Knowledge Representation and Semantic Technologies

Introduction to the Semantic Web

Riccardo Rosati

Corso di Laurea Magistrale in Ingegneria Informatica
Sapienza Università di Roma
2015/2016

What is the Semantic Web?

- "The Semantic Web is a Web of actionable information—information derived from data through a semantic theory for interpreting the symbols."
- "The semantic theory provides an account of 'meaning' in which the logical connection of terms establishes interoperability between systems"

(Shadbot, Hall, Berners-Lee, The Semantic Web revisited, IEEE Intelligent Systems, May 2006)

The Semantic Web: why?

- search on the Web: problems...
- ...due to the way in which information is stored on the Web
- **Problem 1**: web documents do not distinguish between information content and presentation ("solved" by XML)
- **Problem 2**: different web documents may represent in different ways semantically related pieces of information
- this leads to hard problems for "intelligent" information search on the Web

Separating content and presentation

Problem 1: web documents do not distinguish between information content and presentation

- problem due to the HTML language
- problem "solved" by current technology
  - stylesheets (HTML, XML)
  - XML
- stylesheets allow for separating formatting attributes from the information presented

Separating content and presentation

- XML: eXtensible Mark-up Language
- XML documents are written through a user-defined set of tags
- tags are used to express the “semantics” of the various pieces of information

XML: example

HTML:

```
<H1>Seminari di Ingegneria del Software</H1>
  <UL>
    <LI>Teacher: Giuseppe De Giacomo
    <LI>Room: 7
    <LI>Prerequisites: none
  </UL>
```

XML:

```
<course>
  <title>Seminari di Ingegneria del Software
  </title>
  <teacher>Giuseppe De Giacomo</teacher>
  <room>1AI, 1I</room>
  <prereq>none</prereq>
</course>
```

Limitations of XML

XML does not solve all the problems:

- legacy HTML documents
- different XML documents may express information with the same meaning using different tags

The need for a “Semantic” Web

Problem 2: different web documents may represent in different ways semantically related pieces of information

- different XML documents do not share the “semantics” of information
- idea: annotate (mark-up) pieces of information to express the “meaning” of such a piece of information
- the meaning of such tags is shared!

  $\Rightarrow$ shared semantics

The Semantic Web initiative

viewpoint:

the Web = a web of data

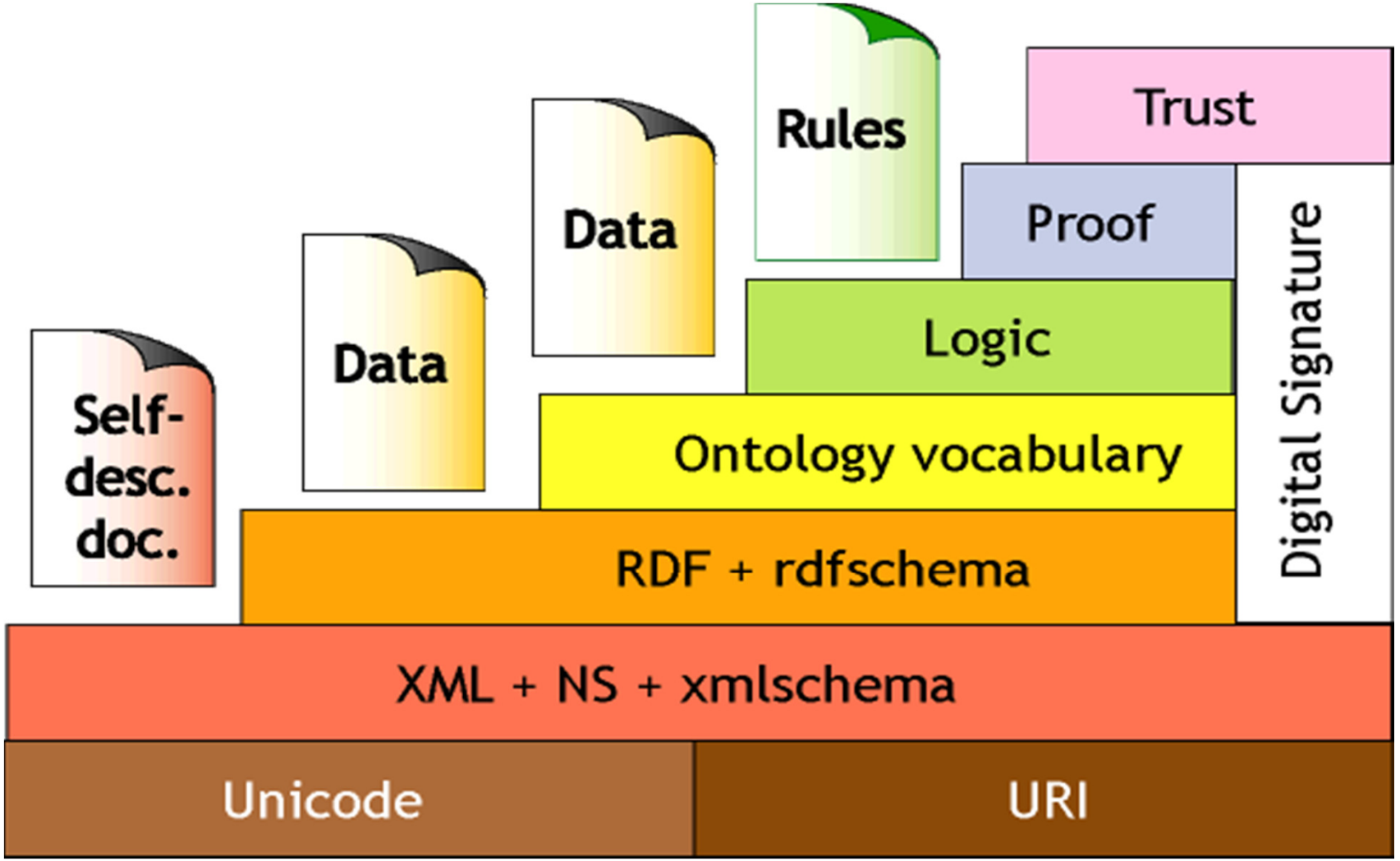
goal:

to provide a common framework to share data on the Web across application boundaries

main ideas:

- ontology
- standards
- “layers”

The Semantic Web Tower



The Semantic Web Layers

- XML layer
- RDF + RDFS layer
- Ontology layer
- Proof-rule layer
- Trust layer

The XML layer

- XML (eXtensible Markup Language)
  - user-definable and domain-specific markup
- URI (Uniform Resource Identifier)
  - universal naming for Web resources
  - same URI = same resource
  - URIs are the “ground terms” of the SW
- W3C standards

The RDF + RDFS layer

RDF = a simple conceptual data model

W3C standard (1999)

RDF model = set of RDF triples

triple = expression (statement)

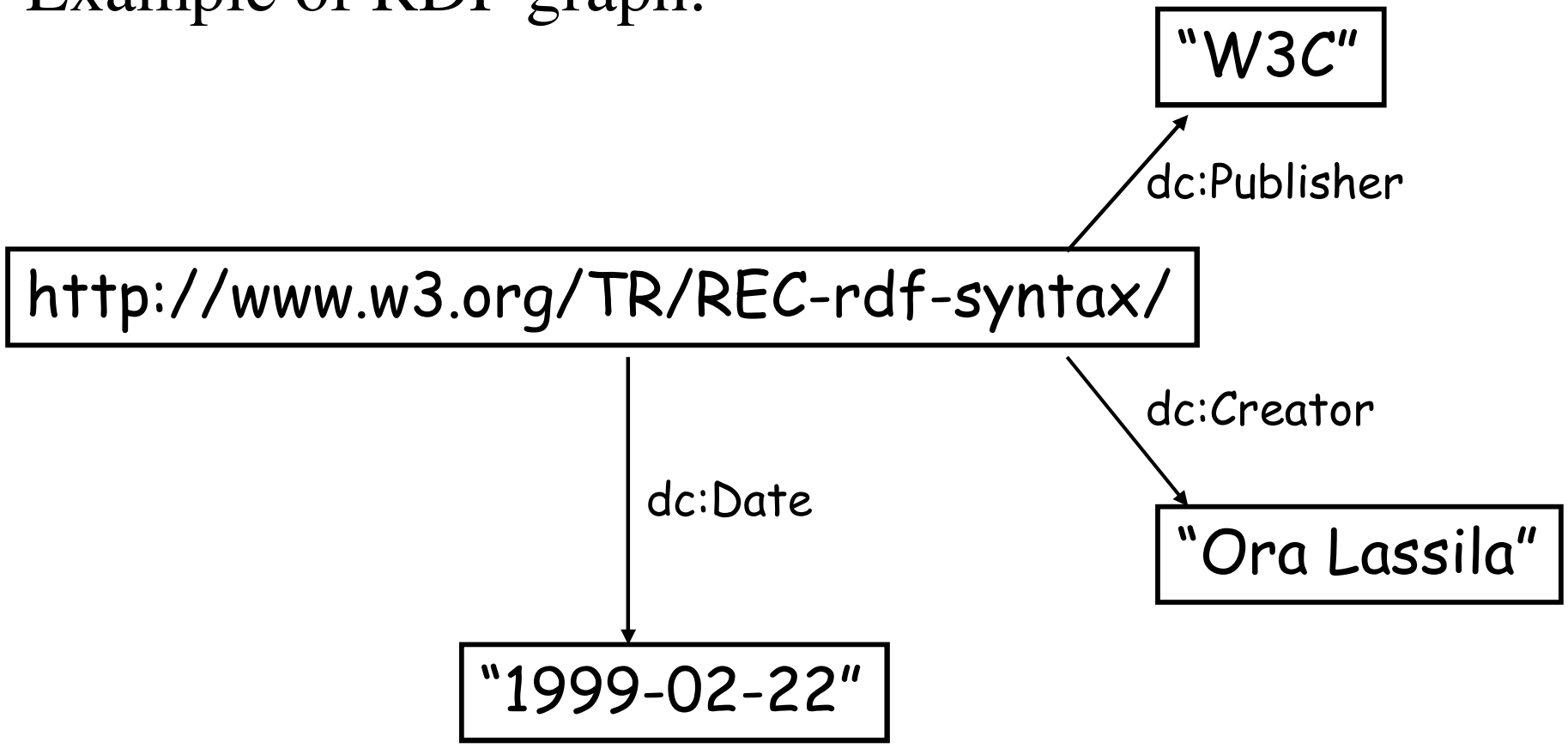
(subject, predicate, object)

- subject = resource
- predicate = property (of the resource)
- object = value (of the property)

=> an RDF model is a **graph**

The RDF + RDFS layer

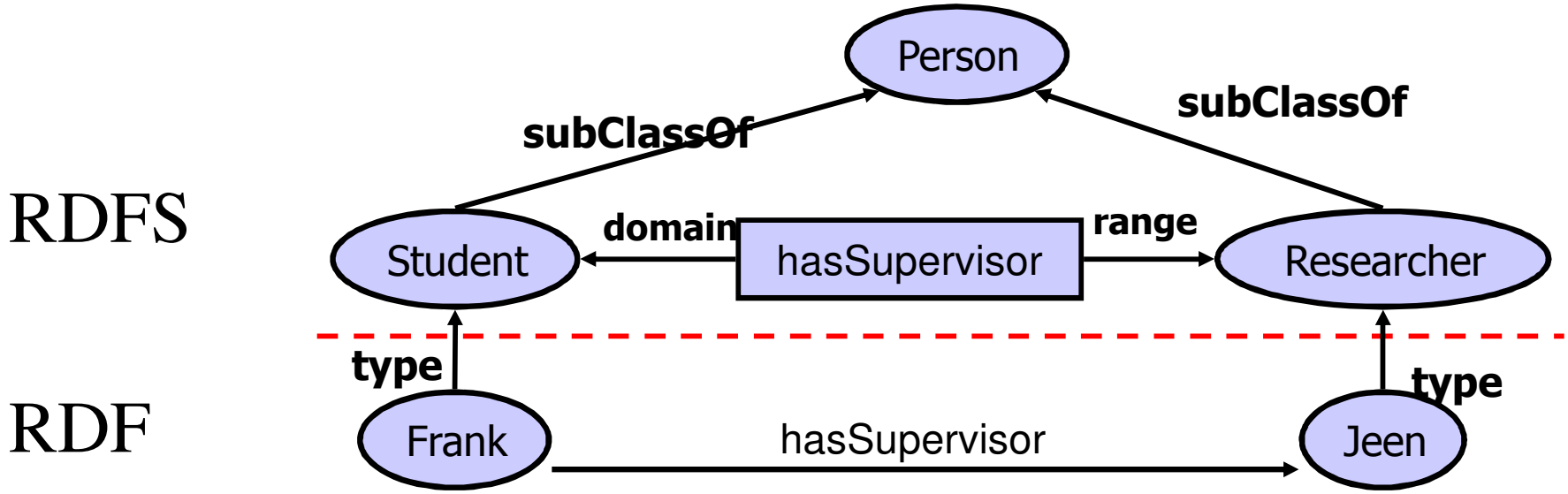
Example of RDF graph:



The RDF + RDFS layer

- RDFS = RDF Schema
- “vocabulary” for RDF
- W3C standard (2004)

example:



The Ontology layer

ontology = shared conceptualization

$\Rightarrow$ conceptual model

(more expressive than RDF + RDFS)

$\Rightarrow$ expressed in a true knowledge representation language

OWL (Web Ontology Language) = standard language for ontologies

The proof/rule layer

beyond OWL:

- proof/rule layer
- rule: informal notion
- rules are used to perform inference over ontologies
- rules as a tool for capturing further knowledge (not expressible in OWL ontologies)

The Trust layer

- SW top layer:
- support for provenance/trust
- provenance:
  - where does the information come from?
  - how this information has been obtained?
  - can I trust this information?
- largely unexplored issue
- no standardization effort

The Semantic Web: main ingredients

- underlying web layer (URI, XML)
  - reusing and extending web technologies
- basic conceptual modeling language (RDF)
- ontology language (OWL)
- rules/proof
- reusing and extending AI technologies
  - knowledge representation
  - automated reasoning
- ...and database technologies
  - data integration

The notion of ontology

- ontology = shared conceptualization of a domain of interest
- shared vocabulary => simple (shallow) ontology
- (complex) relationships between “terms” => deep ontology
- AI view:
  - ontology = logical theory (knowledge base)
- DB view:
  - ontology = conceptual model

Ontologies: example

```
class-def animal                          % animals are a class
class-def plant                           % plants are a class
    subclass-of NOT animal                % that is disjoint from animals
class-def tree
    subclass-of plant                     % trees are a type of plants
class-def branch
    slot-constraint is-part-of            % branches are parts of some tree
        has-value tree
        max-cardinality 1
class-def defined carnivore               % carnivores are animals
    subclass-of animal
    slot-constraint eats                  % that eat any other animals
        value-type animal
class-def defined herbivore               % herbivores are animals
    subclass-of animal, NOT carnivore % that are not carnivores, and
    slot-constraint eats                      % they eat plants or parts of plants
        value-type plant OR (slot-constraint is-part-of has-value plant)
```

Ontologies: the role of logic

- ontology = logical theory
- why?
  - declarative
  - formal semantics
  - reasoning (sound and complete inference techniques)
- well-established correspondence between conceptual modeling formalisms and logic

Ontologies and Description Logics

- OWL is based on a fragment of first-order predicate logic (FOL)
- Description Logics (DLs) = subclasses of FOL
  - only unary and binary predicates
  - function-free
  - quantification allowed only in restricted form
  - (variable-free syntax)
  - decidable reasoning

- DLs are one of the most prominent languages for Knowledge Representation

Ontologies and Description Logics

- expressive abilities of DLs have been widely explored
- reasoning in DLs has been extensively studied
- DL reasoners have been developed and optimized

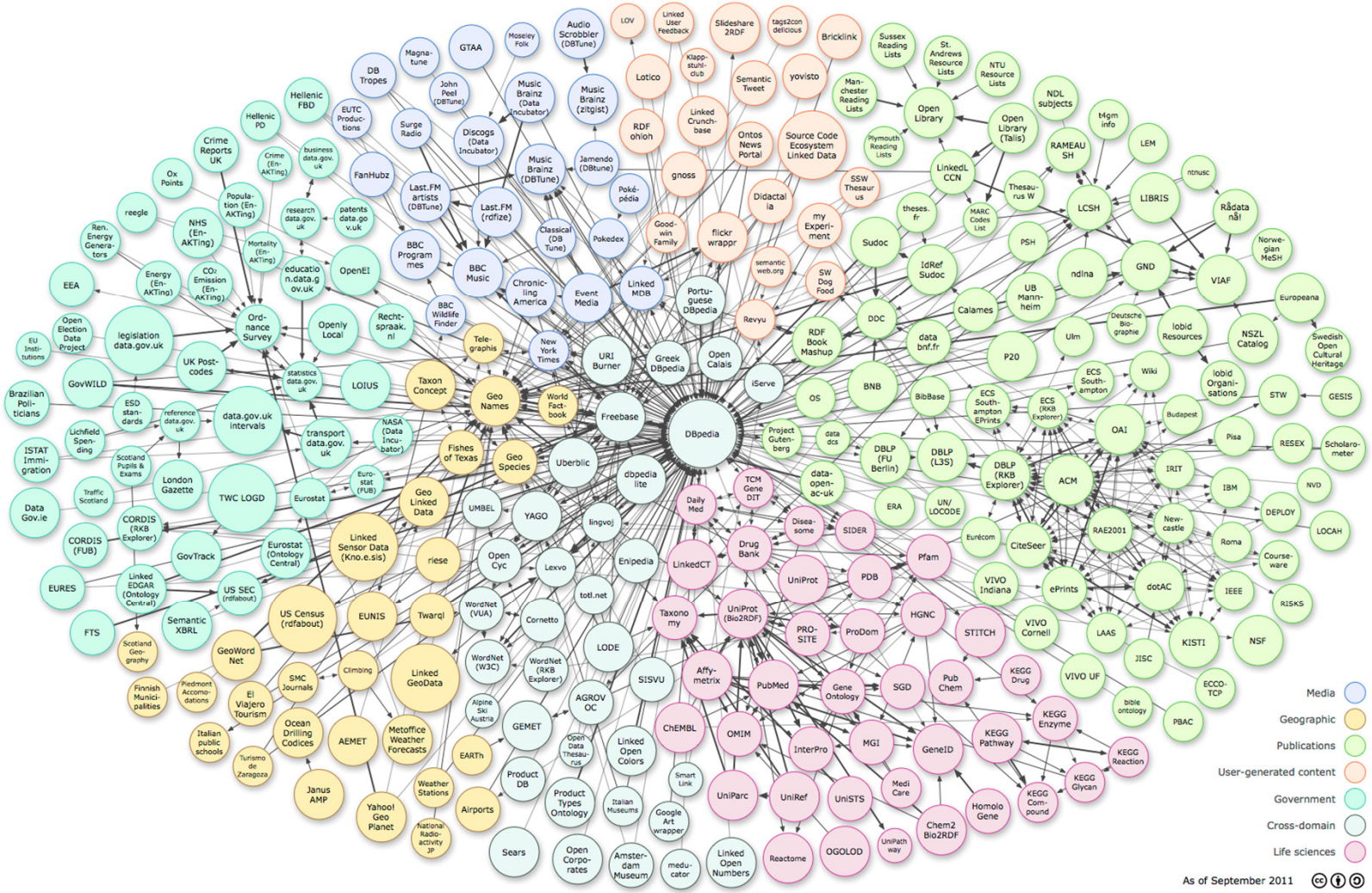
$\Rightarrow$ DLs as a central technology for the SW

Rule-based formalisms

- Prolog
- Logic programming
- Constraint (logic) programming
- Production rules
- Datalog
- ...

RIF (Rule Interchange Format): W3C recommendation (2010)

The Semantic Web in the real world



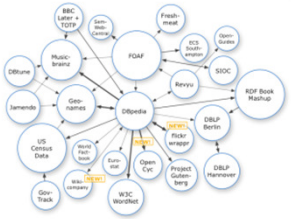
Linking Open Data cloud diagram, 09/2011 (by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/)

Linked Data

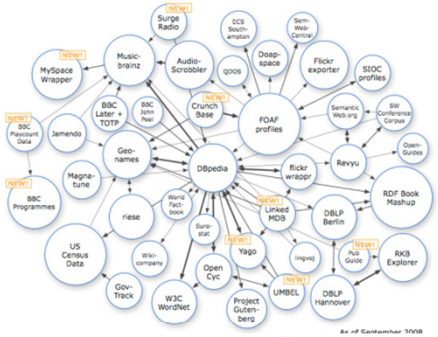
Linked Data: a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF

Linking Open Data (LOD): "The goal of the W3C SWEO Linking Open Data community project is to **extend the Web with a data commons by publishing various open data sets as RDF on the Web** and by **setting RDF links between data items from different data sources**.
RDF links enable you to navigate from a data item within one data source to related data items within other sources using a Semantic Web browser.
As query results are structured data and not just links to HTML pages, they can be used within other applications."

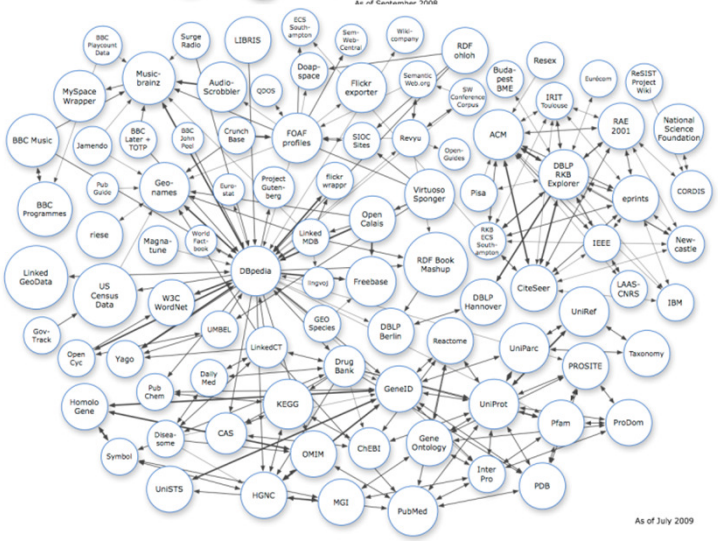
The LOD cloud diagram



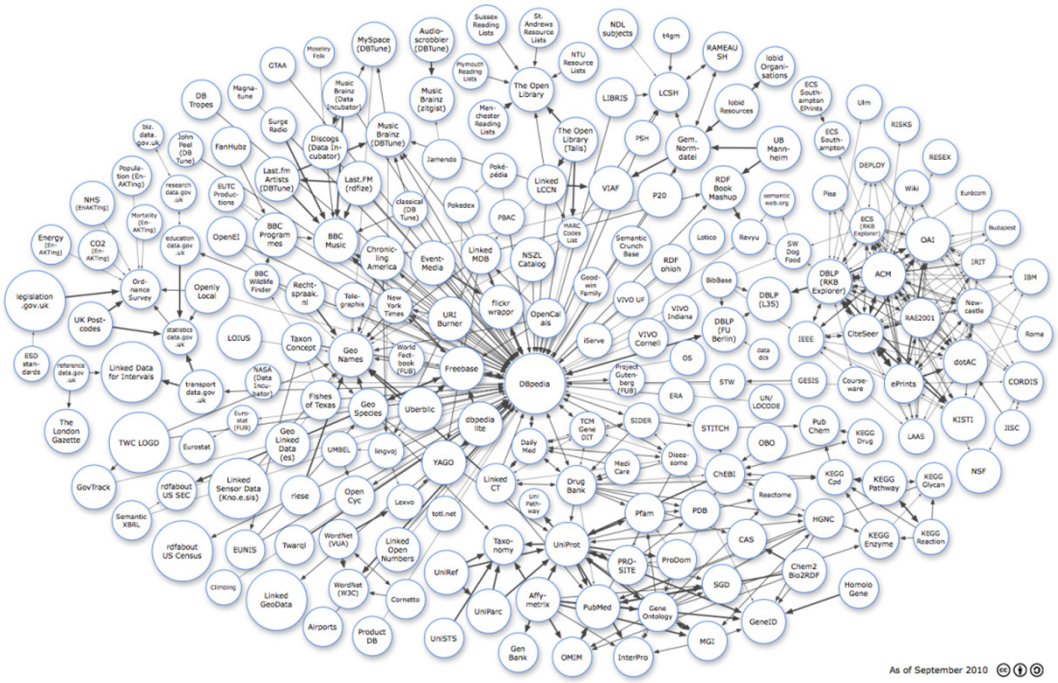
9/2007



9/2008



9/2009



9/ 2010