

# The Semantic Web – an overview

Olle Olsson

Swedish W3C Office

Swedish Institute of Computer Science (SICS)

jFokus2008 – Stockholm, Sweden, Jan 30, 2007

# SICS – Swedish Institute of Computer Science

National research institute

- R&D in information and communication technologies

Objective:

- conduct advanced and focused research in strategic areas of computer science



**Sponsors:**

TeliaSonera, Ericsson,  
Saab Systems,  
FMV (Defence Materiel Administration),  
Green Cargo (Swedish freight railway operator),  
ABB,  
Bombardier Transportation

# Contents

---

The aim: provide an overview of the semantic web.

Main parts of presentation:

1. Semantic Web – motivation and objectives
2. Semantic Web technologies
3. Semantics in action -- examples
4. The larger landscape

# 1. Semantic Web: Motivation and Objectives

Why do we need a complementary approach?

# The web – content for humans

- The web – as we know it:
  - Rich, global source of information
  - Find information by surfing the web
    - Visual inspection of pages, mental understanding
  - Supported by search engines
    - Find pages associated with text fragments
  - Strong support for presentation
- Typical need: combine data on the Web:
  - hotel and travel information may come from different sites
  - searches in different digital libraries
  - etc.
- Humans combine such information easily
  - even if different terminologies are used!

**Common notation:**  
HTML, ...

**Common protocol:**  
HTTP

**Principle:**  
Decentralised

# Web content – machine usable?

- Program access to useful information on the web?
  - Automate routine tasks
  - Smart searches
  - Use online information repositories as *data*
- Automated use of web content is difficult
  - partial information is of limited value
  - difficult to make sense of, e.g., an image
  - automated conclusions from analogies is difficult
  - difficult to combine information automatically
    - is `<bib:creator>` same as `<pubs:author>`?
    - how to combine different XML hierarchies?
  - ...

**Common notation:**  
???

**Common protocol:**  
HTTP, ....

**Principle:**  
Decentralised

# The missing link

Centralised

Decentralised

Human use

Loosely coupled on  
standardised foundation  
(web technologies)

Machine use

Tightly  
coupled

*Missing  
framework?*

# The rationale for the Semantic Web

We have the web of documents  
(text, multimedia, ...)

- Interlinked network of documents
- Provided by independent sources
- Understandable representation and uniform access
- Web browsers make content available to users
- Users make use of content

We need the web of data

- Interlinked network of *data*
  - Provided by independent sources
  - Understandable representation and uniform access
  - *Programs know how to relate data*
  - *Programs* make use of *data*
- Semantic Web (SW) = programmable data web

# The way ahead

## Complications

- Data – stored in databases, applications, web pages, etc.
- Represented in various formats
- Structured in various data models
- Formats and models change over time
- Data interoperability difficult

## Approach:

- Abstract from concrete representations
- Interoperability at abstraction level feasible
- Abstraction captures “semantics”

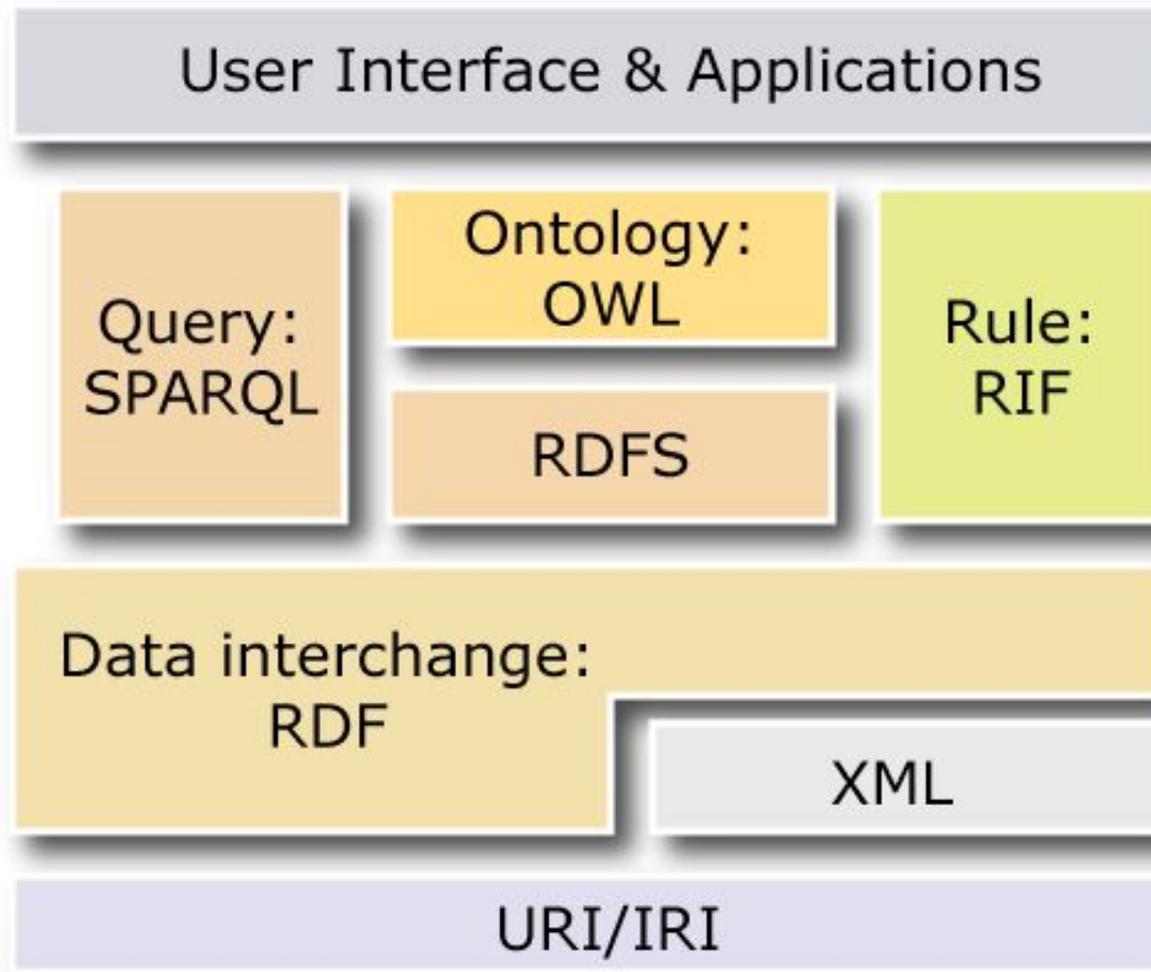
## 2. Semantic Web: Technologies

### Characterizations of Technologies

# What is needed?

- Languages for describing concrete data
  - E.g., “13.50” vs. “Price: 13.50; Currency: Euro”
- Languages for describing types of data (data models)
  - E.g., Price: numeric monetary value; currency; per quantity; ...”
- Methods/tools for mapping data models to data models
  - E.g., ebXML (UN/CEFACT) to eBay
- Methods/tools for searching data
  - Query languages
- Methods/tools for interoperation with other web technologies
  - E.g., with the document web
- Etc. ...

# Semantic Web basic building blocks



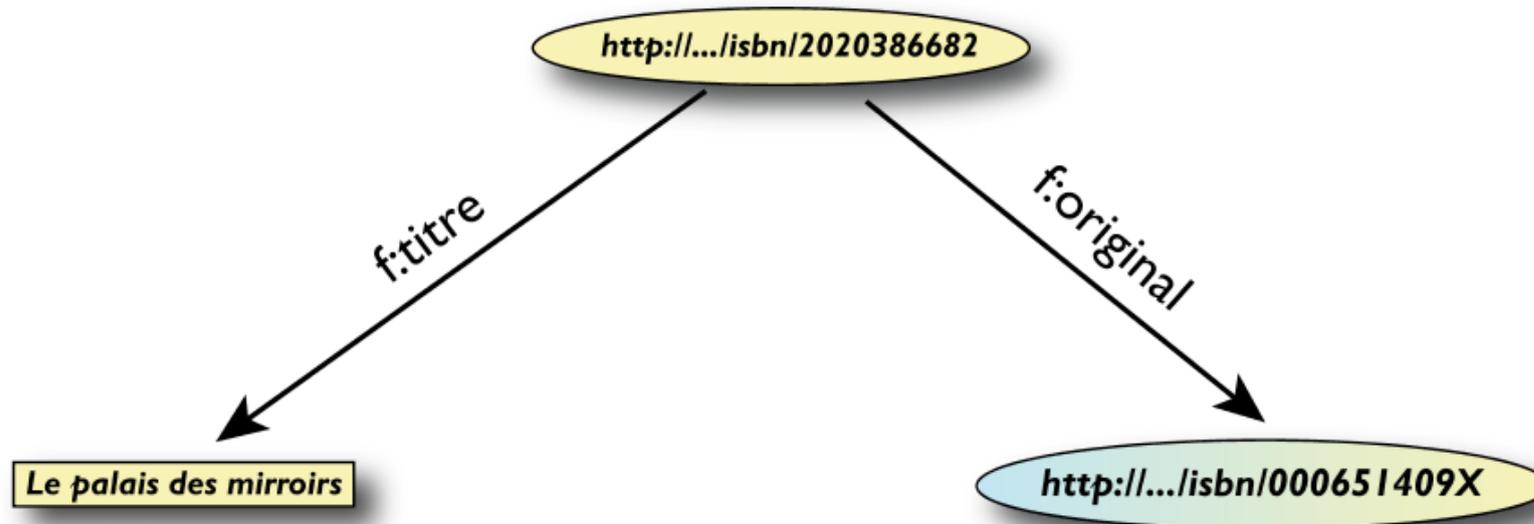
# RDF – Resource Description Framework

- Basic data model – a “triple”
  - triple (s, p, o) is such that:
    - “s”, “p”, and “o” stand for “subject”, “predicate”, and “object”, respectively
    - conceptually: “p” connects, or relates the “s” and “o”
- An example triple:

```
(  
<http://...isbn...6682>,           # "Le palais des miroirs"  
<http://.../original>,           # "is a derivative of the original"  
<http://...isbn...409X>          # "The Glass Palace"  
)
```

- RDF is a general model for such triples
  - Having machine readable formats like RDF/XML, Turtle, n3, RXR, ...
- ... and that's it!

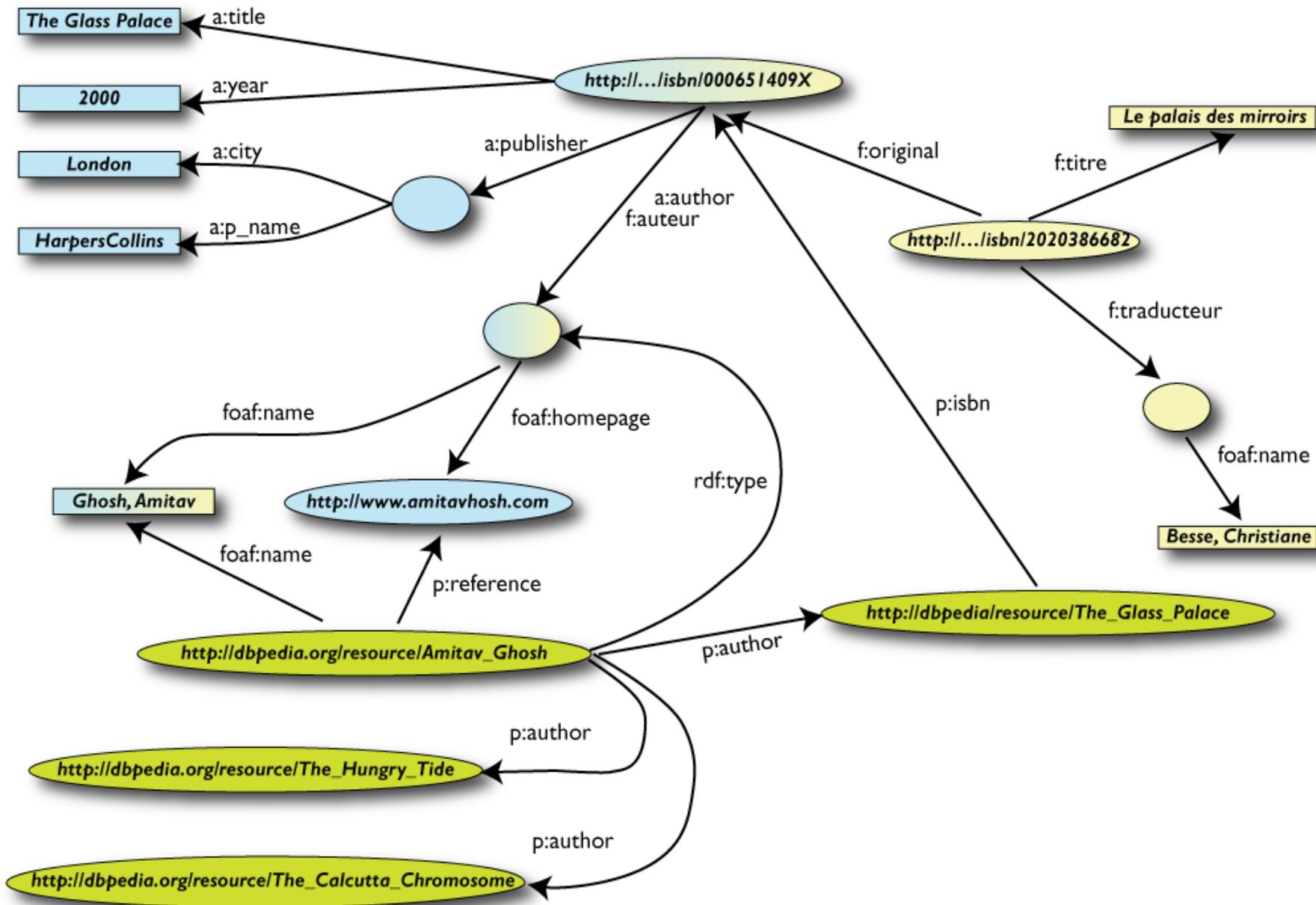
# RDF Example



```
<rdf:Description rdf:about="http://.../isbn/2020386682">  
  <f:titre xml:lang="fr">Le palais des miroirs</f:titre>  
  <f:original rdf:resource="http://.../isbn/000651409X" />  
</rdf:Description>
```

Set of triples form a graph – the *RDF graph*

# RDF graph



# RDFS – RDF Schema

- Simple forms of critical “meta knowledge”:
  - what terms to use
  - what restrictions must apply
  - what other relationships may hold
- RDF Schema
  - officially: “RDF Vocabulary Description Language”
    - the term “Schema” is retained for historical reasons...
- Formalism for defining a schema:
  - Classification: “Class”, “type”
  - Relationships “subClassOf”
  - etc.

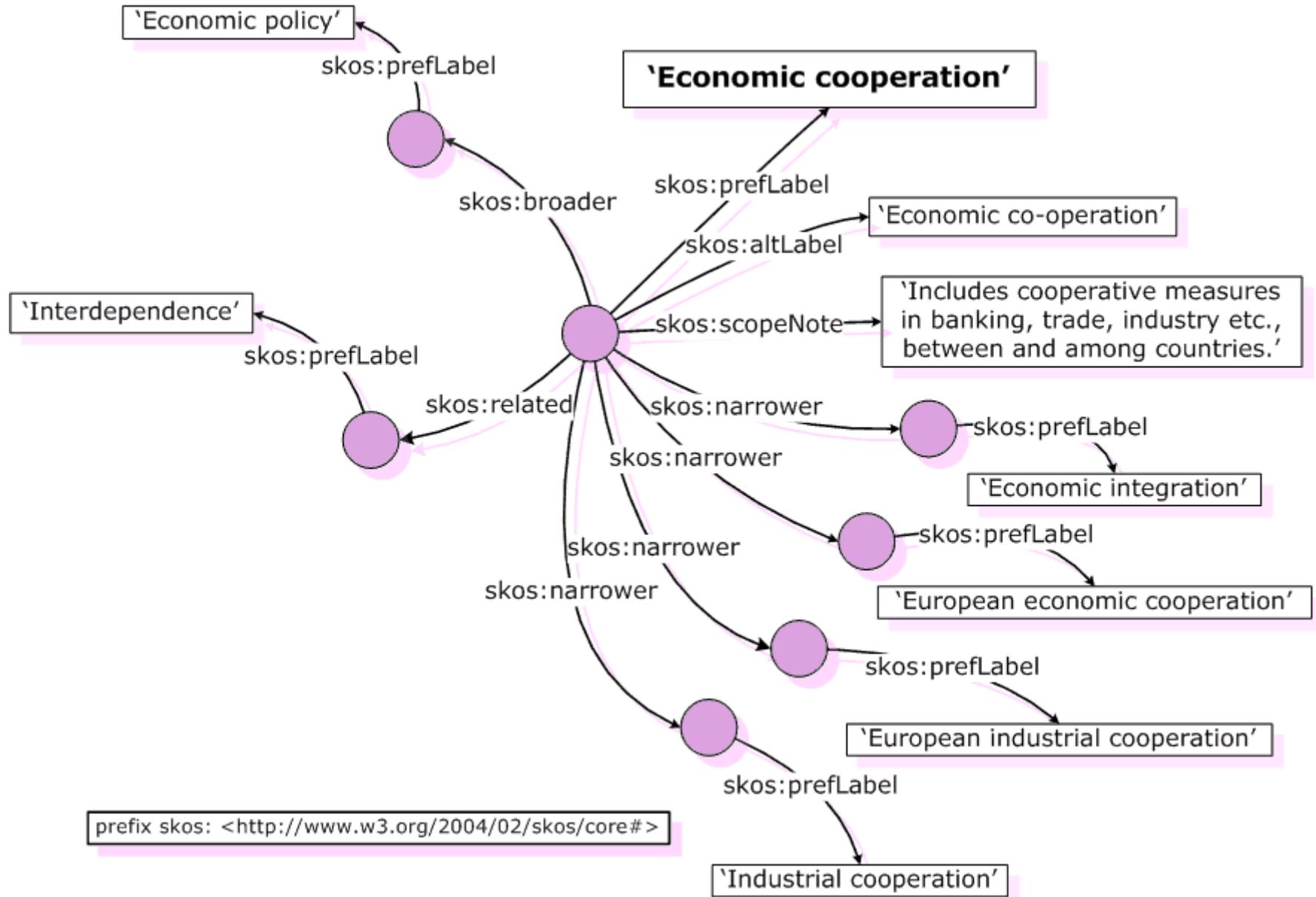
# RDF summary

- RDF basic model
  - Triples <subject, predicate, object>
- RDF syntax
  - Textual representation of sets of triples
  - XML, N3, ...
- RDF semantics
  - Intended meaning of sets of triples
  - Constraints
  - Implications

# SKOS – Simple Knowledge Organisation Systems

- Practical need: simplified representation frameworks
  - for conceptual models
- A system must be simple to allow for a quick port of traditional data
- SKOS is a specialised representation framework
- Suitable for thesauri, classification schemes, subject heading systems and taxonomies
  - [Dewey Decimal Classification](#), Art and Architecture Thesaurus, ACM classification of keywords and terms...
  - DMOZ categories (a.k.a. [Open Directory Project](#))
- Wrapper around RDF
- Adapt/define classes and properties within certain limits

# SKOS example



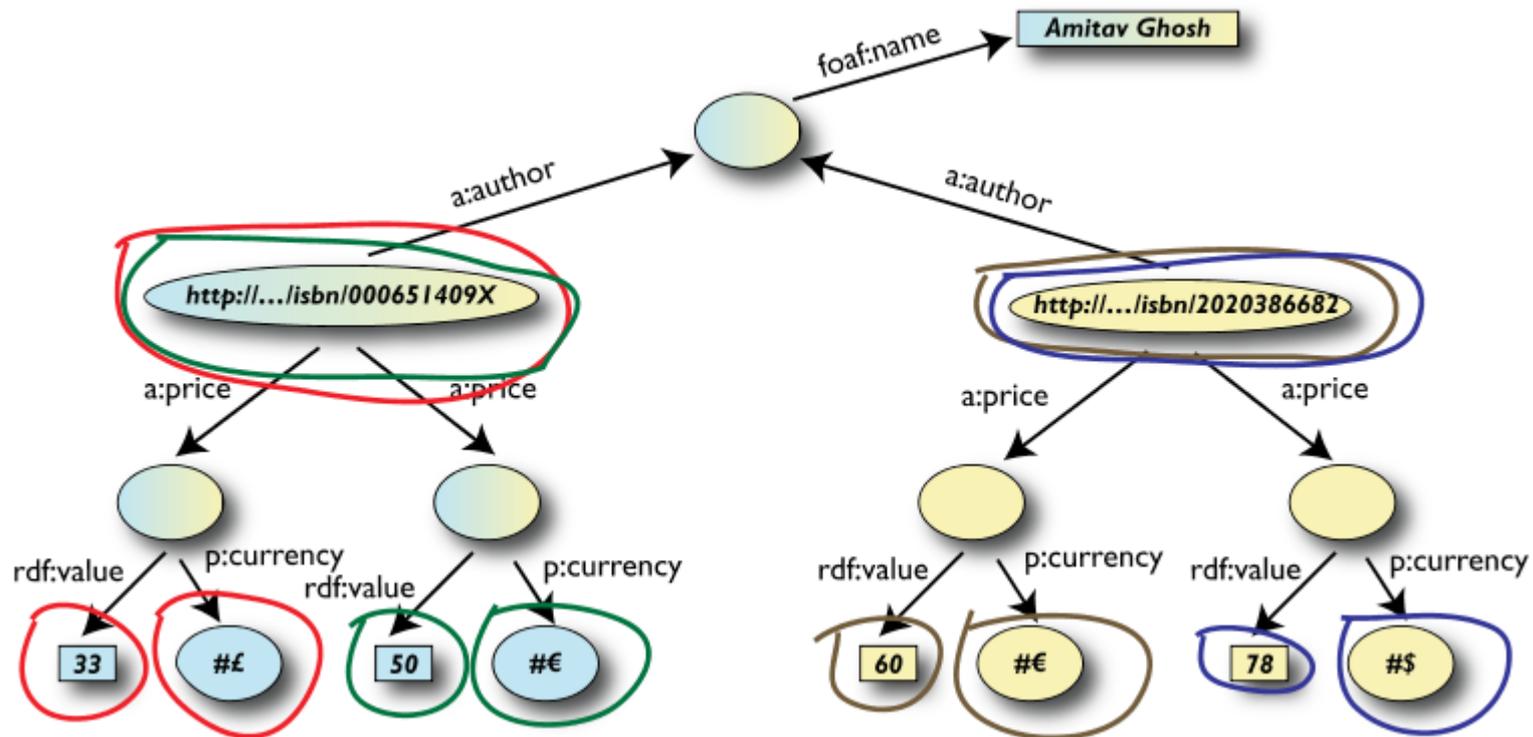
# SPARQL – Query Language for RDF

- How to use data represented in RDF?
- Extract / match / find data in RDF graphs
- Basic need: language for query on RDF graphs
  - example: “give me the (a,b) pair of resources, for which there is an x such that (x parent a) and (b brother x) holds” (ie, return the uncles)
- Queries very important for distributed RDF data!
  - Queries across distributed data bases
- This is the goal of **SPARQL** (Query Language for RDF)
- Compare:
  - SQL: query sets of **tables** of data
  - SPARQL: query **graphs** of data

# SPARQL example

```
SELECT ?isbn ?price ?currency # note: not ?x!  
WHERE { ?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency. }
```

- Returns:  
[[<...49X>,33,£], [<...49X>,50,€],  
[<...6682>,60,€], [<...6682>,78,\$]]



# SPARQL Usage

- Locally, i.e., bound to some programming environments
  - Querying local RDF databases
- Remotely, i.e., over the network
  - separate documents define the protocol and the result format
    - [SPARQL Protocol for RDF](#) with HTTP and SOAP bindings
    - SPARQL results in [XML](#) or [JSON](#) formats
  - big datasets often offer “SPARQL endpoints” for this protocol

# OWL – Web Ontology Language

- RDFS cumbersome for complex use
- Complex applications may want more possibilities:
  - similarity and/or differences of terms (properties or classes)
  - construct classes, not just name them
  - can a program reason about some terms? E.g.:
    - “if «Person» resources «A» and «B» have the same «foaf:email» property, then «A» and «B» are identical”
  - etc.
- OWL – “Web Ontology Language”

# OWL objectives

- A conceptual model describes some domain
- Ontology: formal description of a conceptual model
- OWL is a language for defining ontologies
  - OWL is a meta modelling language – a logical language
- Three layers of OWL are defined: Lite, DL, and Full
  - “OWL Full” is the whole thing
    - Complete logic
  - “OWL DL (Description Logic)” restricts Full in some respects
    - Mechanisable logic
  - “OWL Lite” restricts DL even more
    - Easily implementable

### **3. Semantics in action – examples**

Semantical use of web contents and other cases

# Newsfeeds

## Feeds

- Overview of news items
- RSS format
- Structured description
  - Title
  - Date
  - Abstract
  - etc.

File Edit View Go Bookmarks Tools Help del.jcio.us

http://www.w3.org/2000/08/w3c-synd/home...

Mozilla Stuffs blog-my Com'ty\_Social dictionaries encyclopedias MiscPersonal search se

Export Contact Add to Google Calendar Find with Google Maps Find photos on flickr Find bookmarks

European Sensor Network Architecture - H... http://www.w3.or...3c-synd/home.rss

**World Wide Web Consortium**

*Leading the Web to Its Full Potential...*

<http://www.w3.org/>  
2008-01-29

**W3C Publishes HTML 5 Draft, Future of Web Content**

2008-01-22: W3C today published an early draft of HTML 5, a major revision of the markup language for the Web. The HTML Working Group is creating HTML 5 to be the open, royalty-free specification for rich Web content and Web applications. "HTML is of course a very important standard," said Tim Berners-Lee, author of the first version of HTML and W3C Director. "I am glad to see that the community of developers, including browser vendors, is working together to create the best possible path for the Web." New features include APIs for drawing two-dimensional graphics and ways to embed and control audio and video content. HTML 5 helps to improve interoperability and reduce software costs by giving precise rules not only about how to handle all correct HTML documents but also how to recover from errors. Discover other new features, read the press release, and learn more about the future of HTML. (Permalink)

<http://www.w3.org/News/2008#item8>  
2008-01-22

**Call for Review: Canonical XML 1.1 Proposed Recommendation**

2008-01-29: The XML Core Working Group has published the Proposed Recommendation of Canonical XML 1.1. The specification establishes a method for determining whether two documents are identical, or whether an application has not changed a document, except for transformations permitted by XML 1.0 and Namespaces in XML. Canonical XML 1.1 is a revision to Canonical XML 1.0 designed to address issues related to inheritance of attributes in the

# Microformats

- Annotate HTML content with attributes
- Browser can offer specific actions
  - e.g., via add-ons
- Examples:
  - geo: locations
  - hCard: contact info
  - hCalendar: event info
  - Etc.
- Compare:
  - Skype web page annotations



# GRDDL – Gleaning Resource Descriptions...

- Existing documents may already contain useful descriptions
- Create RDF data by systematic extraction
- GRDDL – Gleaning Resource Descriptions from Dialects of Languages
- GRDDL offers a mechanism for retrieving descriptive data from document
- GRDDL introduces
  - markup for declaring that an XML document includes gleanable data
  - Describing an algorithm, typically represented in XSLT, for gleaning the resource descriptions from the document.

- Enriching web documents by semantical annotations
  - Without disrupting ordinary web document use
  - Enabling extraction of RDF and access as RDF
- RDFa slightly extends (X)HTML by:
  - defining general attributes to add metadata to any elements (c.f. “class” in microformats, but via dedicated properties)
  - provides an almost complete “serialization” of RDF in XHTML
- Similar to microformats approach but with more rigor and fully generic
  - makes it easy to mix different vocabularies (which is difficult in microformats)

# Application trends

- Use of SW technologies
  - Creating added value within user organizations
  - Taking advantage of the investment in XML as a common format
- Used internally in various software tools, e.g.:
  - Configuration descriptions
  - Process descriptions
  - Error reports, etc
- Used for metadata descriptions of digital resources
  - Obvious usage, of course
- Used for data integration
  - Big payoff! Lots of examples.

# SW data begins to accumulate on the Web

- [IgentaConnect](#) bibliographic metadata storage: over 200 million triples
- [Tracking the US Congress](#): data stored in RDF (around 25 million triples)
- [RDFS/OWL Representation of WordNet](#): also downloadable as 150MB of RDF/XML
- [“Département/canton/commune”](#) structure of France published by the French Statistical Institute
- [Geonames Ontology](#) and associated RDF data: 6 million (and growing) geographical features
- [RDF Book Mashup](#), integrating book data from, eg, Amazon
- [“dbpedia”](#): get infobox data of Wikipedia into RDF
- See, for example, the [linked data index](#)

# Find the right experts at NASA

- Expertise locator for nearly 20,000 NASA civil servants using RDF integration techniques over 6 or 7 geographically distributed databases, data sources, and web services...

POPS v.28.3 - Connected to 'POPS on FatDuck' - Using Model 'POPS on FatDuck Model' - Logged in as 'Andy'

**NASA Center (13)**  
ARC  
DFRC  
GRC  
GSFC  
HQ  
IVV  
JSC  
KSC  
LARC  
MAF  
MSFC  
SSC  
WSTF  
Source: x500

**Project (79)**  
Advanced Info Systems technology  
Advanced Studies, Concepts and Tool...  
Advanced Technology Initiatives  
Aeronautics Research Mission Directo...  
Agency Support  
Aqua  
Aquarius  
Cassini  
Center Investment Accounts  
Chandra  
Cloud-Aerosol Lidar and Infrared Pat...  
Constellation Systems Program Support  
Contract Management  
Source: WIMS

**Competency (21)**  
Acquisition Planning  
Acquisition and Contract Management  
Budgeting Management  
Business IT Systems  
Business Management  
Business Work & Team Management  
Contract Formation  
Contract Management and Performan...  
Cost Estimation and Analysis  
Employee & Team Leadership  
Executive Management  
Financial Management  
Governmental Affairs  
Source: CMS

**People (2)**  
Altonell L Mumford  
Michael J Milsted  
Source: x500

**Information Panel**

**Michael J Milsted**

Has Employer: NASA  
firstName: Michael  
phone: +1 202 358-4728  
mailto: michael.j.milsted@nasa.gov  
LDAP DN: cn=Michael Milsted-1,ou=Headquarters,ou=National Aeronautics And Space Administration,c=US  
Has Department: CH1000  
Room Number: 7F15  
Works On Project:  
Space Operations Mission Directorate  
Agency Support  
Has Competency:  
Business Management  
Business Operations Competency Suite  
Business Knowledge Domain  
Budgeting Management  
Financial Operations Competency Suite  
Cost Estimation and Analysis  
Program/Project Analysis

**View Different Social Network's in the POPS Data**

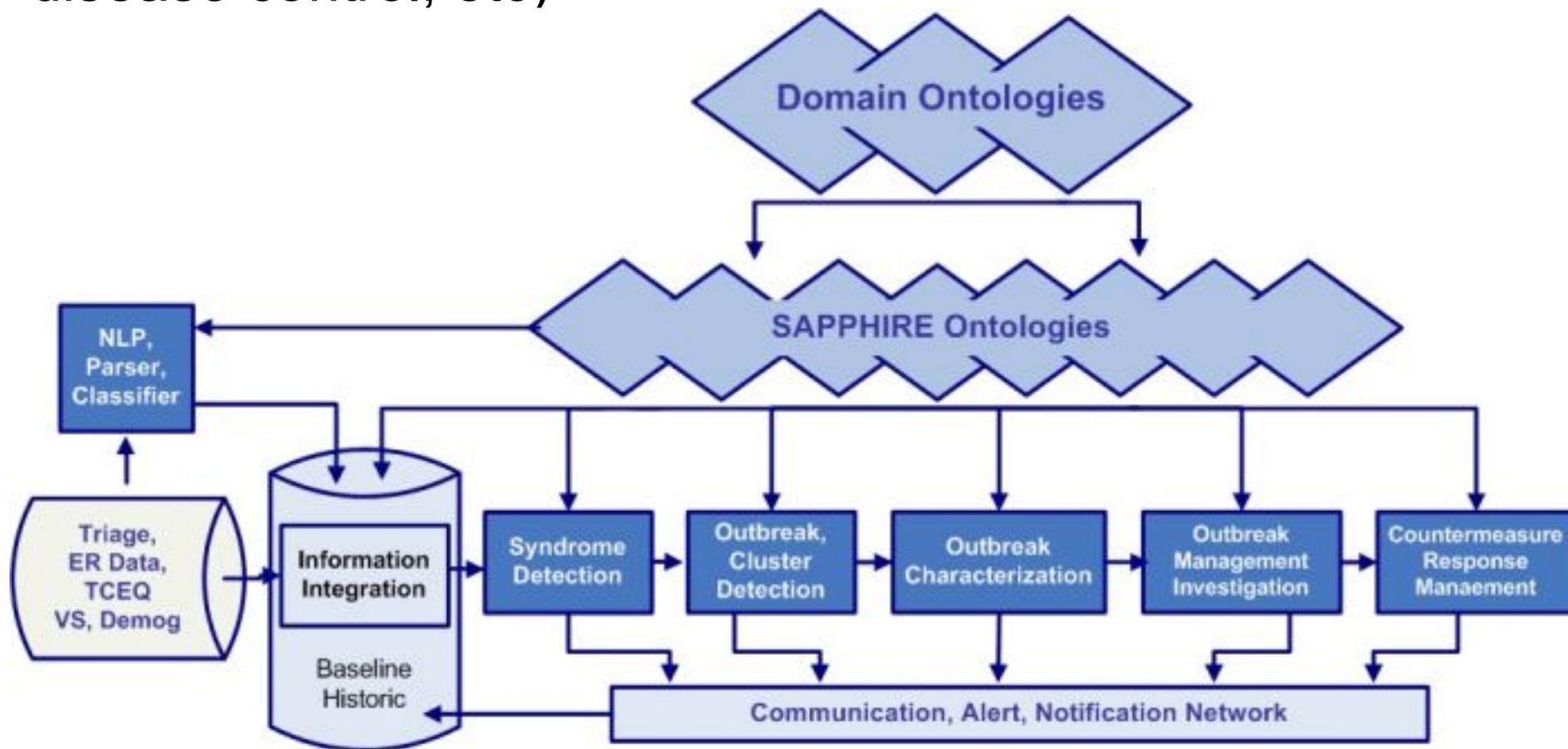
Legend:  
- Same Skill and Same Department (Red)  
- Same Skill and Same Project (Green)  
- Same Skill, Project, and Facility (Blue)  
- Am I Connected? (Experimental) (Purple)

Social Net Query Alternate Paths  
1 of 1

Courtesy of Kendall Clark, Clark & Parsia, LLC

# Public health surveillance

- Integrated biosurveillance system (biohazards, bioterrorism, disease control, etc)



Courtesy of Parsa Mirhaji, School of Health Information Sciences, University of Texas (SWEO Case Study)

# Semantic portal for cultural heritage

Patrimonio Cultural Cantabria Fundación Marcelino Botín

7 de Marzo de 2007

Inicio > Lugares > Cantabria

Buscador:  buscar

▼ Periodos ▼ Personas ▼ Instituciones ▼ Patrimonio ▼ Obras/Monumentos ▼ El Proyecto

ESPECIAL - Titular del especial

**Beato de Liébana**  
Autor de comentarios al Apocalipsis, compone himno litúrgico

Stephen Dedalus watched through the webbed window the lapidary's fingers prove a timedull'd chain. Dust webbed the window and the showtrays. Dust darkened the toiling fingers with their vulture nails.

**Marcelino Sanz de Sautoula y su hija**  
descubren las pinturas de las cuevas de Altamira

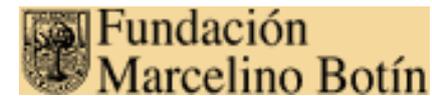
Lugares y Sitios  
Componente de Personas

**Torre medieval de Agüero**  
La torre medieval, de estilo gótico, fue construida en el siglo XIII

Mapa Interactivo TagCloud Linea del Tiempo

Santander Altamira

Santillana del mar Iglesia románica Libro de regla Santa Juliana



¿LO SABIA?

Pulse la BARRA ESPACIADORA para visualizar el siguiente resultado de la búsqueda. Pulse SHIFT-SPACE para volver al resultado anterior.

siguiente

TÉRMINOS RELACIONADOS

- Gobierno de Cantabria
- Universidad de Cantabria
- Comunidad Autónoma de Cantabria
- Caja Cantabria
- Castro Urdiales
- San Vicente de la Barquera
- Norte de España
- Ciudad de Santander

CATEGORÍAS RELACIONADAS

- Negocios (59%)**
  - ↳ Mercado inmobiliario
  - ↳ Guías y directorios
- Viajes y turismo (36%)**
  - ↳ Rural
- Sociedad (4%)**

UBICACIÓN SITIO WEB

- Europa (72%)**
  - ↳ España
- Norteamérica (27%)**

IDIOMA DEL DOCUMENTO

- Español (82%)**
- Inglés (15%)**
- Francés (1%)**

TIPO DE DOCUMENTO

- PDF  DOC  TXT  XLS  SWF
- PPT  RTF

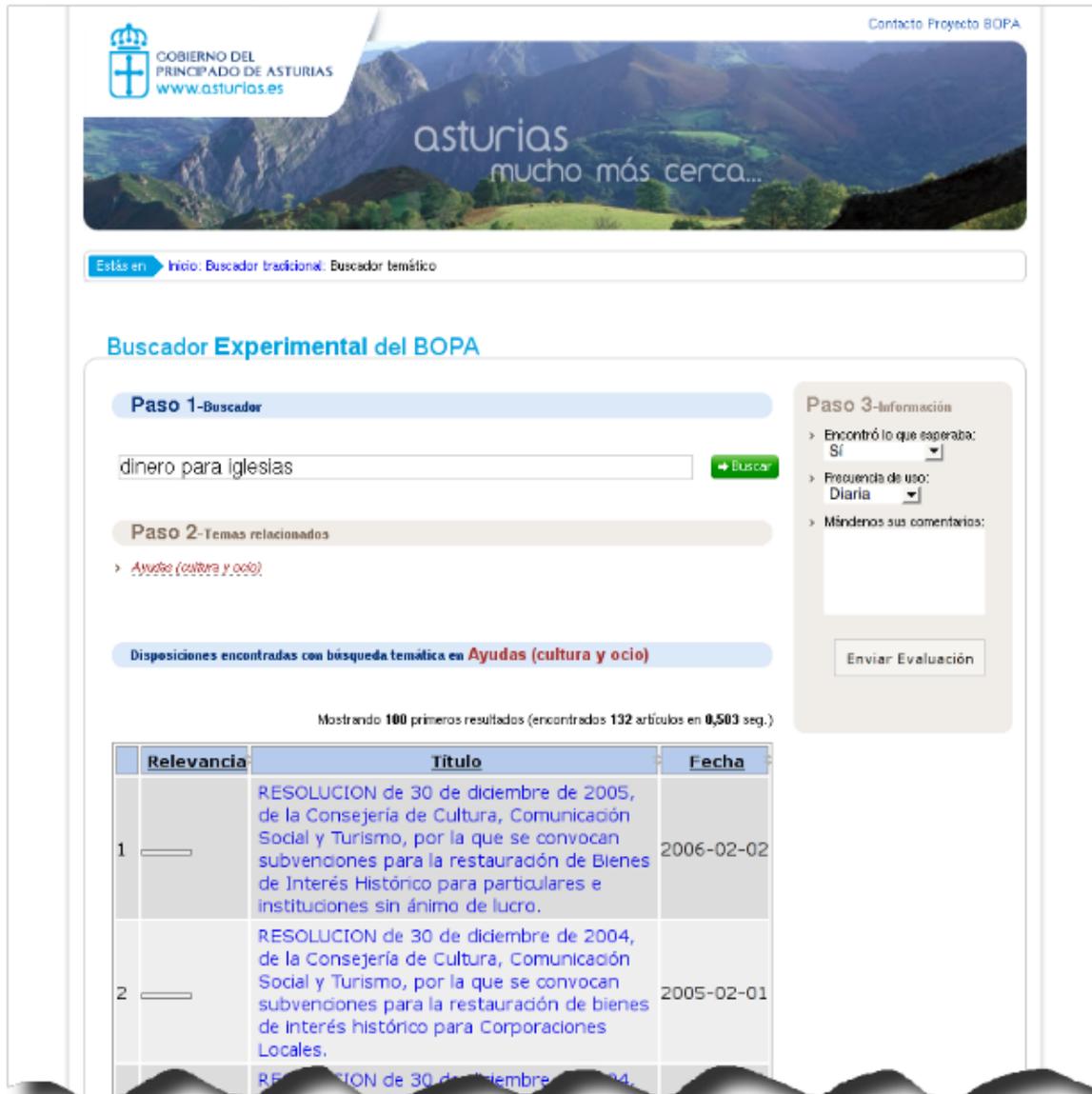
MODIFICAR BÚSQUEDA

- ▶ Búsqueda fonética
- ▶ Buscar en los resultados

Barra de estado

Courtesy of Francisca Hernández, Fundación Marcelino Botín, and Richard Benjamins, iSOCO, (SWEO Case Study)

# Portal to Principality of Asturias' documents



GOBIERNO DEL PRINCIPADO DE ASTURIAS  
www.asturias.es

Contacto Proyecto BOPA

asturias mucho más cerca...

Estás en Inicio: [Buscar tradicional](#): [Buscar temático](#)

### Buscador Experimental del BOPA

**Paso 1-Buscador**

dinero para iglesias

**Paso 2-Temas relacionados**

> [Ayudas \(cultura y ocio\)](#)

Disposiciones encontradas con búsqueda temática en [Ayudas \(cultura y ocio\)](#)

Mostrando 100 primeros resultados (encontrados 132 artículos en 0,503 seg.)

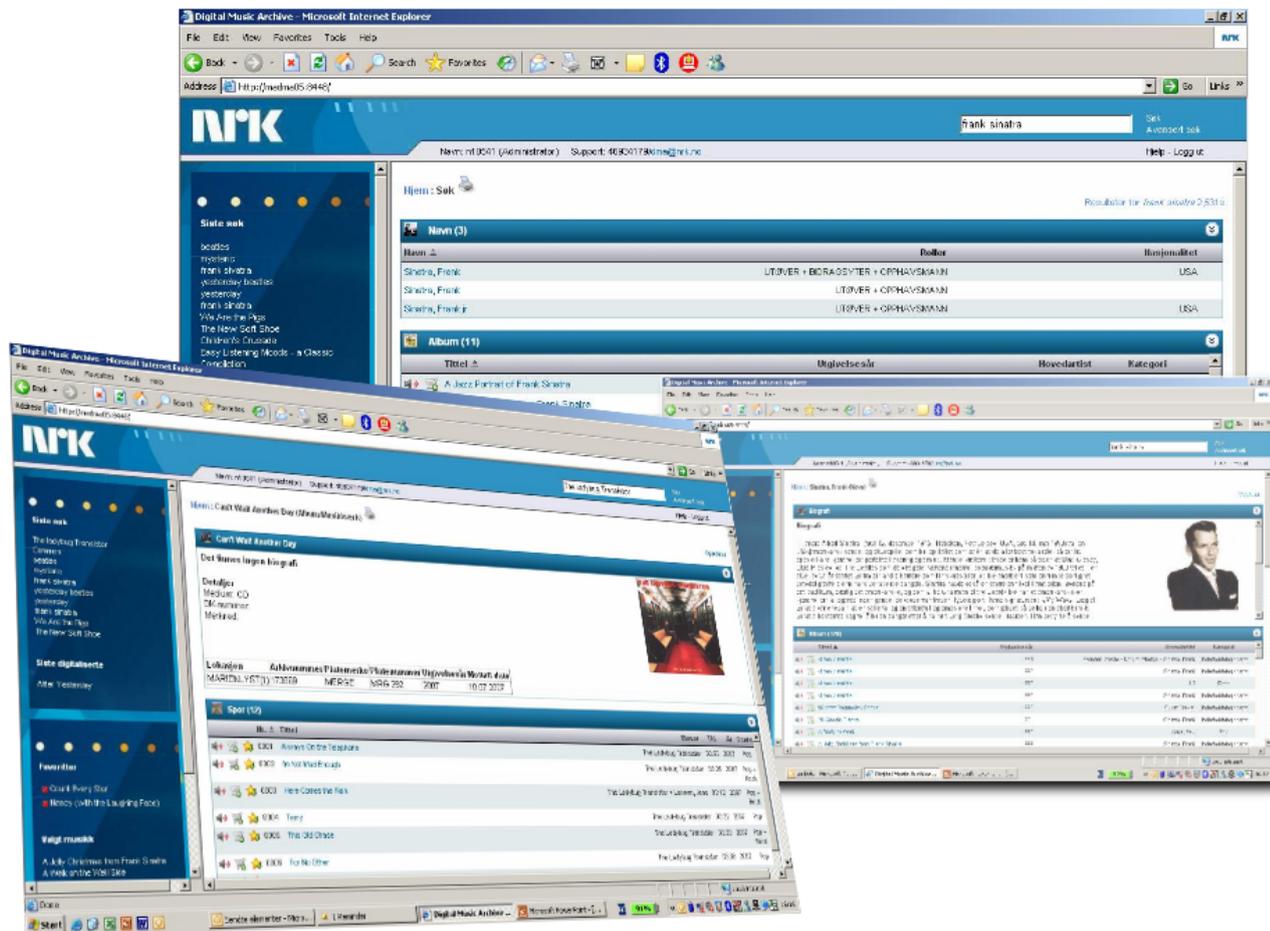
Relevancia	Título	Fecha
1	RESOLUCION de 30 de diciembre de 2005, de la Consejería de Cultura, Comunicación Social y Turismo, por la que se convocan subvenciones para la restauración de Bienes de Interés Histórico para particulares e instituciones sin ánimo de lucro.	2006-02-02
2	RESOLUCION de 30 de diciembre de 2004, de la Consejería de Cultura, Comunicación Social y Turismo, por la que se convocan subvenciones para la restauración de bienes de interés histórico para Corporaciones Locales.	2005-02-01

- Search through governmental documents
- A “bridge” is created between the users and the formal bureaucratic jargon using SW vocabularies and tools

Courtesy of Diego Berrueta and Luis Polo, CTIC, U. of Oviedo, and the Principality of Asturias, ([SWEO Case Study](#))

# Digital music asset portal at NRK

- Used by program production to find the right music in the archive for a specific show



Courtesy of Robert Engels, ESIS, and Jon Roar Tønnesen, NRK (SWEQ Case Study)

# Vodafone live!

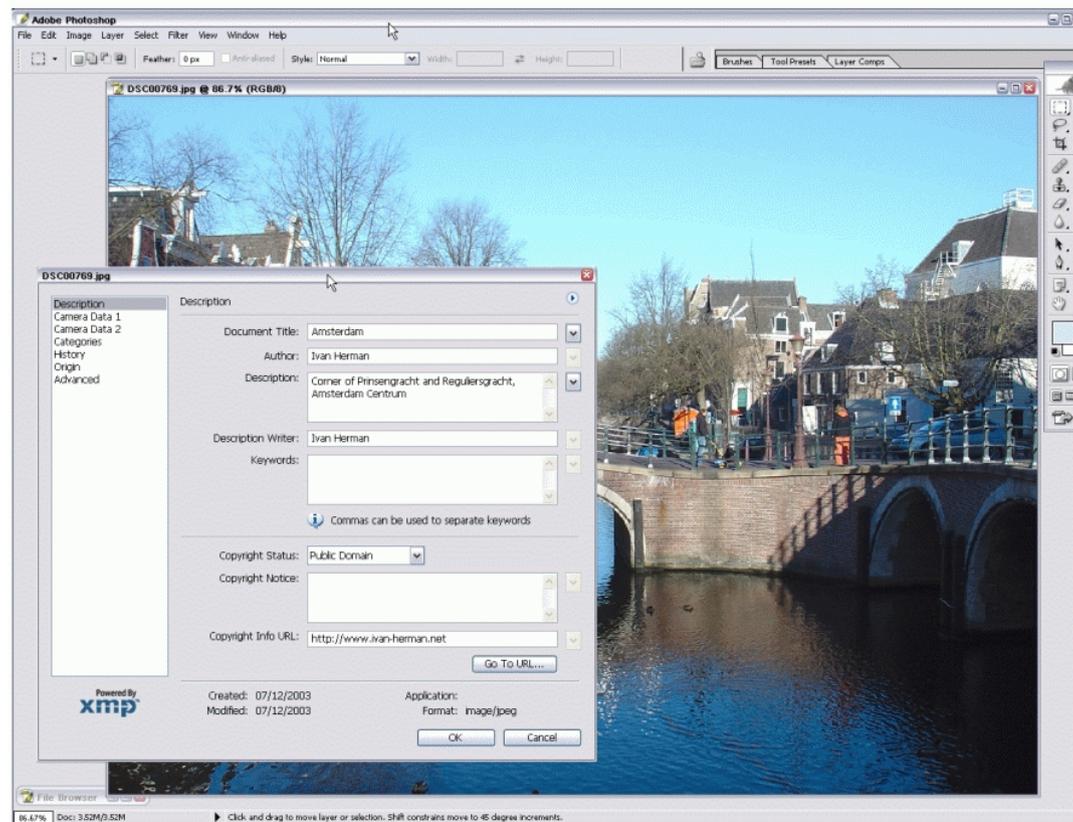
- Integrate various vendors' product descriptions via RDF
  - ring tones, games, wallpapers
  - manage complexity of handsets, binary formats
- A portal is created to offer appropriate content
- Significant increase in content download after the introduction



Courtesy of Kevin Smith, Vodafone Group R&D ([SWEQ Case Study](#))

# Adobe's XMP

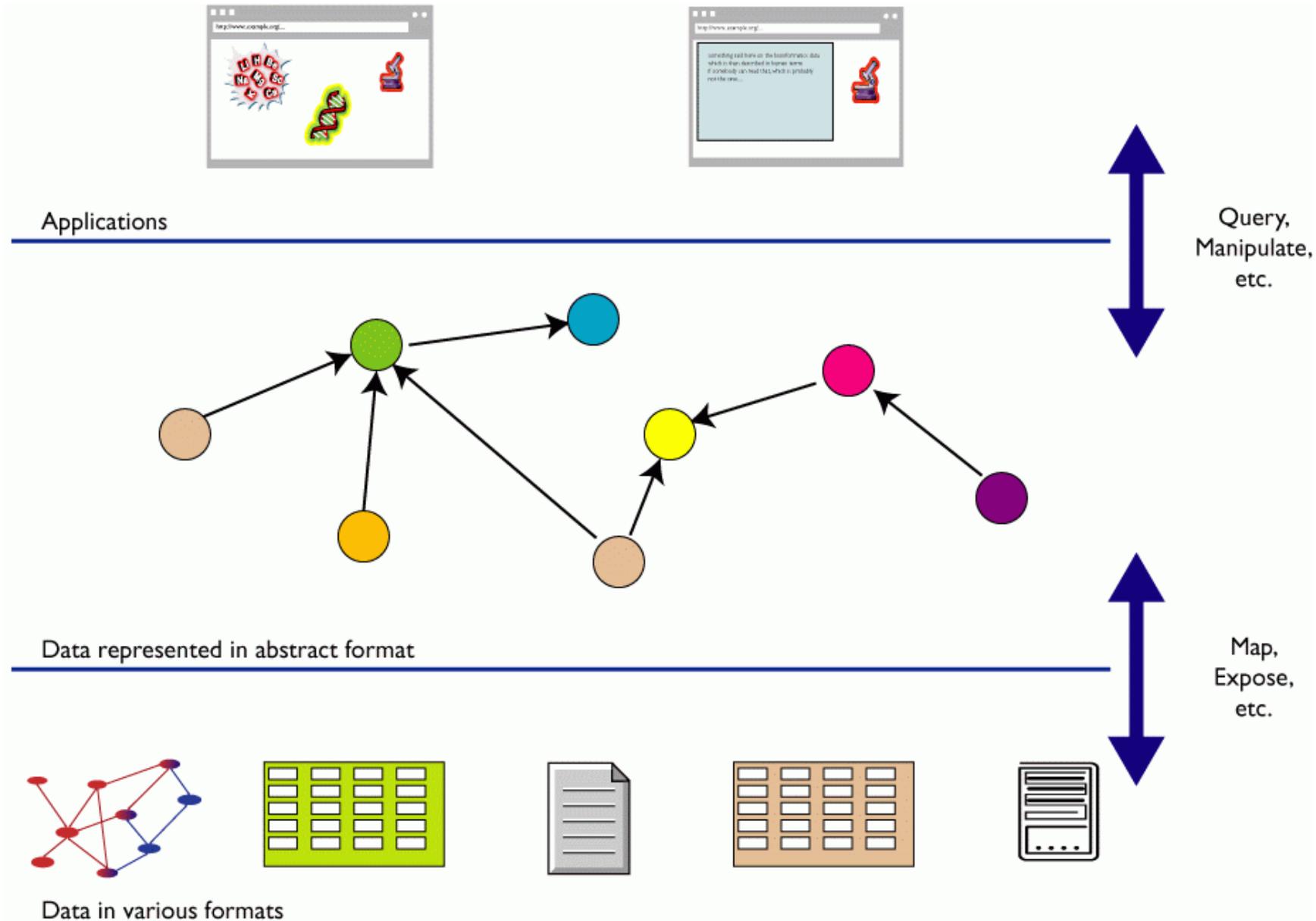
- Metadata is added by, e.g., Photoshop into files in RDF
- **XMP** is a way of embedding + vocabulary + a set of (public) tools (there are also 3<sup>rd</sup> party tools to extract the RDF content)
- Used by a number of platform solutions



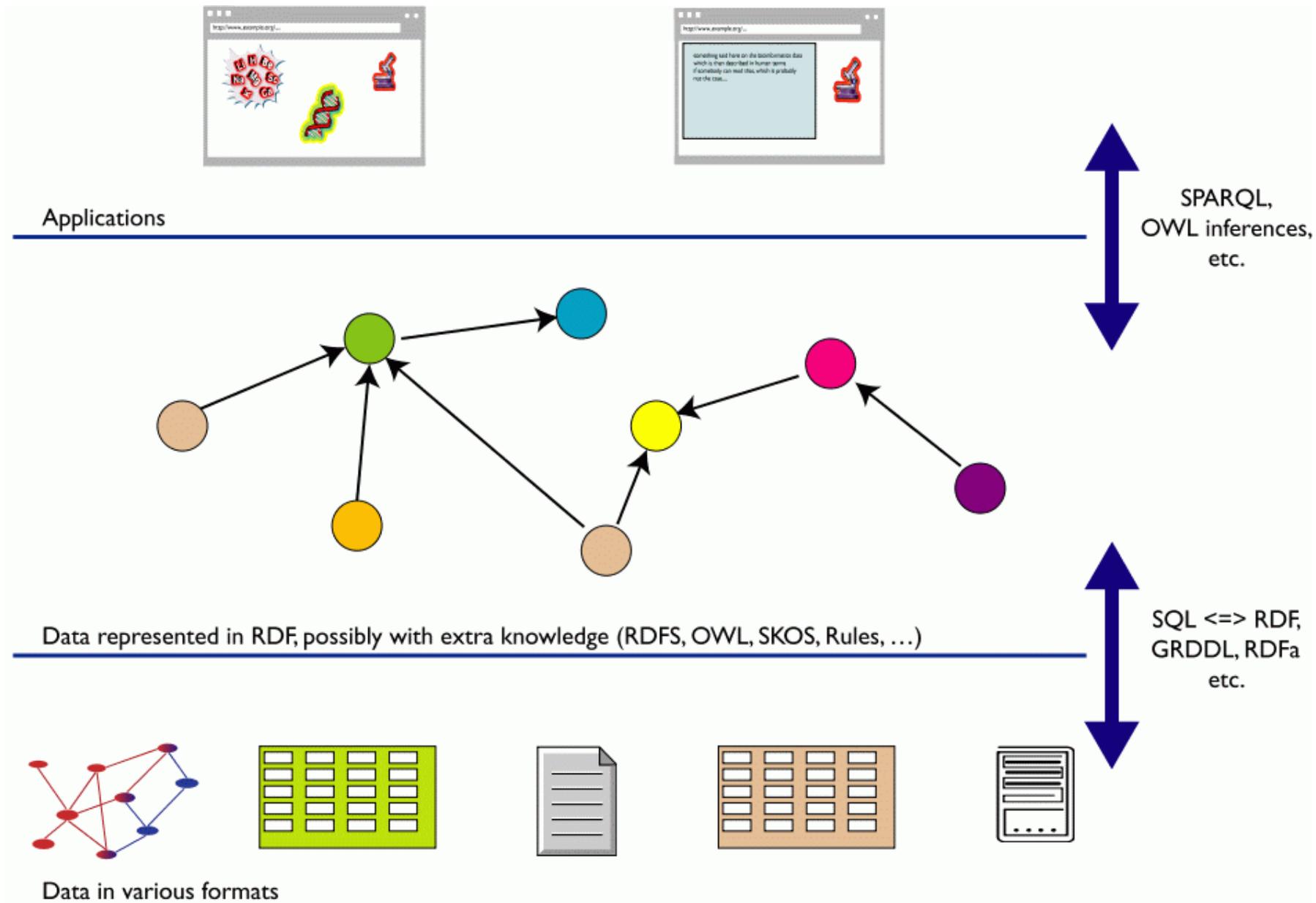
## 4. The larger landscape

How the Semantic Web fits in

# Where do technologies fit in?



# Where do technologies fit in? (cont.)



# Examples of implemented SW Tools

## • Triple Stores

- RDFStore, AllegroGraph, Tucana
- RDF Gateway
- Mulgara, MySQL+SPASQL
- Jena's SDB, D2R Server, SOR
- Virtuoso
- Oracle Spatial 10.2
- Sesame, OWLIM
- Talis Platform
- ...

## • Reasoners

- Pellet, RacerPro, KAON2, FaCT++
- Ontobroker, Ontotext
- SHER
- ...

## • Converters

- flickurl, TopBraid Composer
- GRDDL, Triplr, jpeg2rdf
- ...

## • Middleware

- IODT, Open Anzo, DartGrid
- Ontology Works, Ontoprise
- Oracle Fusion 11g
- Profium Semantic Information Router
- Software AG's EII
- Thetus Publisher, Asio, SDS
- ...

## • Semantic Web Browsers

- Disco, Tabulator, Zitgist
- OpenLink Viewer
- ...

## • Development Tools

- SemanticWorks, Protégé
- Jena, Redland, RDFLib, RAP
- Sesame, SWI-Prolog
- TopBraid Composer
- DOME
- ...

Inspired by "Entreprise Semantic Web in Practice", Jeff Pollock, Oracle. See also [W3C's Wiki Site](#).

# Semantic Web in Document Domain

- Semantic perspective on document web resources
  - Preserve and support the document view of resources
  - Enable semantic access to descriptions embedded in resources
  - Example: RDFa, semantic annotations of XML-based resources
  -
- Document perspective on semantic web resources
  - Preserve and support the semantical access to resources
  - Enable a document view of semantically represented resources
  - Examples: generation of presentation structure and style sheets from data
  -
- Dual perspectives, but different objectives:
  - Supporting methods and tools: RDFa, GRDDL, ...

# Semantic Web in Web Services

- The Web Services (WS) toolbox enables loose coupling between service user and service provider
  - Details in interaction controlled by descriptions – WSDL
- But content-related aspects of service use is not fully supported:
  - Service discovery, service characteristics, etc
- Emerging area: Semantic Web services
  - Semantic descriptions extend core WS descriptions
  - Support for semantic aspects on service use.
- The aim:
  - provide a flexible service framework that addresses the challenges of the web – *the dynamic character of services offered on the web*

# Summary

# Conclusions

- The Semantic Web is here to integrate data on the Web
  - The public web
  - Restricted webs
  - Intranet
- The goal is the creation of a Web of Data
- Core technologies/functionalities are standardised
  - RDF, OWL, SPARQL, ... RDFa, ...
- Additional needs/functionalities in the pipeline
  - By extensions to existing technologies
  - By additional tools and technologies

**Thank You for your attention!**