Two use cases involving Semantic Web Earth Science Ontologies for reservoir modeling and characterization

Jean-François Rainaud[1], Laura Mastella[2], Priscille Durville[3], Yamine Aït Ameur[4], Michel Perrin[2], Sandrine Grataloup[5], Olivier Morel[5]

[1]Institut Français du Pétrole (IFP), [2]Ecole des Mines de Paris (ENSMP), [3]INRIA Sophia Antipolis-Méditerranée, [4]LISI-ENSMA, [5] BRGM Orléans

contact author: J-Francois.Rainaud@ifp.fr

1. INTRODUCTION

At present, oil companies acquire data by measurement surveys during field operation. In order to reduce exploration and production risks, they use various software tools for ensuring the quality of the collected information, carrying out human interpretations and creating new data and new knowledge. However, the additional knowledge corresponding to data processing and interpretation is generally embedded in technical data stores and files or remains hidden in reports or in experts' brains. Moreover, while there exist rough communication standards for exchanging measured data representation, none is available for exchanging and sharing interpretations. The consequence is that interpretation remains in most cases inaccessible for queries or updating and that it is never used in core business knowledge communication. Thus it generally appears impossible to characterize the content of a given interpretation and to determine when, how, why and by whom it was created.

This lack of recorded information on metadata, data and knowledge is becoming critical for companies at the corporate level. Moreover, oil companies now frequently get associated for sharing risks on a given prospect. In this case, the lack of recorded information over interpretations also becomes an obstacle for reporting from one company to the other, the more so as each company, each country has its own description methods and language. Thus the need presently appears for companies to agree on a common way of representing geological entities as well as geological interpretations.

From 2001 on, Institut Français du Pétrole (IFP) and Ecole des Mines de Paris (ENSMP) have developed a new *knowledge-driven* paradigm for reservoir studies based on the belief that geo-model building should not be directly dependant from data (*data-driven*) but rather from geoscientists' interpretations (*knowledge-driven*) [1]. From 2006, this same paradigm is being applied for CO_2 storage studies within a joint research project (e-Wok Hub project[1]) associating professionals and researchers from various French institutions (INRIA, BRGM, IFP, ENSMP, ENSMA, EADS). This multi-disciplinary research group is presently studying solutions for extracting and managing interpretation obtained from documents (reports, presentations, technical files) related to practical users' questions concerning potential CO_2 storage site selection and assessment. For this, we intend to apply solutions using technologies recommended for the Semantic Web.

In the present paper, we will introduce two use cases considered within the e-Wok Hub project, which respectively concern documentary search and subsurface modeling. We will then describe a knowledge-driven methodology based on *semantic annotation* that can be used in both cases, explicit the ontology based solutions that we presently study for operating this methodology, and conclude.

2. TWO USE CASES FOR CO_2 GEOLOGICAL STORAGE SITE STUDIES

We describe here the two use cases studied by the e-Wok Hub consortium.

2.1 Document search to initiate the CO_2 storage prospect

For potential CO_2 storage sites identification, geoscientists consider various data resources, which they eventually interpret using their own expertise. These resources are both numerous and heterogeneous since they do not only consist in research papers and reports but also on various technical data issued from seismic or well drilling surveys or from laboratory studies. In order to be able to manage this

[1] Public website of e-Wok Hub project: http://www.inria.fr/sophia/edelweiss/projects/ewok/

amount of heterogeneous resources, we insert them in a global architecture where semantic plays a central role. This is the solution the e-Wok Hub project is working on. To achieve it, we suggest to develop a set of communicating portals (hubs), offering both: (a) web applications accessible to end-users through online interfaces, and (b) web services accessible to applications through programmatic interfaces [2]. In section 3, we give an overview of the suggested methodology for this scenario. The domain ontology is described in section 4.1.

2.2 Earth modeling for geological site qualification as CO_2 storage

3D and 4D earth models are key tools for geological site identification as CO2 storage or petroleum reservoir. Their setting requires the collaboration of professionals from various fields of earth sciences (geophysicists, geologists, petrologists, reservoir engineers). Each of them is likely to use one or several software applications allowing him/her to provide interpretations of the data related to his/her skill (seismic data, well logs, thin section observations, rock sample analyses). Each of these software tools uses a specific vocabulary and specific formats. This generates an important heterogeneity of representations of the geological objects that constitute the reservoir.

Today, the interpretations carried out during the modeling process are generally not explicit for computer dialog. In the case of seismic data processing, the interpretation is commonly commented by geophysicists by means of notes hand-written over printed cross-sections. However, in an automatic system, such commentaries expressed in natural language are information that cannot be computer-processed, since they are ambiguous and not formalized.

To deal with these interoperability issues at the interpretation level, we propose an approach based on *semantic annotation* of earth models [3]. We use a set of ontologies of geosciences domains to annotate the data with meaningful concepts, and we store data, ontologies and annotations in a same semantic repository. In section 4.2, we give an overview of the domain ontologies for this scenario.

3. PROPOSED METHODOLOGY

In both use cases described above, the main issue is managing the interpretations produced or used by geologists while performing their tasks. We will briefly describe here the methodology that we suggest for achieving this goal and the architecture used for implementing it.

Considering that we have a data repository storing both raw data (data from seismic reflection, well logging or petrophysical tests) and interpreted data (research papers, public reports, GRDECL, RESCUE files), we suggest to use an additional semantic repository (or knowledge base) containing metadata related to the documents recorded into the data repository. In order to be able to populate this additional repository, we need to create at first OWL [4] ontologies for representing the earth science domain and, more precisely, the various fields of earth sciences relevant for reservoir studies. For that purpose, we work on domain texts like project reports or research papers in order to extract significant terms that will be arranged in a hierarchical vocabulary. This term extraction step can be semi-automatically operated using natural language processing tools (like ACABIT[2] or FASTR[3]) but has to be further validated by domain experts. We have developed a collaborative and contextual ontology editor, named ECCO [2], that covers the whole life cycle of such a creation process. The developed ontologies are then stored in the semantic repository (like OntoDB [5], an ontology-based database) or loaded in an inference engine (like CORESE [6]) in order to be requested at runtime.

The second part of our work concerns metadata and metadata generation. For this, we invoke a web service chain calling basic text processes, language detection, linguistic and semantic annotation of texts and then RDF [7] generation. The semantic annotation process relies on concepts/properties or instances detection into texts using NLP[4] tools related with inference engine (CORESE) working on the existing knowledge base. The knowledge base grows up by integrating the newly created annotations. Another way to create annotation is considered when working on raw data. In this case

---

[2] http://www.sciences.univ-nantes.fr/info/perso/permanents/daille/acabit.html

[3] http://www.limsi.fr/Individu/jacquemi/FASTR/

[4] Natural Language Processing

the annotation generation process must be carried out in an interactive way by the user: when the expert produces an interpretation concerning a given data set, the system automatically creates links between the data representation and the ontology concepts and relations that describes his/her interpretation.

After having filled out the knowledge base, we are able to search and retrieve documents by means of an inference engine. Dedicated graphical interfaces dynamically generate SPARQL [8] queries that can be sent to the engine. For now, two dedicated graphical interfaces have been developed: one for geo-localization requests allowing users to select an area on a map, and one for geological dating requests allowing users to choose a geological time period in a graphical scale.

4. DEVELOPED ONTOLOGIES

An ontology is built in reference to a practical goal. For this reason, the aspects of reality that are chosen for encoding in an ontology depend on the task. Intense efforts have been developed by various geological surveys for issuing ontology-based formalizations of the geological knowledge currently represented on geological maps [9-11]. However, our needs for reservoir studies are not the same as those of geological map editors. For instance, the choice made in the NADM model is to carefully store field and sample observation attached to the objects described in geological maps, taking little or no account of genetic considerations. This choice can hardly be ours, since reservoir models built for Oil & Gas E&P first intend to describe the geological history of a prospect with the final goal of quantifying the amount of hydrocarbons produced as a result of this history. For this reason, we have defined specific ontologies related to earth modeling. This meets the traditional definition of Gruber, which stipulates that ontologies allow describing static knowledge attached to a field, by specifying what are the objects that compose the domain and how they are organized [12].

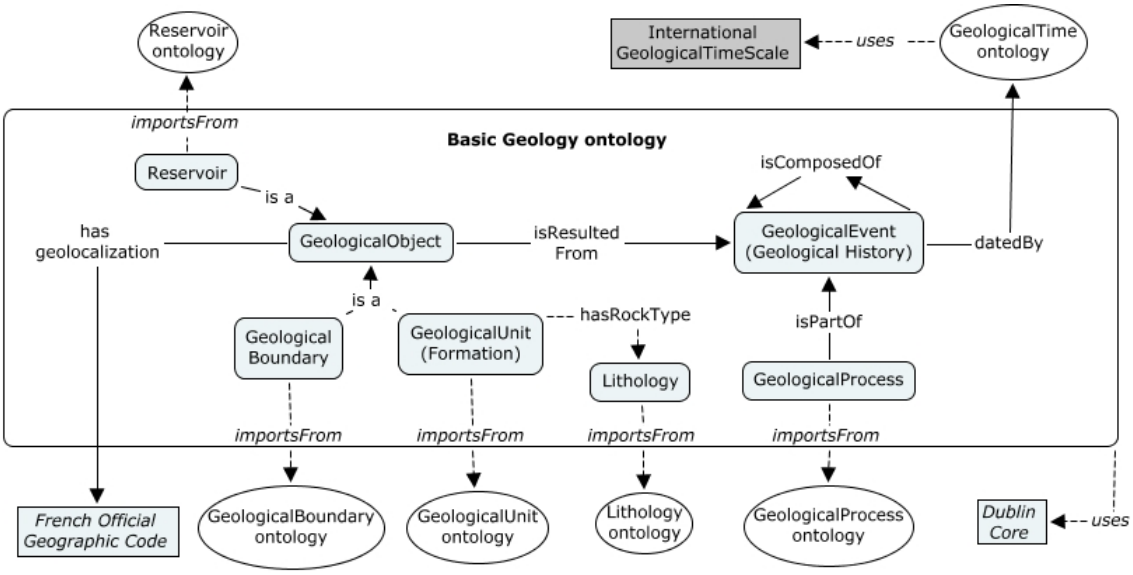

Fig. 1: Overview of the domain ontologies.

4.1 Domain ontologies for document search

At the initial stage of the e-Wok Hub project, domain experts manually extracted from a set of representative text documents, vocabulary relevant to CO_2 storage. We thus created a common vocabulary which can then be refined to describe semantic characteristics in OWL-DL. This vocabulary was tentatively classified resulting in a set of domain ontologies adapted to our needs i.e.:

- an ontology of **geographical terms**, which both rests on administrative nomenclature and on spatial (polygonal) area definition: GeoLocalization ontology;
- an ontology for defining and managing **geological ages**: Geological Time ontology;
- ontologies for describing the **basic geology**: geological units, geological boundaries, geological processes, lithology and reservoir.

These ontologies are described in detail in [13] and an overview is presented in Fig. 1.

4.2 Domain ontologies for earth modeling

Earth models correspond to final representations, which integrate successive steps of interpretation and modeling operated by professionals with various skills (geophysicists, geologists, petrologists, reservoir engineers). The items that are considered by these various specialists are different from one field to another. The data used for building earth models may be subject to various, possibly contradictory interpretations. Since an earth model is the result of a complex chain of successive elementary operations, an important issue is to keep the links between the modeled geo-objects and the various data and interpretations from which they result. We illustrate on Fig. 2 our proposed solution for linking data from Reservoir Prospect to their corresponding interpretations using annotation and ontologies.

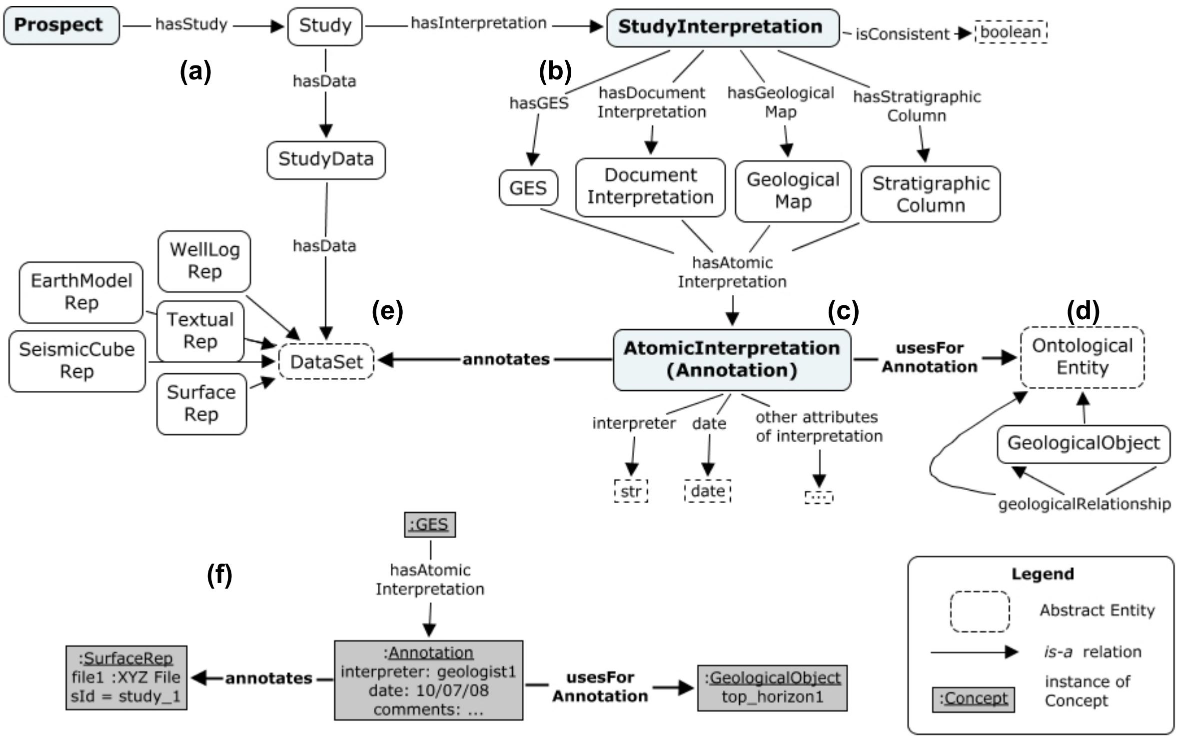

Fig. 2: Graphical representation showing the method proposed for linking data to interpretations using annotation and ontologies.

Various studies can be attached to a prospect **(a)**, each corresponding to a global geological interpretation **(b)**, which can be expressed in several ways (Geological map, Stratigraphic column, Geological Evolution Schema [1]). A global interpretation is composed of various atomic interpretations **(c)**, each linking a geological object **(d)** to a particular set of raw data **(e)**. The geological objects considered are those defined in the ontology for basic geology in section 4.1. In Fig. 2 **(f),** we present an example of this annotation schema, in which a file of type XYZ (an instance of *SurfaceRepresentation*) is interpreted by the geologist as being *top_horizon1*, which is an instance of the concept *Horizon*, a subclass of *GeologicalObject* (that is, an ontological entity). This interpretation is represented by the *annotation* that links the *data* (surface file) to the *geological object* (horizon). And this annotation composes, with several others, the definition of a GES (which represents surfaces and the temporal relations between them).

The ontological entities used to annotate the data sets are the concepts from the multiple ontologies developed for the earth science domains. The most important is the Basic Geology ontology, detailed in the section 4.1, but we also developed *Local Ontologies* (LO) for representing concepts attached to specialized domains such as structural modeling, seismic interpretation, well correlation, and thus corresponding to specific activities in the reservoir modeling chain. Conversely, the concepts from the Basic Geology ontology refer to geological objects, which are used through the whole earth modeling chain, receiving, though, a different characterization in each step.

We see thus geology as the red thread to which all interpretations and representations should be attached and the Basic Geology ontology as the *Global Ontology* (GO) to which all specific ontologies (LO) should be attached. For aligning the concepts of LO and GO, we establish subsumption relationships, notably the *is-a* and the *is-case-of* relations (i.e., LocalConcept *is-case-of* GlobalConcept) [3]. Establishing subsumption relationships between the models is likely to enable us to answer queries that cannot be addressed at present, because we cannot recover the relation between local objects identified in different phases of the geological modeling process.

5. CONCLUSION

We have shown in the present paper that the definition of domain ontologies and the development of a semantic annotation methodology enable to identify and handle raw data, geological objects and geological interpretations according to their semantic contents. This is of particular interest in the case of applications such as document retrieval or earth model building. Considering these particular cases, we have demonstrated the practical interest of knowledge formalization of specialized earth sciences domains for the creation of intelligent tools facilitating the task of end users responsible involved in more or less complex activities.

It appears in view of the two presented use–cases, that semantic web tools and services can provide efficient recording, tracking and knowledge centric prospect evaluation for petroleum E&P companies. In our opinion, the solution proposed in the present paper is a first step towards solving issues that are now becoming of paramount importance for industry, such as those related to memorizing the conditions in which geological interpretations are operated or earth models built.

REFERENCES

1. JF. Rainaud, M. Perrin, and Y. Bertrand. Innovative knowledge-driven approach for Shared Earth Model building. in *Joint SPE/EAGE Conference*. 2005. Madrid.

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

3. L. Mastella, Y. Ait-Ameur, M. Perrin, and JF. Rainaud. Ontology-based model annotation of heterogeneous geological representations. in *4th Int. Conf. on Web Information Systems and Technologies (WEBIST)*. 2008. Madeira, Portugal.

4. D. L. McGuinness and F. v. Harmelen, *OWL Web Ontology Language*. Technical report, W3C Recommendation, 2004. http://www.w3.org/TR/2004/REC-owl-features-20040210/.

5. H. Dehainsala, G. Pierra, and L. Bellatreche. OntoDB: An Ontology-Based Database for Data Intensive Applications. in *Proceedings of 12th Conference Database Systems for Advanced Applications (DASFAA)*. 2007.

6. O. Corby, R. Dieng-Kuntz, and C. Faron-Zucker. Querying the semantic web with the corese search engine. in *16th Eur. Conf. on Artificial Intelligence ECAI'04/PAIS'04*. 2004. Valencia, Spain.

7. F. Manola, E. Miller, and B. McBride, *RDF primer*. Technical report, W3C Recommendation, 2004. http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

8. E. Prud'hommeaux and A. Seaborne, *SPARQL Query Language for RDF*. Technical report, W3C Recommendation, 2008. http://www.w3.org/TR/rdf-sparql-query/.

9. North American Geologic Map Data Model Steering Committee, *NADM Conceptual Model 1.0 - A conceptual model for geologic map information*, in *US Geological Survey Open-File Report*. 2004, North American Geologic Map Data Model. p. 58.

10. R. Raskin, *Development of ontologies for earth system science*, in Geoinformatics: Data to Knowledge, A.K. Sinha, Editor. 2006, Geological Society of America Special Paper 397. p. 195-199.

11. S. Richard, *Geoscience concept models*, in Geoinformatics: Data to Knowledge, A.K. Sinha, Editor. 2006, Geological Society of America Special Paper 397. p. 81-107.

12. T. Gruber, *Toward Principles for the Design of Ontologies Used for Knowledge Sharing.* International Journal of Human and Computer Studies, 1995. **43**(5/6): p. 907-928.

13. L. Mastella, M. Perrin, M. Abel, J.-F. Rainaud, and W. Touari. Knowledge Management for Shared Earth Modelling. in *EAGE Conference & Exhibition*. 2007. London.