CS 7810 - Knowledge Representation and Reasoning (for the Semantic Web)

01 - Introduction

Adila Krisnadhi
Data Semantics Lab,
Wright State University, Dayton, OH

Acknowledgements

- A good number of slides in this presentation are adapted from:
  - Sebastian Rudolph, “Overview and XML”, slides for Foundations of Semantic Web Technologies course, Dresden, April 9, 2014.
  - Pascal Hitzler, “Slides 1 – 01/03/2012”, slides for Knowledge Representation for the Semantic Web course, Winter quarter 2012.

About me and the course

DaSe Lab

- About the course, visit: http://dase.cs.wright.edu/courses/cs-7810-knowledge-representation-and-reasoning-fall-2016

- About me, see: http://dase.cs.wright.edu/people/adila-krisnadhi

About you …

- Who are you?
- Which year?
- Which specialization area, if any?
- Why are you taking this class?

The (World Wide) Web

DaSe Lab

- Pervasive: various aspects of life are web-ized.
  - Social contacts (social networks, blogs, …)
  - Economics (buying, selling, advertising, …)
  - Authorities, administration (government)
  - Education (e-learning, Web as information system, …)
  - Work life (information gathering and sharing)
  - Recreation (games, role play, creativity, …)
- Immensely successful (Why?)
- Generate huge amount of data
- Syntax-standards for transferring structured data
- Machine-processable, human-readable **documents**.

- **But**: content/knowledge and meaning/semantics of transferred data are not accessible by machine.

Problems with the Web (1)

- Huge amount of information, too little structure.
- Emphasis on human consumption.
  - Standard syntax for presenting information to human.

Prof. Dr. Pascal Hitzler

Home | CV | Contact | DaSe Lab for Data Semantics | Publications | Activities | Teaching | Funding | Links

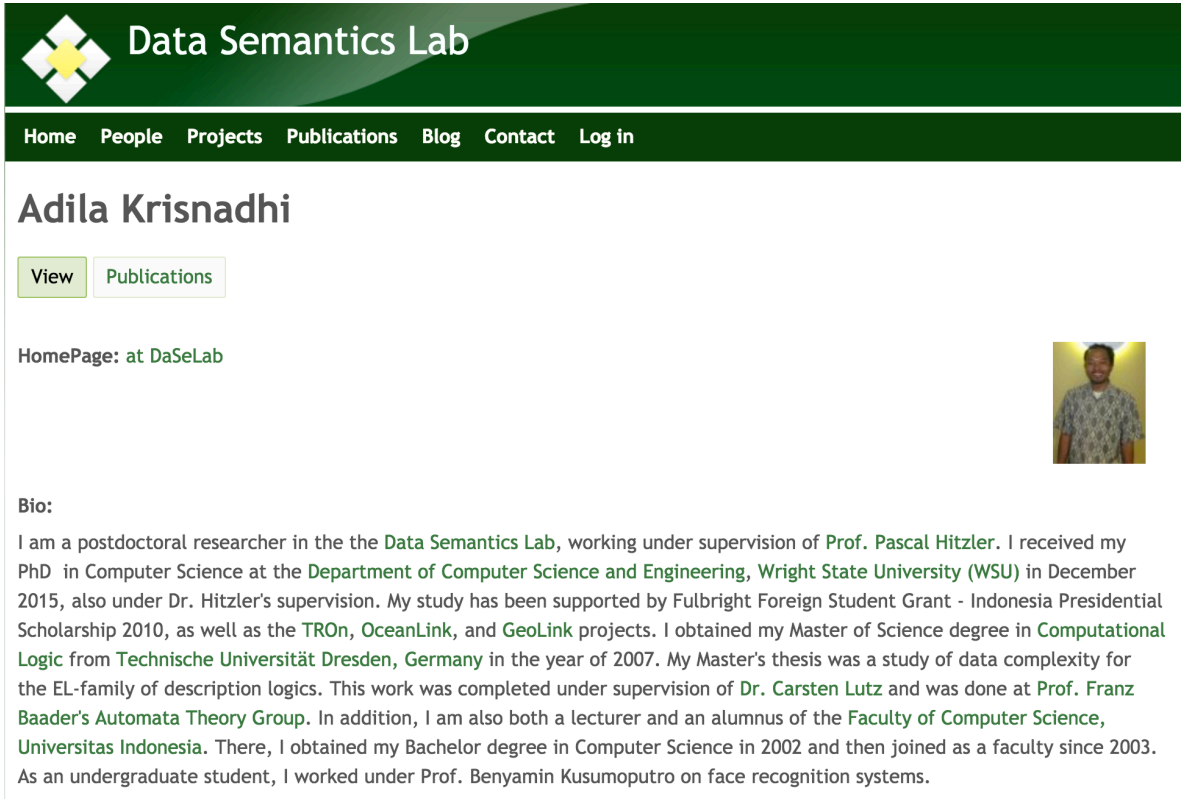
Pascal Hitzler

Professor
Director of Data Science
Director of the Data Semantics (DaSe) Laboratory
Data Science and Security Cluster (DSSC)
Department of Computer Science and Engineering
College of Engineering and Computer Science
Wright State University
Dayton, Ohio, USA

Contact

Problems of the Web (1)

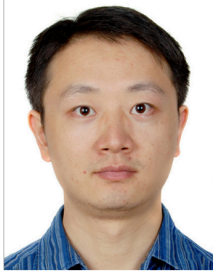
- Huge amount of information, too little structure.
- Emphasis on human consumption.
  - Standard syntax for presenting information to human.



Problems with the Web (1)

- Huge amount of information, too little structure.
- Emphasis on human consumption.
  - Standard syntax for presenting information to human.

主页 | 研究生报考 | 学术活动 | 发表论文 | 链接



漆桂林

教授，博导
东南大学计算机与工程学院

电　　话：+86 (0) 25 52090910
传　　真：跟电话号码相同
电子邮箱：gqi at seu dot edu dot cn

论文 "Conflict-based revision operators in possibilistic logic" 被AAAI'12录取 (full paper).

论文 "A new operator for ABox revision in DL-Lite" 被AAAI'12 录取 (short paper).

论文 "Large scale temporal RDFS reasoning using MapReduce " 被AAAI'12 录取 (short paper).

被邀请为语义Web最权威杂志Journal of Web Semantics的Editorial Board成员

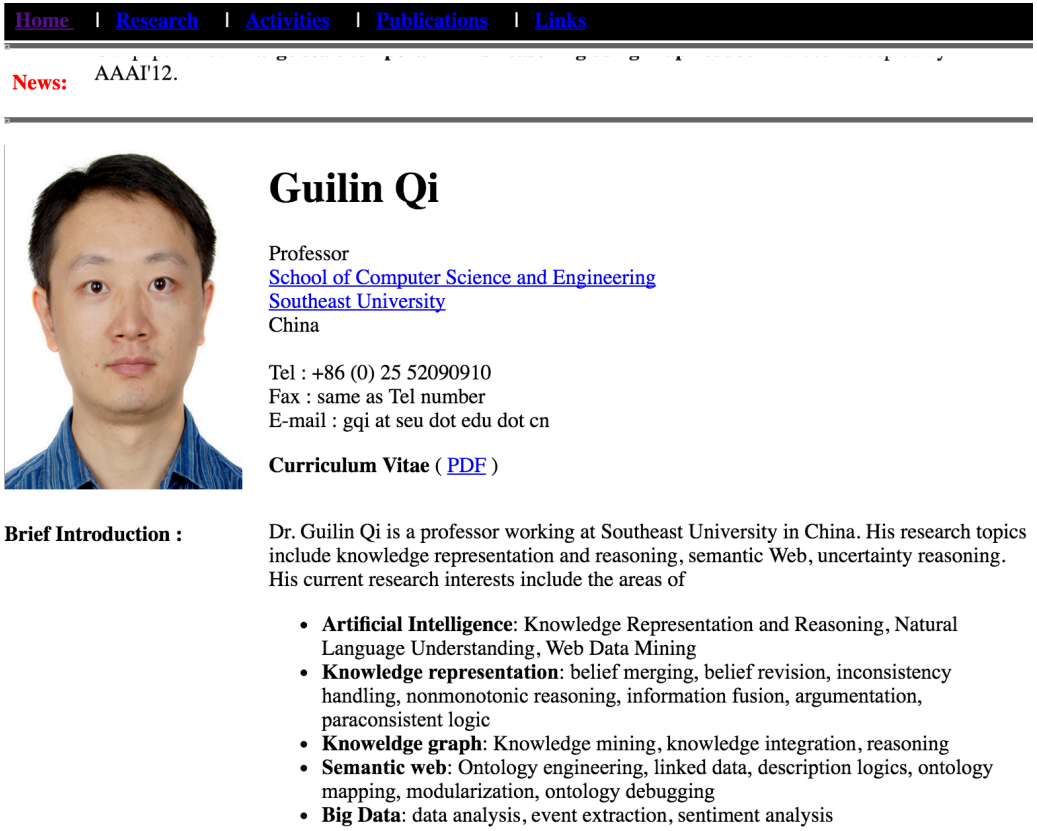
个人简介

漆桂林，东南大学教授，博士生导师。漆桂林是中国计算机协会会员，ACM professional member。现任中国中文信息学会语言与知识计算专业委员会副主任和CCF中文信息技术专委会委员。是澳大利亚Griffith大学访问教授（2011年11月-2012年2月和2013年6月-2013年7月）和法国图卢兹第一大学访问教授（2013年1月-2013年2月）。1998年毕业于宜春学院数学专业，2002年获得江西师范大学数学与信息系硕士学位，2006年获得英国贝尔法斯特女皇大学计算机博士学位。2006年8月至2009年8月在德国Karlsruhe大学AIFB研究所做博士后研究。

漆桂林教授长期从事人工智能、语义Web和知识工程方面的科学研究及教学工作。截至目前,已经发表高质量学术论文100篇多篇,特别是在国际人工智能著名会议IJCAI、AAAI、KR、UAI以及国际语义Web著名会议ISWC、ESWC发表多篇会议文章，在国际著名杂志Information Science和Fuzzy Sets and Systems等发表多篇杂志文章。其中，一篇文章在亚洲语义Web会议上获得最佳论文提名奖。获得欧盟第七框架Marie Curie Actions — International Research Staff Exchange Scheme (IRSES)资助（漆桂林为其中一个workpackage的负责人）。获得多项国家自然科学基金资助，其中两项为项目负责人。2014年获得百度主题研究项目资助。作为主要参与人参加了863大数据“类人类智能”方向的课题。

Problems of the Web (1)

- Huge amount of information, too little structure.
- Emphasis on human consumption.
  - Standard syntax for presenting information to human.



Home | Research | Activities | Publications | Links

News: AAAI'12.

Guilin Qi

Professor
School of Computer Science and Engineering
Southeast University
China

Tel : +86 (0) 25 52090910
Fax : same as Tel number
E-mail : gqi at seu dot edu dot cn

Curriculum Vitae (PDF)

Brief Introduction :

Dr. Guilin Qi is a professor working at Southeast University in China. His research topics include knowledge representation and reasoning, semantic Web, uncertainty reasoning. His current research interests include the areas of

- **Artificial Intelligence**: Knowledge Representation and Reasoning, Natural Language Understanding, Web Data Mining
- **Knowledge representation**: belief merging, belief revision, inconsistency handling, nonmonotonic reasoning, information fusion, argumentation, paraconsistent logic
- **Knoweldge graph**: Knowledge mining, knowledge integration, reasoning
- **Semantic web**: Ontology engineering, linked data, description logics, ontology mapping, modularization, ontology debugging
- **Big Data**: data analysis, event extraction, sentiment analysis

Problems of the Web (2)

DaSe Lab

- Localizing information is challenging.
- Today's search engines are good, but mostly keyword-based.
  - Example: try google for "kohl" images; how many kinds of kohl will you find?

- Desiderata: search for content → semantic search

Problems with the Web (3)

DaSe Lab

- Heterogeneous
  - content
  - structure (e.g., where a particular piece of information can be found on webpages)
  - character encoding (e.g., ASCII vs. Unicode).
- Desiderata: intelligent information integration

Problems with the Web (4)

DaSe Lab

- Deriving new (implicit) information from given pieces of information is possible for human.
  - The current Web can only operate on syntax.

- Desiderata: automated deduction.
  - Formal logics are required.

So what is this “Syntax vs Semantics” about?

DaSe Lab

- Syntax:
  - Greek: σύνταξη → composition, sentential structure
  - denotes the (normative) structure of data
  - characterizes what makes data “well-formed”
- Semantics:
  - Greek: σημαντικός → significant, belonging to the sign
  - denotes the meaning of data
  - characterizes conclusions that can be drawn from data.

- Example:
  - 7+) = (syntactically incorrect
  - 7+3 = 11 syntactically correct, semantically incorrect
  - 7+3 = 10 syntactically and semantically correct

Problems of the Web (Solutions?)

DaSe Lab

- Ad hoc:
  - Employs methods (e.g., NLP techniques, machine learning) to evaluate existing unstructured information on the Web
- A priori:
  - Put structure on the pieces of information prior to publishing on the Web such that later automated deployment and processing are enabled.
- Semantic Web (usually) belongs to the latter.

Basic ingredients for the Semantic Web

DaSe Lab

- Open standards for describing information on the Web
  - Make some knowledge representation languages as standard.
- Methods to obtain further information from such descriptions
  - Reasoning algorithms.

Reasoning

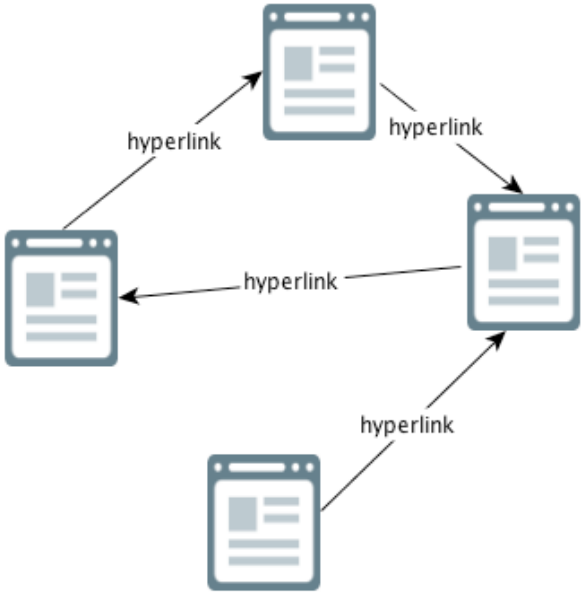
DaSe Lab

- Logical deduction (automated reasoning)
  - Columbus is a state capital.
  - Every state capital is a city.
  - **Then**: Columbus is a city.
- We use logic
  - Predicate logic (for most of the time).
  - The logic determines which conclusions are valid.
  - Algorithms to derive conclusions are needed.

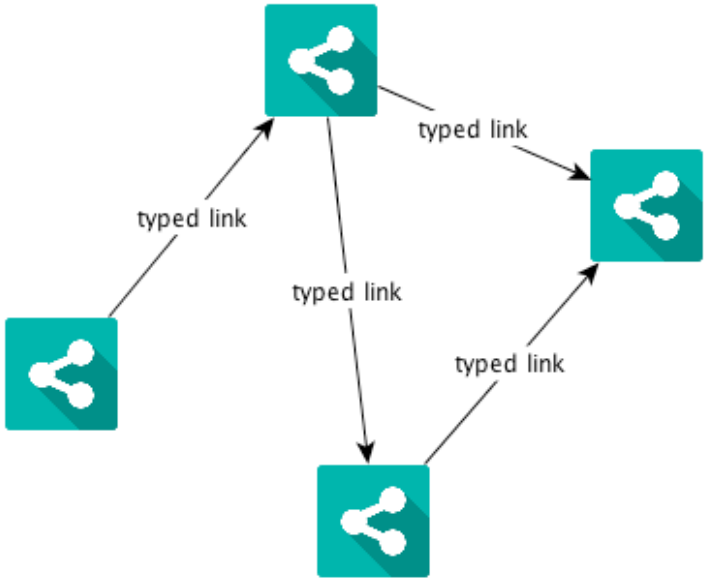
Semantic Web: Basic Idea

DaSe Lab

The Web (of Documents)



The Web (of Data)



Semantic Web = Web of Data + formalized vocabulary
Formalized vocabulary = ontology = collection of terms (used for links and classes) whose semantics are formally defined.

Semantic Web: Basic Idea

DaSe Lab

- Try visualize (at least partially) the Web (of Documents) starting from https://en.wikipedia.org/wiki/Dayton,_Ohio
- Try visualize the Web of Data starting from http://dbpedia.org/resource/United_States
  - Which term would you consider as 'data'?
  - Which term would you NOT consider as 'data'?

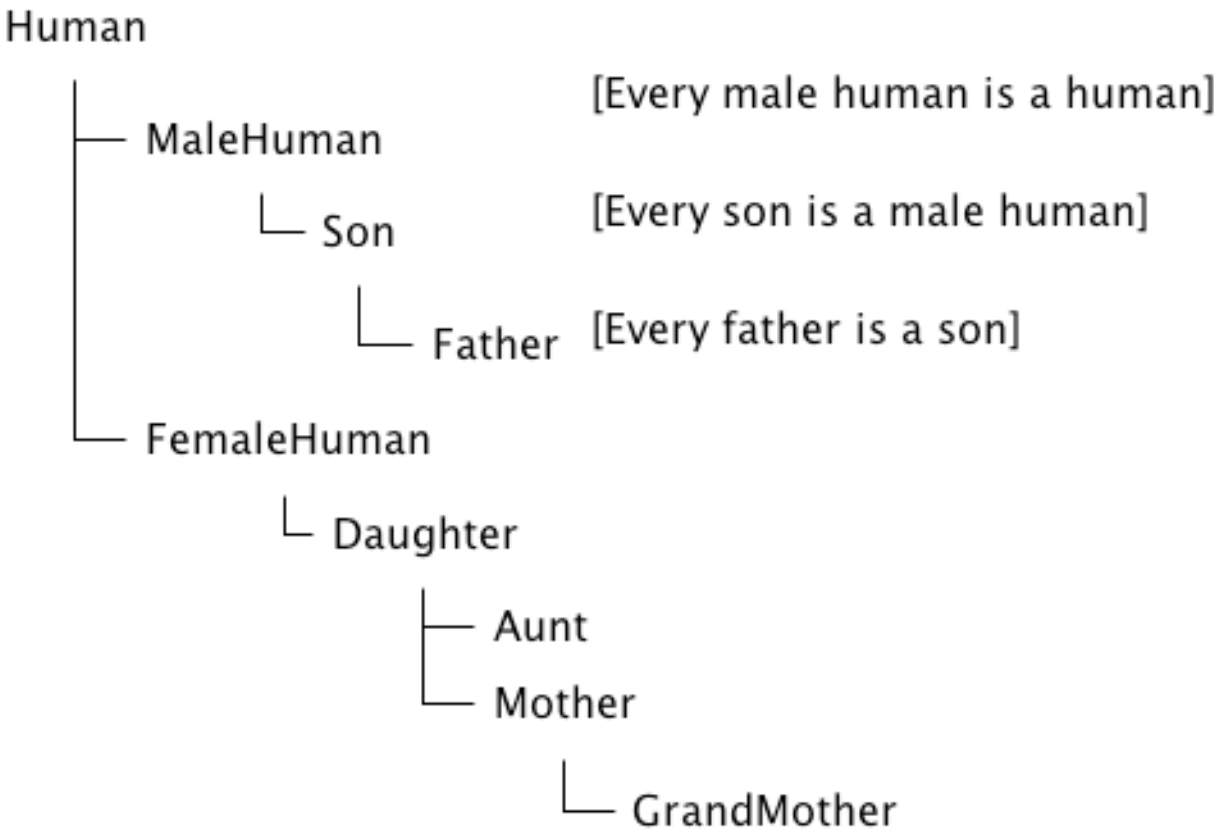
Ontology & Ontology Languages

DaSe Lab

- Vocabulary term: term acting as metadata.
  - For links between data, and class names for typing of instance data
- Ontology: specification containing formal definition of vocabulary terms
  - E.g., “every country is a populated place”
- To specify an ontology, we need an ontology language.
  - Meaning/semantics via logic and automated reasoning
  - Scalability is a challenge: the more expressive an ontology language is, the less scalable it will typically be.

Taxonomy

- Often (not always!), the core of an ontology is a taxonomy -- classes of things arranged in a subclass hierarchy.



Partonomy

- Taxonomy is often confused with partonomy – classes of things arranged in a hierarchy of "part-of" relationships.

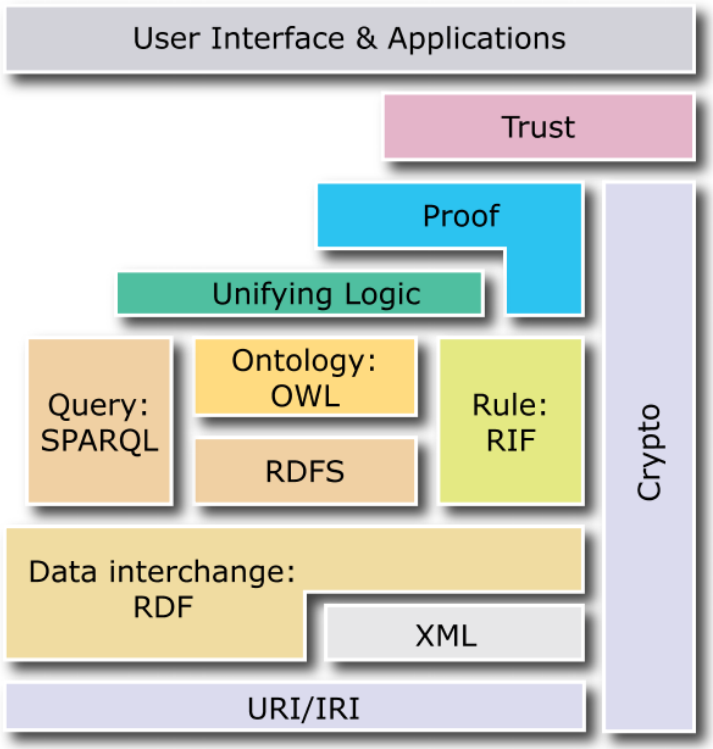
Taxonomy vs Partonomy

DaSe Lab

- Taxonomy: *every* A *is a* B
  - Every father is a man
  - Every horse is a mammal
  - Every university is an organization
  - Every arm is a limb

- Partonomy: A *is a part of* B
  - Arm is a part of body
  - Ohio is a part of USA
  - Engine is a part of car
  - Daselab is a part of Wright State University

Semantic Web Standards



https://www.w3.org/2007/03/layerCake.png

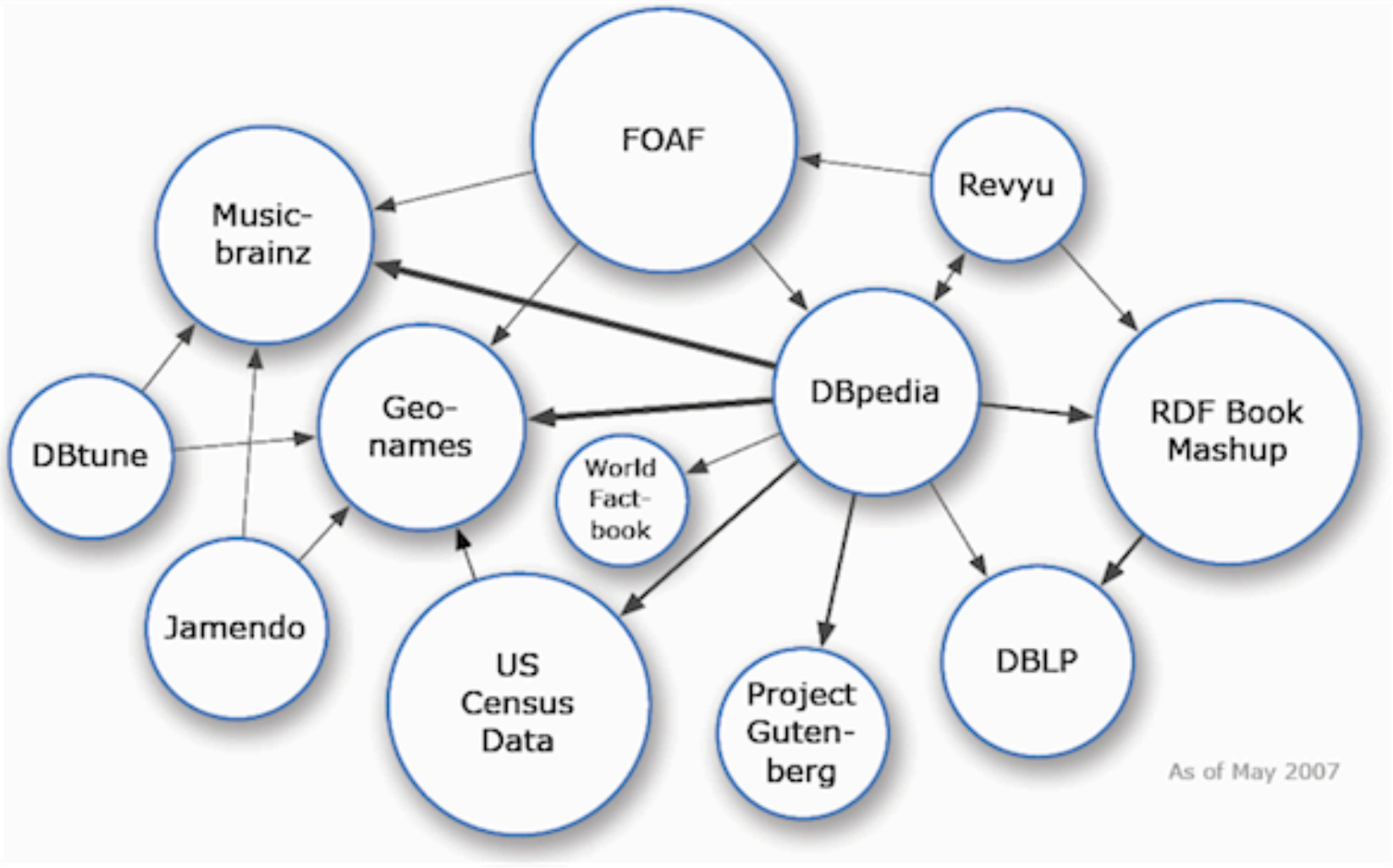
1994 First public presentation of the Semantic Web
1998 Start of standardization of data model (RDF) and a first ontology language (RDFS)
2000 Start of large US & Europe research projects on ontologies (DAML & Ontoknowledge)
2001 Tim Berners-Lee, et al landmark paper.
2002 Start of standardization of new ontology language (OWL)
2004 RDF and OWL standards finalized (RDFS is part of RDF standards)
2008 SPARQL standard finalized
2009 OWL 2 standard finalized
2010 Rule Interchange Format (RIF) standard finalized
2012 OWL 2 standard updated to 2nd edition
2013 SPARQL 1.1 standard finalized, RIF standard updated to 2nd edition
2014 RDF 1.1 standard finalized

Semantic Technologies in the US

DaSe Lab

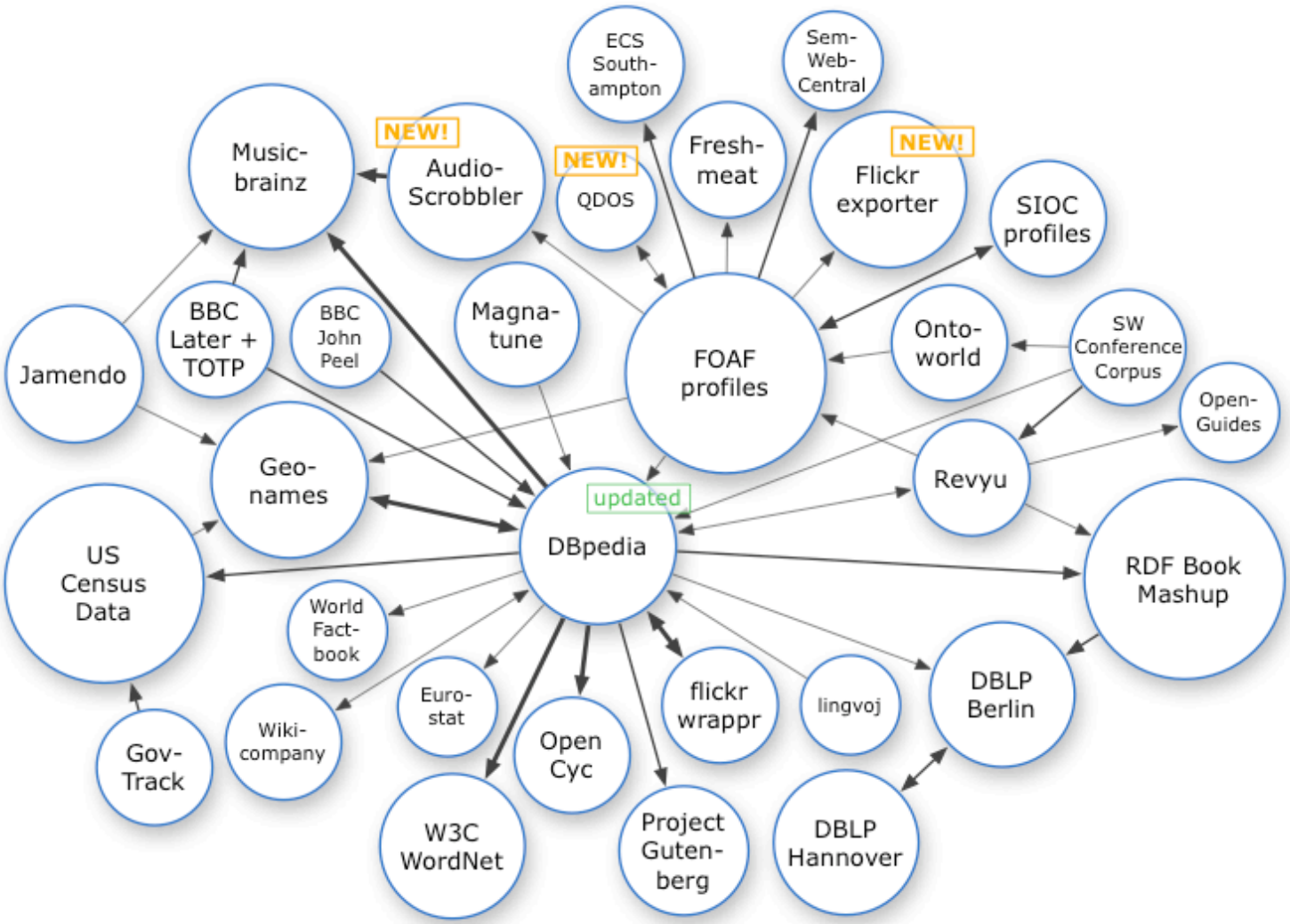
- (Government) research funding via:
  - NIH, NSF, DoD, DoE, AFRL, IARPA, DARPA
- Industrial take-ups:
  - Annual Semantic Technology Conference (since 2015 called Smart Data conference) hosted by Dataversity
  - Investment by major enterprises (Oracle, IBM, HP, Google, Accenture, Siemens, Bosch, GE)
  - Major government contractors (BBN, Lockheed, …)
  - Venture capital (e.g., Vulcan)
  - Structured data on the Web (BBC, nytimes, data.gov, …)

State of Linked Data (2007)



Linking Open Data cloud diagrams, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

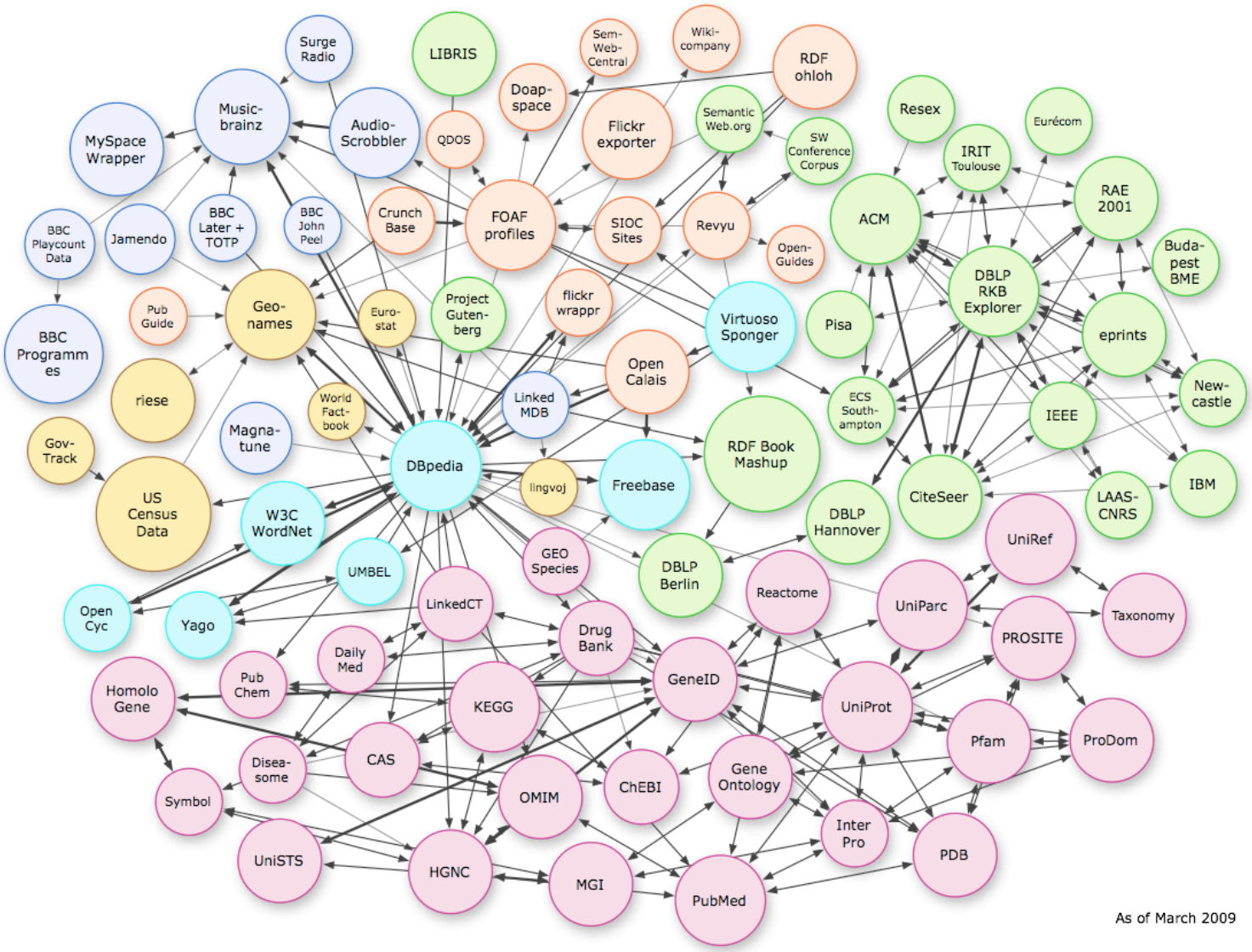
State of Linked Data (2008)



Linking Open Data cloud diagrams, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

State of Linked Data (2009)

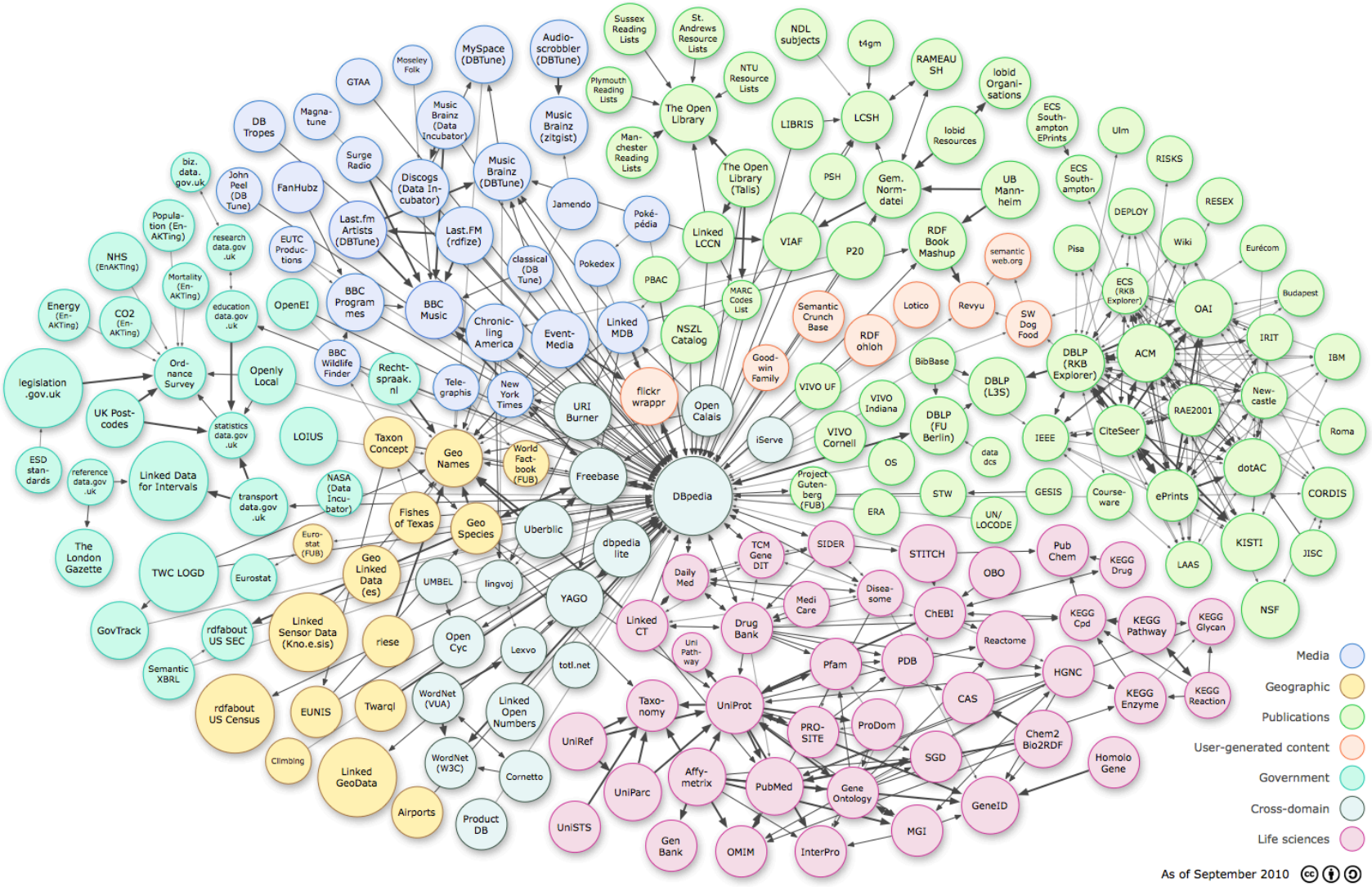
DaSe Lab



Linking Open Data cloud diagrams, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

State of Linked Data (2010)

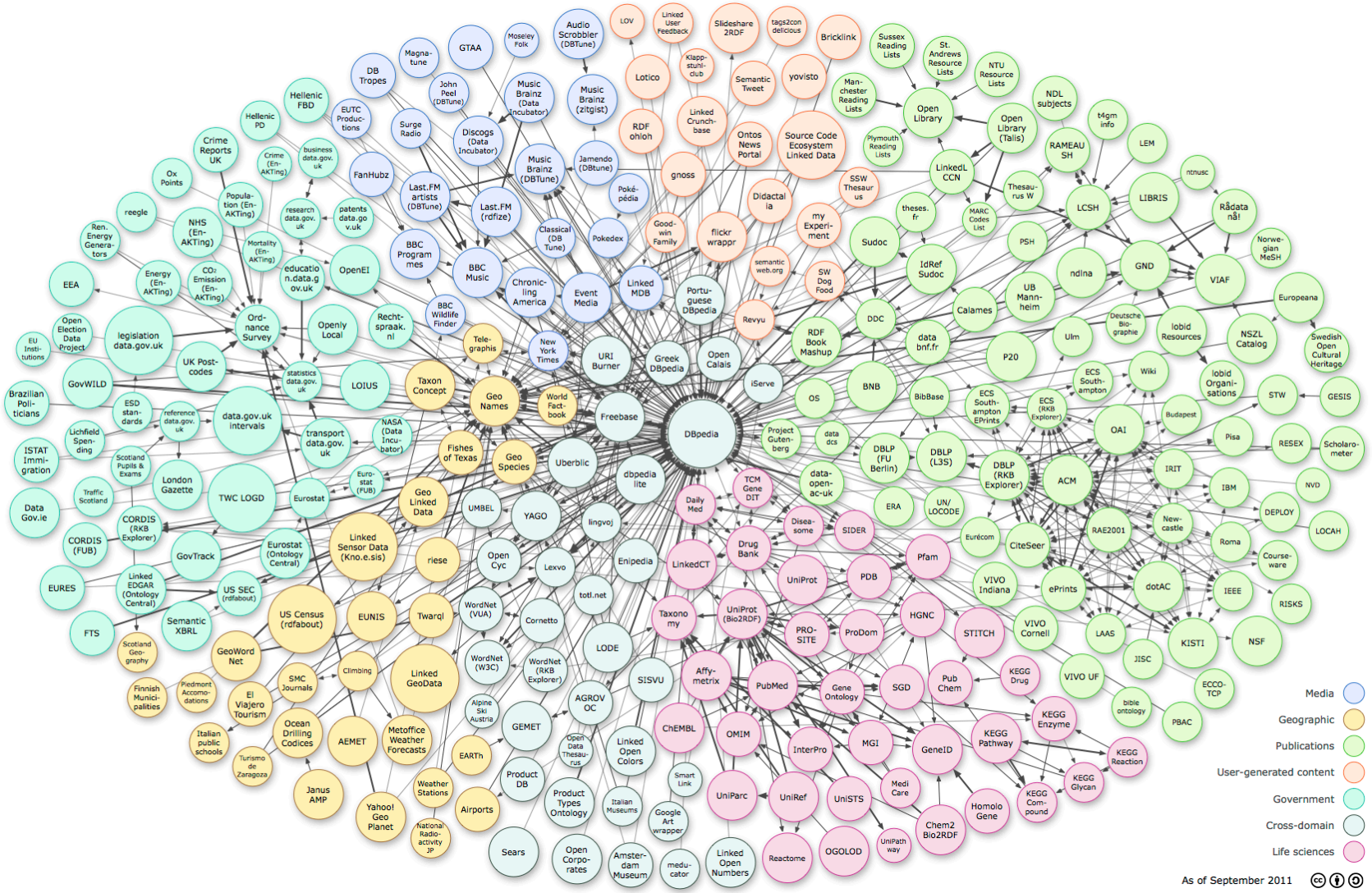
DaSe Lab



Linking Open Data cloud diagrams, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

State of Linked Data (2011)

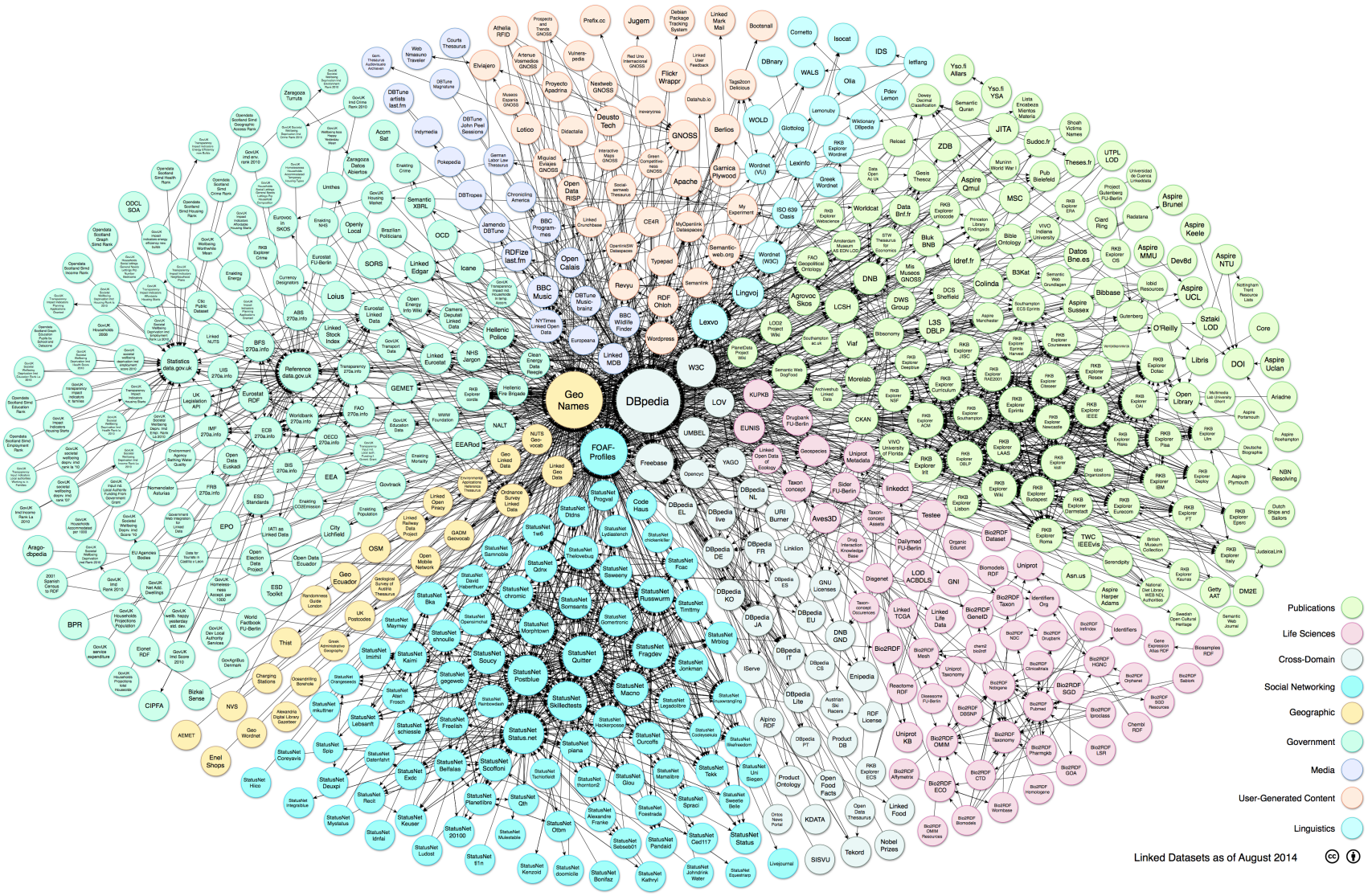
DaSe Lab



Linking Open Data cloud diagrams, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

State of Linked Data (2014)

DaSe Lab



Linking Open Data cloud diagrams, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

Planned Course Content

DaSe Lab

- RDF, RDFS
  - Syntax
  - Semantics
  - Reasoning
- SPARQL
  - Syntax
  - Semantics
  - Reasoning
- OWL, OWL fragments
  - Syntax
  - Semantics
  - Reasoning
- Ontology Engineering with Ontology Design Patterns
- Linked Data Publishing with ODPs

Course Content Discussion

DaSe Lab

- How much do you know about XML?
- How much do you know about predicate logic?

- Next topic: RDF