



# Le Tecnologie Semantiche in S&T DL

Andrea Giovanni Nuzzolese

Aldo Gangemi

*STLab, ISTC-CNR*

# Sommario

- Arricchimento semantico dei contenuti della Library
- Indicizzazione semiotica
- Raccomandazione dei contenuti
- S&T DL Linked Open Dataset
- Semantic Scout

# Arricchimento semantico dei contenuti della Library

# Arricchimento Semantico

- Arricchimento Semantico
  - il processo di associare a dei termini dei metadati semantici
- Obiettivi
  - aggiungere struttura e semantica a contenuti non strutturati, ex. testo
  - abilitare servizi e funzionalità più raffinati, ex. semantic search

# Arricchimento Semantico - come e cosa

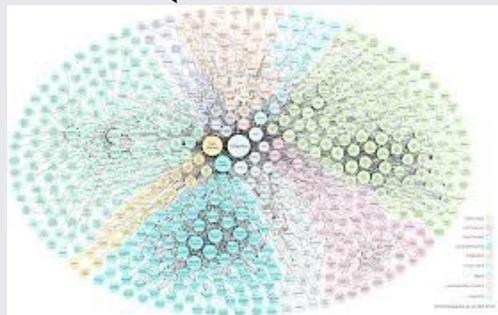
- Tecniche
  - NLP
  - Text mining
  - Machine reading
  - Semantic Web technologies
- Quali contenuti?
  - Titoli
  - Abstract
  - Autori
  - Keyword

# Arricchimento Semantico: esempio

The [Semantic Web](#) Initiative envisions a Web wherein [information](#) is offered [free](#) of [pres](#) change and mixing across [web sites](#) and across [web](#) [Web content](#), few tools [will](#) be [written](#) to consume it; [page](#) appeal to publish [Semantic Web](#) content. To break this [with](#) more flexible information access, we have created a [chick](#) that lets [users](#) make use of [Semantic Web content](#) [web](#) Web. Wherever [Semantic Web](#) content is not available, [with](#) re-structure [information](#) within [web pages](#) into [Piggy](#) [Semantic Web format](#) Through the use of Semantic Web [technologies](#), [Piggy Bank](#) provides

## Semantic Web

The Semantic Web is a collaborative movement led by the World Wide Web Consortium (W3C) that promotes common formats for data on the World Wide Web. By encouraging the inclusion of semantic content in...



Sorgente dati: Linked Open Data

# Apache Stanbol<sup>1</sup>



- Uno stack software di componenti riusabili
  - open-source
  - accessibili attraverso interfacce RESTful
  - risultato del progetto IKS (IP EU FP7)
- Funzionalità
  - servizi di arricchimento semantico per applicazioni che gestiscono contenuti
  - servizi di indicizzazione avanzata dei contenuti



1. <http://stanbol.apache.org>

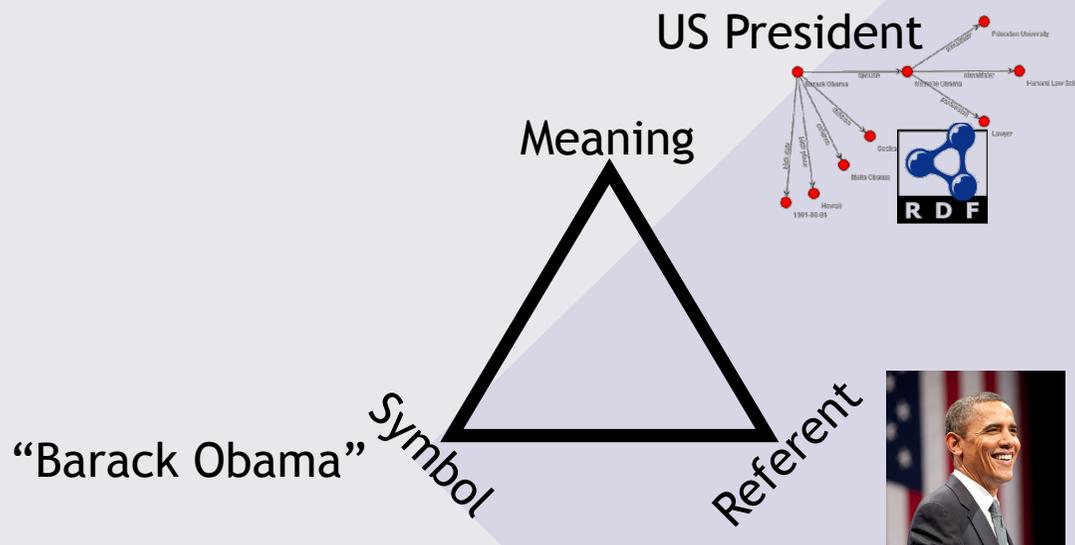
# Indicizzazione semiotica e content recommendation

# Indice semiotico

- Sistema di incizzazione più raffinato che non è limitato al reverse-indexing dei documenti
- Permette di memorizzare le relazioni tra un espressione (ex. “Barack Obama”) ed il tema/topic (ex. “US President”) di tale espressione
- Obiettivi
  - semantic search
  - sistemi di raccomandazione
  - exploratory search

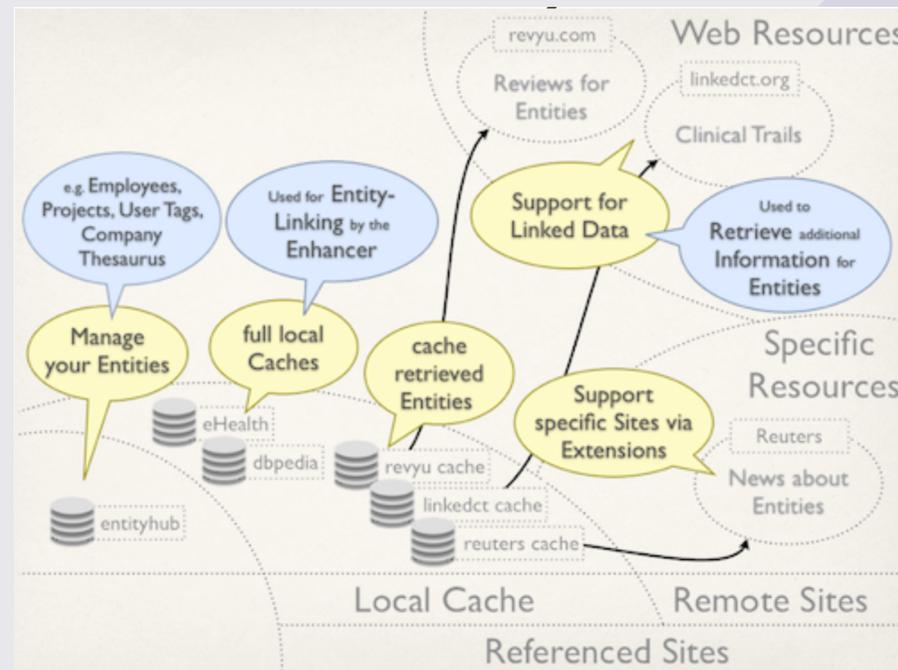
# Indicizzazione semiotica: Esempio

- Per esempio, è possibile associare l'individuo “Barack Obama” (entità) a
  - tutte le rappresentazioni simboliche che fanno riferimento a Barack Obama in un documento (parole)
  - e alle sue possibili interpretazioni (sensi: presidente, afro-americano, democratico, etc.) espresse come RDF



# Content recommendation in S&T DL

- S&T DL implementa un sistema di raccomandazione delle risorse bibliografiche basato su
  - similarity detection attraverso ricerca semantica
  - indice semiotico costruito sui contenuti testuali delle risorse (titoli, autori, abstract, keyword)



RESTful API

# Semantic technologies at work

PRODOTTO

## Cold uniform spherical collapse revisited

Parole Chiave:

N-body simulations, spherical collapse, virialization, Universe, cosmology

Congresso nome: Invisible Universe International Conference

Congresso luogo: Paris, France

Congresso data: 29 June-3 July, 2009

Congresso rilevanza: Internazionale

Organizzazioni:

[Istituto dei sistemi complessi](#)

Sito web: <http://proceedings.aip.org/resource/2/apcpcs/12>

Provenienza: [CNR](#)

Original ID: 204496

ID: r1\_s1-a1-t1\_943f3580ebc4ef11b3b68024730683d7

Vedi anche:

- [Violent and mild relaxation of an isolated self-gravitating uniform sphere](#)
- [Energy ejection in the collapse of a cold spherical self-gravitating sphere](#)
- [A toy model to test the accuracy of cosmological N-body simulations](#)
- [Evolution of isolated overdensities as a control on cosmological structure formation](#)
- [One-dimensional gravity in infinite point distributions](#) (0.14)
- [Causality constraints on fluctuations in cosmology: A study with N-body simulations](#)
- [The complex universe: recent observations and theoretical challenges](#)
- [Gravitational dynamics of an infinite shuffled lattice: Particle collisions and the continuum limit](#) (0.10)
- [Fundamental collapse of the exciton-exciton effective scattering length](#)
- [System-dependent melting behavior of icosahedral anti-Mackay clusters](#)

[Cosmology](#)



Cosmology (from the Greek κόσμος, kosmos "world" and -λογία, -logia "study of"), is the study of the origin, evolution, and eventual fate of the universe. Physical cosmology is the scholarly and scientific study of the origin, evolution, large-scale structures and dynamics, and ultimate fate of the universe, as well as the

Attendibilità: 0.45078

# Semantic technologies at work

PRODOTTO

## Cold uniform spherical collapse revisited

Vedi anche:

- [Violent and mild relaxation of an isolated self-gravitating uniform and spherical cloud of particles](#) (0.69)
- [Energy ejection in the collapse of a cold spherical self-gravitating cloud](#) (0.55)
- [A toy model to test the accuracy of cosmological N-body simulations](#) (0.33)
- [Evolution of isolated overdensities as a control on cosmological N-body simulations](#) (0.20)
- [One-dimensional gravity in infinite point distributions](#) (0.14)
- [Causality constraints on fluctuations in cosmology: A study with exactly solvable one-dimensional models](#) (0.13)
- [The complex universe: recent observations and theoretical challenges](#) (0.12)
- [Gravitational dynamics of an infinite shuffled lattice: Particle coarse-graining, nonlinear clustering, and the continuum limit](#) (0.10)
- [Fundamental collapse of the exciton-exciton effective scattering](#) (0.10)
- [System-dependent melting behavior of icosahedral anti-Mackay nanoalloys](#) (0.07)

dynamics, and ultimate fate of  
the universe, as well as the  
Attendibilità: 0.45078

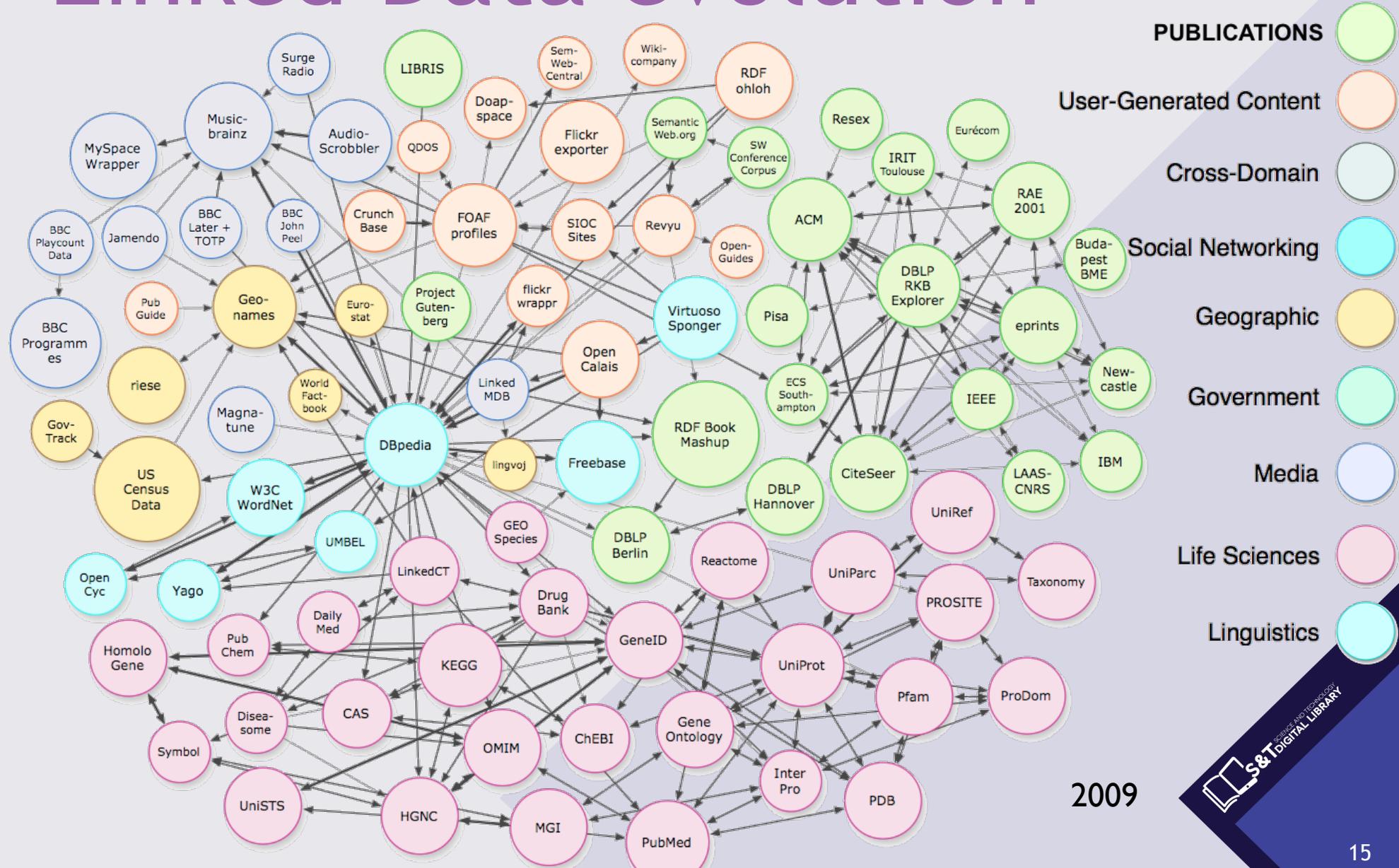
# S&T DL Linked Open Dataset

# Linked (Open) Data – LOD

- Sono un'insieme di buone pratiche per pubblicare, condividere e connettere pezzi di informazione nel Semantic Web utilizzando URIs e RDF
  - abilitano interlinking, interoperabilità e interrogazioni semantiche
  - forniscono contenuti in formati machine-readable

- LOD 5-stars
  - ★ Available on the Web (whatever format), but with an open licence
  - ★ ★ Available as machine-readable structured data (e.g. instead of image scan of a table)
  - ★ ★ ★ As 2 plus non-proprietary format (e.g. CSV instead of excel)
  - ★ ★ ★ ★ All above plus use open standard from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
  - ★ ★ ★ ★ ★ Available the above, plus: Link your data to other people's data to provide context

# Linked Data evolution



PUBLICATIONS

User-Generated Content

Cross-Domain

Social Networking

Geographic

Government

Media

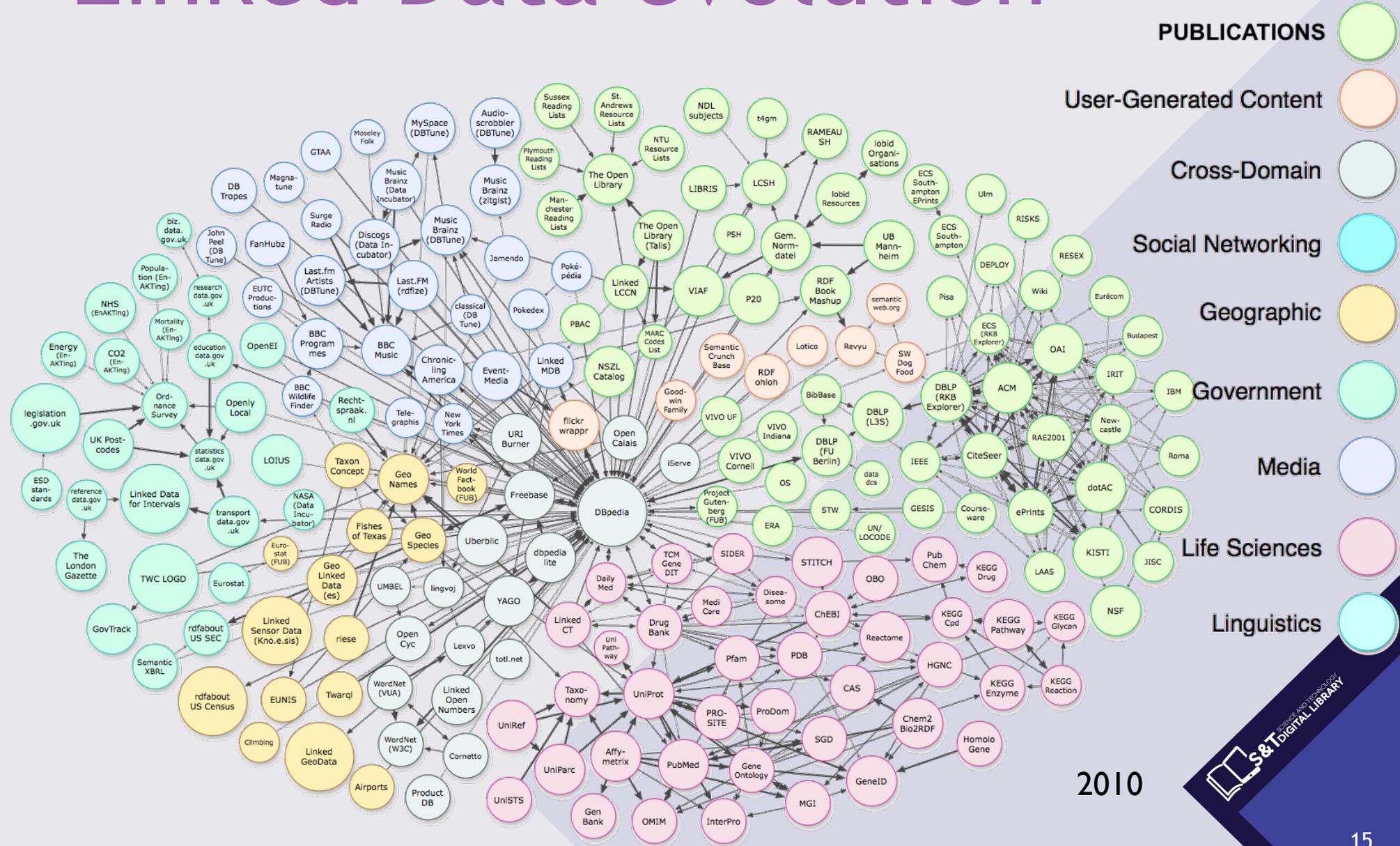
Life Sciences

Linguistics

2009



# Linked Data evolution

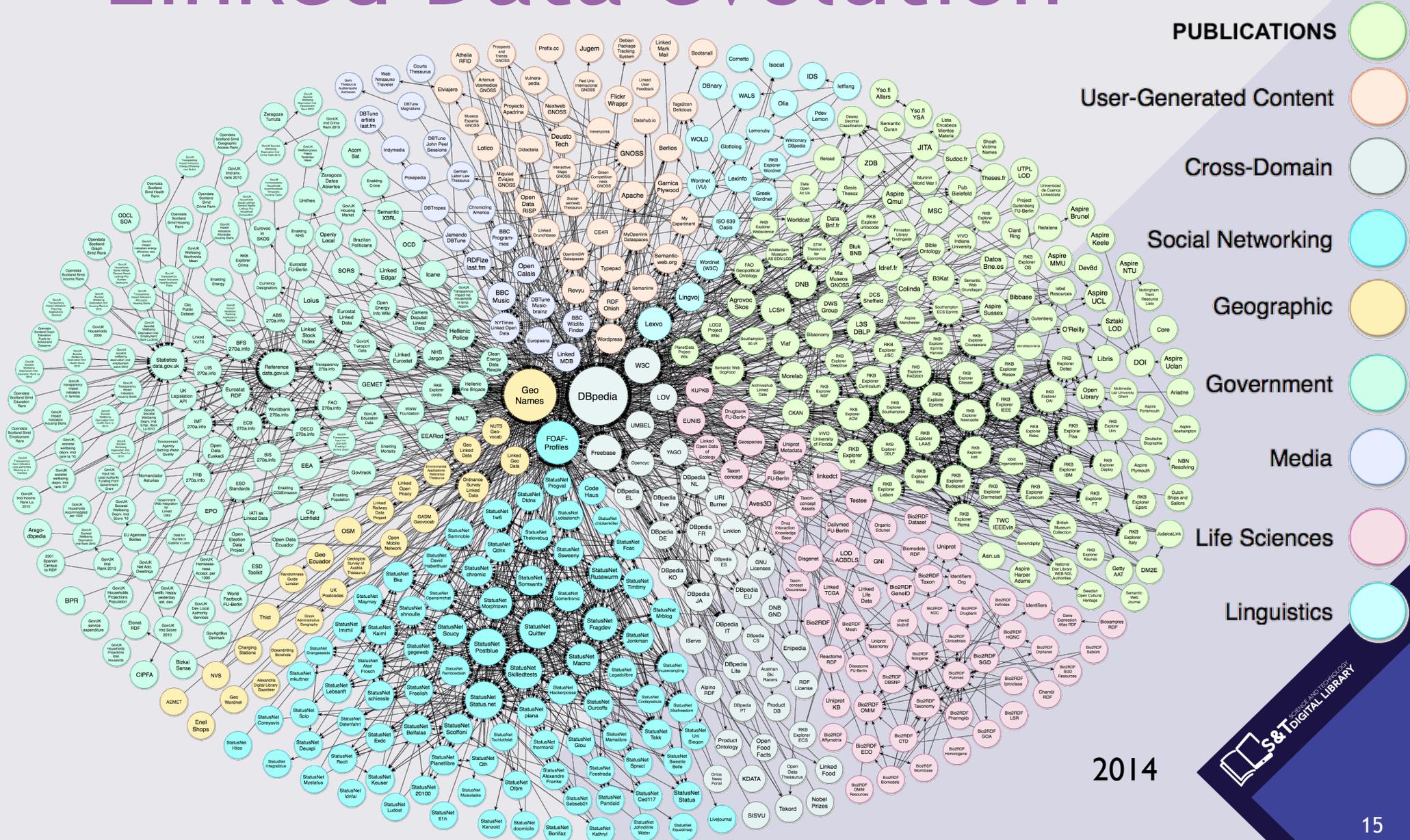


2010





# Linked Data evolution



PUBLICATIONS

User-Generated Content

Cross-Domain

Social Networking

Geographic

Government

Media

Life Sciences

Linguistics

2014



# LOD Stats

**Datasets by topical domain.**

Topic	Datasets	%
<a href="#">Government</a>	183	18.05%
<a href="#">Publications</a>	96	9.47%
<a href="#">Life sciences</a>	83	8.19%
<a href="#">User-generated content</a>	48	4.73%
<a href="#">Cross-domain</a>	41	4.04%
<a href="#">Media</a>	22	2.17%
<a href="#">Geographic</a>	21	2.07%
<a href="#">Social web</a>	520	51.28%
<a href="#">Total</a>	1014	

# Semantic Publishing

- La gestione della conoscenza relativa a prodotti di ricerca e risorse bibliografiche in contesto Semantic Web
- Obiettivi
  - realizzare Linked Data per le risorse bibliografiche e i prodotti della ricerca
  - fornire accesso a risorse bibliografiche arricchite semanticamente
  - abilitare una serie di task semantici come knowledge discovery, knowledge exploration and data integration nel dominio “scholarly”

# S&T DL — LOD

- Modellazione ontologica
  - S&T DL — LOD utilizza le ontologie SPAR<sup>1</sup> e Ontology Design Patterns<sup>2</sup> come modelli concettuali di riferimento per organizzare e rappresentare risorse bibliografiche in RDF
  - Le ontologie sono allineate a standard o best practice, ex. SKOS, Dublin Core, FRBR
- Produzione di Linked Data
  - Reengineering: conversione XSLT delle schede dei prodotti della library da XML in un RDF intermedio: convertite ~150K risorse da PEOPLE, INFN, Archivio Centrale dello Stato e Fondazione Sturzo
  - Refactoring: allineamento a SPAR dei prodotti rappresentati in RDF tramite SPARQL CONSTRUCT (8.340.537 triple RDF)
- Dati e SPARQL endpoint saranno accessibili a [data.stdlib.cnr.it](http://data.stdlib.cnr.it)

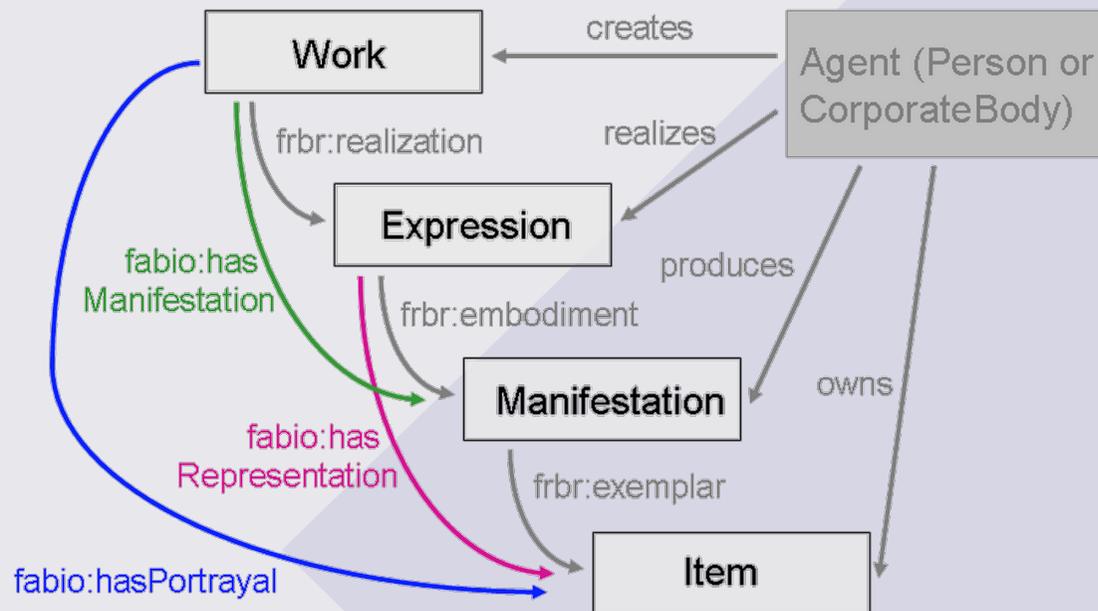
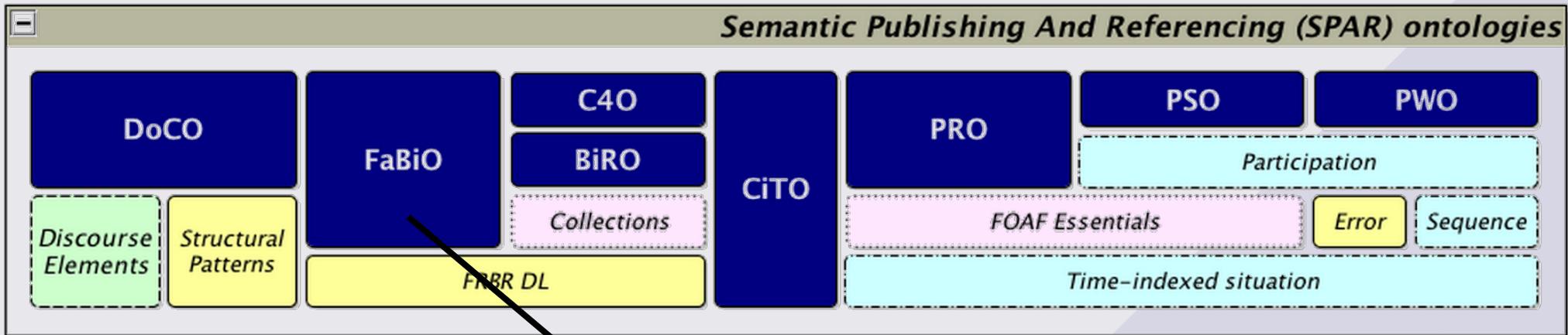
1. <http://www.sparontologies.net/>



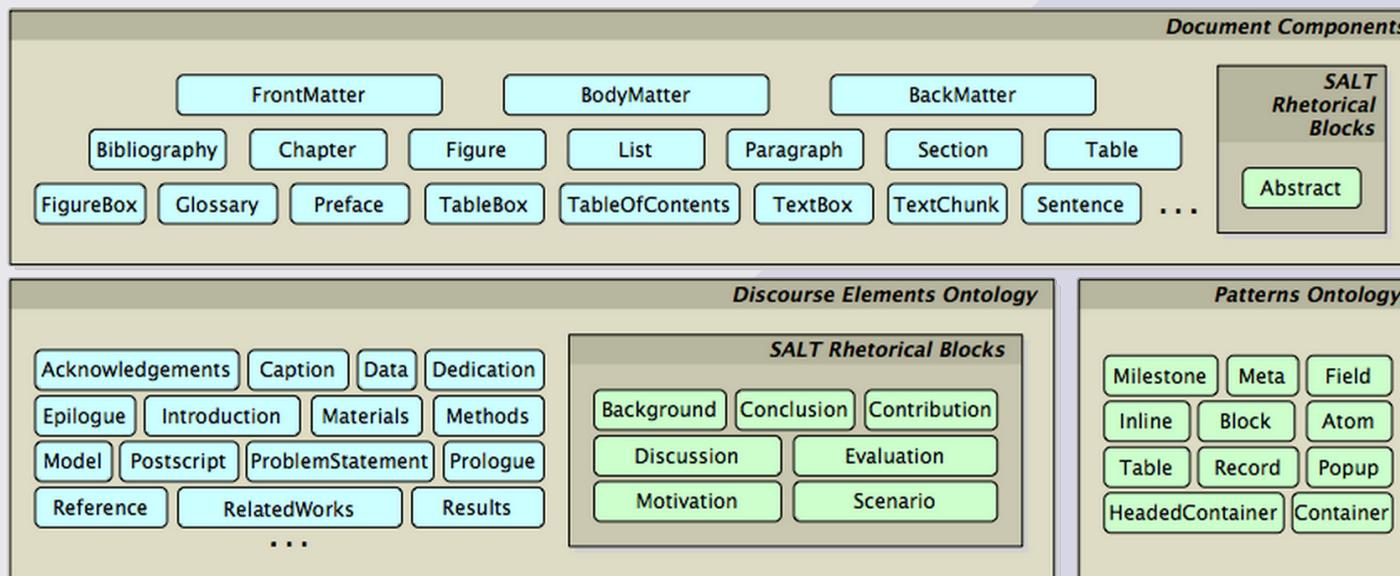
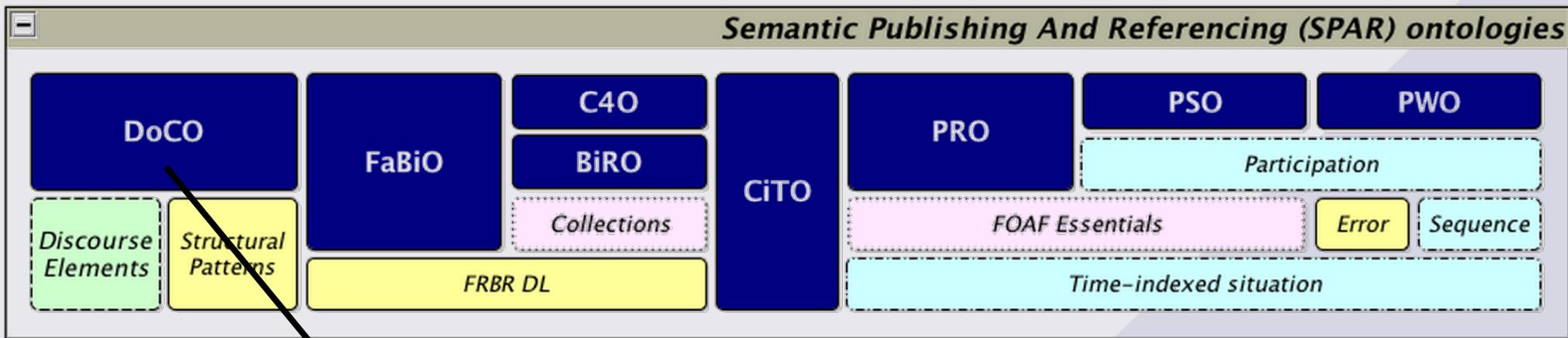
2. <http://www.ontologydesignpatterns.org>



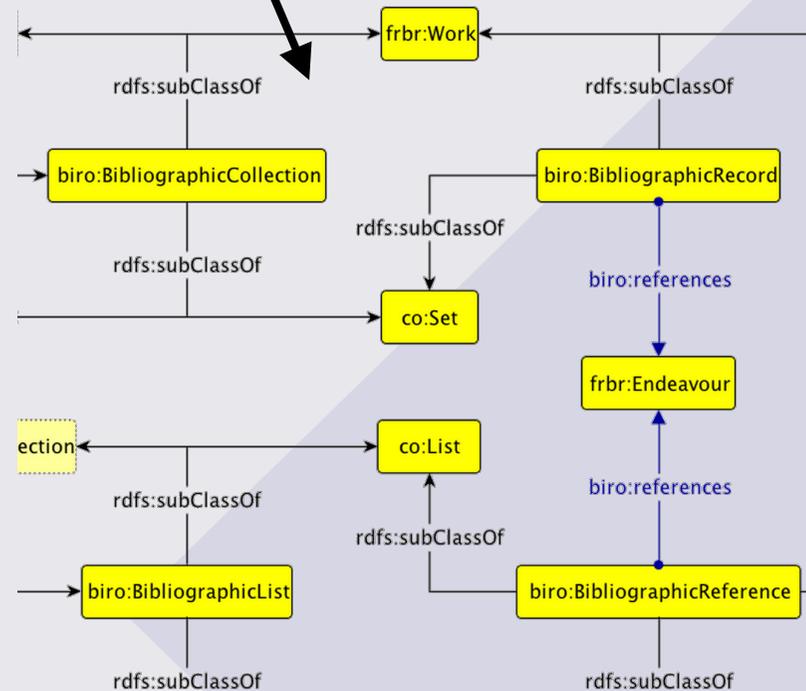
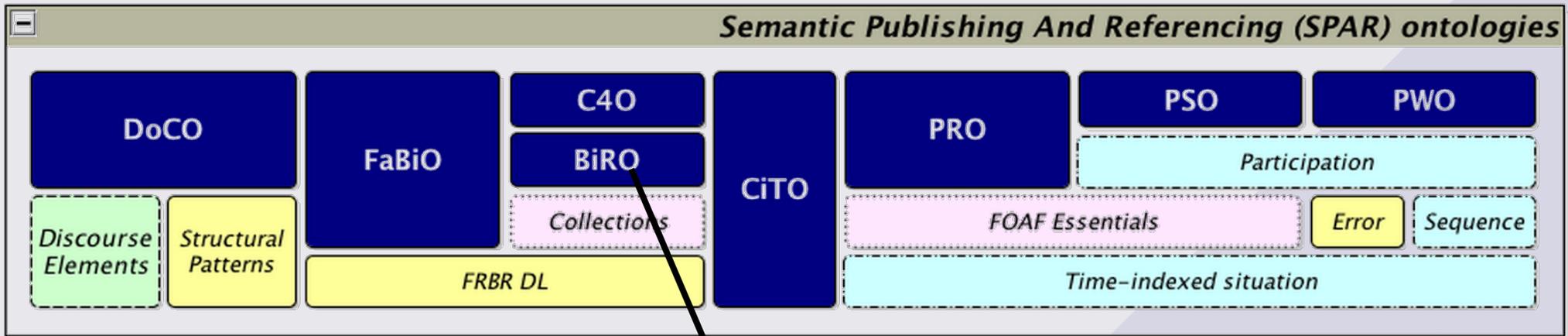
# S&T DL – LOD: Ontology Network



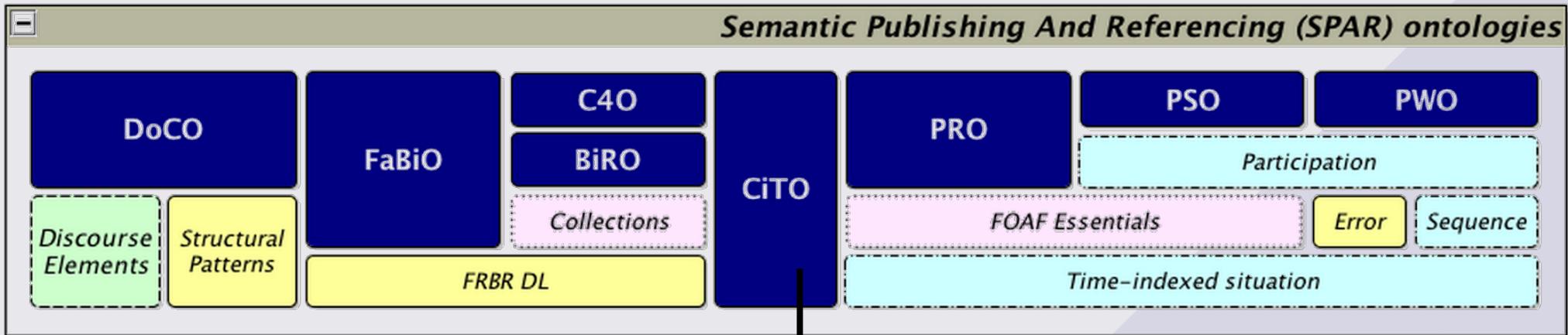
# S&T DL – LOD: Ontology Network



# S&T DL – LOD: Ontology Network



# S&T DL – LOD: Ontology Network

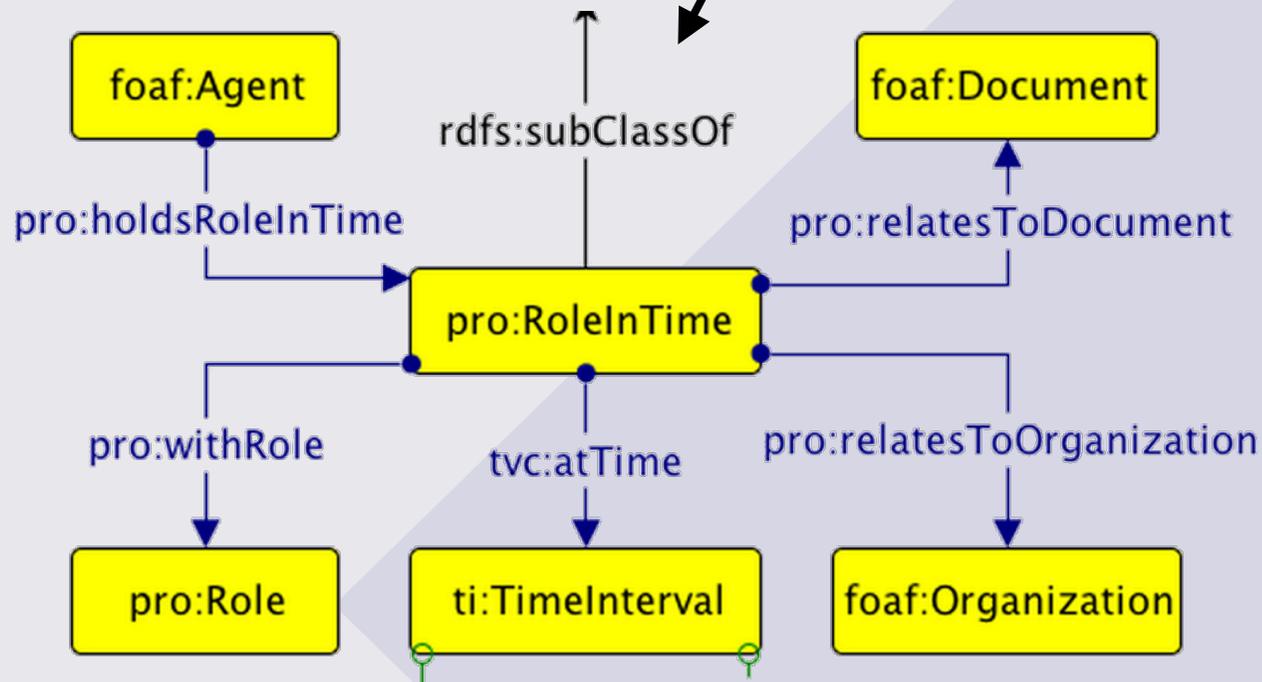
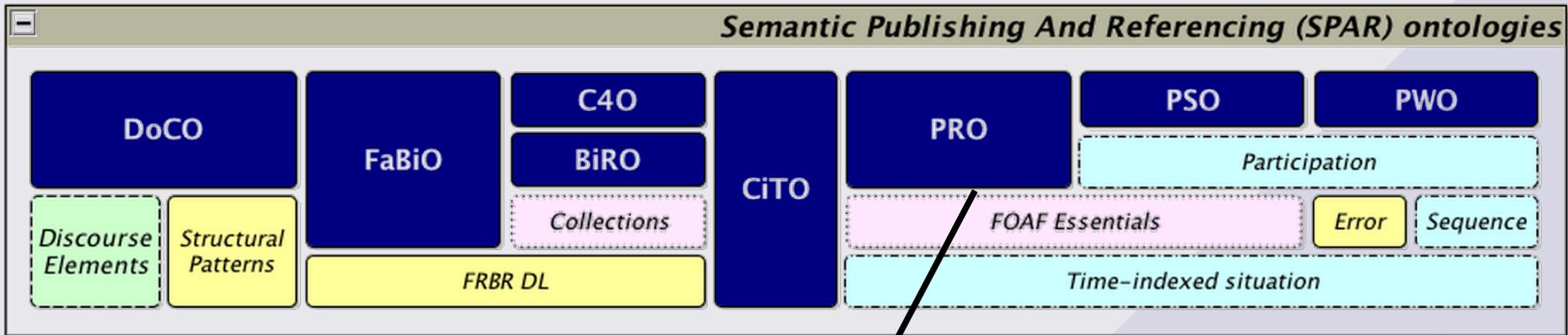


“... For this, I highly recommend X. ...”

The citing entity cites the cited entity as an item of recommended reading. This property can be used, for example, to describe references in a lecture reading list, where the cited references are relevant to the general topic of the lecture, but might not be individually cited within the text of the lecture. Similarly, it could be used to describe items in a 'Suggested further reading' list at the end of a book chapter.

by the following CiTO

# S&T DL – LOD: Ontology Network



# Collegamenti con altri LOD

- [data.cnr.it](http://data.cnr.it)
  - dati del CNR su persone, organizzazioni, progetti di ricerca, prodotti della ricerca e commesse
- [dbpedia.org](http://dbpedia.org)
  - dati di carattere enciclopedico
- dblp
  - dati su pubblicazioni nel dominio computer science
- geonames
  - dati per geolocalizzazione

# Semantic Scout: searching, browsing ed exploring Linked Data

# Semantic Scout

Consiglio Nazionale delle Ricerche

Home | English version | Mappa | Commenti | Sondaggio | Staff | Contattaci | RSS Cerca nel sito

### Semantic Scouting

rdf

Type:  base  latent Lang:  IT  EN

**Elenco completo** | Attività | Persone

- [Intraweb semantic: gestione avanzata dell'informazione in organizzazioni complesse \(ICT.P08.015.001\)](#) (Modulo) sc:0.19703028
- [Model Driven Legislative Drafting \(ICT.P08.009.004\)](#) (Modulo) sc:0.19275072
- [Intraweb semantic: gestione avanzata dell'informazione in organizzazioni complesse \(ICT.P08.015\)](#) (Commessa) sc:0.17240149
- [Gestione della Conoscenza per la Sicurezza \(INT.P01.013.001\)](#) (Modulo) sc:0.13932145
- [XML Technologies For Semantic Web Applications and Secure Workflows \(INT.P01.007.003\)](#) (Modulo) sc:0.123143926
- [Archeologia e società dell'informazione. Metodologie informatiche e modelli formali per una conoscenza arricchita del patrimonio archeologico. \(PC.P06.003.001\)](#) (Modulo) sc:0.123143926
- [Sistemi e Tecnologie Informatiche per la Sicurezza \(INT.P01.013\)](#) (Commessa) sc:0.104491085
- [La rappresentazione delle norme per l'interoperabilità a livello nazionale ed internazionale \(IC.P06.008.003\)](#) (Modulo) sc:0.09851514
- [Risorse e Tecnologie Linguistiche: modelli, metodi di sviluppo, applicazioni, disegno di strategie internazionali \(IC.P02.005.001\)](#) (Modulo) sc:0.09851514
- [Chimica sostenibile e sicurezza ambientale \(PM.P03.014\)](#) (Commessa) sc:0.087075904
- [Chimica sostenibile e sicurezza ambientale \(PM.P05.004.007\)](#) (Modulo) sc:0.086200744
- [Metodi e Sistemi per l' Informazione e la Conoscenza \(ICT.P08.001.002\)](#) (Modulo) sc:0.086200744
- [Chimica sostenibile e sicurezza ambientale \(PM.P03.014.001\)](#) (Modulo) sc:0.086200744
- [Sistemi di supporto all'attività degli organi legislativi e della Pubblica Amministrazione \(IC.P06.008\)](#) (Commessa) sc:0.07388635
- [Archeologia e società dell'informazione. Metodologie informatiche e modelli formali per una conoscenza arricchita del patrimonio archeologico. \(PC.P06.003\)](#) (Commessa) sc:0.07388635
- [Risorse e Tecnologie Linguistiche: standardizzazione, sviluppo, infrastrutture, strategie \(IC.P02.005\)](#) (Commessa) sc:0.061571963

Semantic search

# Semantic Scout



The screenshot displays the 'CNR Semantico' interface. At the top, there is a search bar with the text 'CNR Semantico' and a dropdown menu showing 'DOTT. ALDO GANGEMI'. Below the search bar is a large network graph. The central node is 'DOTT. ALDO GANGEMI'. Surrounding it are various other nodes, including 'Ontology (computer science)', 'Data management', 'Knowledge representation', 'Web development', 'Semantic web', 'World Wide web', 'Istituto ISTC', 'Commissione ICT-P08.015', and 'Modulo ICT-P08.015.01'. The graph also includes names of other researchers such as 'ROBERTA FERRARIO', 'Valentina Presutti', 'LAURE RENVE VIEU', 'MASSIMILIANO CIARAMITA', 'ING. CLAUDIO MASOLO', 'ING. ALBERTO SALVATI', 'EMANUELE BOTTAZZI', 'DOTT. SSA VALENTINA PRESUTTI', 'DOTT. SSA MARIA TERESA SAGRI', 'DOTT. STEFANO BORGIO', 'DOTT. SSA DANIELA TISCORNIA', 'DOTT. NICOLA GUARINO', 'DOTT. ROBERTA FERRARIO', 'DOTT. GERARDO STEVE', 'DOTT. ALBERTO JOSEPH DEHMANN', 'DOTT. DOMENICO MASSIMO MISANELLI', 'CAROLA CATENACCIO', 'ALFIO MASSIMILIANO GILIOZZO', and 'ALESSANDRO ULTRAMARI'. The interface also shows a 'Name' field with 'DOTT. ALDO GANGEMI' and a 'Graph Explorer' label at the bottom.

Browsing

# Semantic Scout



Legenda

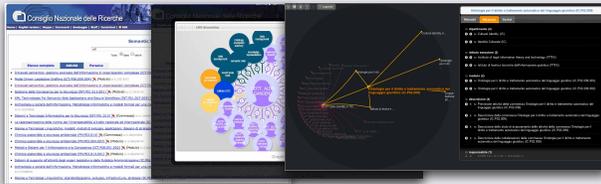
**Ontologie per il diritto e trattamento automatico del linguaggio giuridico (IC.P02.009)**

Metadati
Relazioni
Basket

- dipartimento (2)**
  - 📄 Cultural Identity (IC)
  - 📄 Identità Culturale (IC)
- istituto esecutore (2)**
  - 📄 Institute of legal information theory and technology (ITTIG)
  - 📄 Istituto di teoria e tecniche dell'informazione giuridica (ITTIG)
- modulo (2)**
  - 📄 Ontologie per il diritto e trattamento automatico del linguaggio giuridico (IC.P02.009.003)
  - 📄 Ontologie per il diritto e trattamento automatico del linguaggio giuridico (IC.P02.009.002)
- descrizione (4)**
  - 📄 Previsione attività della commessa Ontologie per il diritto e trattamento automatico del linguaggio giuridico (IC.P02.009)
  - 📄 Descrizione della commessa Ontologie per il diritto e trattamento automatico del linguaggio giuridico (IC.P02.009)
  - 📄 Descrizione dello stato di avanzamento delle attività della commessa Ontologie per il diritto e trattamento automatico del linguaggio giuridico (IC.P02.009)
  - 📄 Descrizione delle collaborazioni della commessa Ontologie per il diritto e trattamento automatico del linguaggio giuridico (IC.P02.009)
- responsabile (1)**
  - 📄 DOTT.SSA DANIELA TIS ...

Relation explorer

# Semantic Scout



Legenda

Ontologie per il diritto e trattamento automatico del linguaggio giuridico (IC.P02.009)

DOTT.SSA DANIELA TISCORNIA

Semantic access to multilingual legal information

Exporting exploration results

# Semantic Scout



## Attensece a

- Institute of cognitive sciences and technologies (ISTC) (Istituto)

## Persona in rapporto

- Rapporto con CNR di DOTT. ALDO GANGEMI (Rapporto con CNR)

## Gestore di

- IntraWeb semantico: gestione avanzata dell'informazione in organizzazioni complesse (ICT.P08.015.001) (Modulo)

## Nome

- ALDO

## Partecipa a commessa

- Partecipazione a Commessa "Ontologia di riferimento per l'integrazione di servizi e organizzazioni nella Pubblica Amministrazione e nelle Aziende" (ICT.P08.007) di DOTT. ALDO GANGEMI nell'anno 2007 (Partecipazione a commessa)

## Semantic Web

### Tema di

- DOTT.SSA MARINA MONTI (Unità di personale interno)
- MARCO ROSELLA (Unità di personale esterno)
- DOTT. ALDO GANGEMI (Unità di personale interno)
- MAURIZIO TESCONI (Unità di personale esterno)
- FRANCESCA BERTAGNA (Unità di personale esterno)
- DOTT.SSA MONICA DE MARTINO (Unità di personale interno)
- ALESSIO BERTONE (Unità di personale esterno)
- DOTT.SSA IRINA RALUCA PRODANOF (Unità di personale interno)
- FRANCESCO RONZANO (Unità di personale esterno)
- RICCARDO ALBERTONI (Unità di personale esterno)
- Valentina Presutti (Unità di personale esterno)
- DOTT.SSA SIMONETTA MONTEMAGNI (Unità di personale interno)
- EMILIANO GIOVANNETTI (Unità di personale esterno)
- MASSIMILIANO CIARAMITA (Unità di personale esterno)
- ELENA CAMOSSÌ (Unità di personale esterno)

# Grazie

[andrea.nuzzolese@istc.cnr.it](mailto:andrea.nuzzolese@istc.cnr.it)