Two use cases involving Semantic Web Earth Science Ontologies

for CO2 reservoir modeling and characterization

Jean-François RAINAUD, IFP
j-francois.rainaud@ifp.fr

ANR : e-Wok_HUB Project
Environmental Web Ontology Knowledge Hub
www.inria.fr/sophia/edelweiss/projects/ewok

R11 Jean François RAINAUD  Norvégian Semandic Days : May 19th  2009
Two use cases involving Semantic Web
Earth Science Ontologies

Agenda

- Introduction
- Two Use Cases for C02 geological storage site studies
  - Document search to initiate the CO2 storage prospect
  - Earth modeling for geological site qualification as CO2 storage
- Proposed Methodology
- Developed Ontologies
  - Domain ontologies for document search
  - Domain ontologies for Earth modeling
- Document searching application
- Project status and conclusion
E-WOK_HUB Project

Introduction: Environmental Web Ontology Knowledge HUB, the team...

A French "Agence Nationale de la Recherche" project

06/2006 - 06/2009

7 partners:

- INRIA (coordination)
- ENSMP-ARMINES
- LISI/ENSMA
- BRGM
- EADS
- Defense & Security Systems
- IFP
- CRITT-Informatique CRCFAO
- Industrials
- Académics

Technologies Transfert Support
Set up interconnected Portals (the e-WOK HUBs), providing

- **WEB applications For Final End Users** with specific GUI
- **WEB services** to set up new applications and workflows
- **Access to Semantic resources**
E-WOK_HUB Project

Introduction: Applicative Objectives

- CO₂ geological storage prospection projects

  - In which are involved people with different profiles
    - geologists, researchers, engineers, project leaders...
  
  - Working on a large variety of available resources
    - Internal or external project reports, scientific articles, databases, larges technical files (SEGY, GRDECL ...) ...
  
  - Producing new knowledge that can be useful for current and future projects improving a better reuse of knowledge.

- Need the discovery, interoperability and integration of these resources
Two Use Cases for C02 geological storage site studies
The Use Cases Objectives

Use Cases for CO2 geological storage site studies

1/ Document search to initiate the CO2 storage prospect

Knowledge resources: docs, papers, reports

1st step: Annotate these resources thanks to a domain ontology
2nd Step: Constitute of a Knowledge Base with these annotations
3rd Step: Offer Semantic Query facilities on this Knowledge Base

Results: This must return accurate documents fragments
Use Cases Objectives

Use Cases for C02 geological storage site studies

1) Earth modeling for geological site qualification as CO2 storage

Knowledge resources: docs, papers, reports, technical structured data, + existing processing business services.

1st step: Annotate these resources thanks to a domain ontology
2nd Step: Constitute of a Knowledge Base with these annotations
3rd Step: Offer Semantic Query facilities on this knowledge base
4th Step: Propose interpretation management along the reservoir characterization process.

Results: This must facilitate interpretation tracking and reuse reservoir characterization processes.
Proposed Methodology
E-WOK_HUB Project

Methodology: IT Environment

Service Oriented Architecture

*Used Standards*

*a hub is a warehouse of semantic business resources*
**E-WOK_HUB** Project

Methodology : IT Environment : The WEB Lab (OpenSource provided by EADS)
Methodology: SOA & Semantic Web Technologies used in the project

- **Semantic web technologies**
  - Ontologies (*RDFS/OWL lite*)
  - Annotations (*RDF*)
  - Rules (*CORESE rules*)
  - Queries (*SPARQL*) invoques CORESE

- **SOA (service oriented architecture)**
  - ESB (*Petals*)
  - Semantic web services (*SAWSDL*)

  *CORESE is used to find possible service compositions*

- **Persistance of the Knowledge Bases on DataStores**
  - Persistance using OntoDB (*use OWL to set up*)
  - Queries using OntoQL (*use SPARQL to query*).
E-WOK_HUB Project

Methodology used during this project (for Use Case 1)

- **Step 1 Define Domain Ontology**

- **Step 2 Annotate semantically the resources**
  - *(A Semantic Annotation is a way of describing the meaning of resources).*
  - Today on our project the semantic annotation is effective on textual documents
  - During the project the semantic annotation will be effective on DB and technical files

- **Step 3 Manage together the annotations and the persistence of the instances of the Domain ontologies**

- **Step 4 : Query semantically the annotations**
Methodology: Step 1 Define Domain Ontology

- Analyse of Resources
  - Free extraction of signifiants words and verbs (FASTR/ACABIT)
  - DataBases and Technical Files examination (DB and XML schémas)

- Conceptualize Domain Ontology
  - Reuse existing Ontologies (ex: Time, DubinCore, French Administrative Geographic organization)
  - Inherits from existing Ontologies (Geon, NADM, GeoSciML)
  - Interact with Expert to define Ontology concepts (C Maps + ECCO) and associate concrete instances

- Translate in OWL.
Methodology: Step 1 Define Domain Ontology: ZOOM on ECCO services (INRIA)

- **ECCO**: ontology editor

- **contextual and collaborative**
- **ontology whole life cycle**
  - term extraction from domain texts
  - vocabulary development
  - hierarchy construction
  - OWL Lite representation edition
  - Track of modifications for annotation evolution
Term extraction from domain texts
Methodology: Step 2 Annotation generation services

- Text only
- Language detection
  - + language Annotation
  - Grammatical analysis
    - + verb, noun, adj, annotation
    - Semantic annotation
      - + domain Annotation
      - Statistic annotation
        - Annotations verb, adj, noun, ...

E-WOK_HUB Project
Methodology: Step 2 Annotation generation services

Text only

Language detection

+ language Annotation

Grammatical analysis

+ verb, noun, adj, annotation

Semantic annotation

+ domain Annotations

Statistic annotation

- Annotations verb, adj., noun, …
“3D stratigraphic geometries of the intracratonic Cenozoic Paris Basin were obtained by sequence stratigraphic correlations...”
An Annotation can be intrusive.

In Documents (non structured)
- A first link is referencing the URI of the document,
- then a second link is established with the instance and attached to a segment of the text

In Technical Information (structured)
- A first link is referencing the URI of the "structured document"
- Then a specific query is attached.
Methodology: Step 3: manage the persistence of Ontologies and annotations.

manage together the persistence of the instances of the Domain ontologies and their annotations in relational databases (extension of PLIB Technologie: ISO 15926 type)
Methodology used during this project (for Use Case 2)

- **Step 1**
  - Use the same Domain Ontology to annotate semantically the resources, structure knowledge database & develop a specific Geological Knowledge Editor (GKE) to complete the automatic phase.

- **Step 2**
  - Annotates semantically the resources and query the annotations to populate the Knowledge database with instances.

- **Step 3 : interactive knowledge adding with the GKE**
  - Annotate Data Files with Domain Ontology and associate it to instances of the knowledge database
  - Complete the knowledge database instances with relationships and properties of instances.

- **Step 4**
  - Query this knowledge database for geological model building
Methodology used during this project (Step 1 for Use Case 2)

GKE architecture
E-WOK_HUB  Project

Geological Knowledge Editor (Step 3 : Associate data and geological Objects)

By Drag and drop the user annotates the ‘Picoref_Faults.xyz’ files with an instance of Fault ‘F3’
E-WOK_HUB Project

Geological Knowledge Editor (Step 3: Business relationship b/wt geological Objects)

1-Drag & Drop of Horizon instance 'TDalleNacrée'

2-Drag & Drop of Horizon Instance 'BMarnesOstrea'
Methodology used during this project (Step 3 for Use Case 2)

The geologist must confirm the interpretation ‘TDalleNacrée IsOlder IsOlder Than BMarnesOstrea’
Domain Ontologies
The « manual vocabulary extraction » operated by experts on the reference documents has enabled us to define the following categories of geological terms, which are relevant in our case:

- Basic geology (units and boundaries),
- Geological Structure
- Geolocalization
- Geological Age
- Geological processes,
- Lithology and mineralogy,
- Hydrogeology and reservoirs
- Earth state and paleogeography concepts,

• These categories do not totally fit with the knowledge models presently available for geology (GEON, NADM, Geoscience ML)

• We have thus defined specific domain ontologies well focused on our needs based on Official dictionary and geological reference thesaurus.
Domain Ontologies: Overview

- Reservoir
  - has geolocalization
  - is a GeologicalObject
    - is issued from GeologicalEvent (Geological History)
      - datedBy
      - isPartOf
      - isComposedOf
      - GeologicalTime
      - uses

- GeologicalObject
  - GeologicalBoundary
    - GeologicalUnit (Formation)
      - has Lithology
          - GeologicalProcess
      - Lithology
      - GeologicalChange

- GeologicalTime
  - International GeologicalTimeScale

- BasicGeology
  - French Official Geographic Code
  - Dublin Core
  - Geological Boundary
  - Geological Unit

© IFP
CMap/Ontology conventions

1. `<class>`
   - `<(synonym)>`
2. `instanceOf`
3. `<instance>`
   - `<attribute>`
     - `<value>`
4. `<property>`
5. `<abstract-class>`
6. `<relation>`
7. `<sub-class>`
   - `<attribute>`
   - `<datatype>`
8. `<ontology-name>`:
   - `<super-class>`
9. `<rule>`

(cardinality [1] by default)

E-WOK_HUB Project

Need of formalisation to Develop Ontologies: "Our Conventions"
Example of one domain Ontology : Basic Geology
Example of Domain ontology: Basic Geology with Instances
Domain Ontology: Geological Boundaries

Geological Boundaries

- GeologicalBoundary
  - GeneticBoundary
    - StratigraphicBoundary (Horizon)
      - ChronostratigraphicBoundary
      - BiostratigraphicBoundary (Biohorizon)
      - LithostratigraphicBoundary (Lithohorizon)
  - NonStratigraphicBoundary
    - FractureBoundary
    - ThrustBoundary
    - MacroFractureBoundary
    - MicroFractureBoundary
    - FaultBoundary
    - DiapiricBoundary
    - ReefBoundary
    - SaltDomeBoundary
    - LithofaciesBoundary
    - IntrusiveBoundary
Domain Ontology: Geological Structure

Geological Structures

- Fold (Pli)
  - hasShape
  - hasType
    - FoldShape
      - Antiform
      - Synform
    - FoldType
      - Concentric Fold
      - Similar Fold
  - OnlapHorizon
- HorizonStructure
- FaultStructure
  - NormalFault
  - ReverseFault
- Diapir
  - Strike&SlipFault
- Unconformable Horizon
  - UnconformableOnLap Horizon
  - ParallelHorizon
Earth Modeling interpretation tracking
Earth modeling interpretation tracking

The geological modelling workflow is considered here as a case study for applying ontology-based integration and model annotation techniques.

We are thinking that The different models should be semantic annotated for being integrated.

And to track the workflows the diverse processus should be semantically annotated also.
Earth modeling interpretation tracking: The interpretation

The interpretation corresponds to an end user decision. How we can express?

**First type of decision:** this data represents a geological Object (for an horizon, it could be a triangulated surface).

**Implementation:** We can create an annotation which associates a geological object to a data (or a part of) and a geological object instance of a Knowledge base.

**Second type of decision:** Define a predicate between geological objects

- e.g. Stuffle_Formation hasinferiorboundary Heiderg_Formation_Top
- e.g. The Fault F3 interrupts the Fault F4

**Implementation:** A typed relationship (property – predicate) will be created between two instances of the Knowledge base and added to the Knowledge base.
E-WOK_HUB Project

Earth modeling interpretation tracking: What we are experimenting

- Formalize the **data-model** of geological models;
- Annotate instances using **local ontologies** (LO);
- Articulate LOs using a **global ontology** (GO).
E-WOK_HUB Project

Earth modeling interpretation tracking: The annotation model

- **DatamodelClass** is an abstract class that is the parent of all classes in the data-models.
- **OntologyConcept** is an abstract class that is the parent of all concepts in the ontologies.
- **AnnotationClass** makes reference both to the data-model class and the ontology concept.
Example of Application: Ontology for Interpretation management

E-WOK_HUB Project

© IFP
Detail on Application ontology for Geological modeling data types
Document searching application

Example of a query

with SparQL using CORESE engine
E-WOK_HUB  Project

Use of Developed Ontologies: Detail on the Capture and Query Process
Dealing with the question:

*Which diageneses have affected the Bathonian formations from the Paris basin?*

- *Diagenesis* is a concept described in the EWOK ontology for geological process. So the documents dealing with diagenesis are referenced as such in the EWOK annotation part of the Knowledge database. Their reference can be gathered by the EWOK system in *REFERENCE SET 1*
Dealing with the question:

**Which diageneses have affected the Bathonian formations from the Paris basin?**

- **Bathonian** is a geological period registered in the International Geological Time Scale and in the EWOK geological time scale ontology. Thanks to semantic annotation, the EWOK system can retrieve references corresponding to documents annotated with the concept **Bathonian** or with related concepts of higher rank such as **Jurassic**, **Secondary**, **Mesozoic**.

- These references will be gathered in **REFERENCE SET 2a**
Dealing with the question:

Which diageneses have affected the Bathonian formations from the Paris basin?

• Another possible option is to extract from the question rather than the word *Bathonian*, the expression *Bathonian formation*.
Dealing with the question:

Which diageneses have affected the Bathonian formations from the Paris basin?

- This expression links two concepts respectively described in the GeologicalTime and in the GeologicalUnit ontologies. They are linked together by means of the ontology for Basic Geology.
Dealing with the question:

**Which diageneses have affected the Bathonian formations from the Paris basin?**

- This expression links two concepts respectively described in the **Geological Time** and in the **Geological Unit** ontologies. They are linked together by means of the ontology for **Basic Geology**.
Dealing with the question:

Which diageneses have affected the Bathonian formations from the Paris basin?

• In this case, the system will retrieve references corresponding to documents annotated with the expression *Bathonian formation* or with any other term that will have been stored as an instance of this expression (for instance *Comblanchien*, which corresponds to a Bathonian formation of the Burgundy region)

• These references will be gathered in *REFERENCE SET 2b*
Dealing with the question:

**Which diageneses have affected the Bathonian formations from the Paris basin?**

- *Paris basin* is a non administrative geographic term, whose synonyms *bassin de Paris, bassin parisien* and which can be described by a polygon. The EWOK system can also identify the various administrative divisions lying inside this polygon. It can thus retrieve references documents annotated with the terms *Paris basin, bassin de Paris, bassin parisien* but also with terms like *ile de France* or *département du Loiret* (or many others corresponding to administrative division with the Paris basin). The corresponding references will be gathered in **REFERENCE SET 3**
Dealing with the question:

**Which diageneses have affected the Bathonian formations from the Paris basin?**

- *Paris basin* is a non administrative geographic term, whose synonyms *bassin de Paris, bassin parisien* and which can be described by a polygon. The EWOK system can also identify the various administrative divisions lying inside this polygon. It can thus retrieve references documents annotated with the terms *Paris basin, bassin de Paris, bassin parisien* but also with terms like *île de France* or *département du Loiret* (or many others corresponding to administrative division with the Paris basin). The corresponding references will be gathered in **REFERENCE SET 3**
Dealing with the question:

**Which diageneses have affected the Bathonian formations from the Paris basin?**

The answer to the question will be a set of references $S$ corresponding to the intersection of **REFERENCE SETS 1, 2 and 3**.

$$S = S_1 \cap (S_{2a} \lor S_{2b}) \cap S_3$$
Conclusion
E-WOK_HUB Project

Conclusion (1/2)

• The e_Wok system will be a possible solution for enabling users to identify and retrieve adequate documentation through internet, in order to solve practical issues such as identifying potential CO2 storage sites.

• The system aims at putting in correspondence semantic contents respectively related to questions asked by users and to various types of documents. It relies on various intercommunicating and cooperating web services.

• Specific goal-oriented ontologies have been developed for formalizing the geological and geographical vocabulary that must be considered in the case of CO2 storage issues. They will be used for complementing documents searched on internet by semantic annotations for allowing their identification, their storage in the system database and their later retrieval.

• Compared with other search methodologies, our approach has the advantage of being goal-oriented and of allowing largely automated document search.
Conclusion (2/2)

• We are in our last year for the e_Wok Hub project

• Domain ontologies have already been defined (2nd version)

• The global architecture of the system has also been defined

Before the end of the project we have:

• to complete the design of user interfaces (Humanization of SparQL)

• to finalize a demonstrator associating annotation on non structured and structured documents.

• To set up the e-Wok_HUBs for operations & validation tests.
Last Publications


Thank you!

j-francois.rainaud@ifp.fr

www.inria.fr/sophia/edelweiss/projects/ewok
• O. Corby, C. Faron-Zucker.  
*RDF/SPARQL Design Pattern for Contextual Metadata*,  
In Proc. of IEEE/WIC/ACM International Conference on Web Intelligence, November 2007, Silicon Valley, USA.

• P. Durville & F. Gandon,  
*SeWeSe : Semantic Web Server*,  

• H. Dehainsala, G. Pierra And L. Bellatreche,  
"*OntoDB: An Ontology-Based Database for Data Intensive Applications*",  
To appear in Proc. of the 12th International Conference on Database Systems for Advanced Applications (DASFAA'07), Bangkok - Thailand, April 2007

• C. Fankam, Y. Ait-Ameur And G. Pierra,  
"*Exploitation of Ontology Languages for both Persistence and Reasoning Purposes : Mapping PLIB, OWL and Flightontology models*",  
To appear in Third International Conference on Web Information Systems and Technologies(WEBIST), 2007

• Stéphane Jean, Yamine Aït-Ameur et Guy Pierra, «Une approche langage pour la gestion de données dans les systèmes de méta-modélisation » dans les actes du XXVème Congrès INFORSID(INFORSID'07), Perros-Guirec, France, 22-25 Mai 2007

Hondjack Dehainsala, Guy Pierra, Ladjel Bellatreche,


• Laura Mastella, Michel Perrin, Mara Abel, Jean-François Rainaud, Walid Touari, Knowledge Management for Shared Earth Modelling, 69th EAGE SPE Joint Conference, London June 11-14 2007, Paper SPE-107152-PP

• Laura S. Mastella, Mara Abel, Luiz F. De Ros, Michel Perrin, Jean-François Rainaud (2007), Event Ordering Reasoning Ontology applied to Petrology and Geological Modelling, IFSA 2007 World Congress Cancun, Mexico, June 18-21, 2007, Special Session 12SS, Soft Computing in Petroleum Applications

• Michel Perrin, Jean-François Rainaud, Laura Mastella, Mara Abel (2007), Knowledge related challenges for efficient data fusion, Data Fusion : Combining Geological, Geophysical and Engineering Data, SEG/AAPG/SPE Joint Workshop, Vancouver 14-18 October 2007
Diapos supplémentaires
Developed Ontologies: Geological Structure

- Geological Structures
  - Fold (Pli)
    - hasShape
      - FoldShape
        - Antiform
        - Synform
        - Concentric Fold
        - Similar Fold
      - FoldType
        - Onlap Horizon
    - hasType
      - Horizon Structure
      - Fault Structure
        - Normal Fault
        - Reverse Fault
      - Unconformable Horizon
      - Parallel Horizon
      - Strike & Slip Fault

© IFP
Developed Ontologies: Geological units and boundaries

Geological Units & Boundaries

- Geological Boundary
- Fluid Filling Boundary
- Isograde
- Tectonic Boundary
- Genetic Boundary

- Geological Object
  - is Constituted By
  - has Superior/Inferior Boundary

- Geological Unit
  - Genetic Unit
  - Sedimentary Unit
  - Intrusive Unit
  - Extrusive Unit
  - Structural Unit
  - Metamorphic Unit

- Water Oil Contact
- Gas Oil Contact
- Water Gas Contact

E-WOK_HUB Project
Developed Ontologies: Sedimentary units

Sedimentary Units

- Eonothem
- Erathem
- System
- Series
- Stage
- Chronozones
- NonStratigraphicUnit
  - MagnetostratigraphicUnit (Magnetozone)
- StratigraphicUnit
  - ChronostratigraphicUnit
  - BiostratigraphicUnit (Biozone)
  - LithostratigraphicUnit
    - hasInferiorBoundary
    - hasSuperiorBoundary

- ChronostratigraphicBoundary
- BiostratigraphicBoundary (Biohorizon)
- LithostratigraphicBoundary (Lithohorizon)
- StratigraphicBoundary (Horizon)
Developed Ontologies: Lithology

Lithology ontology (simplified)

- Lithological nature (lithology): Nature lithologique (lithologie)
  - Exogenous rocks (exogenous, exogenic) Roches Exogènes
  - Endogenous rocks (endogenous rocks) Roches Endogènes
  - Sedimentary rocks Roches sédimentaires
  - Residual rocks Roches résiduelles
  - Metamorphic rocks Roches métamorphiques
  - Igmatic rocks Roches magmatiques
  - Detrital rocks Roches détritiques
  - Biogenic and physico-chemical rocks Roches biogènes et physico-chimiques
  - Volcanic rocks Roches volcaniques
  - Plutonic rocks Roches plutoniques
  - Hydrothermal rocks Roches hydrothermales

- Rock architecture
  - Minerals Association (Paragenesis)
  - PrimaryMineralogy
  - Metamorphic Paragenesis
  - Diagenetic Paragenesis
  - SupergeneAlteration Paragenesis
  - Hydrothermal Alteration Paragenesis

- Texture

- Minerals

- Rock

- Mineral Association (Paragenesis)
Developed Ontologies: Reservoir

Reservoir

- If contains oil or gas or water, then is a
  - reservoir réservoir

GeologicalObject

- If contains water, then is either or
  - unsaturated zone (vadoze zone) zone non saturée (zone vadose)
  - saturated zone zone saturée
### ECCO: Collaborative Vocabulary Edition

<table>
<thead>
<tr>
<th>Property</th>
<th>Concept</th>
<th>Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligocene</td>
<td>Lutetian</td>
<td>Tertiary</td>
</tr>
<tr>
<td>Jurassic</td>
<td>Subsidence rate</td>
<td>Cretaceous</td>
</tr>
</tbody>
</table>

**Metadata added on each term:**

**Collaborative vocabulary edition:**

Elaboration and history are kept as metadata.
Service annotation creation process

WSDL → Semantic repository → SAWSDL → RDF

© IFP
R11 Jean François RAINAUD  Norvégian Semandic Days : may 19 th 2009
How to use service annotation?

- WSDL
- SPARQL query
- Semantic repository
- CORESE
- BPEL activities
- Orchestra
E-WOK_HUB Project

Developed Ontologies: The Earth Modeling Use case

- Models (Meta data & data files)
- Data Bases
- Reports
- Previous Information System

Knowledge Extraction
- Data Extraction
- Knowledge Extraction

Knowledge Addition
- Knowledge Completion
- Updating & Persistence

Knowledge Exploitation in Applications
- Updated Representations
- Updated Data Bases
- Updated Information System

Knowledge Extraction
- Data Base
- Knowledge Base
The Earth Modeling Use case

E-WOK_HUB Project

Ontology (OWL)

1/ Annotate

Text Files
Word/pdf

Structured data
XML/binary

2/Associate

Annotation Data Base

Knowledge Data Base

© IFP
E-WOK_HUB Project

Developed Ontologies Interpretation management

- **e1:** Etude
  - **has**
  - **attributedBy**
    - **Untel**
  - **uses**
    - **c1:** Choice
      - **uses**
        - **:SurfaceRep**
          - **file1:** XYZ File
          - **annotates**
            - **:Annotation**
              - **interpolator:** geologist1
              - **date:** 10/07/08
              - **comments:** ...
    - **uses**
      - **c2:** Choice
      - **uses**
        - **:GES**
          - **hasAtomicInterpretation**
            - **usesForAnnotation**
              - **:GeologicalObject**
                - **top_horizon1**