

Database Development and Design (CPT201)

Lecture 8b: Introduction to Semantic Web and RDF

Dr. Wei Wang
Department of Computing

Learning Outcomes

- Introduction to the following
 - Semantic Web
 - Resource description framework
 - Ontology
 - Linked Open Data
 - Query the Web of Data

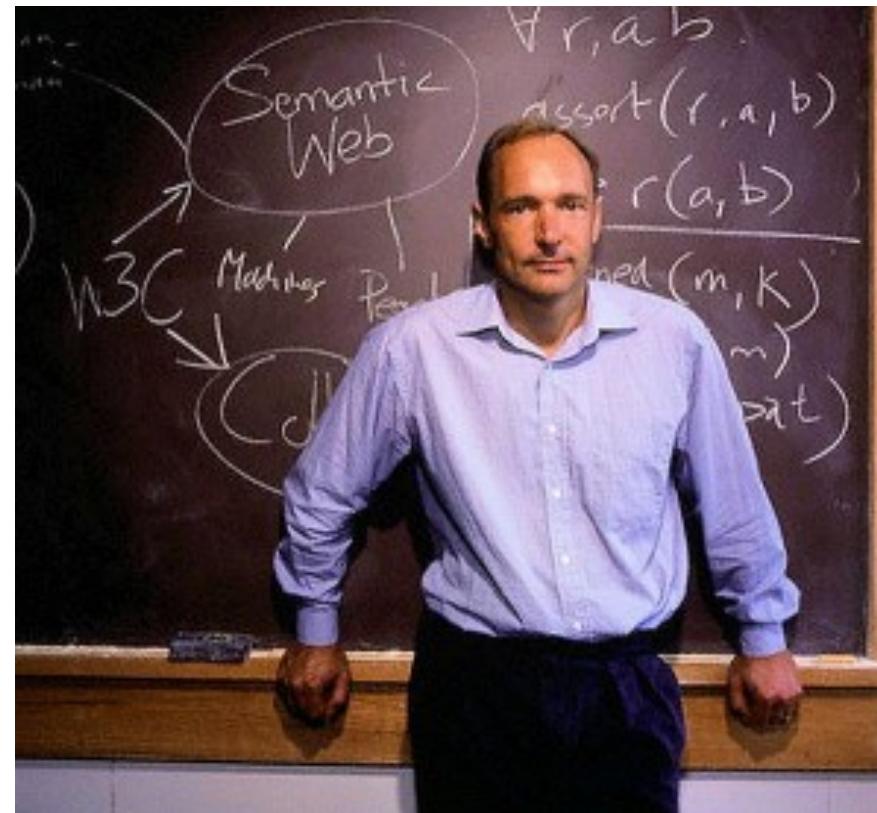
Semantic Web



"The Semantic Web" is an

- ... extension of the current web in which
- ... information is given well-defined meaning,
- ... better enabling computers and people to work in cooperation."

- *The Semantic Web, Tim Berners-Lee, James Hendler and Ora Lassila, Scientific American, May 2001*

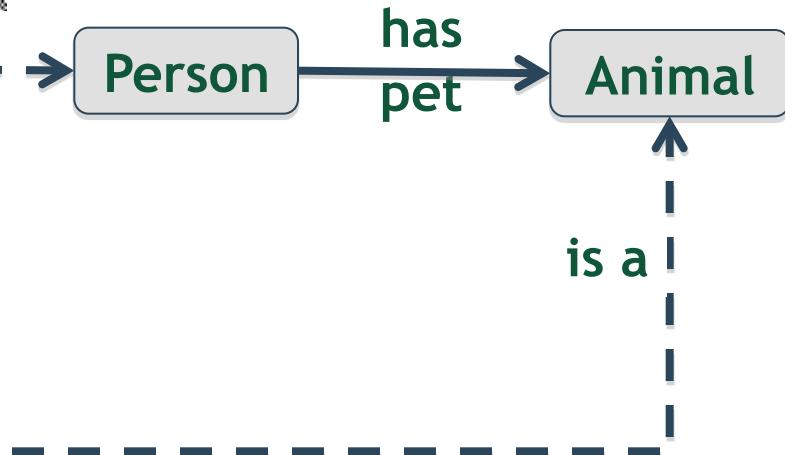
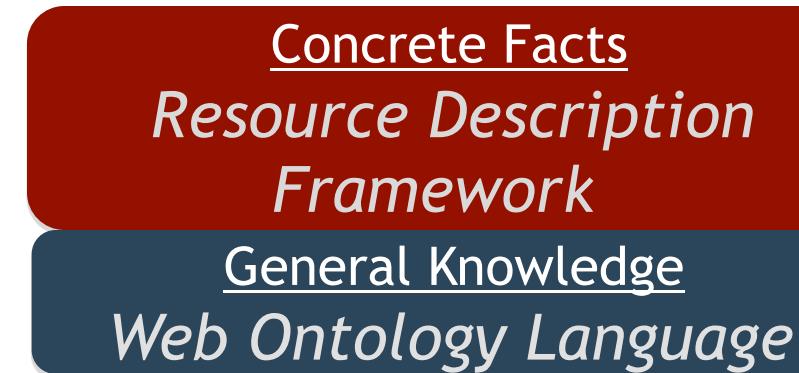
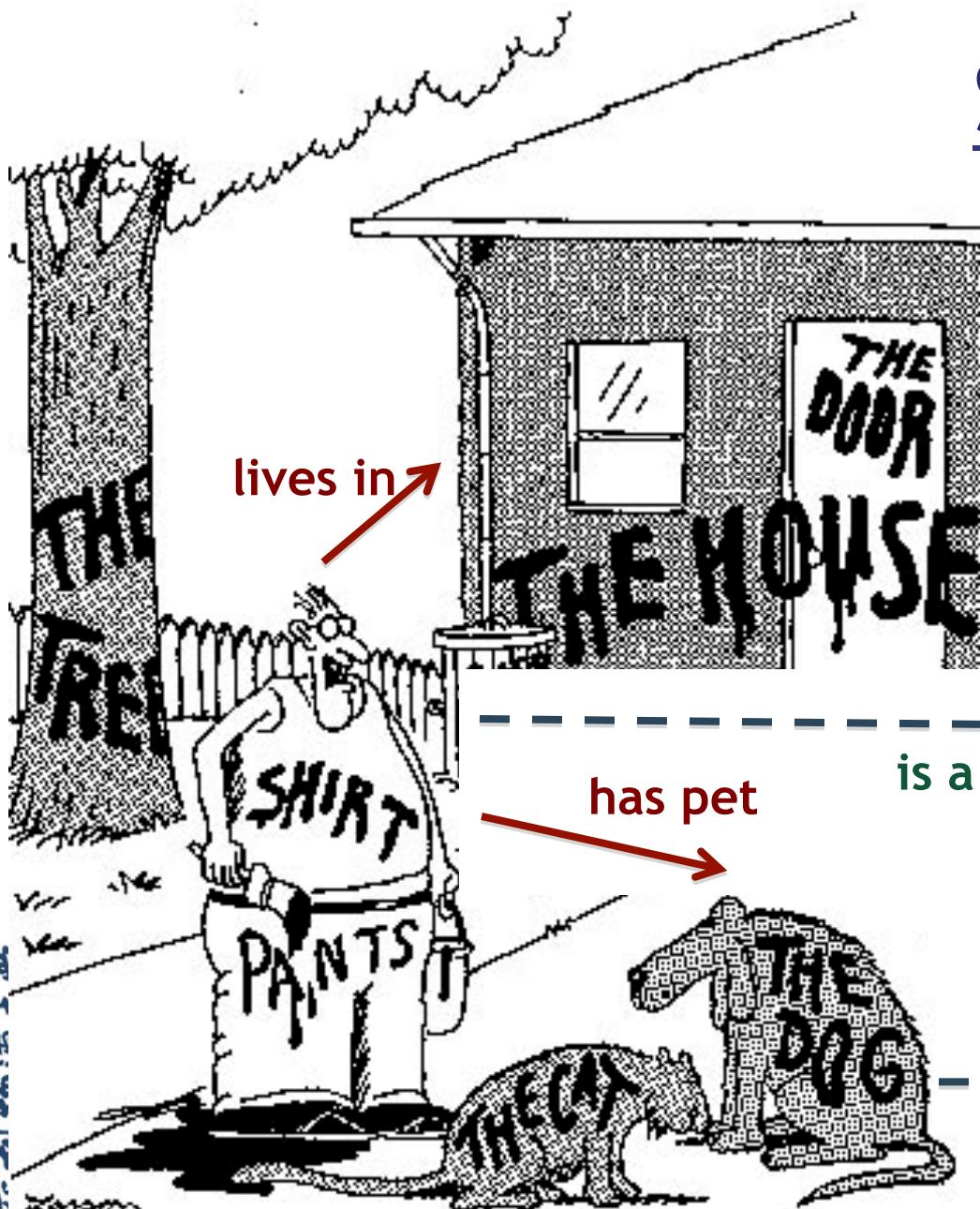


Why Semantic Web?

- Tasks often require combining data across the Internet, e.g.:
 - Integrating data across the enterprise, e.g. hotel, transport, meeting, personal info come from different sites
 - Mining data from biochemical, genetic, pharmaceutical, patient databases
 - Cross-referencing disparate digital libraries
- Humans understand how to combine this information...
 - But not always easy (different vocabularies, languages, formats)
 - Machines aren't smart enough...

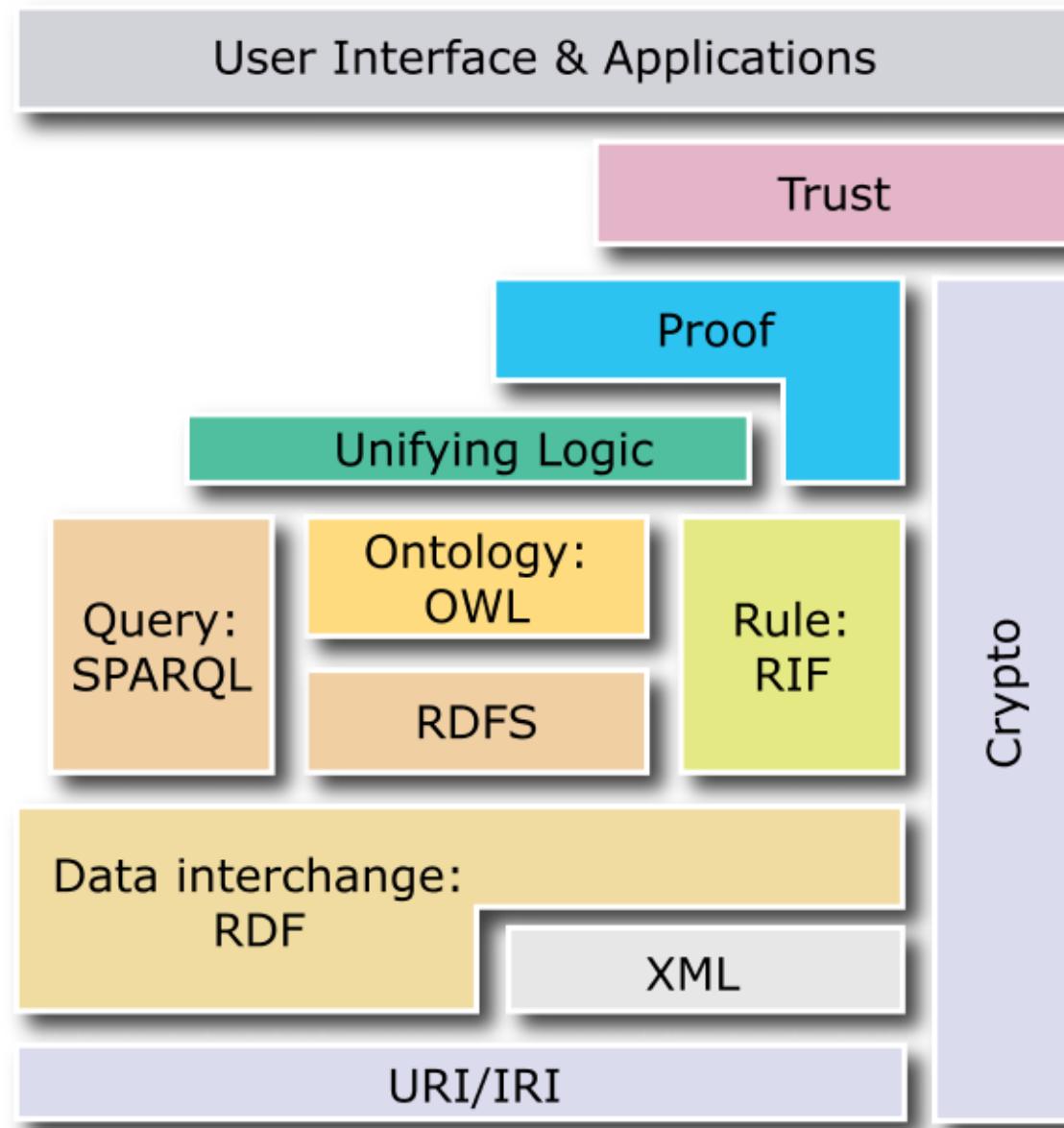
Semantic Web

(according to Farside)

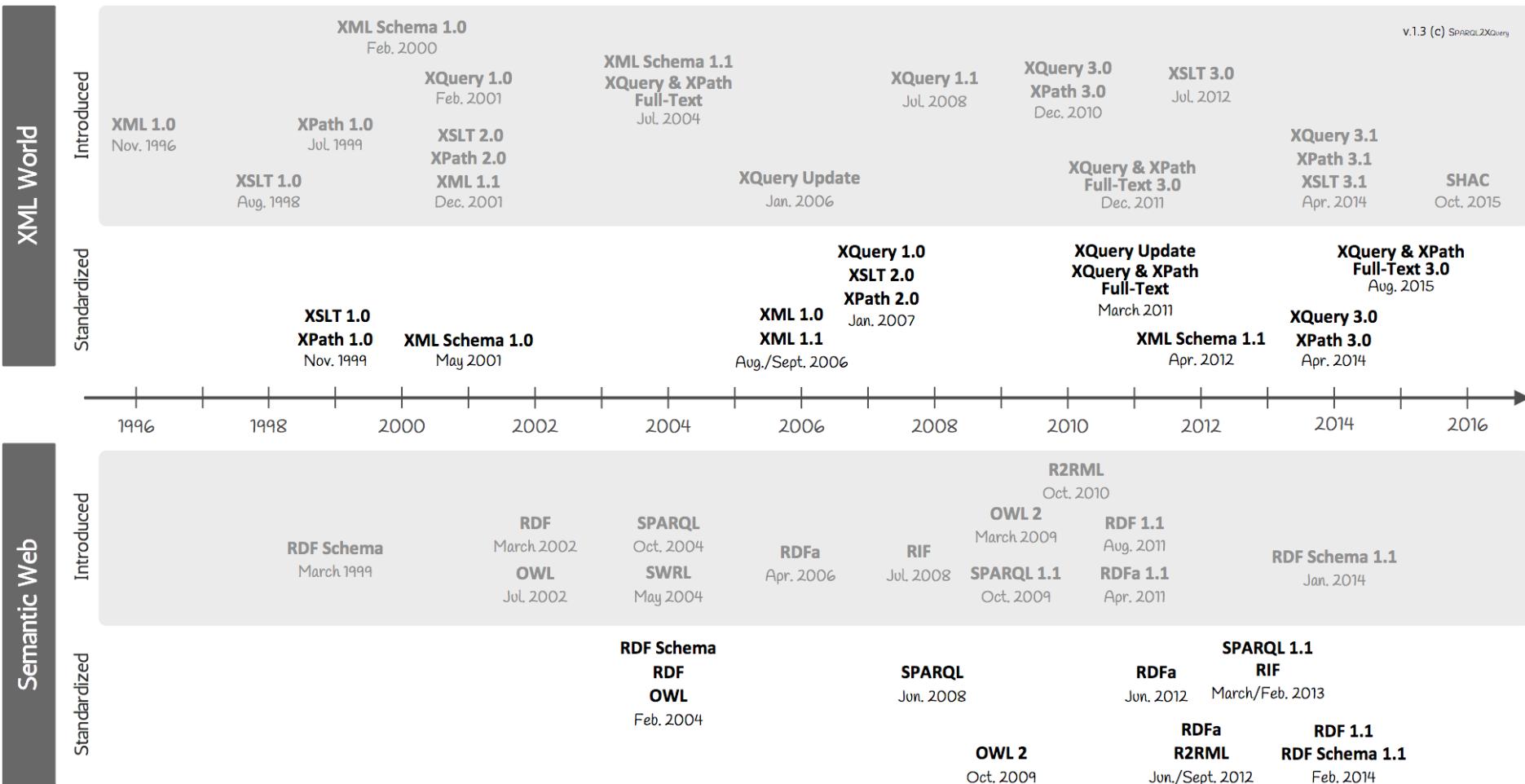


"Now! – That should clear up a few things around here!"

Semantic Web Stack



Evolution of Technologies



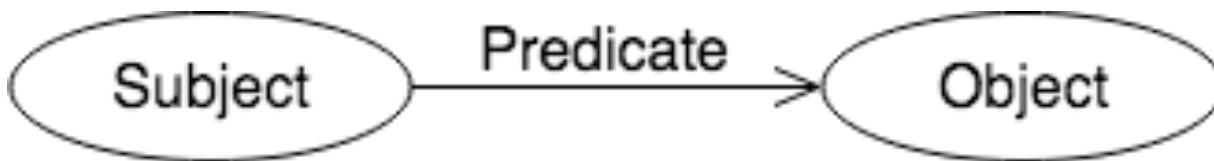
This work is available under a CC BY-SA license. This means you can use/modify/extend it under the condition that you give proper attribution.
Please cite as: Bikakis N., Tsinarakis C., Gioldasis N., Stavrakarakis I., Christodoulakis S.: "The XML and Semantic Web Worlds: Technologies, Interoperability and Integration. A Survey of the State of the Art" In Semantic Hyper/Multi-media Adaptation: Schemas and Applications, Springer 2013.

Semantic Web Today and En Route

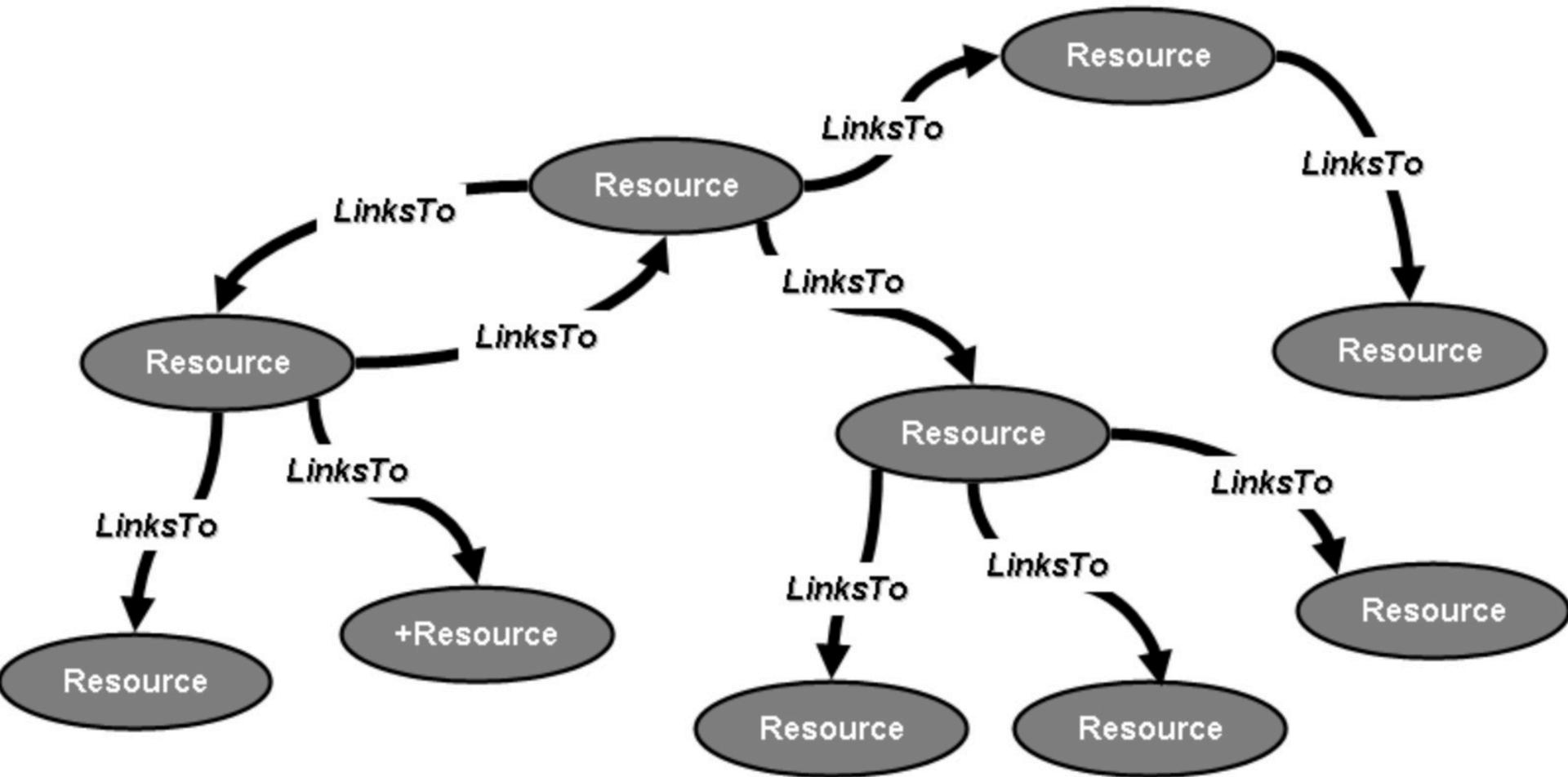
- Semantic Web helps build a technology stack to support a “**Web of data**”.
- The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network.
- The term “Semantic Web” now refers to W3C’s vision of the **Web of linked data**.
- Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data.

Resource Description Framework (RDF)

- RDF is a framework for representing information in the Web.
- Facilitate data merging even if the underlying schemas differ.
- The core structure of the abstract syntax is a set of **triples**, each consisting of a **subject**, a **predicate** and an **object**.
- A set of such triples is called an RDF **graph**. Each triple is represented as a node-arc-node link.

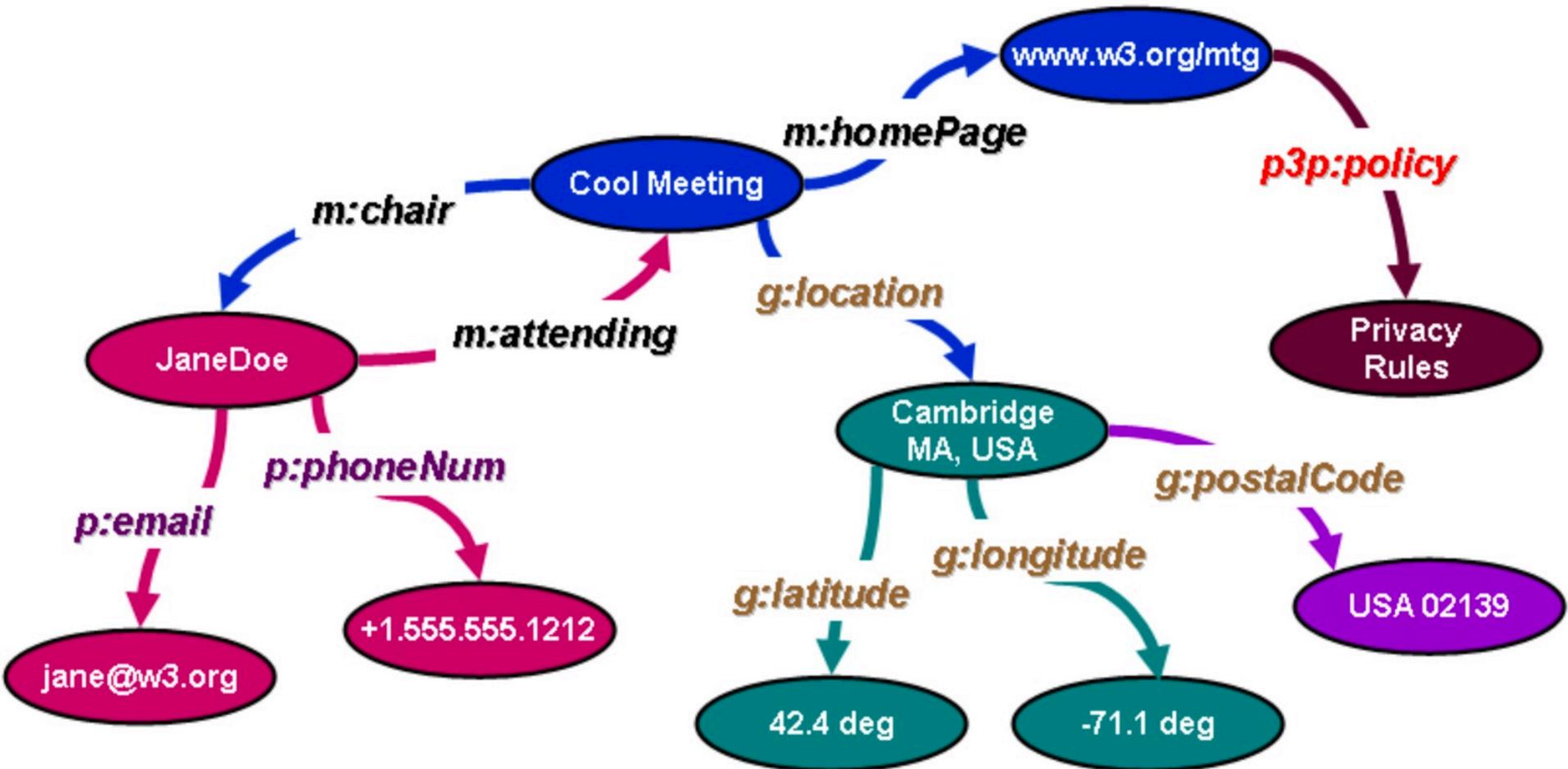


Most of the Current Web



<https://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/>

With RDF Data



<https://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/>

RDFS and OWL 2

- RDFS: Resource Description Framework Schema
 - Provides basic capabilities for describing RDF vocabularies
- OWL: Web Ontology Language
 - Built on top of RDFS and RDF
 - Provides additional capabilities in knowledge representation
- OWL is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things.
 - a computational logic-based language
 - knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit.
 - RDF and OWL documents, known as **ontologies**

Ontology

- “An ontology is an explicit specification of a conceptualisation.”
- “While a conceptual schema defines relations on data, an ontology defines terms to represent knowledge.”
 - Data: ground atomic facts
 - Knowledge: expressible in logical sentences with existentially and universally quantified variables.
- In the context of Semantic Web, ontology and knowledge base sometimes are used interchangeably (but with some subtle differences).

Linked Data

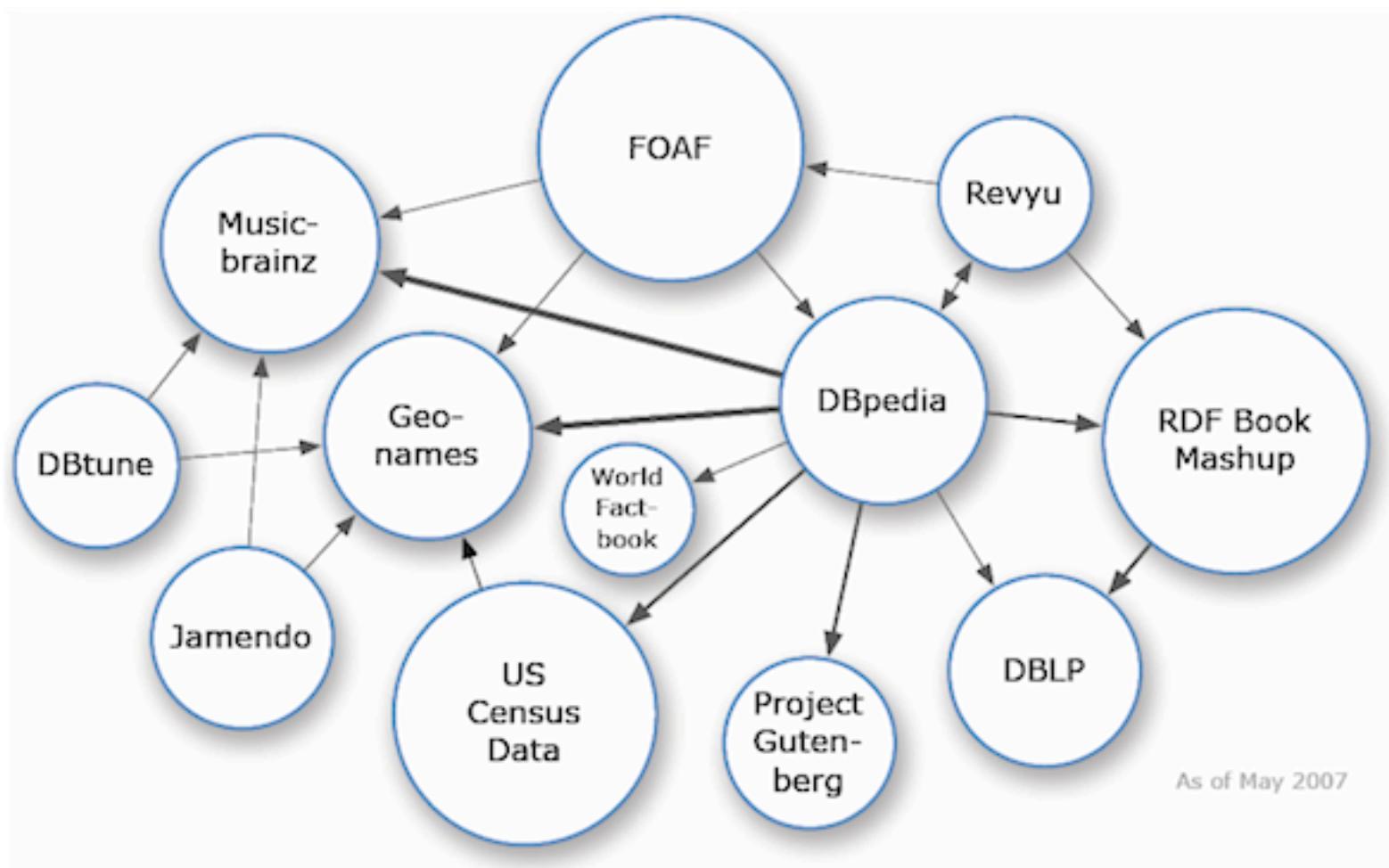
- huge amount of data on the Web available in a standard format
- reachable and manageable by Semantic Web tools
- relationships among data should be made available
- collection of interrelated datasets on the Web is also referred to as **linked data**.
- makes easy either conversion or on-the-fly access to existing databases (relational, XML, HTML, etc).
- setup query endpoints to access that data more conveniently
 - W3C provides a palette of technologies (RDF, GRDDL, POWDER, RDFa, the upcoming R2RML, RIF, SPARQL) to get access to the data.

Linked Data - Connect Distributed Data across the Web

- Linked data design principles:
 - Use URIs as names for things
 - Use HTTP URIs so that people can look up those names.
 - When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
 - Include links to other URIs, so that they can discover more things.
- *Tutorials on publishing linked data available in the references at the end of lecture.*

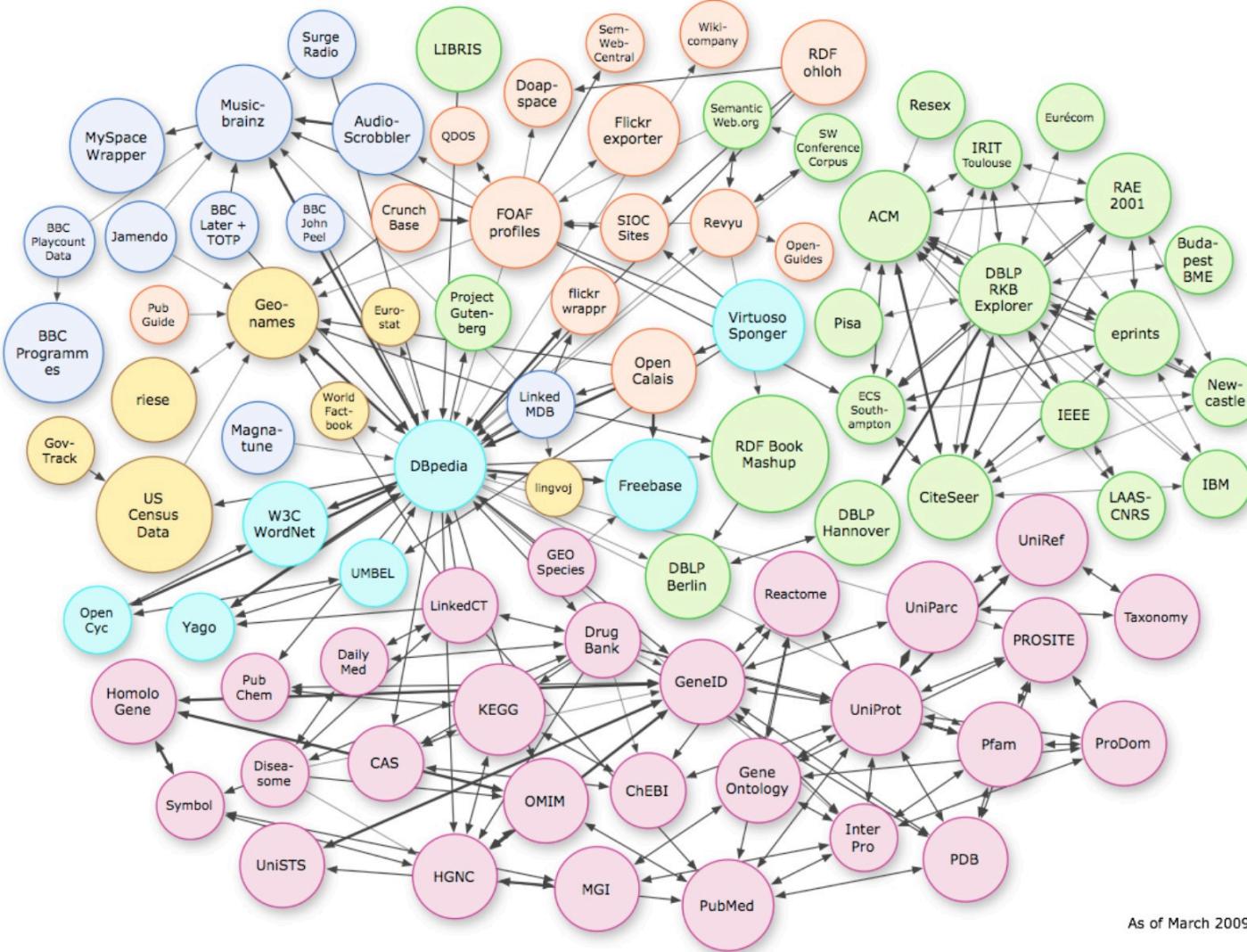
<https://www.w3.org/DesignIssues/LinkedData.html>

Linked Open Data by 2007



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

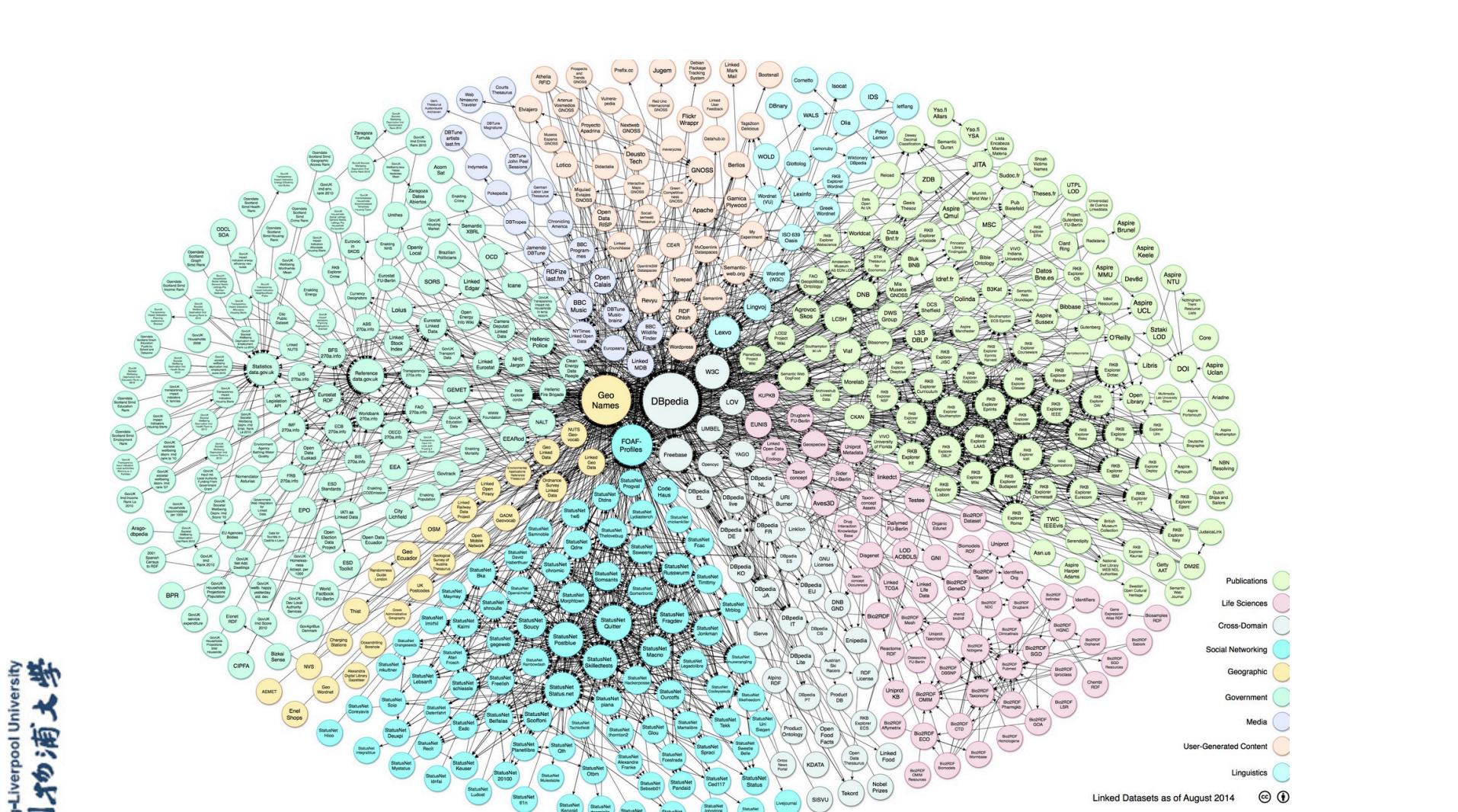
Linked Open Data by 2009



As of March 2009

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

Linked Open Data by 2014



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

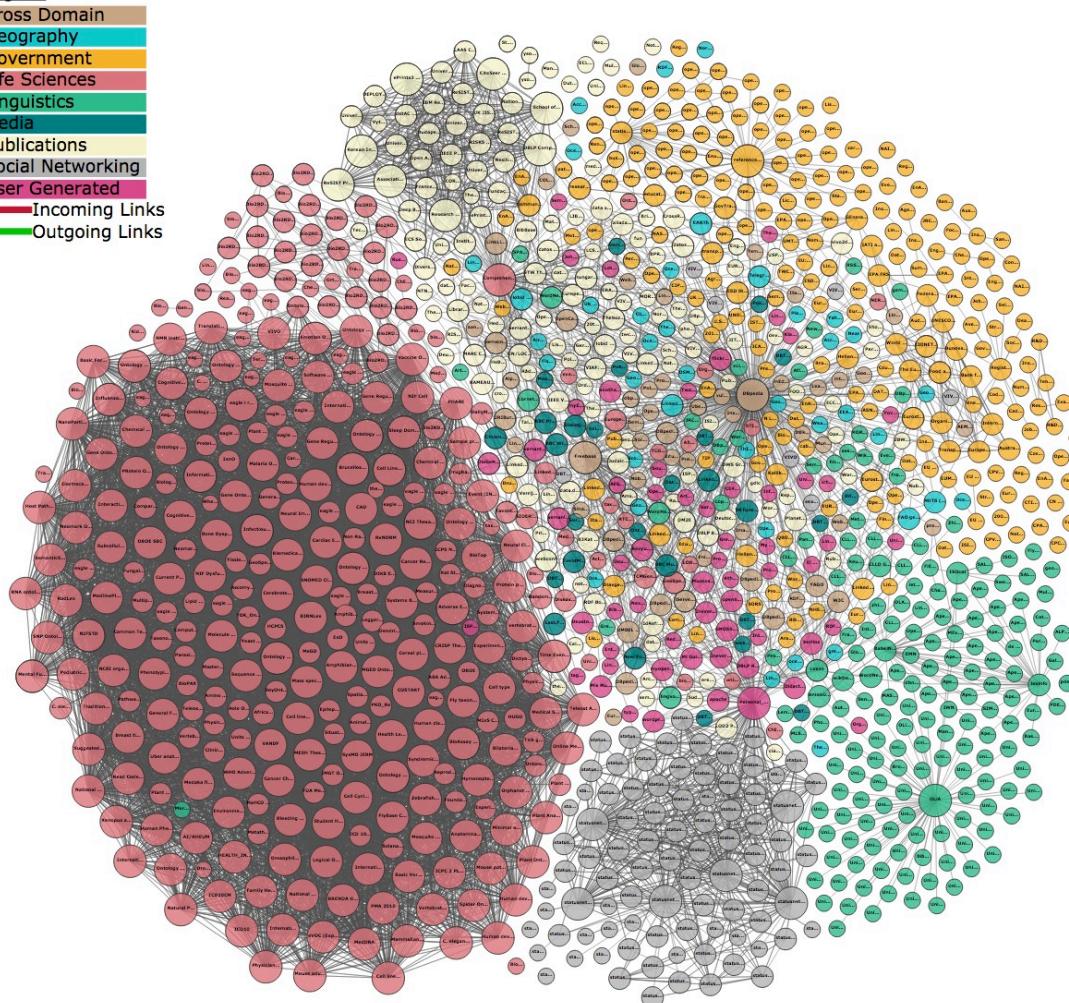
Linked Datasets as of August 2014



Linked Open Data by 2017

Legend

Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
— Incoming Links
— Outgoing Links



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

SPARQL: Query the Web of Data

- SPARQL can be used to express queries across diverse data sources
 - whether the data is stored natively as RDF or viewed as RDF via middleware.
- SPARQL contains capabilities for querying required and optional graph patterns
 - along with their conjunctions and disjunctions.
- SPARQL also supports extensible value testing and constraining queries by source RDF graph.
- The results of SPARQL queries can be results sets or RDF graphs.

SPARQL: Query the Web of Data cont'd

- SPARQL queries are based on (triple) patterns.
 - RDF can be seen as a set of relationships among resources (i.e., RDF triples);
 - SPARQL queries provide one or more patterns against such relationships.
 - These triple patterns are similar to RDF triples, except that one or more of the constituent resource references are variables.
- A SPARQL engine would return the resources for all triples that match these patterns.
 - A tutorial for querying semantic Web data:
<http://www.linkeddatatools.com/querying-semantic-data>

Using SPARQL to query the DBpedia 1

- DBpedia SPARQL endpoint:

<http://dbpedia.org/sparql>

- Retrieve country instances:

SELECT ?country

WHERE

{?country rdf:type dbo:Country}

} LIMIT 100

Using SPARQL to query the DBpedia 1 - Results

dbpedia.org/sparql?default-graph-uri=http%3A%2F%2Fdbpedia.org&query=%0D%0A%09
[http://dbpedia.org/resource/Salian dynasty](http://dbpedia.org/resource/Salian_dynasty)
[http://dbpedia.org/resource/Severan dynasty](http://dbpedia.org/resource/Severan_dynasty)
[http://dbpedia.org/resource/Shang dynasty](http://dbpedia.org/resource/Shang_dynasty)
[http://dbpedia.org/resource/Shiva \(Judaism\)](http://dbpedia.org/resource/Shiva_(Judaism))
[http://dbpedia.org/resource/Song dynasty](http://dbpedia.org/resource/Song_dynasty)
[http://dbpedia.org/resource/South-West Africa](http://dbpedia.org/resource/South-West_Africa)
[http://dbpedia.org/resource/Stone Age](http://dbpedia.org/resource/Stone_Age)
<http://dbpedia.org/resource/Syldavia>
[http://dbpedia.org/resource/Tang dynasty](http://dbpedia.org/resource/Tang_dynasty)
<http://dbpedia.org/resource/Tenochtitlan>
[http://dbpedia.org/resource/Thirteen Colonies](http://dbpedia.org/resource/Thirteen_Colonies)
<http://dbpedia.org/resource/Triassic>
[http://dbpedia.org/resource/Umayyad Caliphate](http://dbpedia.org/resource/Umayyad_Caliphate)
[http://dbpedia.org/resource/United Kingdom of the Netherlands](http://dbpedia.org/resource/United_Kingdom_of_the_Netherlands)
[http://dbpedia.org/resource/United Nations](http://dbpedia.org/resource/United_Nations)
[http://dbpedia.org/resource/United Nations Interim Administration Mission in Kosovo](http://dbpedia.org/resource/United_Nations_Interim_Administration_Mission_in_Kosovo)

Using SPARQL to query the DBpedia 1 – Results cont'd



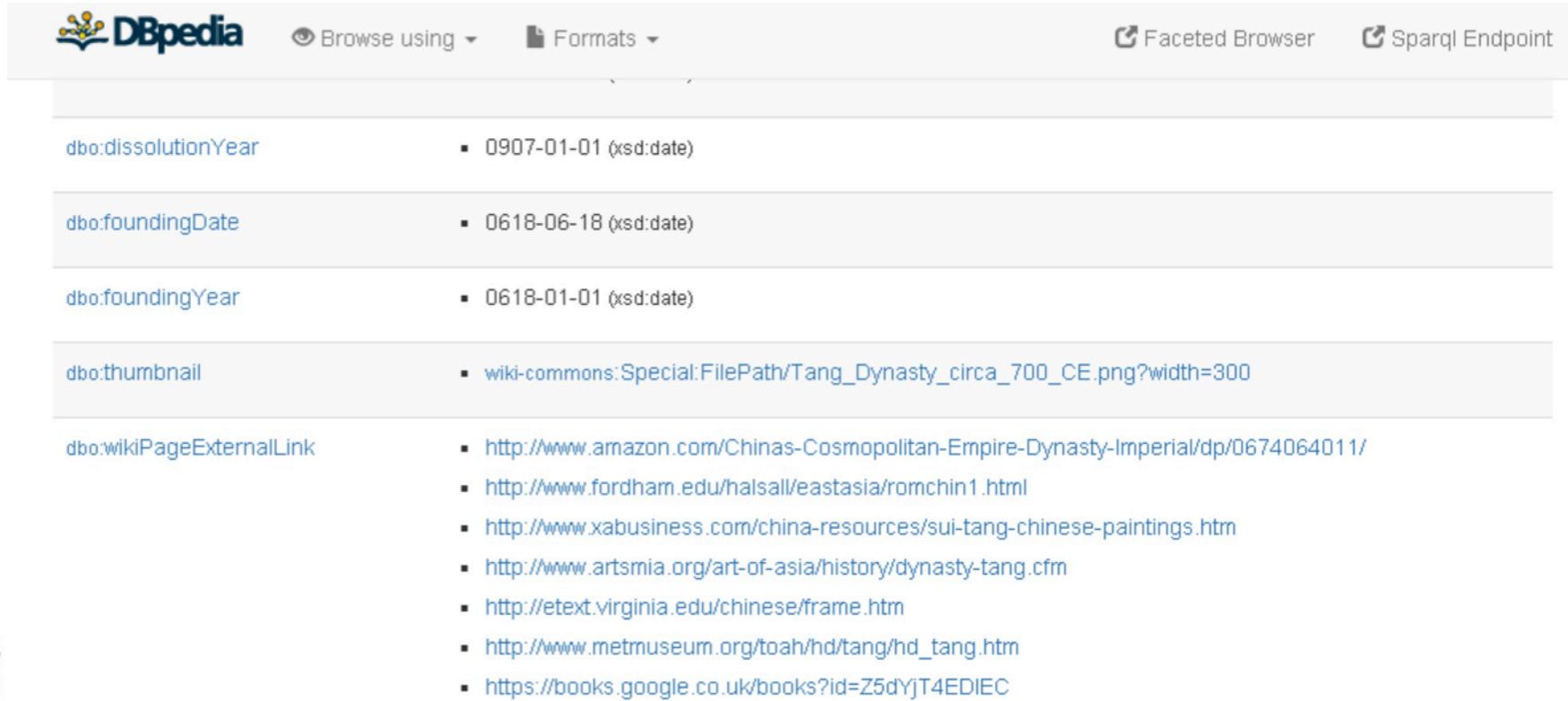
The screenshot shows the DBpedia homepage. At the top, there's a navigation bar with the DBpedia logo, a "Browse using" dropdown, a "Formats" dropdown, a "Faceted Browser" link, and a "Sparql Endpoint" link.

About: 唐朝

An Entity of Type : [populated place](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

“唐”重定向至此。關於唐和唐朝的其他意思，詳見唐 (消歧義)和唐朝 (消歧義)。唐朝（618年-907年）共歷289年，21位皇帝。由唐高祖李淵所建立，與隋朝合稱隋唐。唐室出身自關隴世族，先祖李虎在南北朝的西魏是八柱國之一，封為唐國公。其後代李淵為隋朝晉陽（在今山西太原西南）留守，在隋末民變時出兵入關中以爭奪天下，於618年受隋恭帝楊侑禪位建國唐朝，在唐朝統一戰爭中統一了天下。唐朝定都長安（今陝西西安）。並設東都洛陽、北都晉陽等陪都。唐朝的疆域廣大但時常變動，630年就超過隋朝極盛時的版圖。唐朝也是自秦漢以來，第一個不使用前朝所築長城及不築長城的統一王朝。其鼎盛時為7世紀，當時中亞的綠洲地帶受唐朝支配。其最大範圍南至羅伏州（今越南河靜）、北括玄闕州（今俄羅斯安加拉河流域）、西及安息州（今烏茲別克斯坦布哈拉）、東臨哥勿州（今吉林通化）的遼闊疆域，國土面積達1076萬平方公里。中唐後漠北、西域的領地相繼失去，到晚唐時衰退到等同中國本土的大小，但仍然保有河套地區及河西走廊。唐代天寶十三年（754年）戶口統計為五千二百八十八萬四百八十八人，不過許多學者考慮到當時統計不嚴，存在大量沒有計入統計的瞞報戶

Using SPARQL to query the DBpedia 1 – Results cont'd



The screenshot shows a search results page for the DBpedia query. The top navigation bar includes links for 'Browse using' (with a magnifying glass icon), 'Formats' (with a document icon), 'Faceted Browser' (with a magnifying glass icon), and 'Sparql Endpoint' (with a gear icon). The main content area displays the following data:

dbo:dissolutionYear	▪ 0907-01-01 (xsd:date)
dbo:foundingDate	▪ 0618-06-18 (xsd:date)
dbo:foundingYear	▪ 0618-01-01 (xsd:date)
dbo:thumbnail	▪ wiki-commons:Special:FilePath/Tang_Dynasty_circa_700_CE.png?width=300
dbo:wikiPageExternalLink	<ul style="list-style-type: none">▪ http://www.amazon.com/Chinas-Cosmopolitan-Empire-Dynasty-Imperial/dp/0674064011/▪ http://www.fordham.edu/halsall/eastasia/romchin1.html▪ http://www.xabusiness.com/china-resources/sui-tang-chinese-paintings.htm▪ http://www.artsmia.org/art-of-asia/history/dynasty-tang.cfm▪ http://etext.virginia.edu/chinese/frame.htm▪ http://www.metmuseum.org/toah/hd/tang/hd_tang.htm▪ https://books.google.co.uk/books?id=Z5dYjT4EDIEC

Using SPARQL to query the DBpedia 2

- Retrieve artist instances whose birth place is a country contains the string of “United Kingdom”.

PREFIX dbprop: <<http://dbpedia.org/property/>>

PREFIX dbpedia-owl: <<http://dbpedia.org/ontology/>>

SELECT ?artist ?place

WHERE

{

?artist rdf:type dbpedia-owl:Artist.

?artist dbprop:birthPlace ?place.

?place rdf:type dbpedia-owl:Country.

?place rdfs:label ?label.

FILTER regex(?label, "United Kingdom").

}

Using SPARQL to query the DBpedia 2 – results

artist	place
http://dbpedia.org/resource/Gary_Frank	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/The Etherington Brothers	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Reg Smythe	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Ed Furness	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Rich Johnston	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Alan Davis	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Ho Che Anderson	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Rhydian Vaughan	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Simon Oliver	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Mike McMahon (comics)	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Nick Landau	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Steve MacManus	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Steve Moore (comics)	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Alan Martin (writer)	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Rufus Dayglo	http://dbpedia.org/resource/United_Kingdom
http://dbpedia.org/resource/Jason Chan Chi-san	http://dbpedia.org/resource/United_Kingdom

Using SPARQL to query the DBpedia 2 – results cont'd



Browse using ▾

Formats ▾

Faceted Browser

dbo:birthDate	<ul style="list-style-type: none">▪ 1972-11-21 (xsd:date)
dbo:birthPlace	<ul style="list-style-type: none">▪ dbr:Gloucester
dbo:imdbId	<ul style="list-style-type: none">▪ 1262679
dbo:nationality	<ul style="list-style-type: none">▪ dbr:British_citizenship
dbo:thumbnail	<ul style="list-style-type: none">▪ wiki-commons:Special:FilePath/Rich_Johnston,_2007.jpg?width=300
dbo:wikiPageExternalLink	<ul style="list-style-type: none">▪ http://www.dynamicforces.com/htmlfiles/tommy.html?showhistory=ok▪ http://www.sequart.com/interviews/index.php?interview=740▪ http://www.comicbookresources.com/?page=column&id=11▪ http://www.richandmark.com▪ http://www.2000adreview.co.uk/features/interviews/2006/johnston/rich-johnston.shtml▪ http://www.bleedingcool.com

End of Lecture

- Summary
 - Semantic Web, ontologies
 - RDF, OWL
 - Linked Open Data
 - Query the Web of Data
- Reading
 - See references next slides

References

- Berners-Lee, T.; Hendler, J.; Lassila, O. (2001). "The Semantic Web". *Scientific American*. 284(5): 34.
- Nigel Shadbolt; Wendy Hall; Tim Berners-Lee (2006). "The Semantic Web Revisited" (PDF). *IEEE Intelligent Systems*. Retrieved April 13, 2007.
- Lee Feigenbaum (May 1, 2007). "The Semantic Web in Action". *Scientific American*. Retrieved February 24, 2010.
- RDF premier, <https://www.w3.org/TR/rdf11-concepts/>
- Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199-220.
- Linked data, <https://www.w3.org/standards/semanticweb/data>
- Semantic Web use cases and case studies, <https://www.w3.org/2001/sw/sweo/public/UseCases/>
- Linked data design issues, <https://www.w3.org/DesignIssues/LinkedData.html>
- Tom Heath and Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space*, <http://linkeddatabook.com/>
- Chris Bizer, Richard Cyganiak, How to Publish Linked Data on the Web (Tutorial), <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/>
- SPARQL, <https://www.w3.org/2001/sw/wiki/SPARQL>
- Querying Semantic Data, <http://www.linkeddatatools.com/querying-semantic-data>