Two Earth Science use cases involving Semantic Web Ontologies

for 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

Two Earth Science use cases involving Semantic Web 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

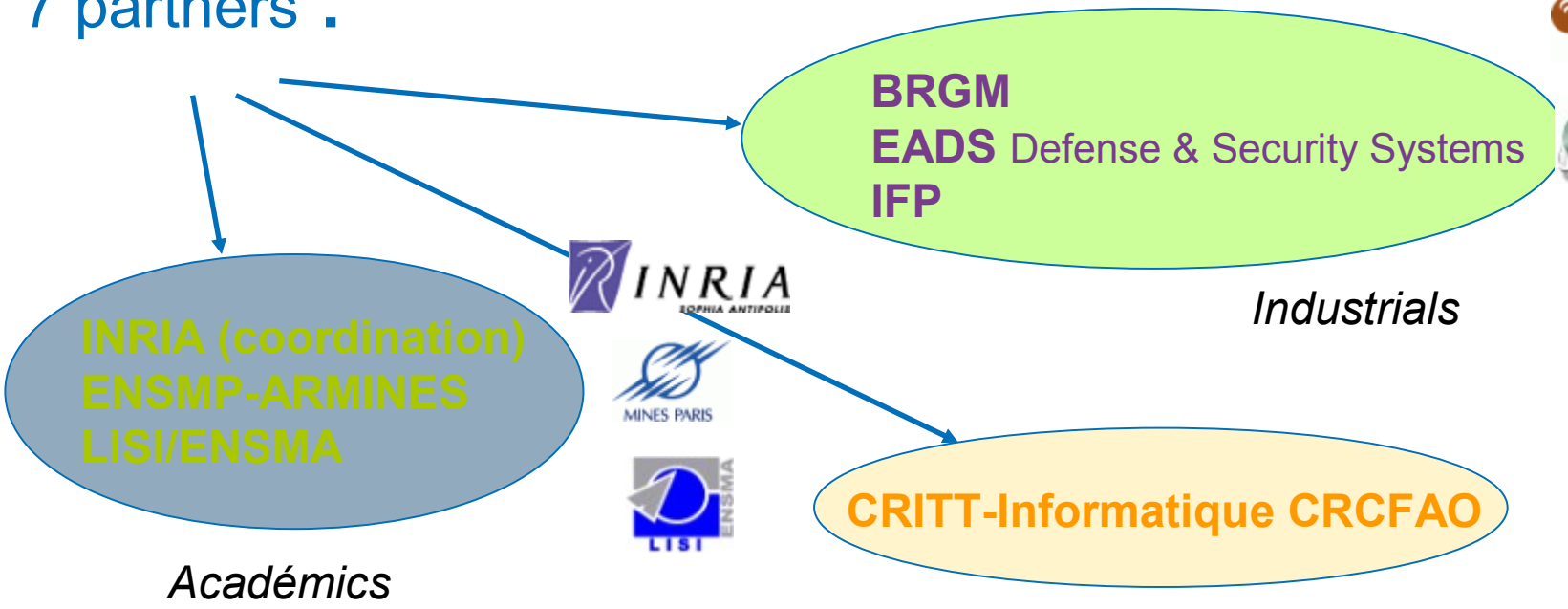
E-WOK_HUB Project

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

A French "Agence Nationale de la Recherche" project

06/2006 - 06/2009

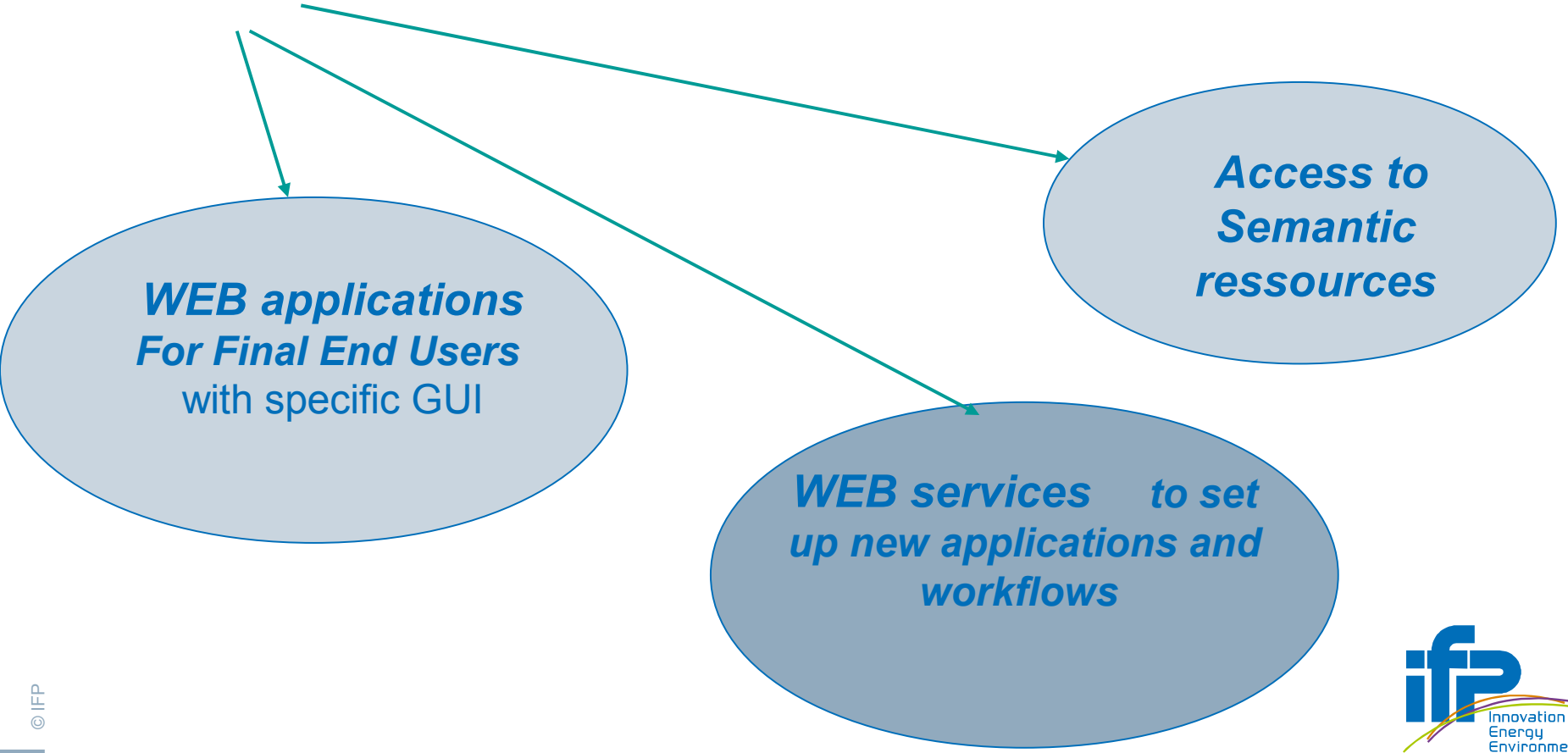
7 partners :



E-WOK_HUB Project

Introduction : Scientific Objectives

Set up interconnected Portals (the e-WOK_HUBs), providing



E-WOK_HUB Project

Introduction : Applicative Objectives BRGM/IFP

- CO_2 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 usefull 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

1/ Unstructured data

- Use Cases for C02 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

Innovation Energy Environment

E-WOK_HUB Project

The Use Cases Objectives

- Use Cases for C02 geological storage site studies

2/ Structured and Unstructured data

  - 2/ 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

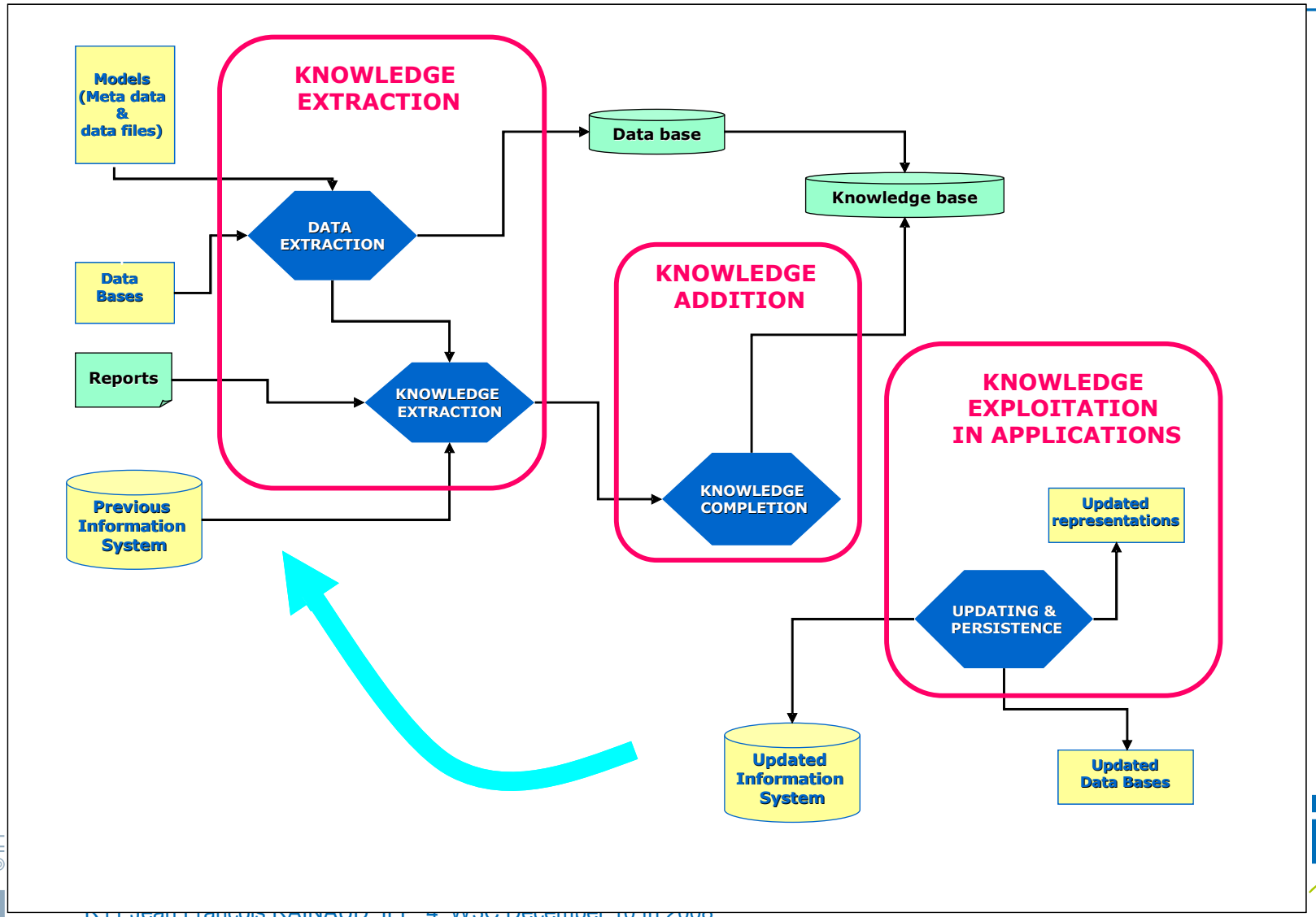
3rd Step : Offer Semantic Query facilities on this knowledge base

4th Step : Propose interpretation management along the reservoir characterization process.

Results : This must facilitate interpretation tracking and reuse of res[illegible] characterization processes.

E-WOK_HUB Project

Developped Ontologies : The Earth Modeling Use case



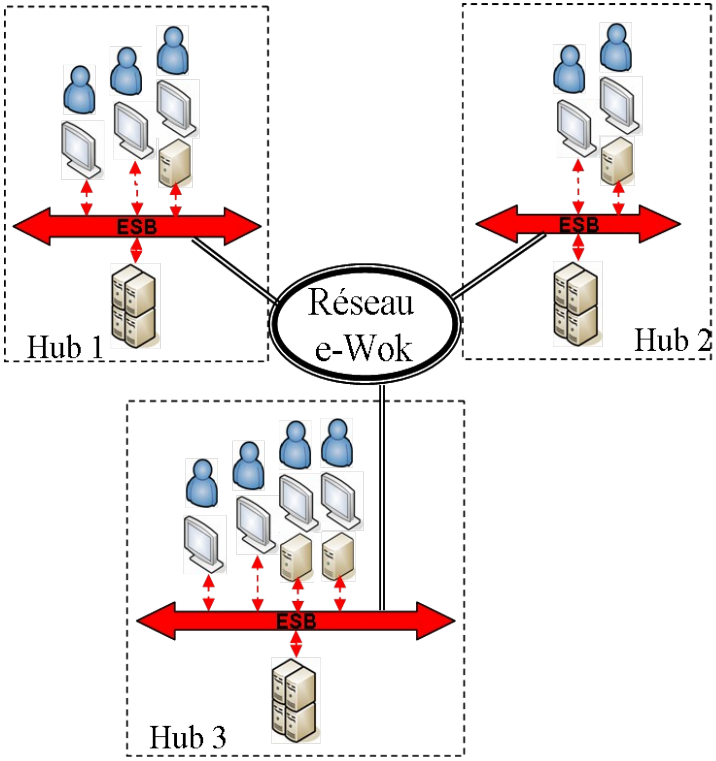
Innovation Energy Environment

Proposed Methodology

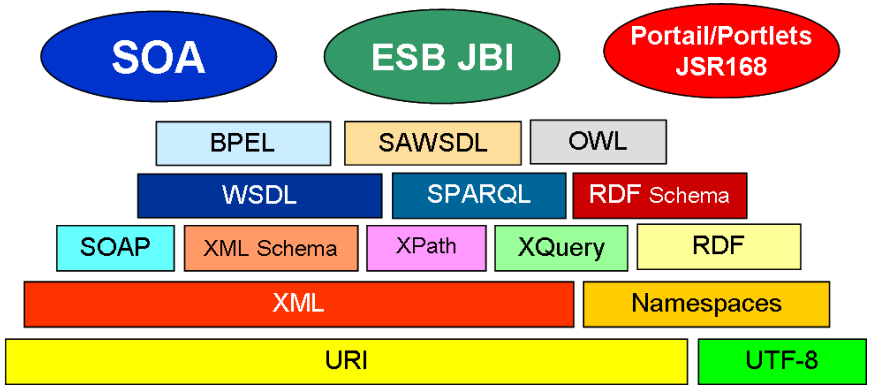
E-WOK_HUB Project

Methodology : IT Environment

Service Oriented Architecture



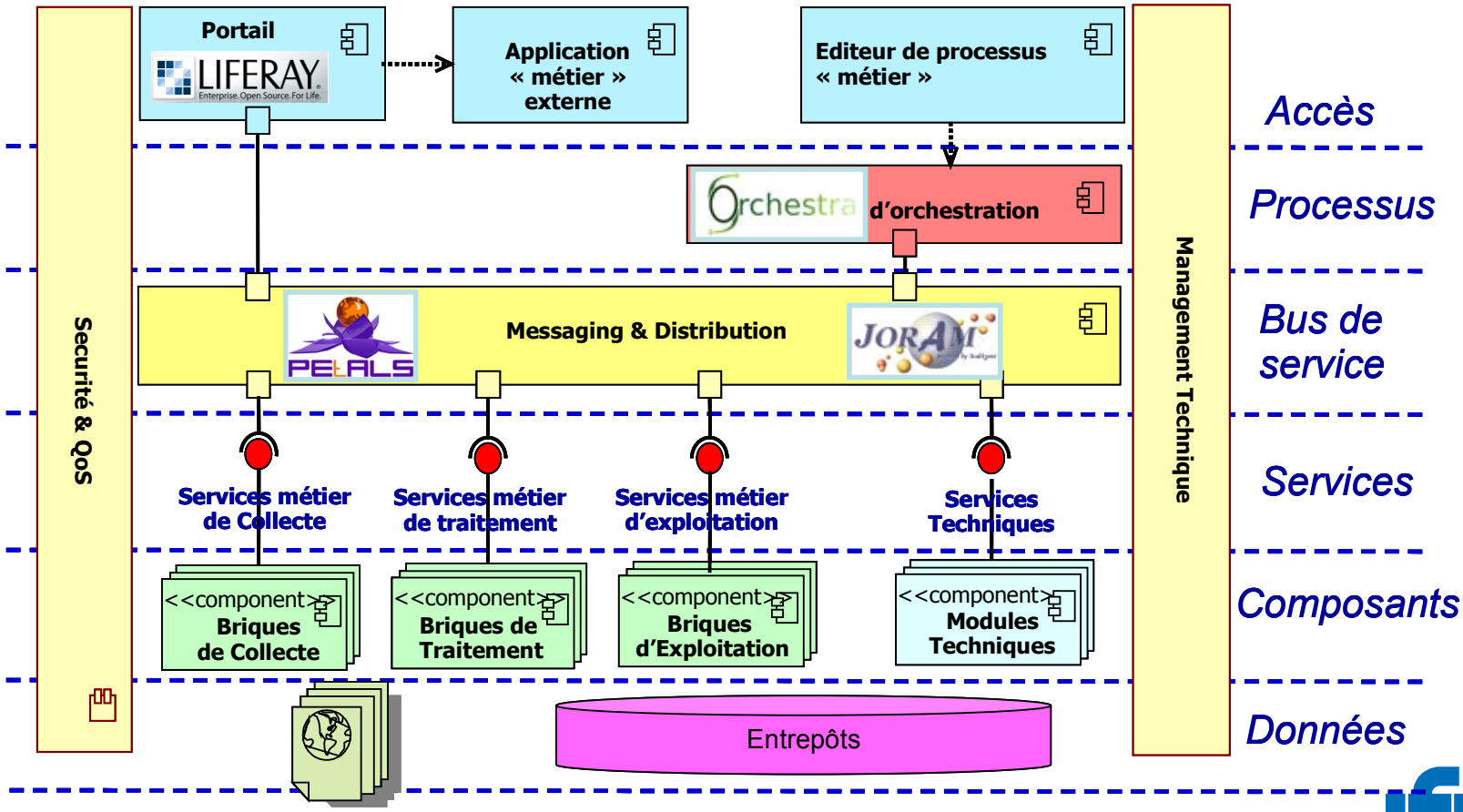
Used Standards



a hub is a warehouse of semantic business resources

E-WOK_HUB Project

Methodology : IT Environment : The WEB Lab (OpenSource provided by EADS)



E-WOK_HUB Project

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) ..*** PORTLET technology ***(Liferay)***
  - 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 (demo on unstructured data)

- 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 persistence of the instances of the Domain ontologies and the annotations
- Step 4 : Query semantically the annotations

E-WOK_HUB Project

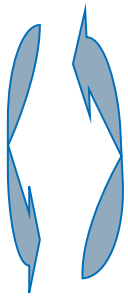
Methodology : Step 1 Define Domain Ontology

- 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)

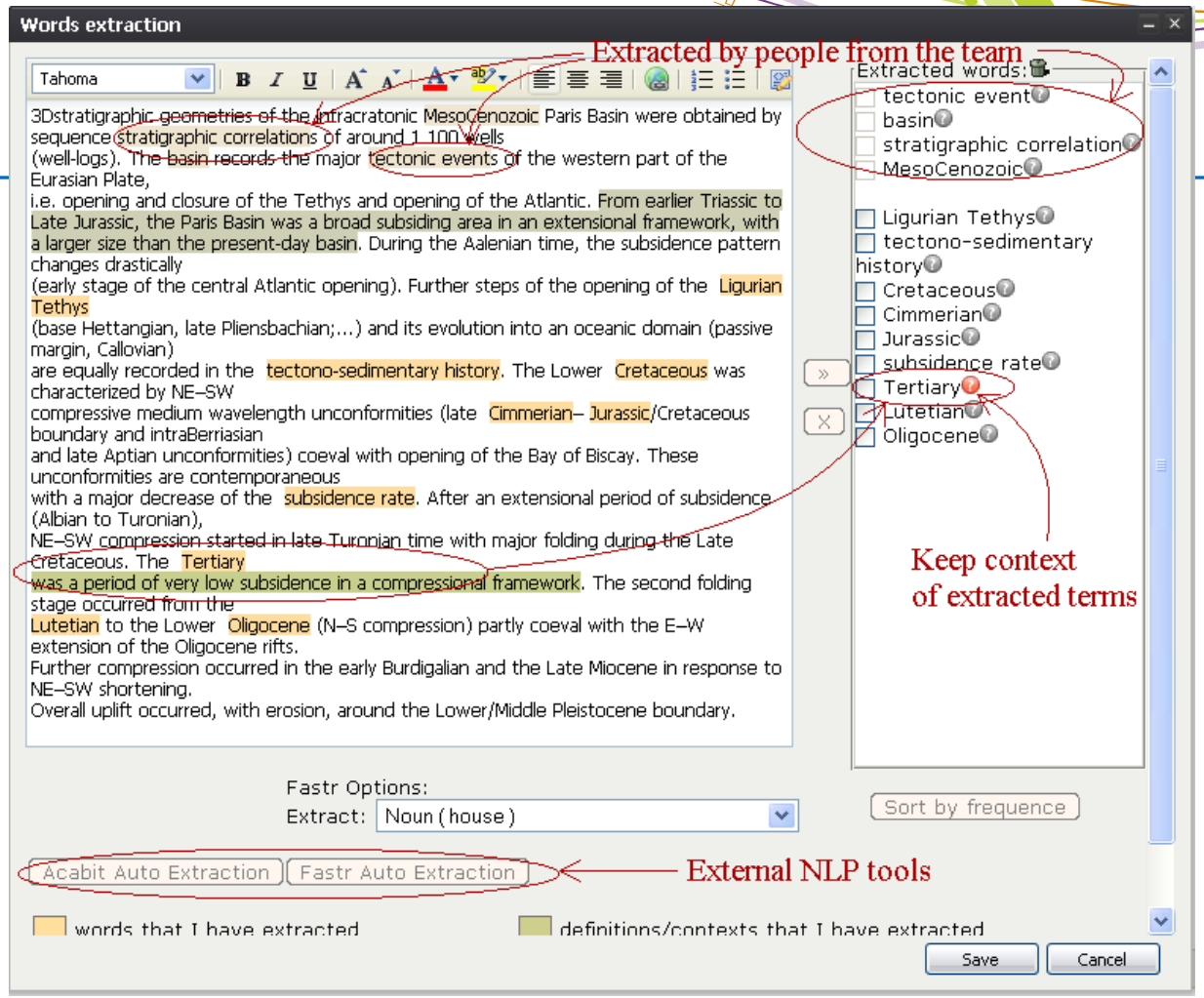
ECCO : **ontology editor**

- **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**



ECCO:

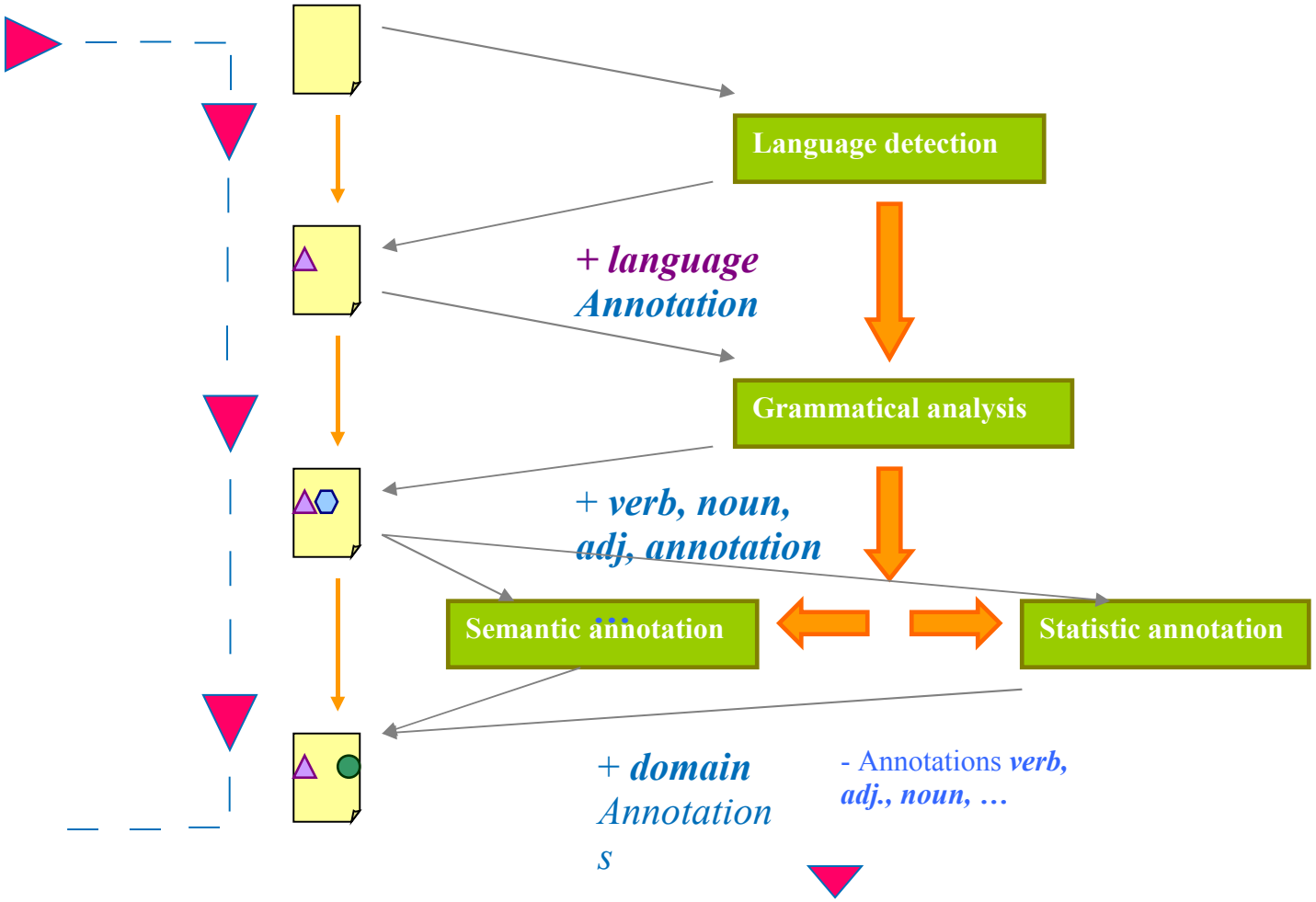
Term extraction from domain texts



Innovation Energy Environment

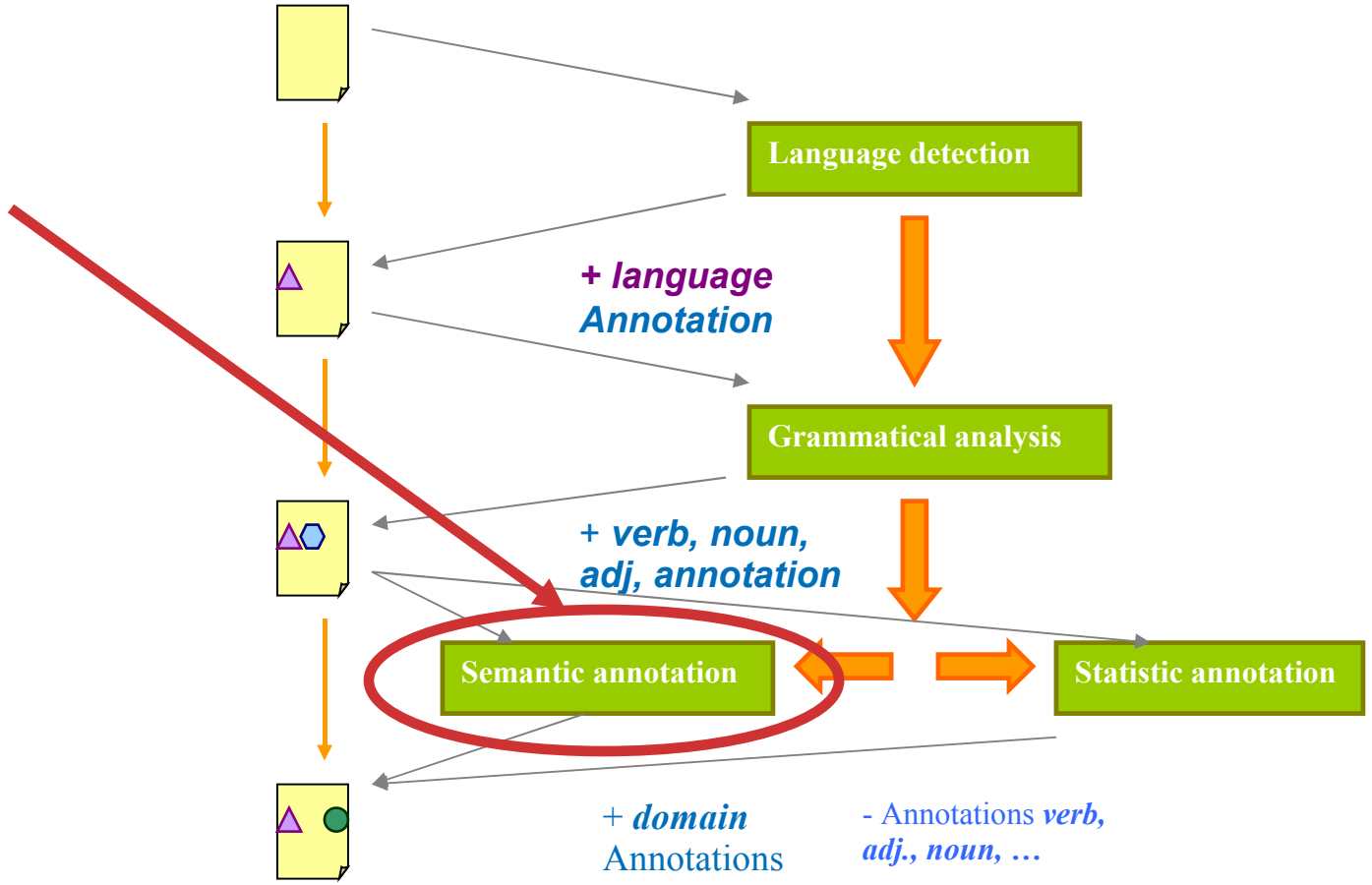
E-WOK_HUB Project

Methodology : Step 2 Annotation generation services



E-WOK_HUB Project

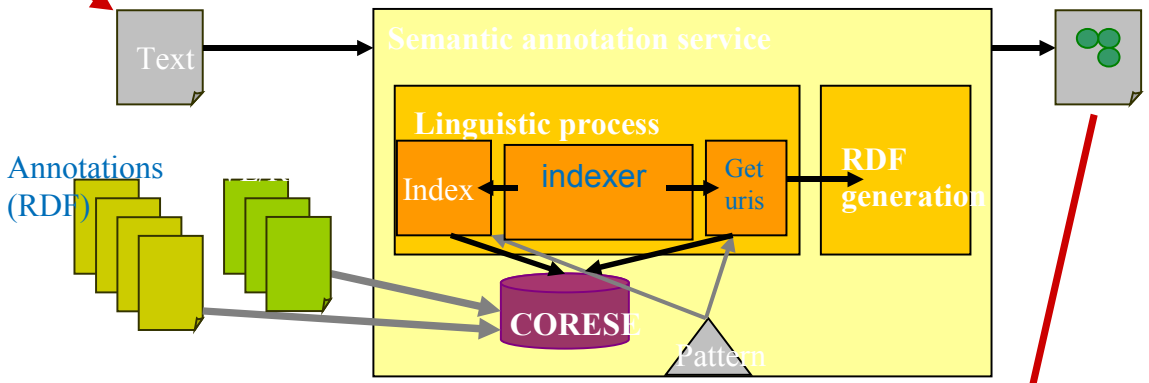
Methodology : Step 2 Annotation generation services



E-WOK_HUB Project

Methodology : Step 2 : Annotation Service (INRIA)

"3D stratigraphic geometries of the intracratonic Cenozoic Paris Basin were obtained by sequence stratigraphic correlations…"




```
<rdf:RDF xmlns:geo="http://rdf.inria.fr/geo/"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:wl="http://model.core.weblab.eads.com#"
         xmlns:cos='http://www.inria.fr/acacia/corese#'
         cos:graph="weblab://myWS/myDocument/1_inria#">
<wl:Segment rdf:about= « weblab://myWS/myDocument#seg1 »>
   <geo:datedAt rdf:resource="http://www.owl-ontologies.com/geoTime.owl#Cenozoic" />
</wl:Segment>
</rdf:RDF>
```

E-WOK_HUB Project

Methodology : Step 2 : Annotation Service : just a comment .

- An Annotation can be intrusive.

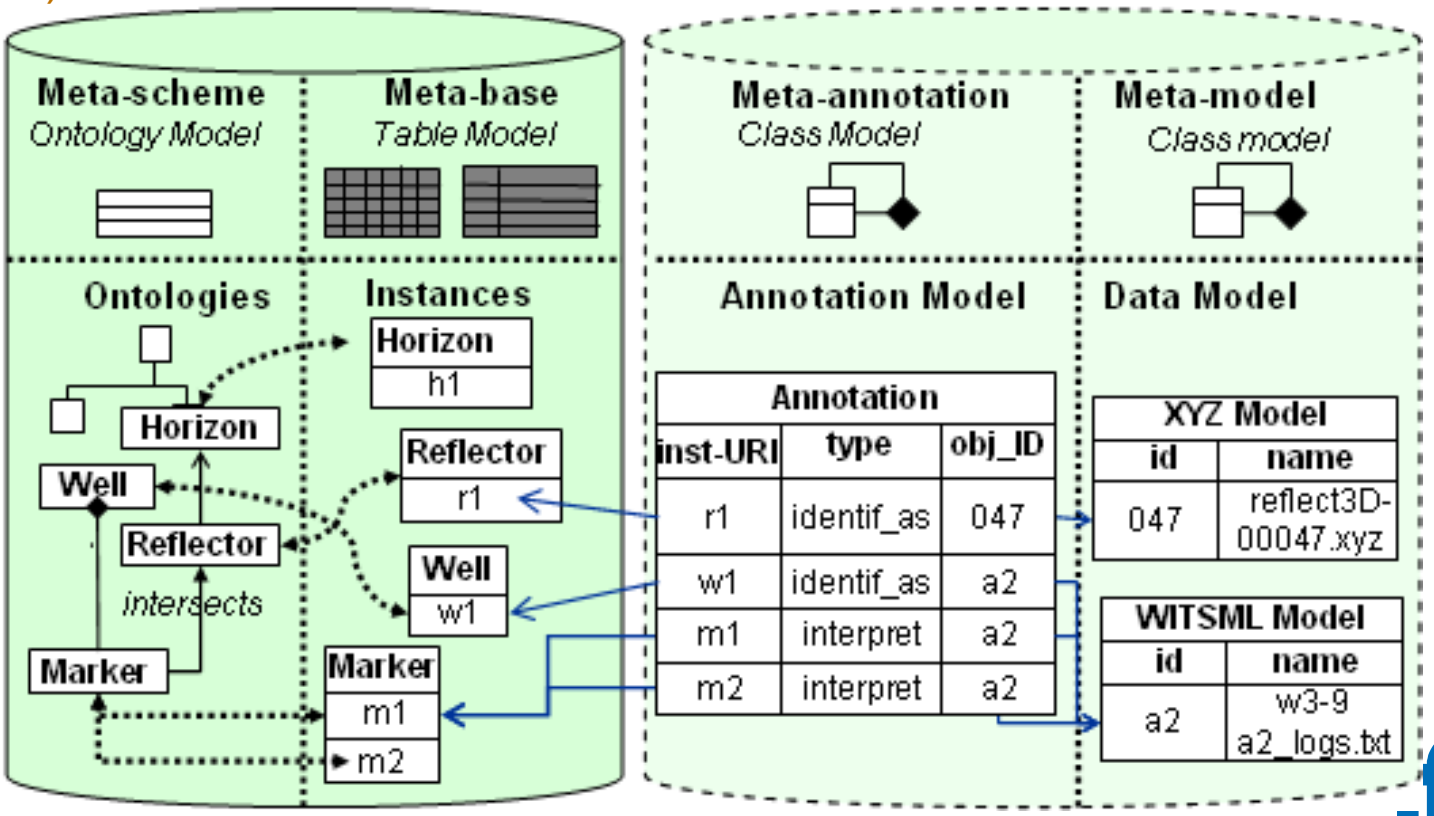
- In Documents (non structured)
  - A first link is referencing the URI of the document,
  - then a second link is established with the instance and attached to a segment of the text

- In Technical Information (structured)
  - A first link is referencing the URI of the "structured document"
  - Then a specific query is attached.

E-WOK_HUB Project

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)



Developped Ontologies

E-WOK_HUB Project

Developped Ontologies : Overview

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:

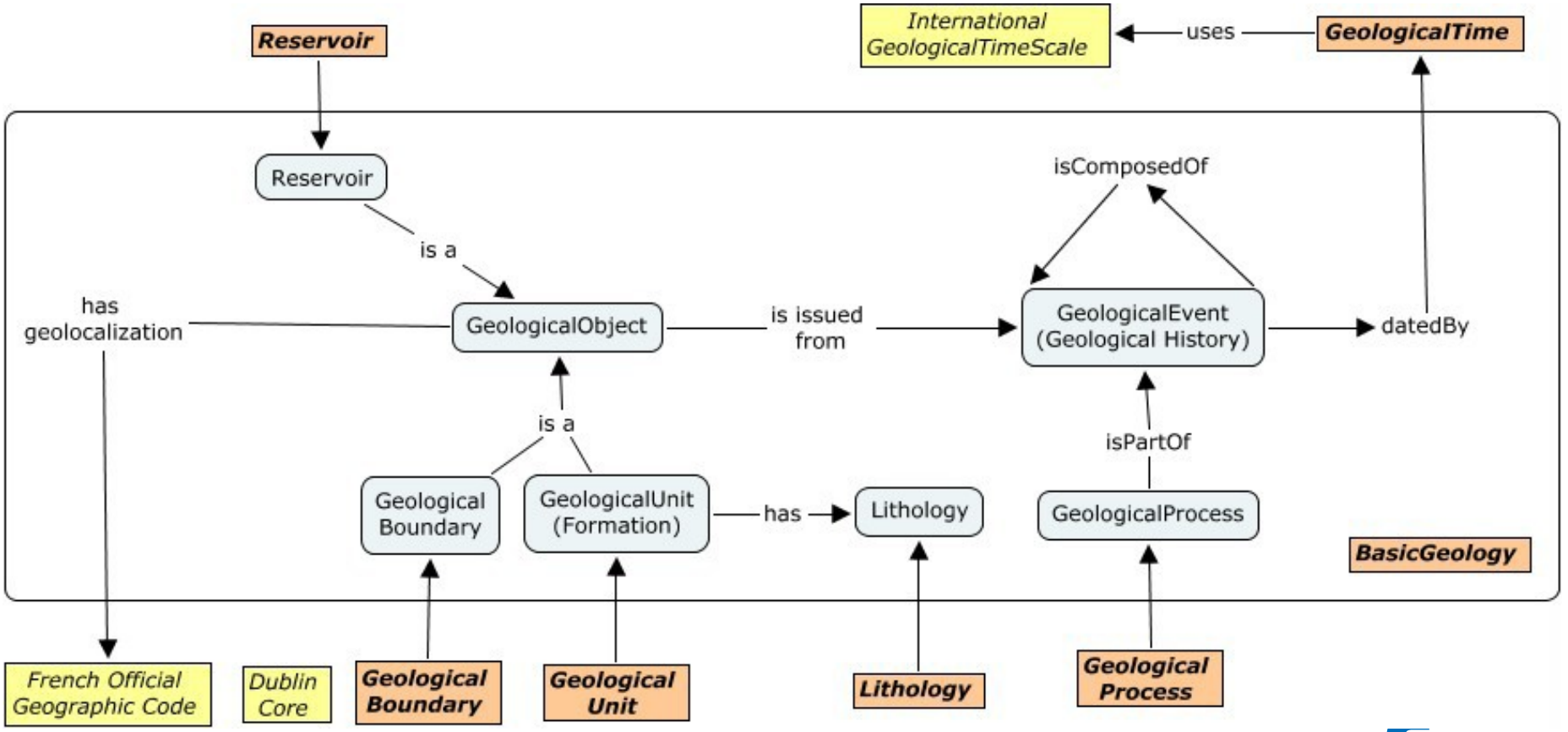
- *Basic geology (units and boundaries),*
- *Geological Structure*
- *Geolocalization*
- *Geological Age*
- *Geological processes,*
- *Lithology and mineralogy,*
- *hydrogeology and reservoirs*
- *Earth state and paleogeography concepts,*

- 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 [illegible] based on Official dictionnary and geological reference thesaurus.

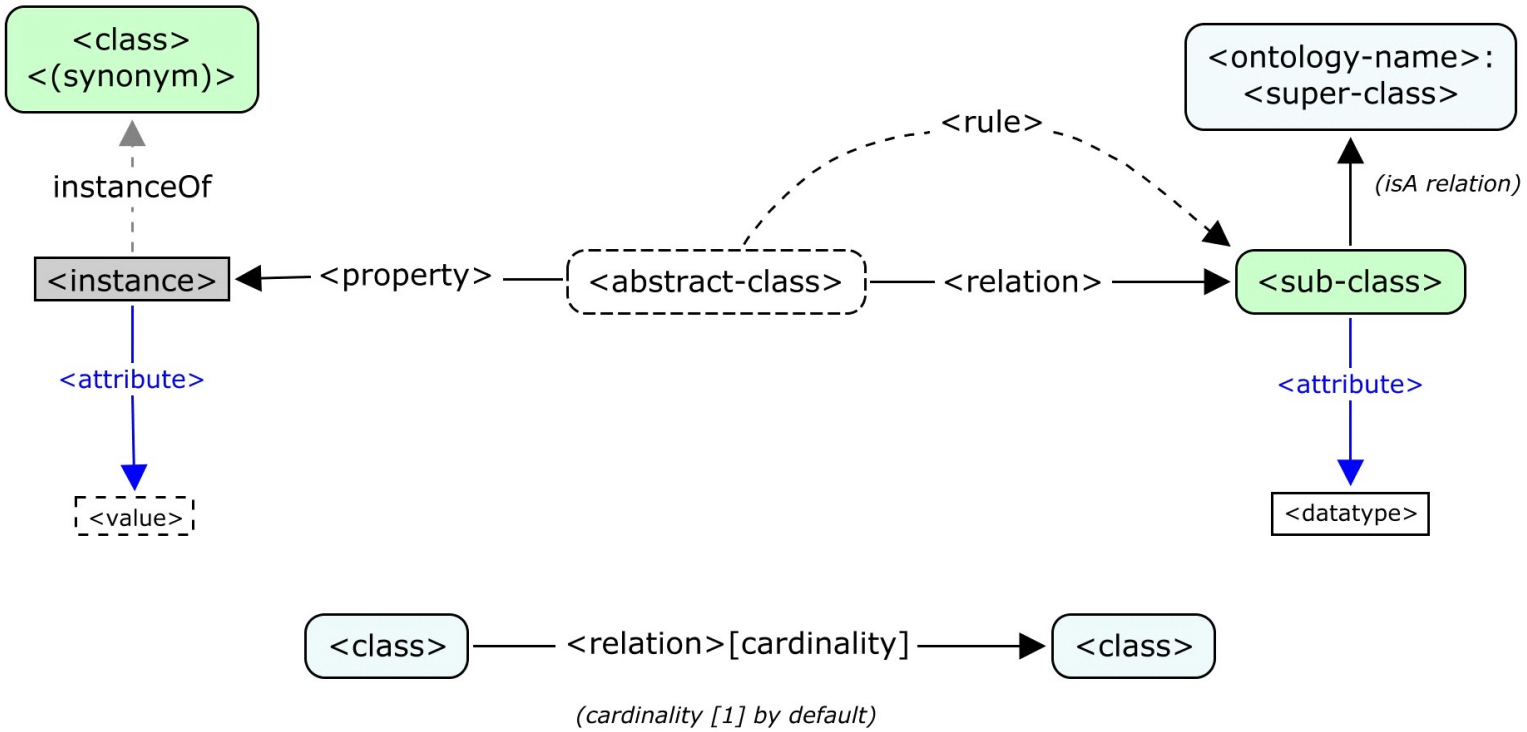
E-WOK_HUB Project

Developped Ontologies : Overview



Developped Ontologies : Conventions

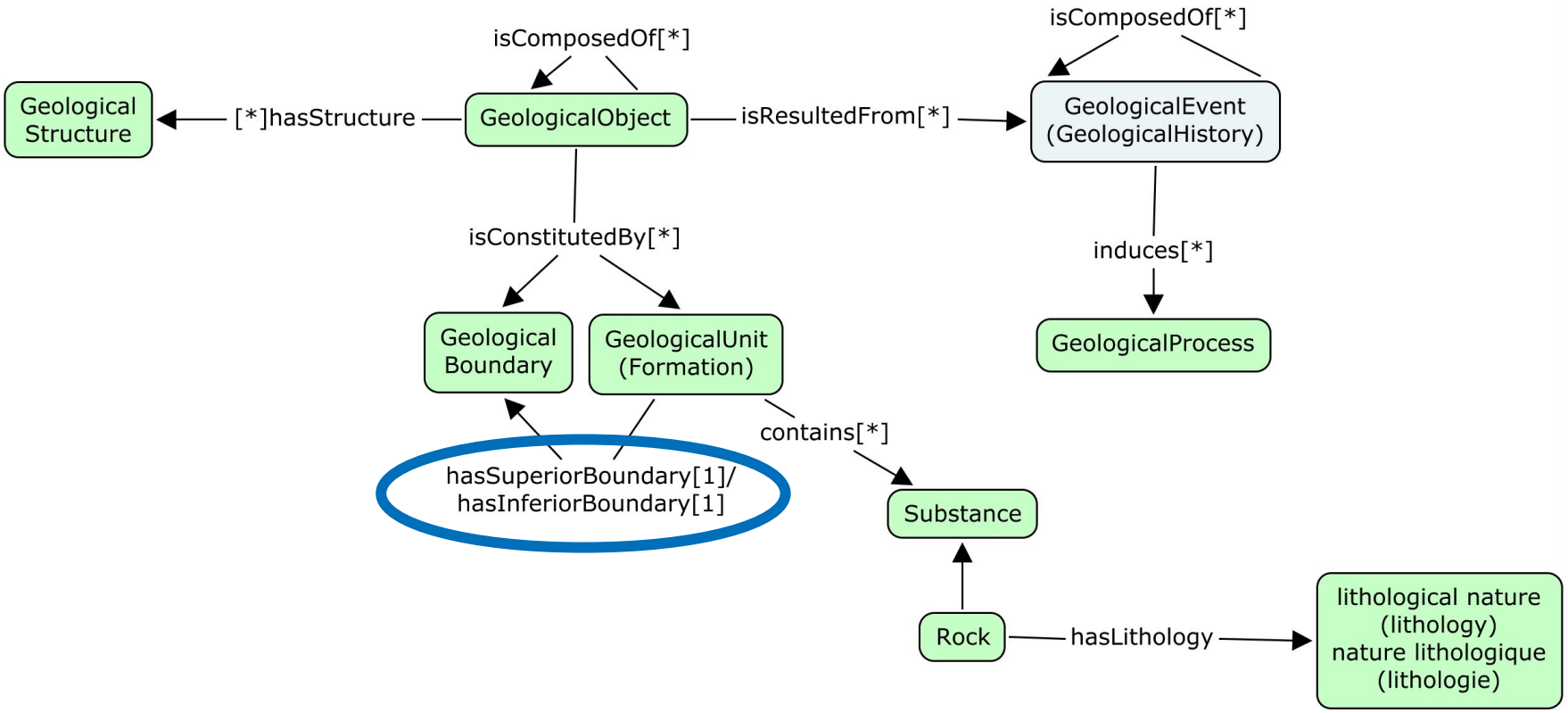
CMap/Ontology conventions



E-WOK_HUB Project

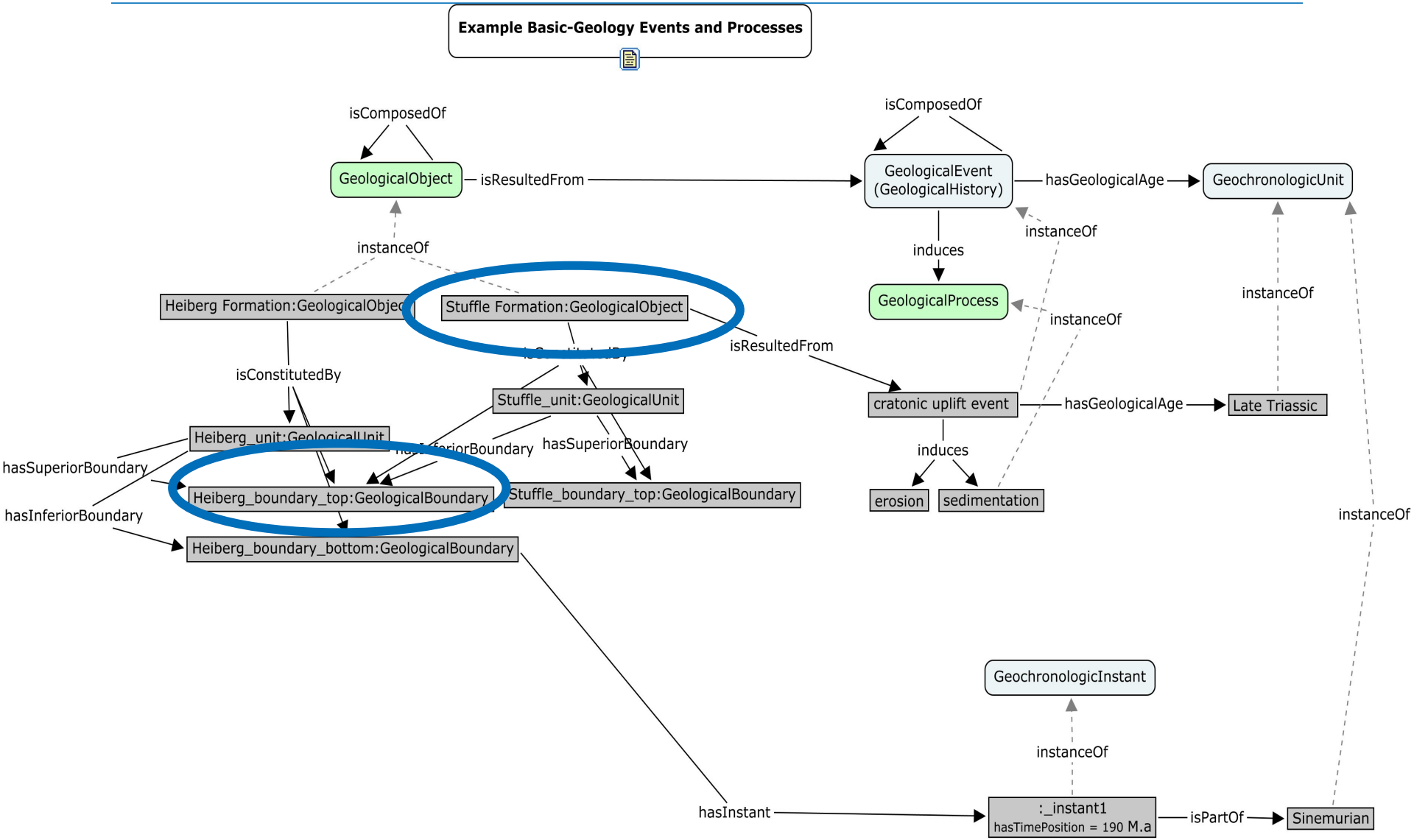
Developped Ontologies : Basic Geology

Basic Geology ontology



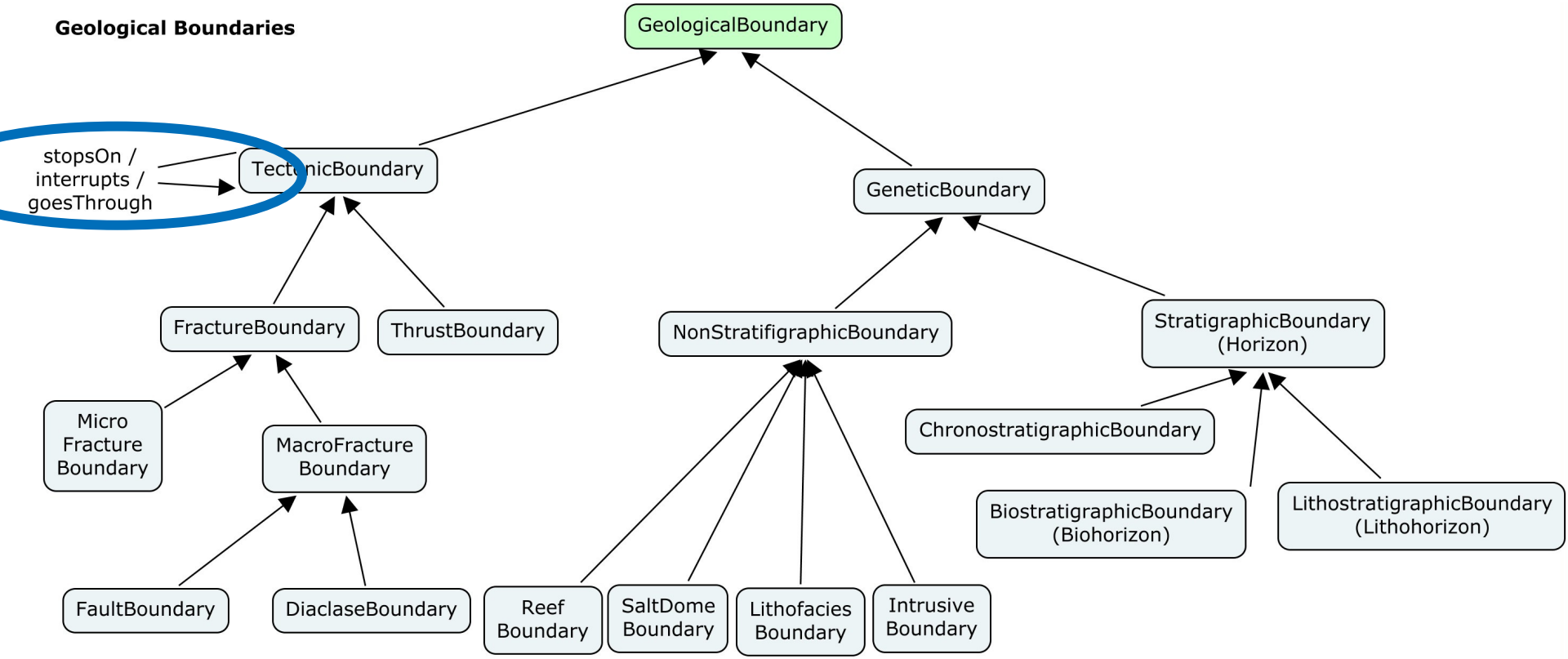
E-WOK_HUB Project

Developped Ontologies : Basic Geology with Instances



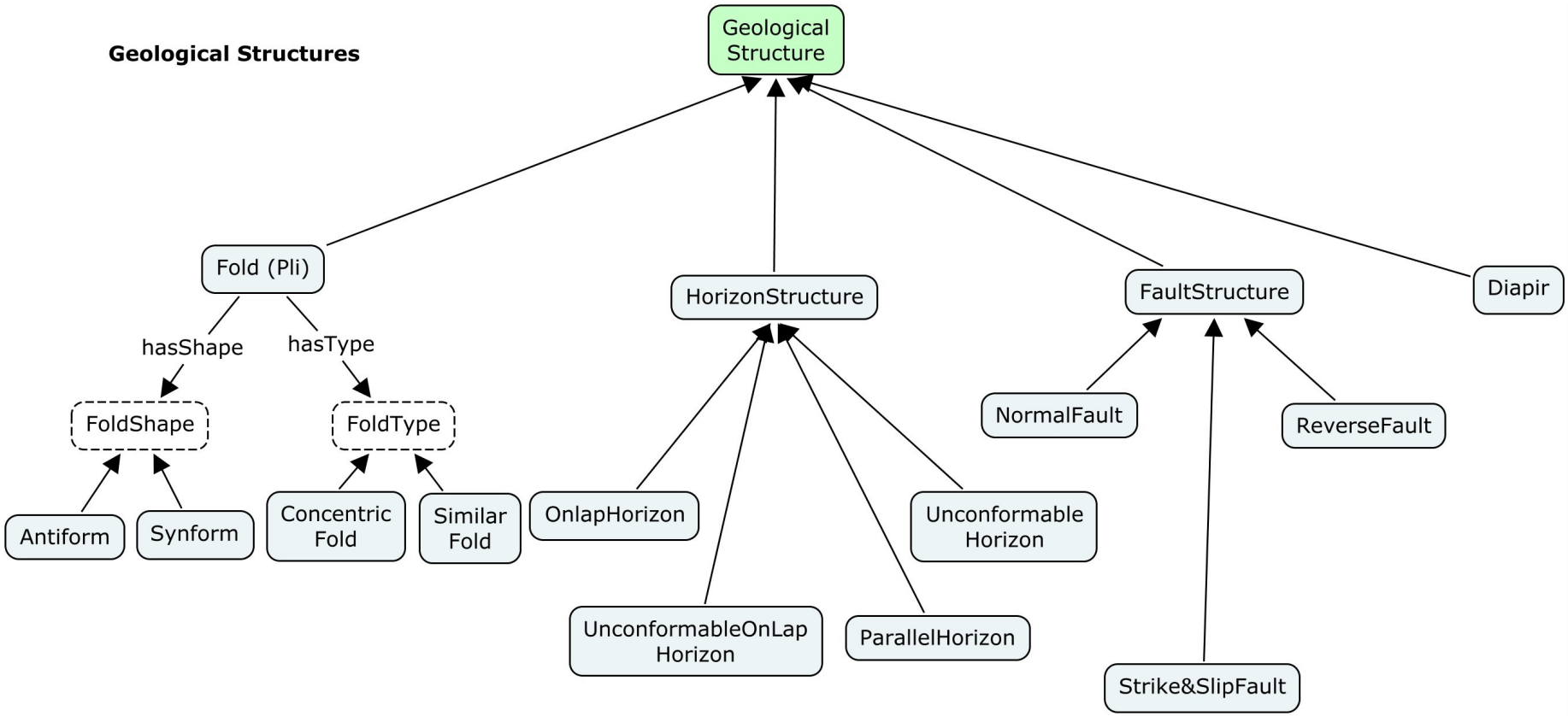
E-WOK_HUB Project

Developped Ontologies : Geological Boundaries



E-WOK_HUB Project

Developped Ontologies : Geological Structure



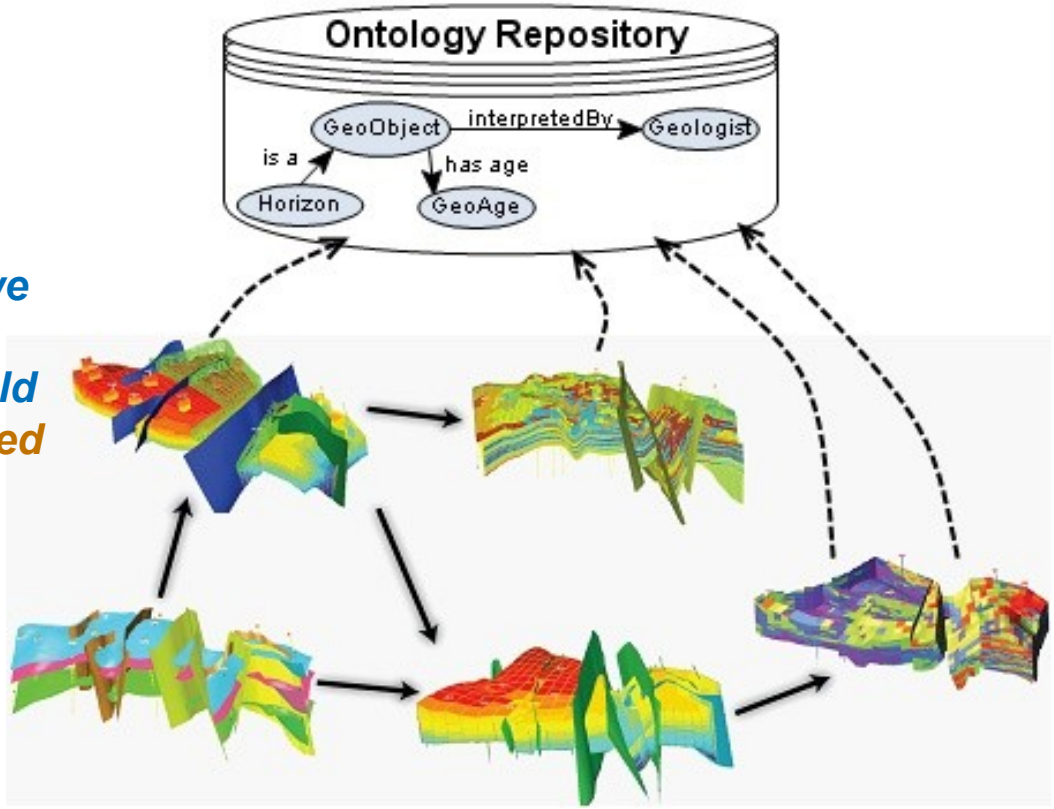
Earth Modeling interpretation tracking

E-WOK_HUB Project

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.*



to track the workflows we are also tinking taht the diverse processus should be semantically annotated too

E-WOK_HUB Project

Earth modeling interpretation tracking: The interpretation

The interpretation corresponds to a 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 partof) 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

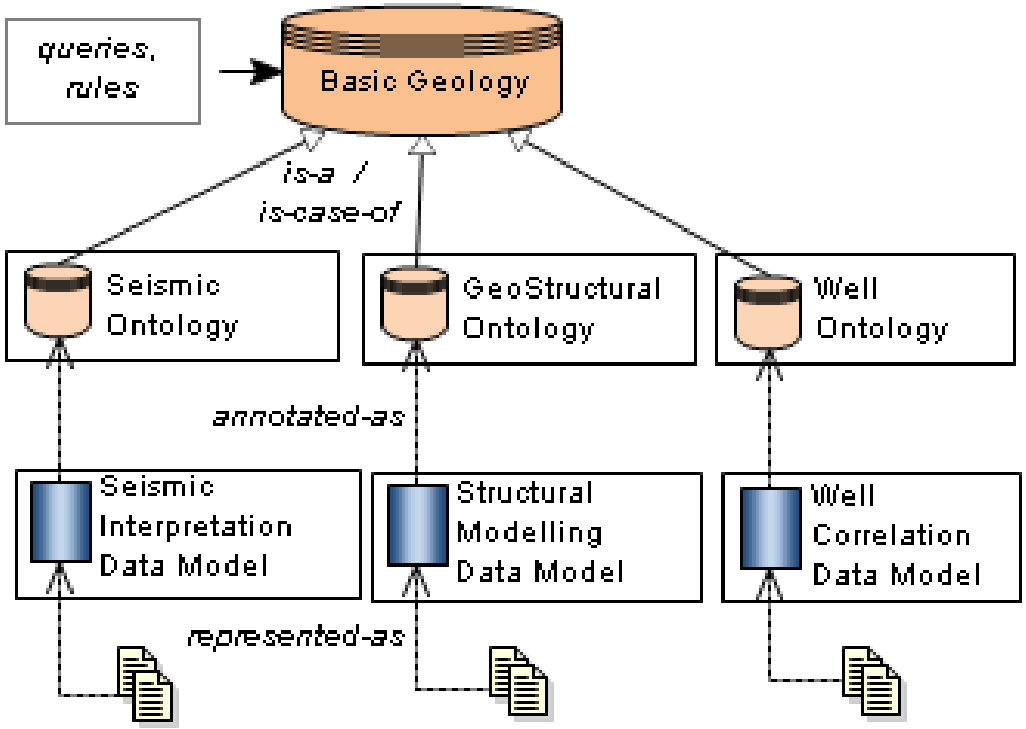
- *e.g. The Fault F3 interrupts the Fault F4*

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

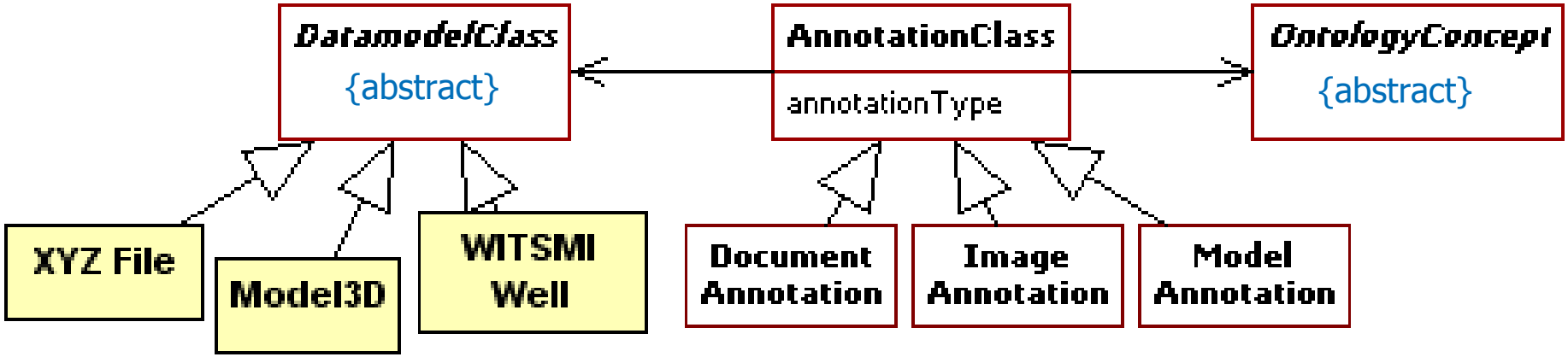
- Formalize the data-model of geological models;
- Annotate instances using local ontologies (LO)
- Articulate LOs using a global ontology (GO)
- .



E-WOK_HUB Project

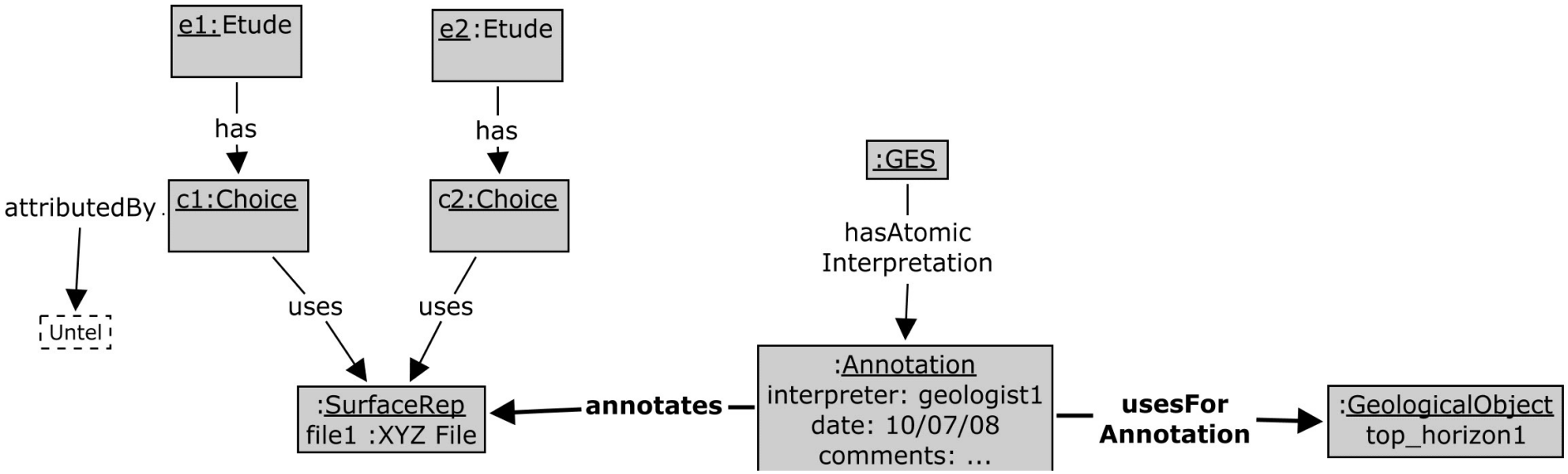
Earth modeling interpretation tracking: The annotation model

- 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.



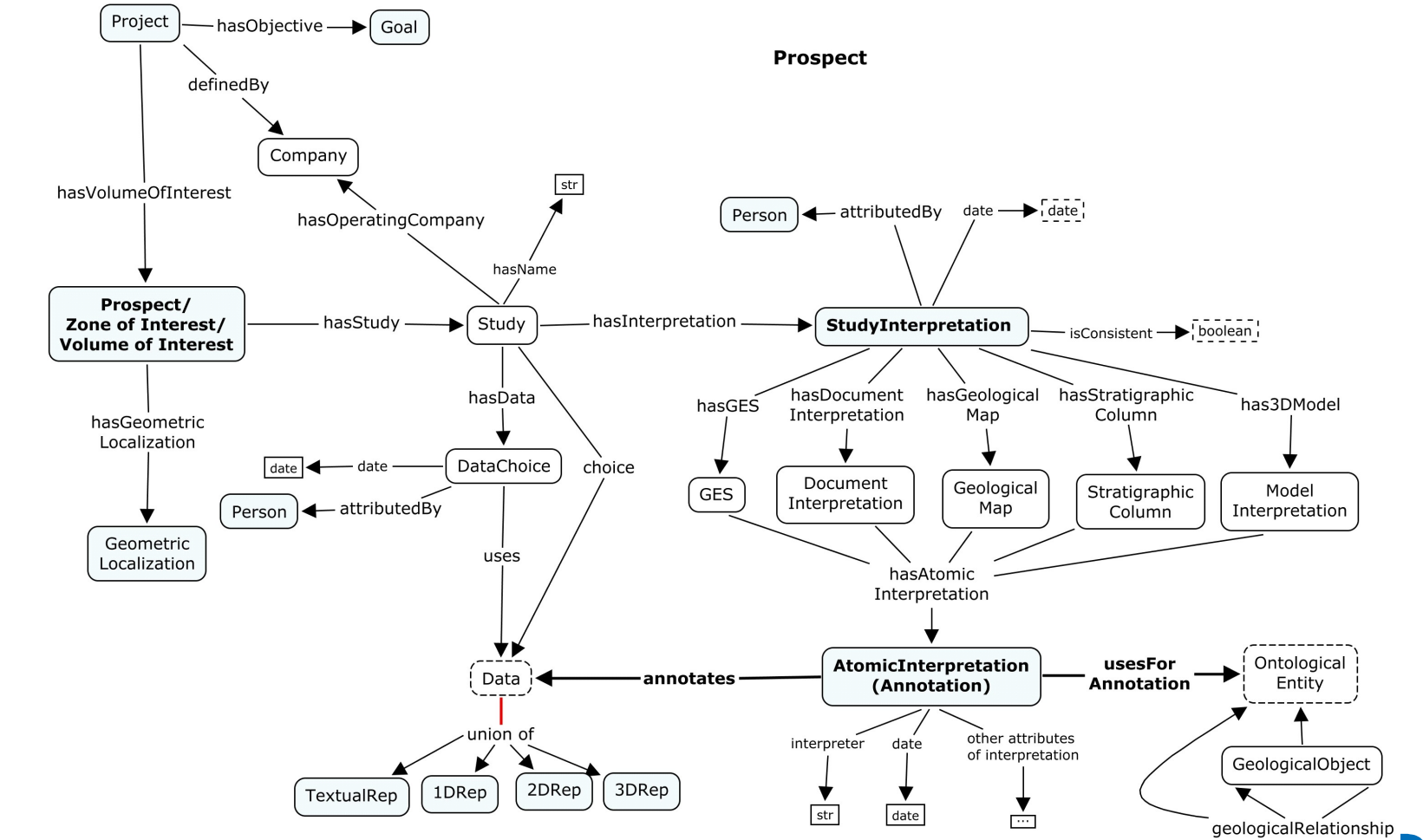
E-WOK_HUB Project

Developped Ontologies Interpretation management



E-WOK_HUB Project

Developped Ontologies Interpretation management



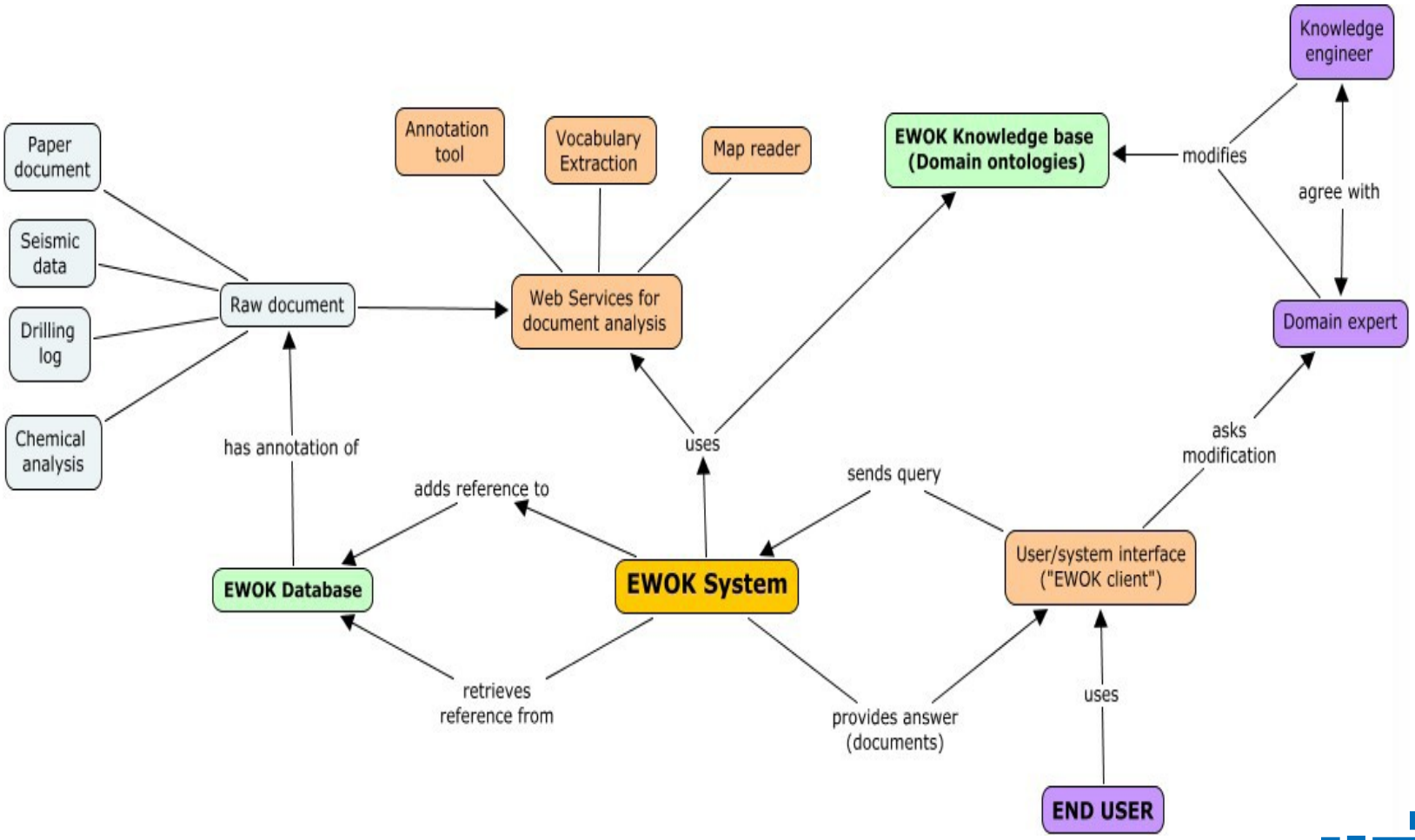
Document searching application

Example of a query

with SparQL using CORESE engine

E-WOK_HUB Project

Use of Developped Ontologies : 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*

Innovation Energy Environment

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 sytem 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 ?

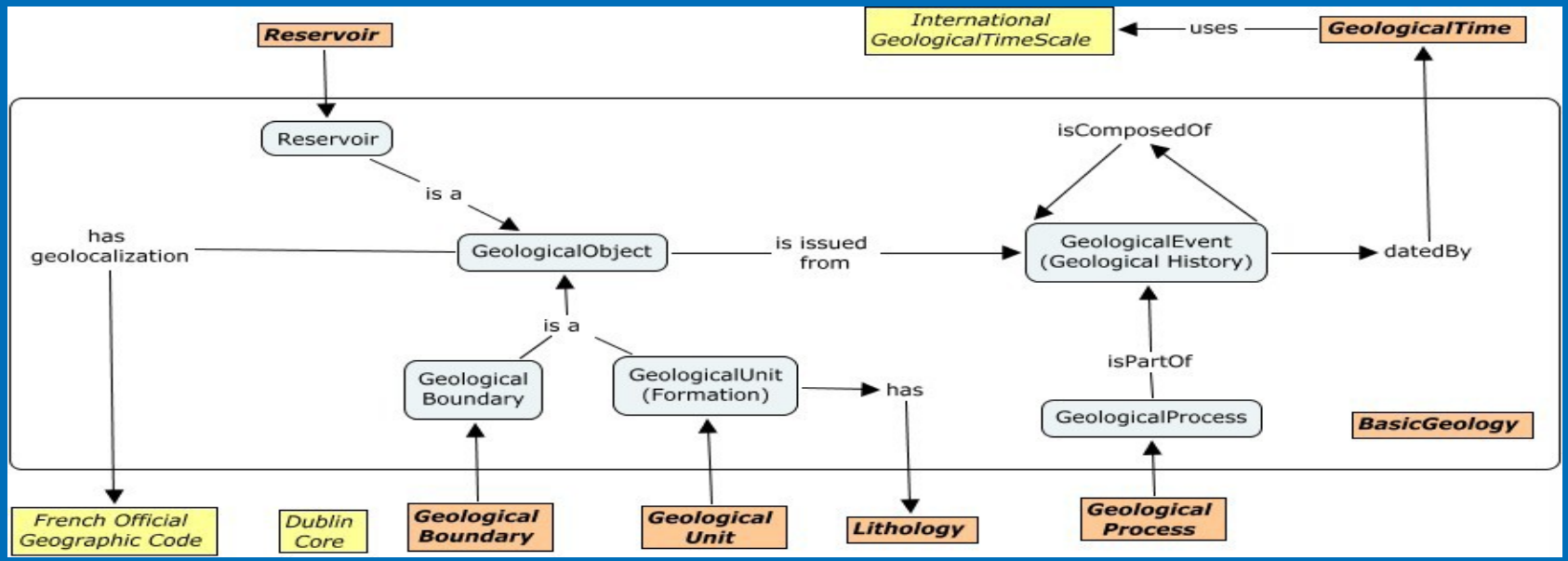
- Another possible option is to extract from the question rather than the word *Bathonian, the expression Bathonian formation*.

Innovation Energy Environment

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 .

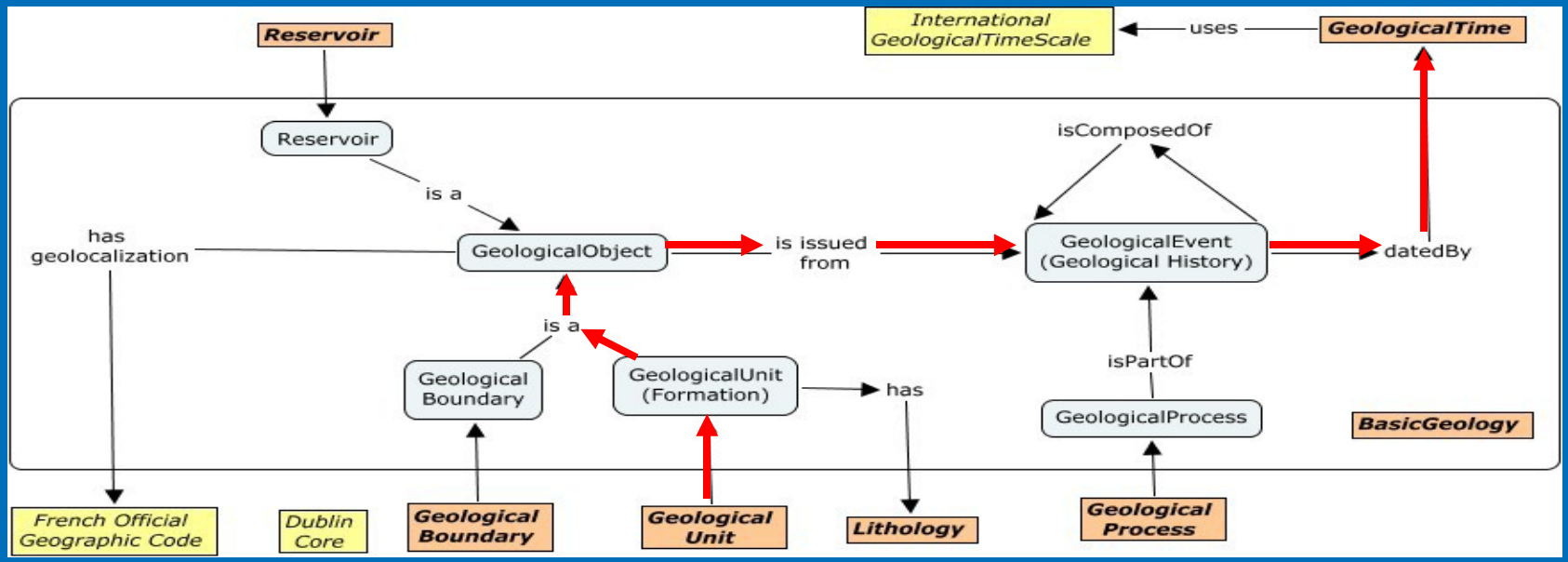


Innovation Energy Environment

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 .



Innovation Energy Environment

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 synomyms *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 synomyms *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 ?

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

$$S = S1 \cap (S_{2a} \text{ or } S_{2b}) \cap S3$$

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**.

E-WOK_HUB Project

Conclusion (2/2)

- We are entering 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 tests.

Innovation Energy Environment

Last Publications

e-WOK_HUB Consortium, *Semantic Hubs for Geological Projects*, ESWC'2008 Workshop on Semantic Metadata Management and Applications (SeMMA'2008), June 2, 2008, Teneriffe, Spain.

Perrin, M, Durville,P, Grataloup, S, Mastella, L.,Lions, S, Morel, O, and Rainaud, J.-F. *Knowledge issues for automatic identification of CO2 storage sites by means of Semantic Web technology.* EAGE Workshop on CO2 Sequestration, Budapest Hungary, 28-29 September 2008.

Alain Giboin, Priscille Durville, Fabien Gandon, *Ingénierie ontologique participative : essai de mise en oeuvre avec l'éditeur collaboratif ECCO,* Atelier IC 2.0, joint aux IC2008, 19èmes Journées Francophones d'Ingénierie des Connaissances, Nancy

Mastella, L., Ait-Ameur, Y., Perrin, M, Rainaud, J.-F and Abel M., *Formalizing Geological Knowledge through Ontologies and Semantic Annotation,* 70 th EAGE (European Association of Geoscientists and Engineers) conference and exhibition incorporating SPE-EUROPEC Expanded abstract. Rome, Italy 9-12 June 2008.

Mastella, L., Ait-Ameur, Y., Perrin, M. and Rainaud, J.-F., *Ontology-Based Model Annotation of Heterogeneous Geological Representations.* 4th Conference on Web Information Systems and Technologies (WEBIST), Madeira, Portugal, 4-7 May 2008.

http://www.inria.fr/sophia/edelweiss/projects/ewok/publications/publis.html

Thank you !

Questions ?

More publications

Publications (1)

• 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,
WWW'2007, Developers track, Banff, 2007.

• 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

Publications (2)

• S. Jean, Y. Aït-Ameur And G. Pierra,
"Querying Ontology Based Databases Using OntoQL (an Ontology Query Language)",
Proc. Ontologies, DataBases,
and Applications of Semantics (ODBASE'2006), in Lecture Notes in Computer Science vol. 4275, Springer 2006, pp. 704-721

•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

• Nabil Belaid, Ladjel Bellatreche, Yamine Ait Ameur and Guy Pierra, « Intégration de sources à base ontologique : architecture en réseau VS architecture en étoile » Plate-Forme AFIA: Atelier Thématique GDR I3 sur Ontologies et Gestion de l'hétérogénéité sémantique (OGHS), pp. 9-20, Juillet, 2007

Publications (3)

•Ladjel Bellatreche and Guy Pierra, « OntoAPI: An Ontology-based Data Integration Approach by an a Priori Articulation of Ontologies » 4th International Workshop on P2P Data Management, Security and Trust (PDMST'07), pp. 799-803, IEEE Computer Society Press, September, 2007
Hondjack Dehainsala, Guy Pierra, Ladjel Bellatreche,

•Yamine Aït Ameur, « Conception de bases de données à partir d'ontologies de domaine. Application aux bases de données du domaine technique » A apparaître dans les actes des 1ères journées francophones sur les ontologies, Octobre 2007, Sousse-Tunisie.

•Stéphane Jean, Yamine Ait Ameur and Guy Pierra, « An Object-Oriented Based Algebra for Ontologies and their Instances », To appear in LNCS proceedings of Advances in Dabase and Information systems, ADBIS 2007, Sept 30th – Oct 3rd 2007, Varna Bulgaria

Publications (4)

•P.-H. Luong, R. Dieng-Kuntz, A. Boucher,
Évolution de l'ontologie et gestion des annotations sémantiques inconsistantes.
Actes des 7èmes journées d'Extraction et de Gestion de Connaissances, EGC'07, Namur, Belgium, January 23-26, 2007.

• P.-H. Luong & R. Dieng-Kuntz.
A Rule-based Approach for Semantic Annotation Evolution.
The Computational Intelligence Journal, 23(3):320-338. Blackwell Publishing, Malden, MA 02148, USA.

Publications (5)

• 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