

# Informatica Applicata

## 3.1 Introduzione al Web Semantico

**Antonella Poggi**

Anno Accademico 2012-2013

DIPARTIMENTO DI SCIENZE DOCUMENTARIE  
LINGUISTICO FILOLOGICHE  
E GEOGRAFICHE



**SAPIENZA**  
UNIVERSITÀ DI ROMA

## 3.1 Introduzione al Web Semantico\*

- **motivazioni**
- l'iniziativa del Web Semantico
- l'architettura
- i Linked Open Data

\* Le slide sul Web Semantico sono tratte da una sezione dedicata del corso “Complementi di Software e Servizi”, tenuta dal prof. Rosati nell'ambito del corso di Laurea Magistrale in Ingegneria Informatica

## Cos'è il Web Semantico?

- “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”

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

## Il Web Semantico: perché?

- La ricerca delle informazioni nel web è molto meno semplice ed immediata di quanto si vorrebbe
- I problemi sono dovuti alla maniera in cui l'informazione è memorizzata nel Web
- **Problema 1:** i documenti web non distinguono tra il contenuto delle informazioni e la presentazione delle stesse (“risolto” da XML)
- **Problem 2:** diversi documenti possono rappresentare in diversi modi pezzi di informazioni tra loro semanticamente collegate
- problemi difficili che ostacolano la ricerca “intelligente” di informazioni nel Web

# Separare la presentazione dal contenuto

- **Problema 1:** i documenti web non distinguono tra il contenuto delle informazioni e la presentazione delle stesse (“risolto” da XML)
- dovuto al linguaggio HTML
- “risolto” dalla tecnologia attuale
  - XML: le etichette sono definite dall’utente e permettono di esprimere proprietà logiche e strutturali dell’informazione presentata
  - fogli di stile (sia per HTML che per XML): permettono di esprimere proprietà di formattazione

# Limiti di XML

XML non risolve tutti i problemi!

- documenti legacy HTML
- diversi documenti XML possono esprimere informazioni con lo **stesso significato** utilizzando **etichette diverse**, così come informazioni con **significato diverso** utilizzando le **stesse etichette**
  - i DTD/schemi XML non permettono di specificare la “semantica” ovvero il significato delle etichette, ma solo la *struttura* ammessa per i documenti

## L'esigenza di un Web “Semantico”

- **Problema 2:** diversi documenti possono rappresentare in diversi modi pezzi di informazioni tra loro semanticamente collegate
- i documenti XML non condividono la “semantica” dell'informazione
- idea: **annotare**, attraverso un linguaggio di marcatura apposito, pezzi di informazione per specificare il “significato” di tale informazione
  - ⇒ significato del nuovo linguaggio di marcatura condiviso
  - ⇒ **semantica condivisa!**

## 3.1 Introduzione al Web Semantico

- motivazioni
- **l'iniziativa del Web Semantico**
- l'architettura
- i Linked Open Data

\*



# L'iniziativa del Web Semantico

- Iniziativa condotta dal W3C
- Visione: il Web = una rete di informazioni
- Obiettivo: “progettare e sviluppare un framework comune per condividere e riutilizzare informazioni sul Web, superando i confini delle applicazioni, delle imprese e delle comunità”
- Principali idee:
  - promuovere l’inclusione di informazioni semantiche sulle pagine web, attraverso l’uso di formati di dati comuni, i.e. di **standard**
  - promuovere l’uso di **ontologie**
  - architettura a **“livelli”**

## Il Web Semantico: basi di partenza

- Tecnologie dell'intelligenza artificiale
  - rappresentazione della conoscenza
  - ragionamento automatico
- Tecnologie delle basi di dati
  - integrazione di dati

# Un pò di storia...

- Il termine viene coniato dal suo inventore, Tim-Berners Lee nel 2001, nell'articolo "The Semantic Web", *Scientific American Magazine*
  - definito come "una rete di informazioni che possono essere processate direttamente o indirettamente dalle macchine"
  - previsto come l'evoluzione dell'esistente Web
- Dal 2001, il W3C vigila sulle definizioni degli standard collegati al Web Semantico
- Diversi progetti (e finanziamenti) internazionali sono stati e sono a tutt'oggi dedicati alla sua realizzazione
- 2006: Tim Berners-Lee et al. ammettono: "questa semplice idea...è rimasta ancora ampiamente inesplorata"
- 2006: Tim Berners-Lee, in una sua nota, propone delle regole per rendere i dati del Semantic Web fra loro collegati (Linked Data), secondo il principio che una rete di dati collegati rende gli stessi dati maggiormente fruibili
- *Chi lo critica*: non è realizzabile!
- *Chi lo sostiene*: diverse applicazioni nell'ambito della biologia, delle scienze umane, dei social network (...) ne hanno già dimostrato la validità

## 3.1 Introduzione al Web Semantico

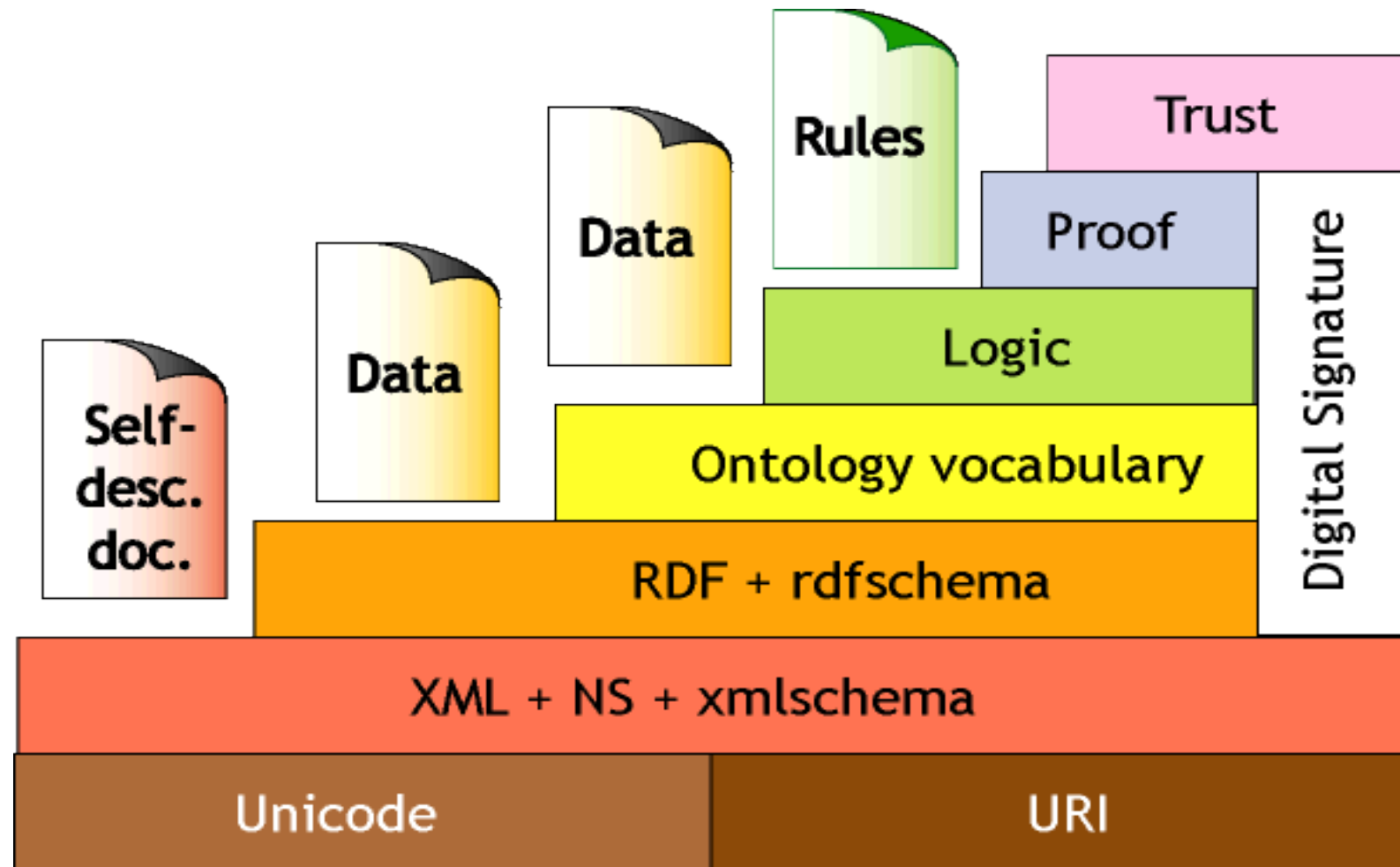
- motivazioni
- l'iniziativa del Web Semantico
- **l'architettura**
- i Linked Open Data

\*

# I linguaggi del Web Semantico: i requisiti

- Forma: hanno bisogno di avere una sintassi *ben definita*
  - per permettere agli strumenti software di determinare che costrutti usano
- Significato: hanno bisogno di una semantica *ben definita*
  - per permettere agli strumenti software di determinare cosa significano i costrutti che usano
- Importante: Il Web Semantico è rivolto alle macchine => **l'informazione deve essere “processabile” in maniera automatica!**

# The Semantic Web Tower



## Web Semantico: le basi sintattiche

- gli indirizzi **URI** e il sistema di codifica **Unicode** rappresentano modalità standard di denotare entità e simboli di scambio
- **XML** è il meta-linguaggio standard usato per la marcatura
- **XML Schema** è il linguaggio standard usato per definire la grammatica dei documenti XML legali
- il meccanismo dei **namespace** (NS) permette di combinare documenti con vocabolari eterogenei

## Web semantico: i livelli

- La torre del Web Semantico prevede più linguaggi standard, con potere espressivo crescente
  - Il livello di RDF e RDF Schema (RDFS) – per esprimere semplici metadati, ovvero informazioni semantiche di base
  - Il livello dei linguaggi per la specifica di ontologie (OWL)
  - Il livello dei linguaggi per la specifica di prove e regole
  - Il livello dei linguaggi che tengono conto dell'autorevolezza delle sorgenti di informazioni
- **N.B.** Per le scelte che sono state fatte sino ad oggi, il livello dei linguaggi basati sulla logica è stato di fatto inglobato nel livello dei linguaggi per le ontologie, in quanto OWL è basato lui stesso sulla logica

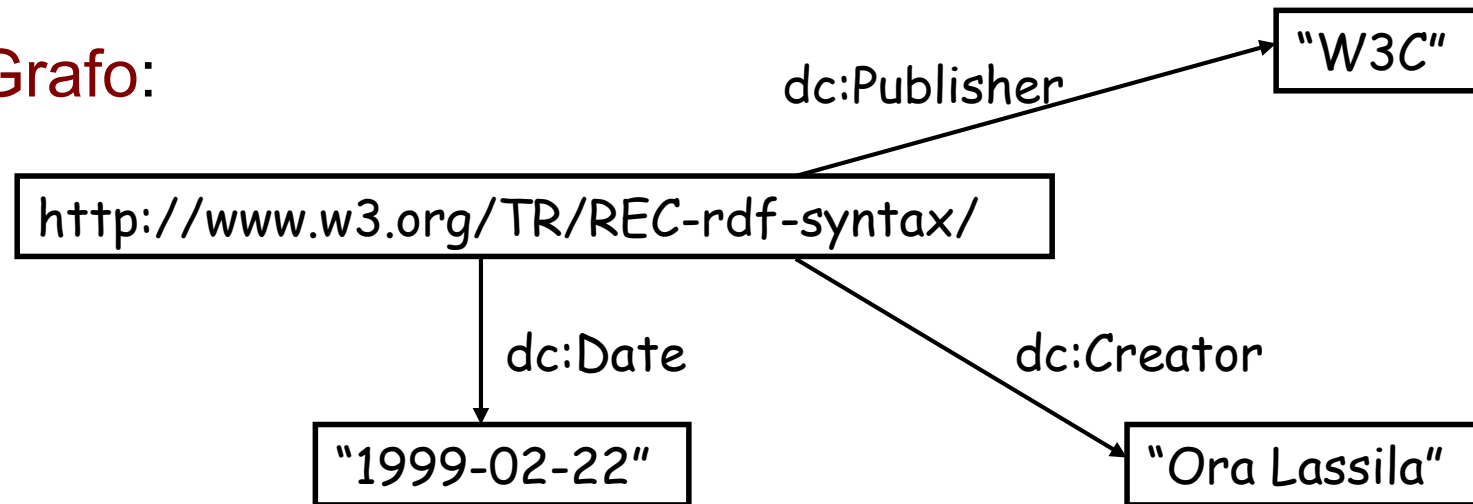


## Il livello di RDF e RDFS: RDF

- Resource Description Framework, Standard del W3C (1999)
- semplice linguaggio per la modellazione concettuale
- **Tripla RDF**: espressione (asserzione) nella forma  
(soggetto, predicato, oggetto)
  - **soggetto**: risorsa
  - **predicato**: proprietà (della risorsa)
  - **oggetto**: valore (della proprietà)
- **Modello RDF**: insieme di triple RDF
  - un modello RDF è un **grafo**!

## Modello RDF: esempio

- Grafo:

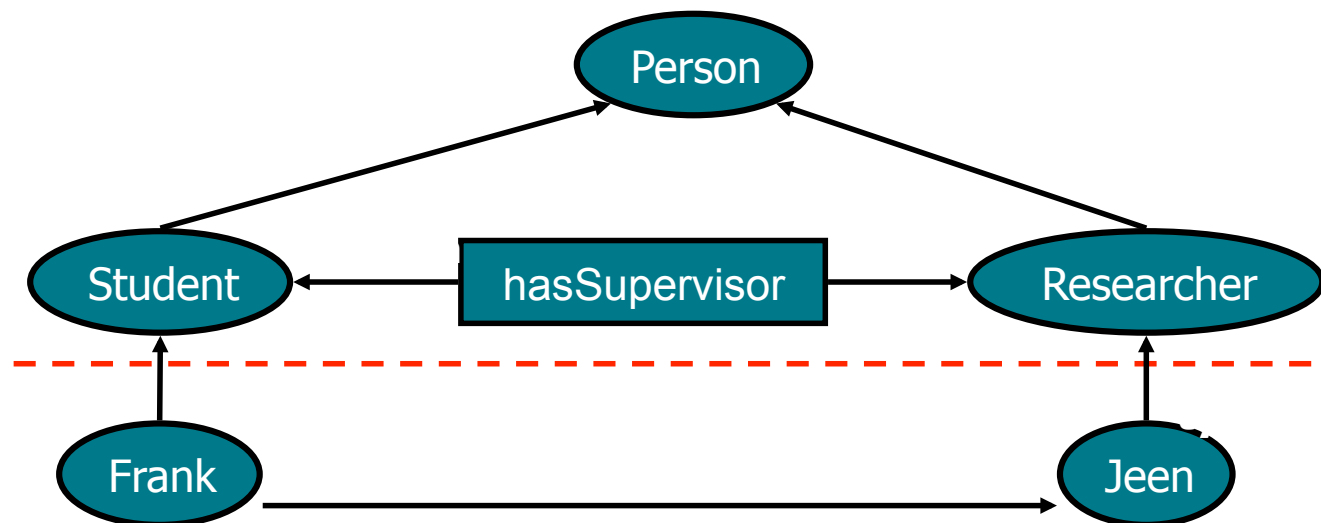


- Insieme di triple:

- (<http://www.w3.org/TR/REC-rdf-syntax/>, `dc:Publisher`, `"W3C"`)
- (<http://www.w3.org/TR/REC-rdf-syntax/>, `dc:Creator`, `"Ora Lassila"`)
- (<http://www.w3.org/TR/REC-rdf-syntax/>, `dc:Date`, `"1999-02-22"`)

## Il livello di RDF e RDFS: RDFS

- RDF Schema, Standard del W3C (2004)
- Permette di definire un “vocabolario”, ovvero uno **schema** per un modello RDF



# Il livello dei linguaggi per la specifica di ontologie

- **Ontologia:**
  - concettualizzazione condivisa, esplicitazione dei concetti rilevanti nel dominio di interesse e delle loro relazioni
    - il primo a parlare di ontologie è stato Aristotele (384 a.c.)!
  - espressa in un vero linguaggio per la rappresentazione della conoscenza (più espressivo di RDF e RDFS)
  - visione nell'ambito dell'Intelligenza Artificiale: ontologia = teoria logica (base di conoscenza)
  - visione nell'ambito delle basi di dati: ontologia = modello concettuale
- **Ontology Web Language (OWL):** linguaggio standard raccomandato dal W3C per la definizione di ontologie

# Ontologia: esempio

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

# Ontologie: il ruolo della logica

- ontologia= teoria logica
- perché?
  - deve essere espressa in un linguaggio dichiarativo
  - deve essere caratterizzata da una semantica formale
  - deve permettere di fare ragionamento automatico, ovvero di usare tecniche di inferenza **corrette** e **complete**
- esiste una corrispondenza ben consolidata tra logica e formalismi per la modellazione concettuale

# OWL e le logiche descrittive

- OWL è basato su un frammento della logica del primo ordine (logica dei predicati)
  - Logiche descrittive (Description Logics, DL): sottoclassi della logica del primo ordine
    - linguaggi maggiormente usati per la rappresentazione della conoscenza
    - capacità espressive delle DL sono state ampiamente studiate
    - tecniche di ragionamento nelle DL sono state altresì studiate a fondo
    - strumenti di ragionamento (ragionatori) sulle DL sono stati sviluppati e ottimizzati
- => le DL sono oggi al centro del Web Semantico!

## Il livello dei linguaggi per la specifica di prove e regole

- Linguaggi che vanno oltre ad OWL
- **Regole**
  - nozione informale
  - utilizzate per fare inferenza sulle ontologie
  - strumento per catturare ulteriore conoscenza, non esprimibile in OWL
- **Prova**
  - serie di esecuzioni di regole (inferenze) che portano a nuova conoscenza



# Formalismi basati sulle regole

- Diversi linguaggi sono basati sulle regole
- Nessun linguaggio è stato però ad oggi standardizzato
- *Work in progress*

## **Il livello dei linguaggi che tengono conto dell'autorevolezza delle sorgenti di informazioni**

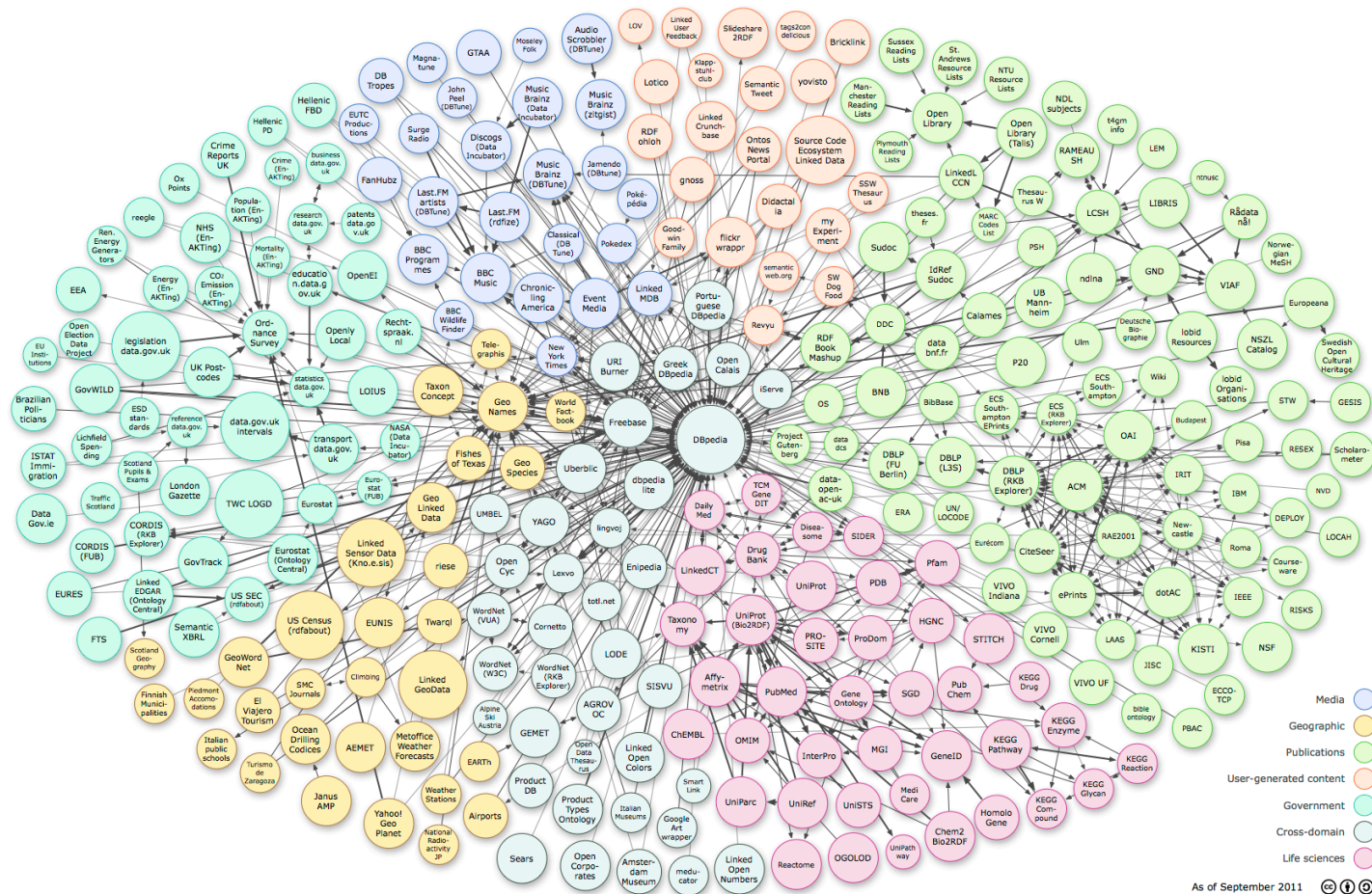
- Supporto per la gestione della provenance/  
autorevolezza
  - da dove viene questa informazione?
  - come è stata ottenuta questa informazione?
  - posso ritenere autorevole questa informazione?
- Problema largamente inesplorato
- Non esistono ancora sforzi di standardizzazione

## 3.1 Introduzione al Web Semantico

- motivazioni
- l'iniziativa del Web Semantico
- l'architettura
- **i Linked Open Data**

\*

# Il Web Semantico nel mondo reale

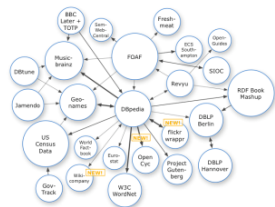


**Open Data** cloud diagram, 09/2011 (by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>)

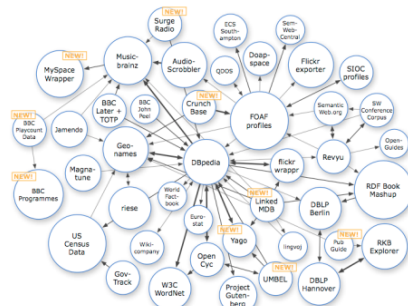
# Linked Open Data: cosa sono

- **Linked Data**: best practice per pubblicare, condividere e collegare dati, informazioni e conoscenza nel Web Semantico, usando le URI
- **Linked Open Data (LOD)**: risultato di una serie di progetti ancora oggi dedicati a pubblicare sul Web insiemi di dati “aperti” al pubblico, nel formato RDF, usando le URI e definendo collegamenti tra dati appartenenti a sorgenti diverse
  - i collegamenti RDF permettono di navigare a partire da dati nell’ambito di una sorgente dati a dati ad essi collegati nell’ambito di altre sorgenti dati
  - poiché i risultati delle interrogazioni sono dati strutturati (e non dei link a pagine HTML) possono essere usati nell’ambito di applicazioni

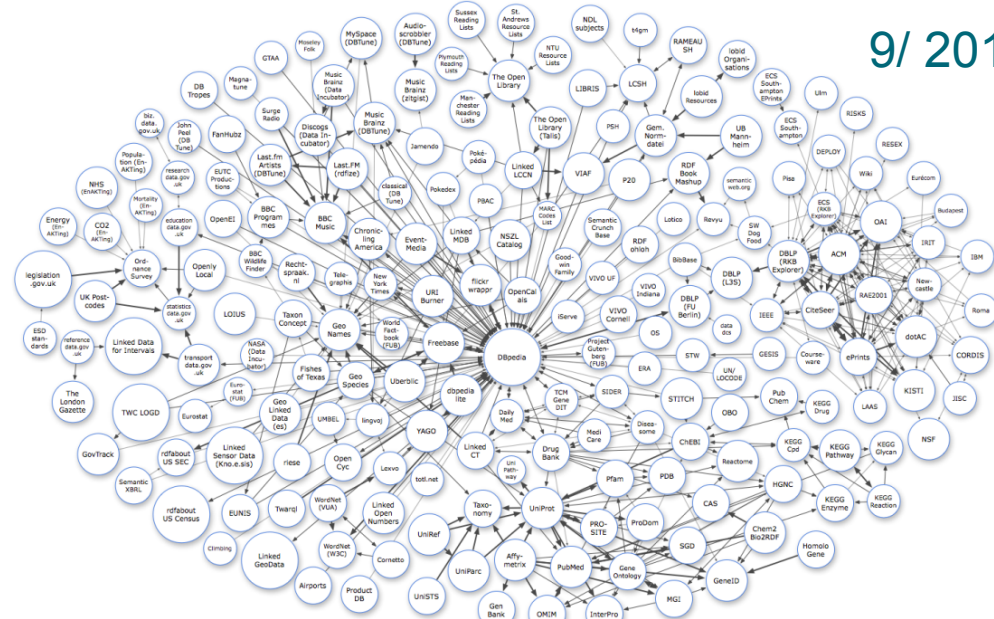
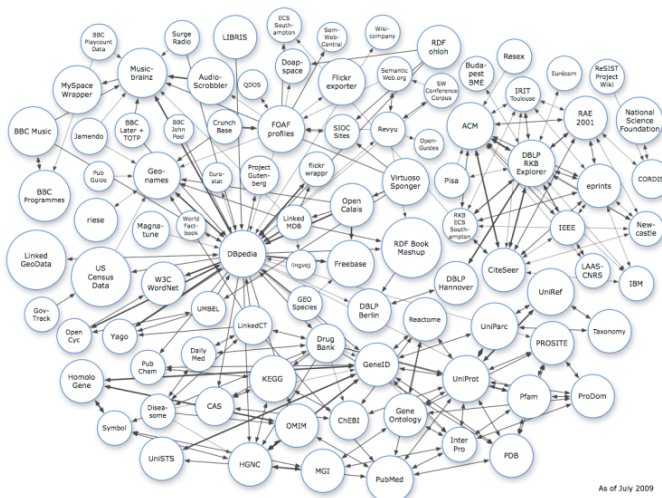
# I diagrammi del LOD cloud



9/2007



9/2008



9/2010

9/2009

As of September 2010 © i f @