Adding Semantic to Web Data and Services
Part 4 – Use Cases

Doctoral School, St Etienne January 2009

Alain Léger FT R&D Orange Labs Research
DR Knowledge Processing (KRR)
Manager Industry Area IST NoEs OntoWeb et Knowledgeweb (2000 -2007)
Associated DR CNRS Lyon I - LIRIS

Plan Cours 1 (5 janv 09 13:30 – 17:15 / 6 janv 09 8:15 – 12:00)

- Why adding semantics to the Web ? (1h30)
  - Introduction
  - Take Away and References

- Foundations of Semantic Web (2h15)
  - Introduction to Logics and Description Logics
  - Standards and non-standards Inferences

- From XML, RDF to OWL (2h45)
  - XML, RDF, RDF-S
  - OWL

- Applications and Roadmap (1h00)
  - Application Scenarios
  - Visions prospectives et verrous technologiques
Plan

- Semantic Web: The Big picture
  Hype or Reality?
- Accelerate the take-up
  Realize Business Use Cases!
  Make them profitable!
- And next …
  Relax 😊

Industry permeates the Research
Fast take up in industry is key!!

- **Showing the value to Business Units**
  - Do not oversell the technology (AI syndrom ...)
  - Convincing benefits on Not toy scenarios!
  - Fast ROI

- Hiding the complexity of technology to all users
- Focusing the research effort on key Industry Roadblocks
- Making available tools and compliant Frameworks
- Sharing the knowledge and theoretical skill with industry
- Standardizing on key elements

Do not realize the full picture at once!

Applications

Technologies, Innovations

Deployed useful applications

Research &
Main Goal

- Promote greater awareness and faster take-up of SWS technology by industry
  - in close and permanent synergy with research area

**Industry**

“bridge the gap” between industry and research
(bi-directional process!)

**Academia**

- How? create a virtuous feedback loop between industry and research, including education and training

---

**NoE Knowledge Web Industry Area process**

1. **Specific Industry Needs**
   - Use Cases & Knowledge components

2. **Technology Suitability Evaluation**
   - Interoperable Framework & Ontology content

3. **Ontology Content Recommendation**
   - certifying, and serving validated ontologies
   - OOA Alliance

4. **Promotion, Deployment**
   - Technology RoadMaps, Success stories, Technology show cases, Best Practices

5. **Cross-Network cooperation** - joint program of activities & Joint Education, competence centers

6. **Kweb Portal**
Industry Support

Alain Léger
Robert Meersman
Chairs of Industry Area Knowledgeweb

Industry Board members

Including:
- Client Industry
- Technology push companies
- Ontology content partners
Industry members

- Publishing, Content
  - Merck, Rock International Ltd
  - Office Core Engineering
- Manufacturing
  - MOTION
- Banking
  - Allied Bank
  - Information and Financial
  - Citibank
- Telecom
  - Ericsson Telecom
  - Nokia Telecom
  - Samsung Telecom
- Energy
  - SFR (nat network)
- Transport
  - Europabus
- Health care and Biomedicine
  - BNP
  - IT Language and Computing
  - Synthetica
  - Hostens
- Space
  - Space Act
- Automotive
  - PSA
  - FCA
- Media
  - Radio
- Space Defense
  - Airbus
  - EADS
  - DASA
  - DGA
  - MINDEF
  - Deimos Space

Industry Partners facts

- 50 Industry partners on Board (Including OOA)
- 50 prospects
- 15 nationalities
- 14 economic sectors
- Focused on areas
  - With rapid promising results
    - Health Care and Life Sciences
    - Telecommunications
    - Human Resources management
    - Culture Museums
Contract Alliance

- Knowledge Web Early Access Policy
  - A panel of industrial corporations working as a "laboratorial" target market
  - Where all can
    - Study industrial requirements,
    - Test the industrial value of their ideas,
    - Collect inspiration for new research opportunities.
- Early Knowledge transfer
  - "Early Releases"
    - Results and deliverables from KW technical, are to be made available to Panellists at the respective owners' discretion;
    - Early Releases have to be treated as confidential by the Panellist,
  - The Panelists agree
    - To test the Early Releases diligently and to communicate to the Contractors sufficient progress and error reports,
    - To provide inputs such as but not limited to, recommendations, or industrial requirements, feedback to the Contractors in relation to its own use of the Semantic Web services & technology.
- Industry managers responsibility
  - Ensuring timely achievement of the Panellists' tasks,
  - Ensuring feedback and information for the benefit of all Contractors

Ontology Outreach Authority

- From the POV of an enterprise, Semantic Web is an intrusive and a disruptive technology. Need also to pay attention to roles of legacy systems, methodology, and resulting research and education challenges
- Ontologies are very difficult to standardize. Ontologies usually are application-dependent; subjective; general evaluation criteria lacking

We suggest "standardization lite" by recommending ontologies:

Evaluate an ontology whether it is in concordance with the claims of its developers.
Goals, requirements, scope, reusability, usability, etc.

First sectors addressed:
- Human Resources and Employment
- Healthcare and Life Sciences
Semantic Web Use Cases

KnowledgeWeb : Industry Area – Research Area
Alain Léger, Lyndon Nixon, Malgorzata Mochol,
Roberta Cuel, François Paulus, Mustafa Jarrar.
Heraklion, 3 June 2005

Use case collection process

We send a short questionnaire to each industry board member asking for input
Use case collection process

We follow up with a face-to-face meeting to collect further details and write up the use case in consultation with industry.

Use Cases: Results so far

1. Recruitment
2. Multimedia content analysis and annotation
3. Peer-to-peer eScience portal
4. News aggregation service
5. Product lifecycle management
6. Data warehousing in healthcare
7. B2C marketplace for tourism
8. Digital photo album management
9. Geosciences project memory
10. R&D support for coffee
11. Co-ordination of Real Estate Management
12. Hospital Information System
13. Agent-based system for insurance
14. Daimler Chrysler Semantic Web Portal
15. Specialized Web Portals for Businesses
16. Integrated access to Biological data
Knowledge processing tasks

- Extract research challenges from use case
- Direct research to industrial requirements

Analysis of use cases

Technology locks

Integration of Biological Data Repositories use case:
Summary – An unified point of access to different biological data repositories accessible through the Internet
Technology locks – Generation and extraction of knowledge from biological data...
Using standards.. providing a unified entry point to different biological data repositories (ontology-based reasoning)
### Task typology

<table>
<thead>
<tr>
<th>Knowledge processing tasks</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Translation</td>
<td>Wrapper</td>
</tr>
<tr>
<td>Ontology Management</td>
<td>Reasoning</td>
</tr>
<tr>
<td>Matching</td>
<td>Reasoning</td>
</tr>
<tr>
<td>Matching Results Analysis</td>
<td>Reasoning</td>
</tr>
<tr>
<td>Content Annotation</td>
<td>Reasoning</td>
</tr>
<tr>
<td>Semantic Query Processing</td>
<td>Reasoning</td>
</tr>
<tr>
<td>Composition of Web Services</td>
<td>Reasoning</td>
</tr>
<tr>
<td>Results Reconciliation</td>
<td>Query Processor</td>
</tr>
<tr>
<td>Semantic Query Processing</td>
<td>Planner</td>
</tr>
<tr>
<td>Composition of Web Services</td>
<td>Planner</td>
</tr>
<tr>
<td>Results Reconciliation</td>
<td>Results Reconciler</td>
</tr>
<tr>
<td>Schema/Ontology Merging</td>
<td>Ontology Manager</td>
</tr>
<tr>
<td>Producing Explanations</td>
<td>Match Manager</td>
</tr>
<tr>
<td>Personalization</td>
<td>Profiler</td>
</tr>
<tr>
<td>Directory Management</td>
<td>Directory Manager</td>
</tr>
</tbody>
</table>

Most tasks repeat over cases, suggesting a stable typology containing core tasks stipulated by industry needs.

### Drilling down into Use Cases
Overview

- Overview
- D1.1.4: Use Case Summaries
  - Use Case 1: Recruitment, Worldwidejobs
  - Use Case 2: B2C portal, FT
  - Use Case 3: News Aggregation, Neofonie
  - Use Case 4: Product Lifecycle Management, Semtation
  - Use Case 5: Managing Knowledge at Trenitalia
  - Use Case 6: Access to Biological data, Robotiker
  - Use Case 7: Needs in Petroleum industry, IFP
  - Use Case 8: Hospital information systems, L&C global
  - Use Case 9: Multimedia Processing

Conclusions
Use Case 1: Recruitment

Job Portal

Employer

Job Posting annotated using controlled vocabularies

Semantic Matching of applicant’s profile with job postings

Recommended open positions

Applicant

Interview Recommendations

Semantic Matching of applications with position’s requirements

Recommended positions

Automated Preselection

Job Application annotated using controlled vocabularies

RDF Repository

Crawler

Semantic Matching Engine

Semantic Portal

Net API

Information Integration + Matching

RDF-annotated Websites

Non-RDF HR System

Information Providers

Information Consumers
Use Case 1: Recruitment

<table>
<thead>
<tr>
<th>Semantic Matching</th>
<th>Match Manager</th>
<th>Improved matching: weighting, measures…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage&amp; retrieval</td>
<td>Directory Manager</td>
<td>Scalability, retrieval reliability&amp;performance</td>
</tr>
<tr>
<td>Knowledge Extraction</td>
<td>Wrapper; annotater</td>
<td>Mapping non-RDF, semi-automation</td>
</tr>
<tr>
<td>Ontology managing</td>
<td>Ontology manager</td>
<td>Guidelines, change tracking (versioning)</td>
</tr>
<tr>
<td>Trust</td>
<td>Match Manager</td>
<td>Reputation as ranking criteria</td>
</tr>
</tbody>
</table>

Use Case 2: B2C portal

Customer

Query plan generation
Discovery and composition

Offers

Automated Knowledge fusion

Next two weeks I'm going to Heraklion. Could you propose me a personalized leisure package and a concise guide in Spanish?
Use Case 2: B2C portal

<table>
<thead>
<tr>
<th>Semantic Matching</th>
<th>Match Manager</th>
<th>Adapted matchers, for DB and services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local schemas mappings</td>
<td>Ontology Manager</td>
<td>Semi-automated ontology Merging and mappings</td>
</tr>
<tr>
<td>Knowledge Extraction</td>
<td>Automated Wrapper</td>
<td>Mapping non-RDF, semi-automation, ease of integration of new services / resources</td>
</tr>
<tr>
<td>Ontology managing</td>
<td>Ontology manager</td>
<td>Guidelines, change tracking (versioning), tool and methodology</td>
</tr>
<tr>
<td>Trust</td>
<td>Match Manager</td>
<td>Reputation as ranking criteria</td>
</tr>
</tbody>
</table>

Use Case 3: News Aggregation

Provision of an aggregated news service that is able to provide business clients with accurate search, thematic clustering, classification of news stories, and e-mail notification of stories of interest.

http://www.newsexpress.de

The current solution is a semi-automatic approach consisting of two phases:
- the manual creation by a source expert of a XSLT template for each news source,
- the automatic processing of that news source through a thematic clustering algorithm (NLP) and classification with category mappings.
Use Case 3: News Aggregation

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Annotation manager</th>
<th>Tool integration, automation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontology Mapping</td>
<td>Wrapper</td>
<td>Dynamicism, performance</td>
</tr>
<tr>
<td>Ontology Development</td>
<td>Ontology manager</td>
<td>Extraction from text, fuzziness</td>
</tr>
<tr>
<td>Search</td>
<td>Query, reasoner, results reconcile</td>
<td>Rich, efficient query, matching</td>
</tr>
<tr>
<td>Security, Trust</td>
<td>Annotation, wrapper, results</td>
<td>Usage rights, assess trust</td>
</tr>
</tbody>
</table>

Use Case 4: Product Lifecycle

There is a high cost associated with the development and maintenance of product catalogues throughout the product lifecycle. Expensive and complex tools are used due to non-standardized terminology of the structure of product catalogue data, causing difficulties in development and maintenance of that data.

A tool supports the representation of product configurations as Visio diagrams, in which the diagram components are tied to descriptions using a common terminology. The current tool could be extended to support Semantic Web standards such as OWL in the future.
## Use Case 4: Product Lifecycle

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontology development</td>
<td>Ontology manager</td>
</tr>
<tr>
<td>Personalisation</td>
<td>Profiler</td>
</tr>
<tr>
<td>Logic / rules</td>
<td>Reasoner</td>
</tr>
</tbody>
</table>

## Use Case 5: Managing Knowledge at Trenitalia

**Institutional K**

- Unique access
- A common taxonomy

**Portale**

**KBase**

- A central Repository of Content Management
- A unique Workflow of K contribution

**Distributed KM: Research and Development**

- Peer to peer document sharing and reviewing systems
- Semantic information retrieval tools
- Expert maps
- Social technologies that allow communities exchanging processes
- Documents upload on official repositories
Use Case 5: Managing Knowledge at Trenitalia

<table>
<thead>
<tr>
<th>Semantic Matching</th>
<th>Match Manager</th>
<th>Improved matching: weighting, measures…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage&amp; retrieval</td>
<td>Directory Manager</td>
<td>Scalability, retrieval reliability&amp;performance</td>
</tr>
<tr>
<td>Queries</td>
<td>Wrapper; annotater</td>
<td>Mapping non-RDF, semi-automation</td>
</tr>
<tr>
<td>Ontology managing</td>
<td>Ontology manager</td>
<td>Guidelines, change tracking (versioning)</td>
</tr>
<tr>
<td>Trust</td>
<td>Match Manager</td>
<td>Reputation as ranking criteria</td>
</tr>
<tr>
<td>Collaborative and social creation and reviewing of documents</td>
<td>Collaborative writing systems, and evaluation ranking</td>
<td>Contribution and position within community as ranking criteria</td>
</tr>
</tbody>
</table>

Use Case 6: Integrated access to biological data

Biologist Expert

a researcher wants to compare the result of an experiment with the genome annotation database

Ontology merging and mapping

Existing Ontologies → Semi-Automated merging and mapping → Annotations and wrappers generation

Request understanding → Semantic best Match → Selection of matching combinations

Automated Knowledge extraction

New knowledge generation → Knowledge mining and fusion → Selected Data resources
Use Case 6: Integrated access to biological data

Information
Integration + Matching
Knowledge fusion

Biology
Integration + Matching
Knowledge fusion

Semantic Query
Matching Engine

Automated Ontology
Merging and Mapping

Semantic Mediations

Biological Data
Repositories
(Nucleotide Sequences, amino acid sequences, ...;
corporate databases, results of experiments (DNA-chips),
health cards, medical literature sites,...)

Internet

Automated wrapper
generation

Non-RDF
BD System

RDF Repository

Non-RDF
BD System

Non-RDF
BD System

Non-RDF
BD System

RDF-annotated Websites

Use Case 6: Integrated access to biological data

<table>
<thead>
<tr>
<th>Semantic Matching</th>
<th>Match Manager</th>
<th>Adapted matchers for very heterogeneous DB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local schemas mappings</td>
<td>Ontology Manager</td>
<td>Semi-automated ontology Merging and mappings</td>
</tr>
<tr>
<td>Knowledge Extraction</td>
<td>Automated Wrapper</td>
<td>Mapping non-RDF, semi-automation, ease of integration of new services / resources</td>
</tr>
<tr>
<td>Ontology managing</td>
<td>Ontology manager</td>
<td>Guidelines, change tracking (versioning), tool and methodology</td>
</tr>
<tr>
<td>Knowledge Generation</td>
<td>Knowledge miner</td>
<td>Knowledge mining over ontology mediated sources</td>
</tr>
</tbody>
</table>
Use Case 7: Geoscience semantic memory

In a KM context to give semantic access to multi projects documents and data (software, subsurface models) to practitioners

Geoscience Experts

Automated annotation

90% Automated deep annotation with available taxonomy/ontology

Deeply Annotated Projects documents

Query plan generation

Request understanding

Semantic best Match

Selection of matching combinations

Automated Knowledge summary

Best available Information / knowledge

Knowledge extraction and summary

Selected Doc / Data Extraction

Use Case 7: Geoscience semantic memory

Information Integration + Matching
Knowledge summary

Complex Multimedia Documents Data Repositories

Geoscience experts

Semantic Mediations

Semantic Query Matching Engine

Automated Ontology Annotation

Automated summarization

Intra/Extranet

Non-RDF BD System

Non-RDF BD System

Non-RDF BD System
Use Case 7: Geoscience semantic memory

<table>
<thead>
<tr>
<th>Semantic Matching</th>
<th>Match Manager</th>
<th>Selection of the best (part-of) documents and data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated annotation</td>
<td>Annotator</td>
<td>Automated on-the-fly of complex documents / DB / SW</td>
</tr>
<tr>
<td>Knowledge Summary</td>
<td>Summarizer</td>
<td>Summarizing complex documents and data table</td>
</tr>
<tr>
<td>Ontology managing</td>
<td>Ontology manager</td>
<td>Large domain Ontology building and maintenance</td>
</tr>
<tr>
<td>Navigation</td>
<td>Ontology navigator</td>
<td>Ontology-driven navigation</td>
</tr>
</tbody>
</table>

Use Case 8: Hospital Information System

Dealing with the issues of database integration in the domain of healthcare

I want to access all diagnoses and treatment information about my patients, from all of these heterogeneous resource easily and quickly
Use Case 8: Hospital Information System

**Identified Technology needs**

- How to build a good ontology
- Foundational challenges in ontology modeling
- How to aggregate with other patient data:
  - Echograph
  - Cardiograms
- How to deal with (free text) medical reports:
  - Extraction,
  - Indexing,
  - Annotating

---

**Use Case 8 : Hospital Information System**

<table>
<thead>
<tr>
<th>Automated annotation</th>
<th>Annotator</th>
<th>To automatically annotate medical reports and DB.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated mapping</td>
<td>Merger</td>
<td>To automatically merge medical facts about a patient.</td>
</tr>
<tr>
<td>Knowledge Extraction</td>
<td>Extractor</td>
<td>Knowledge extraction from medical reports</td>
</tr>
<tr>
<td>Ontology Engineering</td>
<td>Methodology</td>
<td>Modeling methodology that guide ontology builders to have a high quality (and reusable) ontology content, i.e. facing foundational modeling challenges...</td>
</tr>
</tbody>
</table>
Use Case 9: Multimedia processing (1)

- Knowledge structures (ontology modeling) for multimedia resources
  - Tools for enriching ontologies with multimedia information
  - Construction of domain ontologies for knowledge-assisted analysis
- Knowledge-assisted multimedia content analysis tools to support concept detection and tracking
  - Ontology learning and knowledge-assisted analysis
  - Person / face detection and recognition, mood detection
  - Ontological text analysis

Use Case 9: Multimedia processing (2)

- Algorithms for high-level semantic reasoning for multimedia content
- User query analysis tools and intelligent search, retrieval, ranking and relevance feedback mechanisms
  - User query processing
  - Visual / conceptual hybrid search; relevance feedback

Give me the names of the players that made the score of this football match and give me their rating in the 2003 worldcup?
Other prominent applications!

- Bibster – A semantics-based Bibliographic P2P
  - [http://bibster.semanticweb.org](http://bibster.semanticweb.org)
- CS AKTive space – Semantic data integration
  - [http://cs.aktivespace.org](http://cs.aktivespace.org) (Winner 2003 SemWeb challenge)
- Flink: SemWeb for analysis of Social Networks
- Museum Finland: Sem Web for cultural portal
- ScienceDesk collaborative knowledge management system in NASA
  - [http://sciencedesk.arc.nasa.gov/](http://sciencedesk.arc.nasa.gov/) (3rd prize 2004 SemWeb challenge)
- Also see Applications and Demos at W3C SWG BPD
  - [http://esw.w3.org/mt/esw/archives/cat_applications_and_demos.html](http://esw.w3.org/mt/esw/archives/cat_applications_and_demos.html)

Industry-Research (some) Challenges

- Key areas for semantic solutions are **search** and **data integration**
- Industry wants to better find and use the (legacy) databases (migration!)

- Key technology locks are:
  - **development** of ontologies i.e. modelling of business domains, authoring, best practices and guidelines, re-use of existing ontologies and simple tools!
  - **knowledge extraction** i.e. the population of ontologies by finding knowledge within legacy data
  - **mapping** i.e. overcoming heterogeneity (use of different ontologies) by determining how one ontology can be expressed in terms of another
  - **Scalability**: approximation, modularization, distribution
  - **Matching**: exact vs. fuzzy matching
  - **Web services**: where are they really needed?
  - **Language extensions**: what aspects are missing
    - E.g. data types, expressiveness of rules, context, …
Transfer to industry in action

- European Semantic Web research is working hard
  - to meet industrial requirements underway
- Industry members provide a concrete testbed for testing and evaluating research results
  - tools, ontologies, components and methodologies underway
- It is also important to support technology migration & industrial training
  - adapted courses to practitioners underway
- First phase of results
  - concrete technology transfer to industry already in 2006

Who have real Business today?

Early Adopters!!
**Key Technology announced August 2005!**

**What Oracle Brings to Semantic Technologies?**

- Web-Medication: Providing information about Web resources and services that require (context rating, capabilities, others)
- Line of Business Applications: Share knowledge bases that can rely on simple technologies to full-fledged ontologies
- Social Network Applications: Friend of a friend applications, social network tracking and navigation common in security and intelligence applications
- Semantic Information Integration: Define simple, centrally business information model to support information sharing across applications
- Semantic Web: Enable automated processing of Web information by software agents
- Portals and Wikipedia applications: Ability to query large amounts of metadata

**RDF Data Model**

Oracle’s Spatial RDFNet2 introduces the industry’s first space, scalable, secure and reliable data management platform for RDF-based applications. New ontologies have been defined to manage RDF data in Oracle. Based on a graph data model, RDF types are partitioned, instead of rigidly typed, similar to other relational database systems. The Oracle Spatial RDF database ensures that application developers benefit from the scalability of the Oracle database to deploy scalable semantic-based inference applications.

**XML Data Management & Tools**

XML serves as a syntactic foundation for semantic technologies. Semantic technologies like RDF and OWL are built upon XML to guarantee a base level of interoperability.

---

**NoE Knowledge Web : 3 Pillars**

- **Industry**
- **Research**
- **Education**

**Horizontal integration**

- Ontology recommendation body (OBA)
- Virtual research centre
- Virtual education Platform

- KnowledgeWeb is an EU FP6 Network of Excellence
  - Runs from 2004 to 2007 (4 years)
  - Budget of approx €7 million
  - 18 academic partners and research centers from 11 countries
Selected references

- The Semantic Web: research and Applications
  - LNCS 2342 (ISWC 2002)
  - LNCS 2870 (ISWC 2003)
  - LNCS 3053 (ESWS 2004)
  - LNCS 3298 (ISWC 2004)
  - LNCS 3532 (ESWS 2005)
- Journal of Web semantics (Elsevier)
- Thematic portal http://www.semanticweb.org
- WWW conferences
  - Yearly semantic web applications challenge http://challenge.semanticweb.org
  - SIGsemis semantic web information systems journal
    - http://www.sigsemis.org
    - http://www.w3.org/2001/sw/
  All the subgroups
  - http://www.w3.org/2004/12/rules-ws/report/ (Rules working draft)
  - http://www.w3.org/2004/12/sw-es.html (Sem Web in Life sciences)
  - http://www.w3.org/2001/sw/Europe/ (Transferring W3C standards to Practitioners)
  - http://knowledgeweb.semanticweb.org/
  - Education portal
  - Industry portal