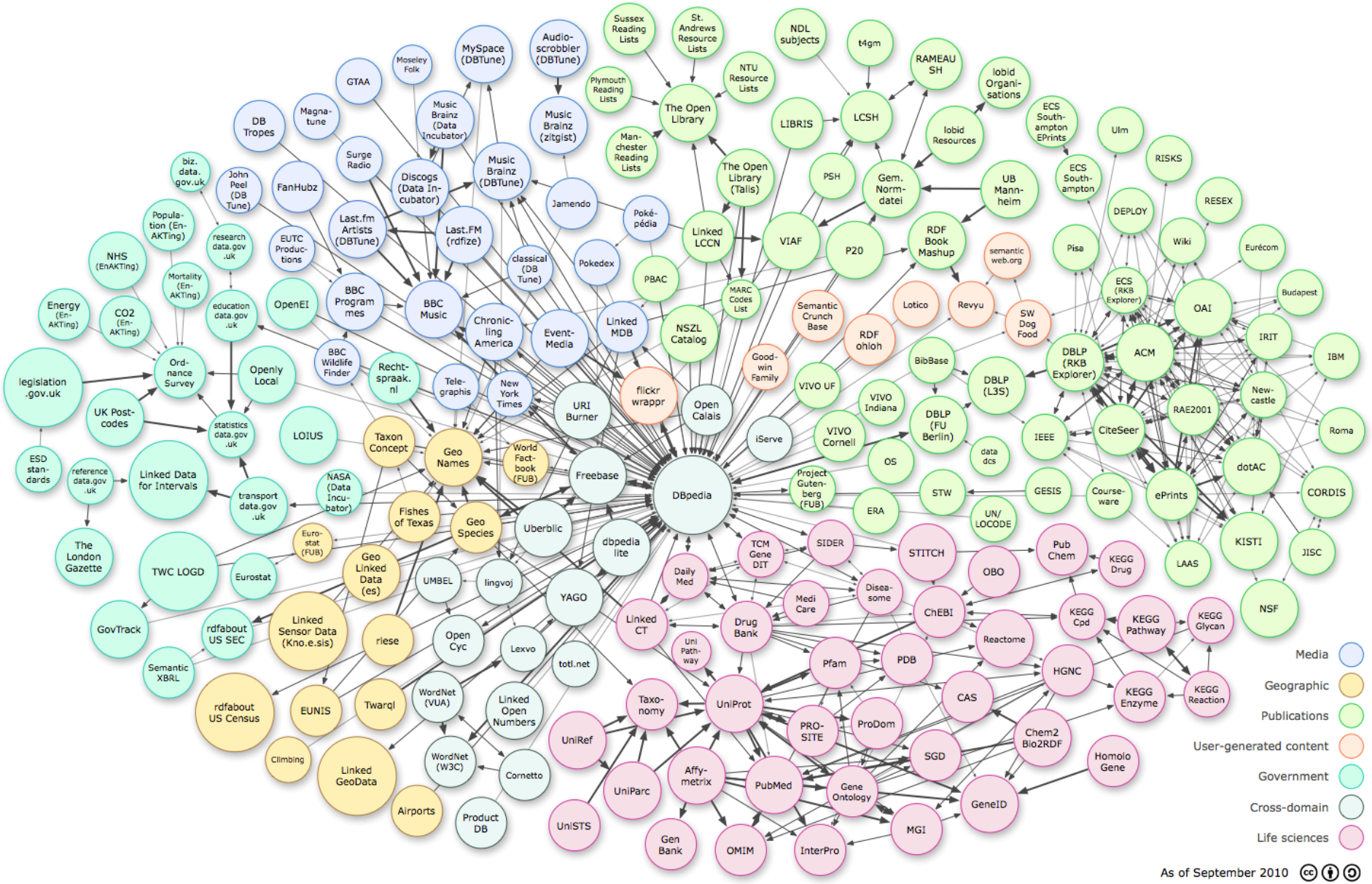
Linked Open Data 2009



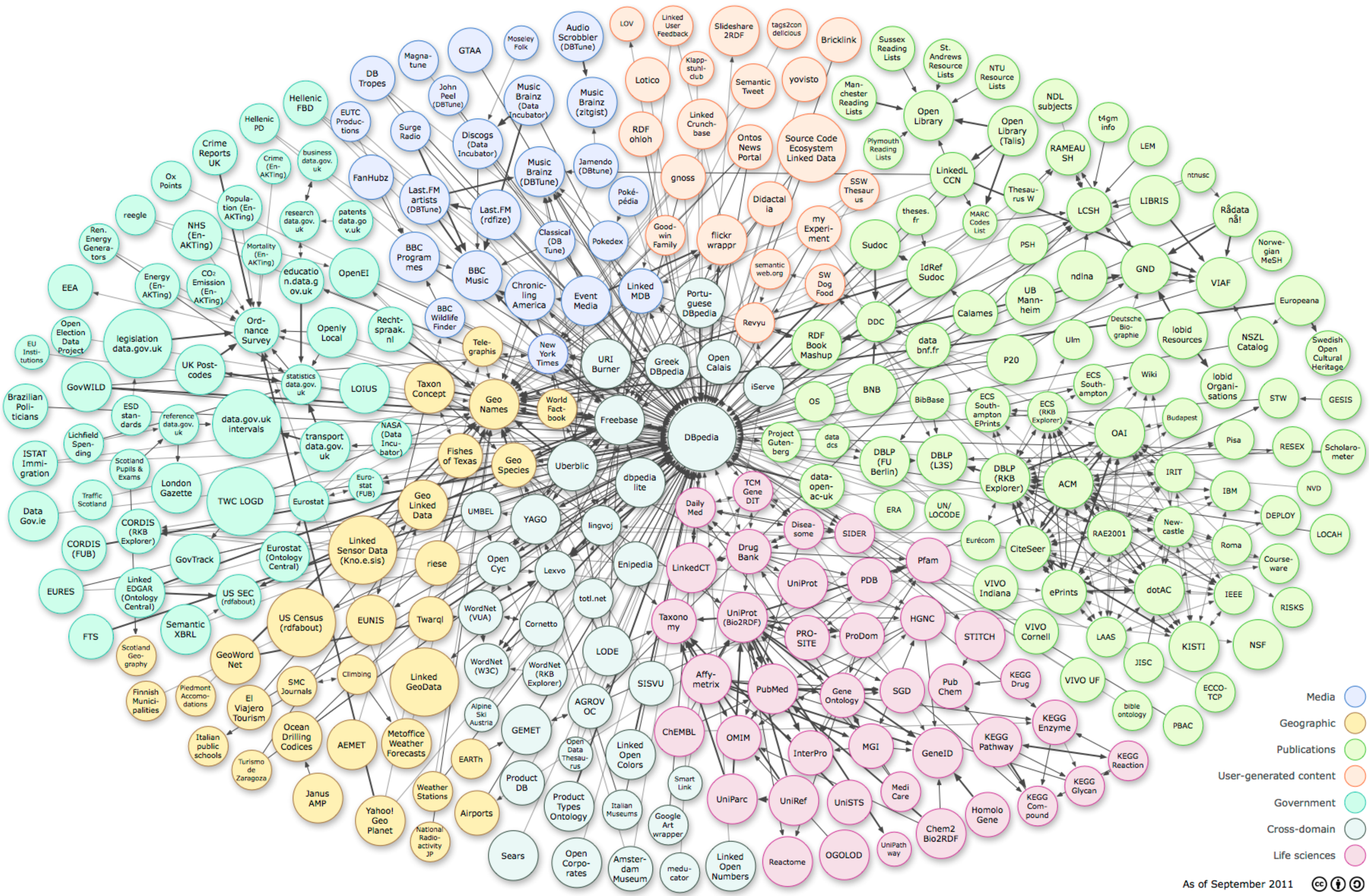
As of July 2009

Linked Open Data 2010

kno.e.sis



Linked Open Data 2011



Linked Open Data

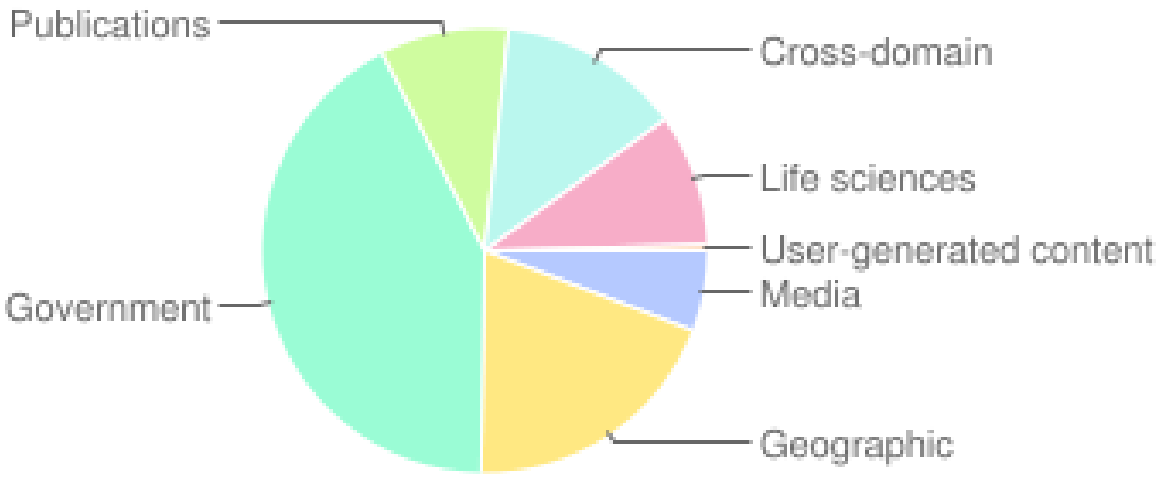
Number of Datasets

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/

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

Course Project

- **Throughout the course, each of you will create an ontology**
- **We'll do this step by step**

- **The goal of the project is to learn "hands-on" about ontology modeling**

- **You'll be given incremental assignments to work on your ontology.**

First project assignment

- Select a domain which you want to model (e.g., “family”).
  - use a domain you have good knowledge about
  - use a domain which is accessible for others (e.g. do not choose Algebraic Topology)
  - you will be stuck with this domain until the end of the project!
  - send name of the domain to me by 01/04/2010 9pm.
- Select approx. 20 class names for an initial taxonomy.
  - Taxonomy must be at least 3 nodes deep.
- Create the taxonomy and write it up.
  - choose your own representation method – but it must be generally understandable (or give an explanation)
  - send taxonomy to me by 01/8/2010
- Write a few sentences why you chose this domain.
  - send this to me by 01/8/2010

Course Project

- **Deadlines are hard.**
- **The more “difficult” domains are the more interesting ones ;)**
- **You will be able to revise your taxonomy later.**
- **Do this first part without discussion with other students.**
- **Expect that, at some stage, other students will have access to your ontology. I may also discuss examples explicitly in class or on slides. So whatever you write up will be “public.”**
- **Don’t worry (yet!) too much about “correctness” of your modeling.**

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

Course contents overview

- Resource Description Framework (RDF)
  - Syntax
  - Semantics
  - Automated Reasoning
- Web Ontology Language (OWL)
  - Syntax
  - Semantics
  - Automated Reasoning

If time permits:

- SPARQL
- OWL and Rules
- Applications

Course contents discussion

- **What's your knowledge about XML?**
- **What's your knowledge about predicate logic?**

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

Organizational matters

- **Course Website: http://www.semantic-web-book.org/page/KR4SW-12**
- **Slides will usually be posted the evening before class.**
- **If you have not yet received a class email from me by tomorrow morning, let me know asap.**
- **Class meetings:**
  - **Tuesday 2:10pm to 3:50pm in Russ 355**
  - **Thursday 2:10pm to 3:50pm in Russ 355**
  - **Class starts 3rd of January 2012 and ends 8th of March 2012**
- **"Office hours:**
  - **Thursdays 1pm to 2pm and by appointment. I usually have to leave right after class.**
  - **Please use email as main means of communication with me (besides talking with me in or after class).**

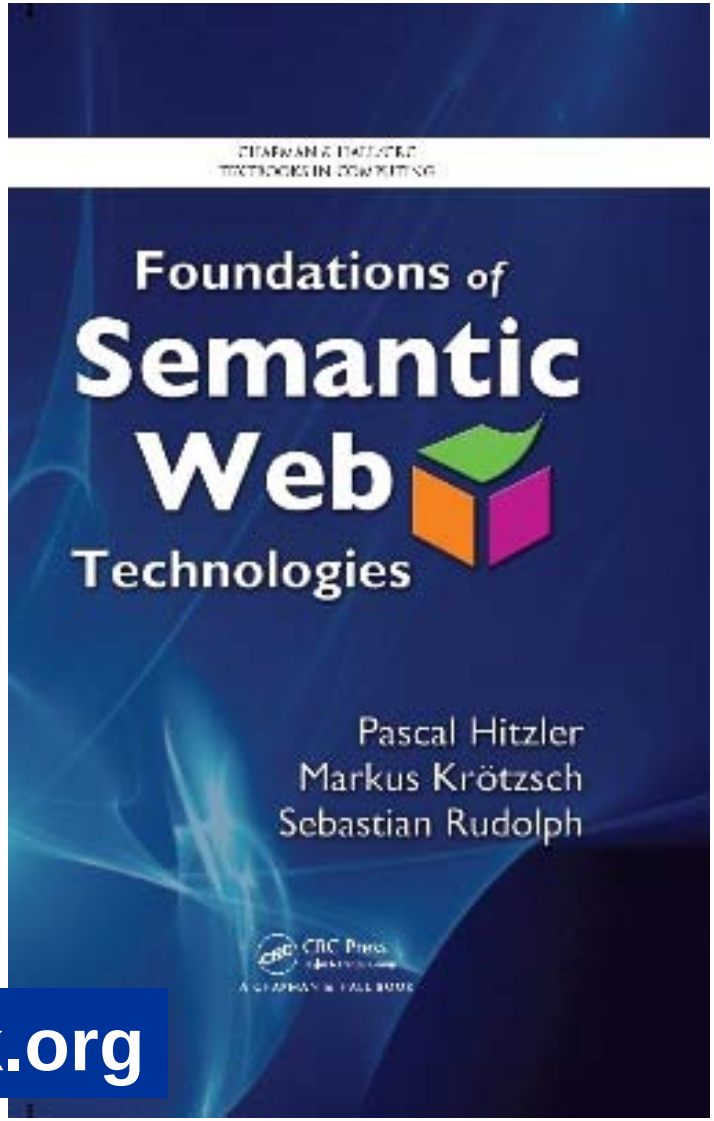
Textbook (required)

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)，2011，to appear

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

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

Grading

Homework 10%, Project 30%, Class Presentation 30%, Final Exam 30%

- **Homework: Students will take turns in presenting completed homework in class, and grading will be done by evaluating the participation in the tutorial sessions.**
- **Project: The project will be an ontology modeling project which will have several parts and run over the whole quarter. Students are expected to write an experience report on each part.**
- **Class Presentations: The class presentations will consist of original research papers.**
- **Final Exam: The final exam will be oral, i.e. in the form of a short interview (20 minutes), with the examiner asking questions and the student answering.**

Class Presentations

- **Presentations of original research papers extending beyond the material covered in this lecture.**
- **They will all be concerned with issues related to ontology reasoning/formal semantics.**
- **If you know a paper which you'd like to present, please let me know asap.**
- **I'll make some suggestions later in the class.**

Any further questions or open issues?

Topic next Thursday: RDF Syntax

First project assignment

- Select a domain which you want to model (e.g., “family”).
  - use a domain you have good knowledge about
  - use a domain which is accessible for others (e.g. do not choose Algebraic Topology)
  - you will be stuck with this domain until the end of the project!
  - send name of the domain to me by 01/04/2010.
- Select approx. 20 class names for an initial taxonomy.
  - Taxonomy must be at least 3 nodes deep.
- Create the taxonomy and write it up.
  - choose your own representation method – but it must be generally understandable (or give an explanation)
  - send taxonomy to me by 01/8/2010
- Write a few sentences why you chose this domain.
  - send this to me by 01/8/2010