

The Semantic Web – an overview

Olle Olsson
Swedish W3C Office
Swedish Institute of Computer Science (SICS)

jFokus2008 – Stockholm, Sweden, Jan 30, 2007

© 2007 W3C

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



Olle Olsson: "The Semantic Web - an overview", jFokus2008

(3)

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
 - Find pages associated with text fragments
- 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!



Olle Olson: "The Semantic Web - an overview", jFokus2008
© 2008 W3C

(5)

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

Common
notation:
HTML, ...

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?*



Olle Olsson: "The Semantic Web - an overview", jFokus2008 (7)

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

- 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”



Olle Olsson: © 2008 W3C

(9)

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

User Interface & Applications

Query:
SPARQL

Ontology:
OWL

RDF:
RDFS

Rule:
RIF

Data interchange:
RDF

XML

URI/IRI

RDF – Resource Description Framework

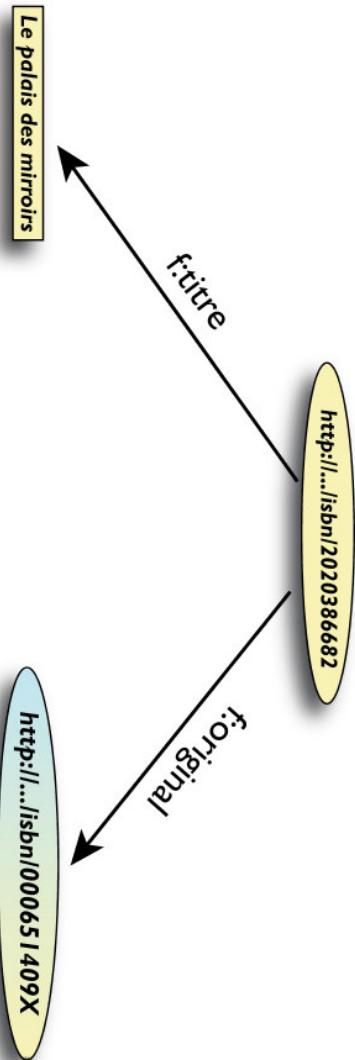
- Basic data model – a “triple”
 - triple (**s**, **p**, **o**) is such that:
 - “**s**”, “**p**”, and “**o**” stand for “subject”, “predicate”, and “object”, respectively
 - conceptually: “**p**” connects, or relates the “**s**” and “**o**”
- An example triple:

```
(  
  <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,
 - ... and that's it!

W3C WORLD WIDE WEB
c o n s o r t i u m
Svenska W3C kontoret
Olle Olsson: “The Semantic Web - an overview”, jFokus2008
© 2008 W3C (13)

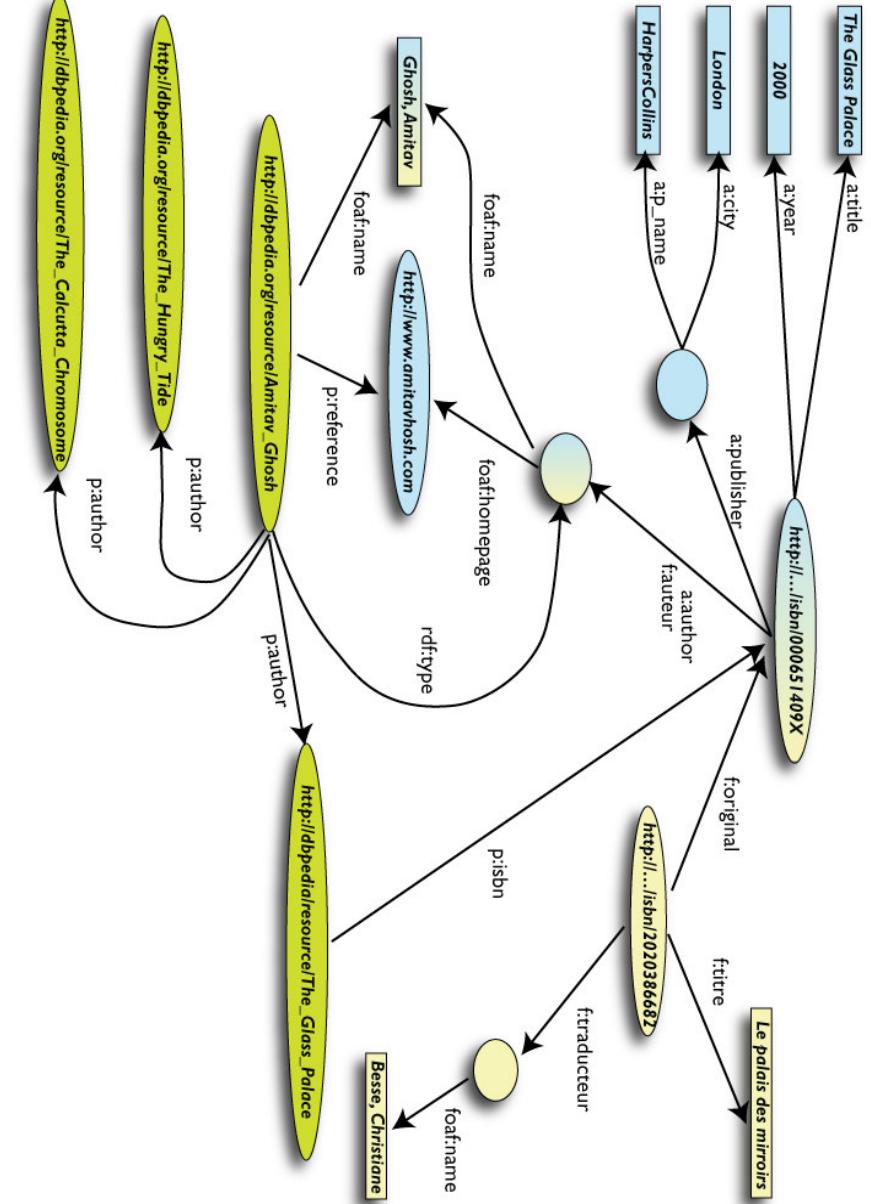
RDF Example



```
<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

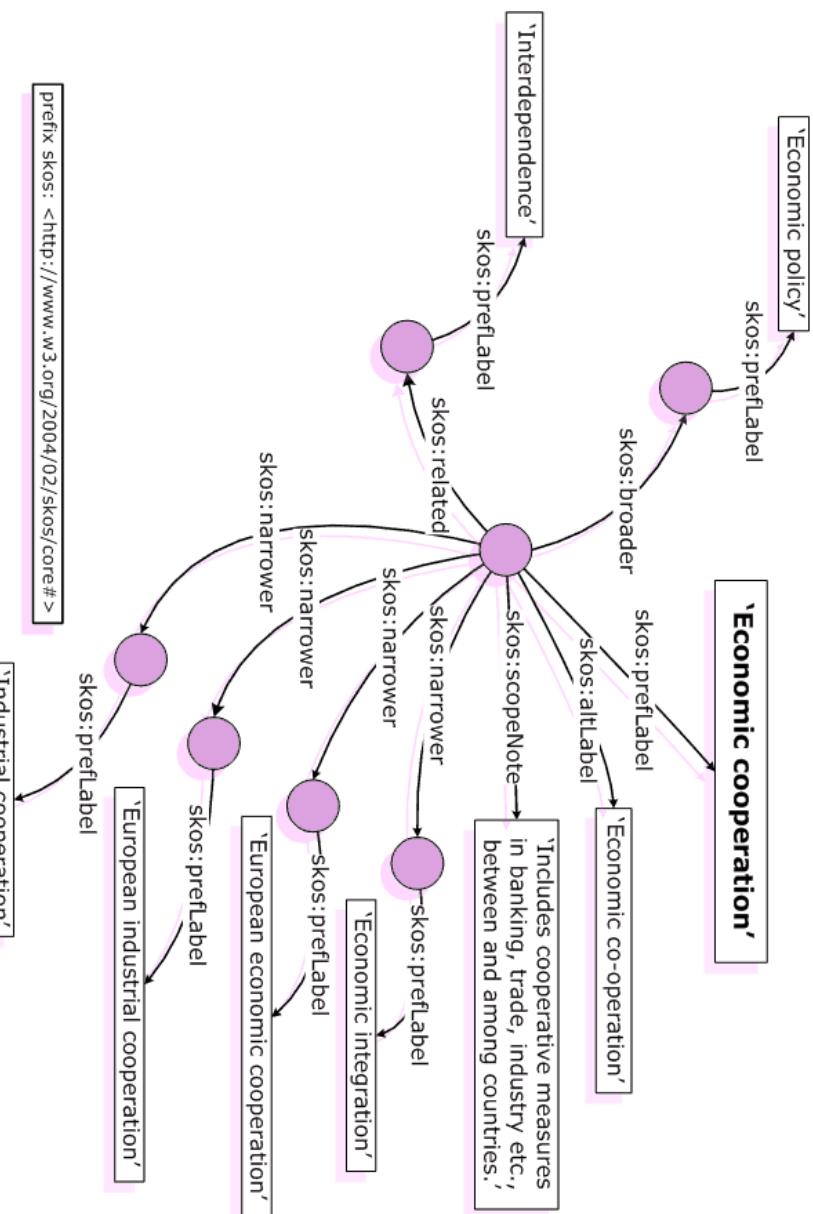


Olle Olsson: "The Semantic Web - an overview", jFokus2008
© 2008 W3C (17)

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



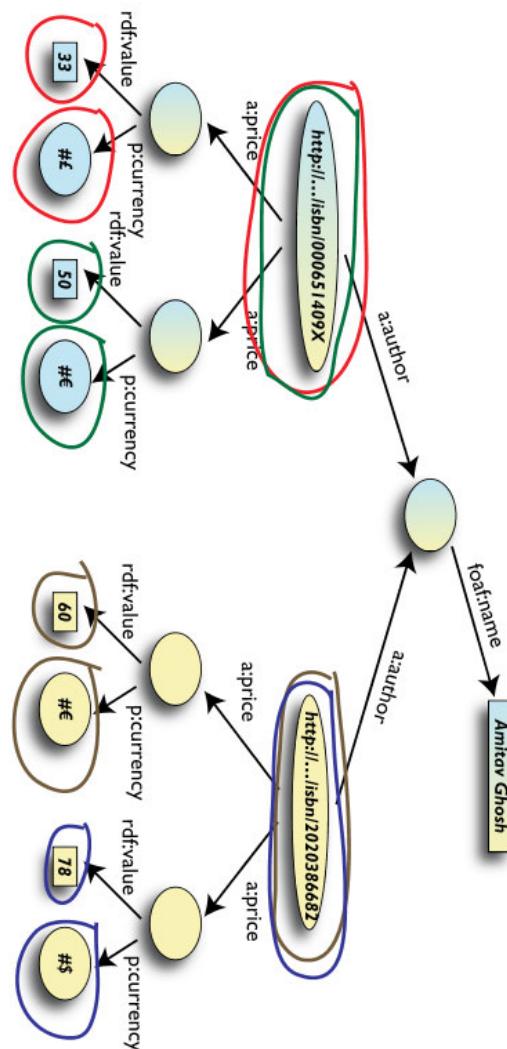
SPARQL – Query Language for RDF

- 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,\$]]



W3C WORLD WIDE WEB
Svenska W3C kontoret
Olle Olsson: "The Semantic Web - an overview", jFokus2008
© 2008 W3C (21)

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”



Olle Olsson: “The Semantic Web - an overview”, jFokus2008
© 2008 W3C

(23)

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



Olle Olsson: "The Semantic Web - an overview", jFokus2008 (25)

Newsfeeds

- Feeds
- Overview of news items

A screenshot of a web browser window. The title bar says "World Wide Web Consortium". The menu bar includes "File", "Edit", "View", "Go", "Bookmarks", "Tools", "Help", and "delicious". Below the menu is a toolbar with icons for back, forward, search, and other functions. The main content area shows a news feed with several items. One item is highlighted with a blue border. The URL "http://www.w3.org/2000/08/w3c-synd/home.rss" is visible at the bottom of the feed area.

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 ways 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

- RSS format
- Structured description
 - Title
 - Date
 - Abstract
 - etc.

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
- A screenshot of a web browser window. The address bar shows 'Great Barr - Wikipedia, the free...'. The page content is about 'Great Barr' in the West Midlands, England. Annotations are overlaid on the page:
 - A yellow box highlights the word 'Great' with the label 'article' below it.
 - A blue box highlights the word 'Barr' with the label 'geo (1)' below it.
 - A green box highlights the word 'Wikipedia' with the label 'hCard' below it.
 - A red box highlights the word 'Free' with the label 'hCalendar' below it.
 - A grey box highlights the word 'encyclopedia' with the label 'geo (1)' below it.Annotations also include a link to 'Find with Google Maps' and a 'Find with Yahoo! Maps' button.
- ## 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.

RDFa

- 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)



Svenska W3C kontoret
© 2008 W3C

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

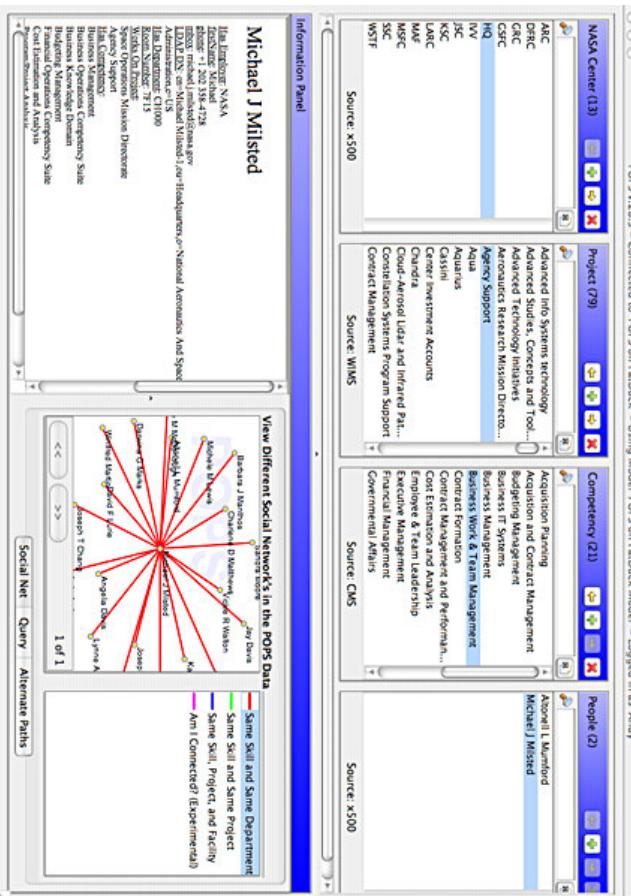


Olle Olsson: “The Semantic Web - an overview”, jFokus2008

(31)

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...



Courtesy of Kendall Clark, Clark & Parsia, LLC



WORLD WIDE WEB
c o n s o r t i u m

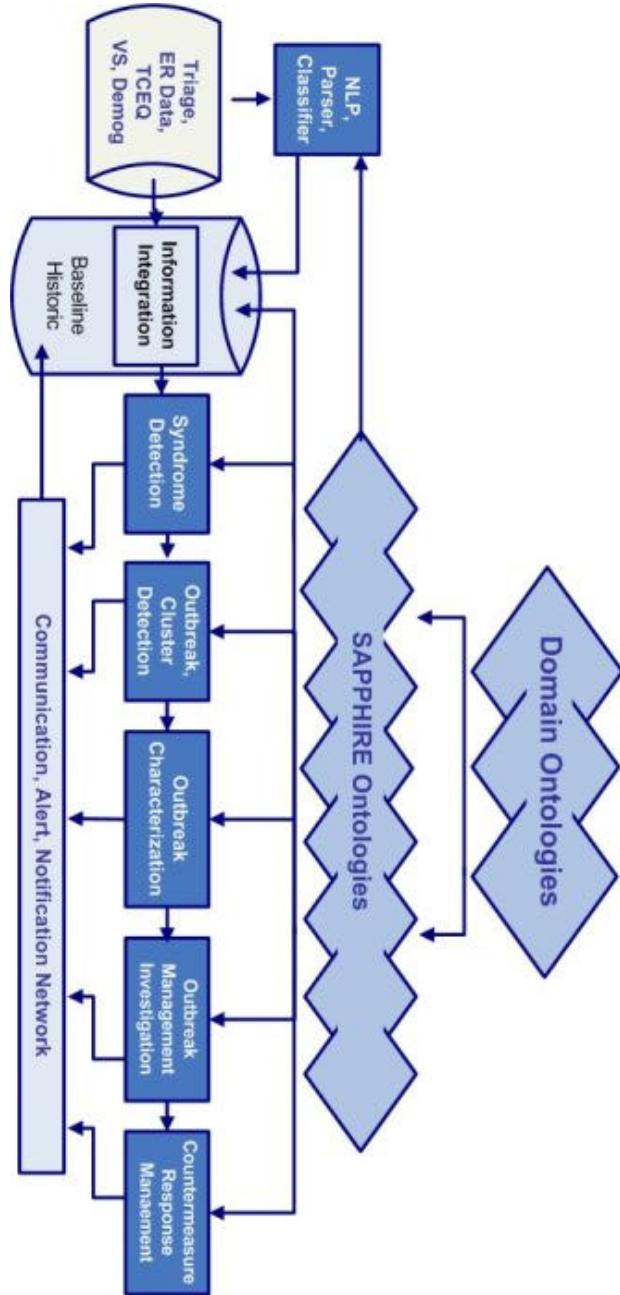
Svenska W3C kontoret

© 2008 W3C

(32)

Public health surveillance

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



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

W3C WORLD WIDE WEB
c o n s o r t i u m
Svenska W3C kontoret
Olle Olsson: "The Semantic Web - an overview", jFokus2008 (33)

Semantic portal for cultural heritage

Fundación
Marcelino Botín

This screenshot shows a semantic web interface for cultural heritage. At the top, there's a navigation bar with links like 'Proyectos', 'Personas', 'Instituciones', 'Entornos', 'Otras/Normativas', and 'El Proyecto'. Below the navigation is a search bar with placeholder text 'Búsqueda:' and a 'Buscar' button. The main content area features several cards:

- ESPECIAL - Titular del especial**: A card showing a painting of a church interior with the text: "Beatriz de Liebana". It includes a link to "7 de Marzo de 2007".
- Mapa interactivo**: A map of Cantabria with various locations marked.
- TopCloud**: A cloud of terms related to the site.
- Lunes del tiempo**: A weather forecast for Cantabria.
- Santillana del mar**: A card with a photo of a building and text about the town.
- Torre medieval de este**: A card with a photo of a tower and text about its history.
- Marcelino Sanz de Sautuola y sus hijos**: A card with a photo of a painting and text about the artist.
- Agüero La torre medieval, de este**: A card with a photo of a tower and text about its history.
- Almendra**: A card with a photo of a building and text about the town.
- Loreto de regla Santa María**: A card with a photo of a building and text about the town.

At the bottom, there's a map of the Iberian Peninsula with specific locations highlighted, and a legend for symbols used in the map.

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

W3C WORLD WIDE WEB
c o n s o r t i u m
Svenska W3C kontoret

Olle Olsson: "The Semantic Web - an overview", jFokus2008
© 2008 W3C

Portal to Principality of Asturias' documents



Este en [http://www.asturias.es](#) Buscador turístico

PRINCIPAL DEL ASTURIAS
www.asturias.es

asturias mucho más cerca...

Paso 1-Buscar destino

dinero para jefes

Paso 2-Temas recomendados

Ayudas (culturayocio)

Disposiciones encantadas con búsqueda temática en [Ayudas \(culturayocio\)](#)

Mostrando 100 artículos resultados encontrados 123 artículos en 48 páginas

Relevancia	Título	Fecha
1	RESOLUCIÓN 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	RESOLUCIÓN 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

RESOLUCIÓN de 30 de diciembre

Paso 3-Información

Encuentro lo que estás buscando

Forzar

Próximamente

Diariamente

Miércoles sus comentarios

Enviar Evaluación

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

W3C WORLD WIDE WEB Svenska W3C kontoret © 2008 W3C

Olle Olsson: "The Semantic Web - an overview", jFokus2008 (35)

■ Search through governmental documents

■ A “bridge” is created between the users and the formal bureaucratic jargon using SW vocabularies and tools

Digital music asset portal at NRK

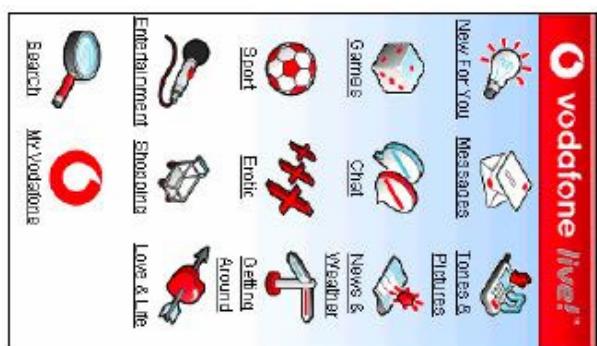
- Used by program production to find the right music in the archive for a specific show

A screenshot of a semantic web-based digital music asset portal. It shows several search results for "NRK" and "NRK.no". Each result includes a thumbnail image, a title, and some descriptive text. The interface includes various buttons and links typical of a web application, such as "Nyhet", "Søk", "Logg ut", and "Logg inn". The overall design is clean and modern, reflecting the Semantic Web's integration of structured data and user interaction.

Courtesy of Robert Engels, ESI/S, and Jon Roar Tønnesen, NRK (SWEO 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 ([SWEO Case Study](#))

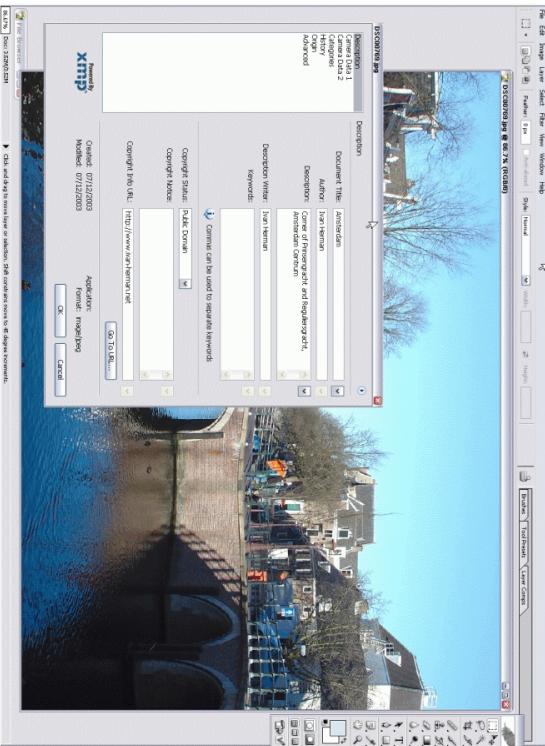
W3C WORLD WIDE WEB
c o n s o r t i u m
Svenska W3C kontoret

Olle Olsson: "The Semantic Web - an overview", jFokus2008

(37)

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 3rd party tools to extract the RDF content)
- Used by a number of platform solutions



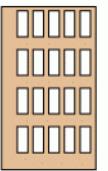
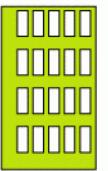
4. The larger landscape

How the Semantic Web fits in

Where do technologies fit in?

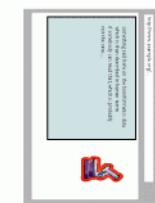
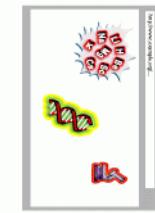
W3C WORLD WIDE WEB c o n s o r t i u m Svenska W3C kontoret
Olle Olsson: "The Semantic Web - an overview", jFokus2008
© 2008 W3C (39)

Data represented in abstract format

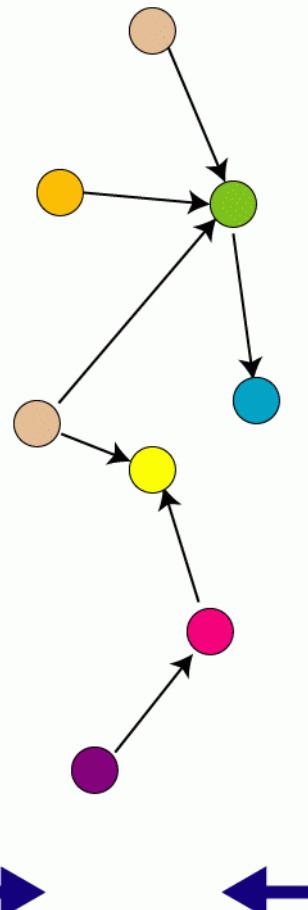


Map,
Expose,
etc.

Applications

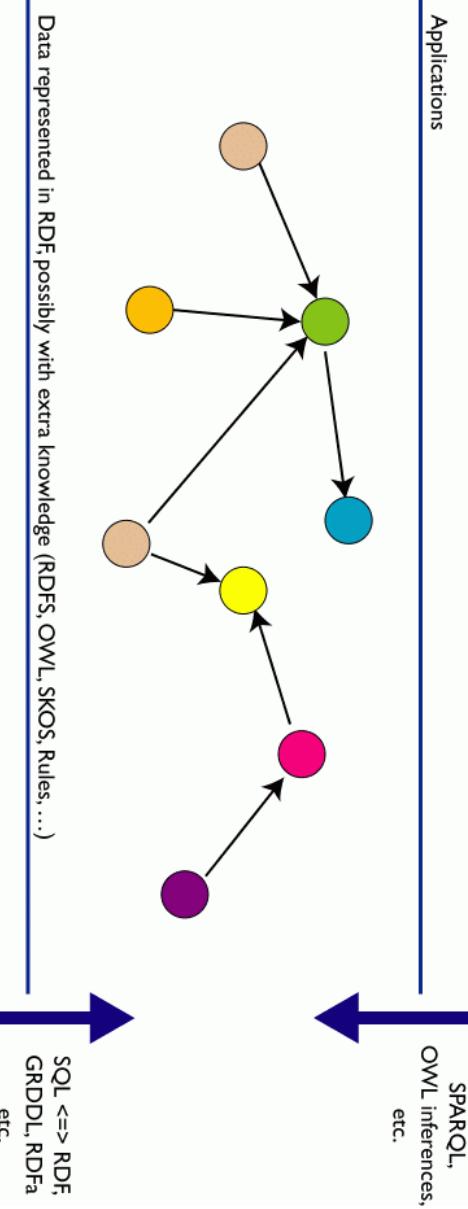


Query,
Manipulate,
etc.



Data in various formats

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
 - ...
- **Converters**
 - flickr!, TopBraid Composer
 - GRDDL, Triplr, jpeg2rdf
 - ...
- **Middleware**
 - IODT, Open Anzo, DartGrid
 - Ontology Works, Ontoprise
 - Oracle Fusion 11g
 - Profium Semantic Information Router
 - Software AG's Ell
 - 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 "Enterprise 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, ...



Olle Olsson: "The Semantic Web - an overview", jFokus2008
© 2008 W3C (43)

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!



Olle Olsson: "The Semantic Web - an overview", jFokus2008

(47)

© 2008 W3C