SEMANTIC SOLUTIONS FOR OIL & GAS: ROLES AND RESPONSIBILITIES

Jeremy Carroll, Ralph Hodgson, {jeremy,ralph}@topquadrant.com

This paper is submitted to The W3C Workshop on Semantic Web in Energy Industries focus area 3 "Pragmatic impact of Semantic Web technologies and standards for O&G"

Position/Abstract

Successful deployment of Semantic Web solutions in the oil and gas industry depends on identifying the different roles and responsibilities in such a deployment, and allocating the tasks appropriately. We sketch four generic roles: Semantic Web specialist, ontologist, domain expert, data entry. The most cost-effective manner to provide the first of these is by using advanced tools from specialist providers, such as TopQuadrant. Appropriate tools enhance the productivity of users in the three other roles by abstracting away some of the nitty-gritty aspects of the Semantic Web Recommendations. Much of the initial work in domain-specific ontology design, data integration and configuring data sources is done as a combined task between ontologists, who specialize in structuring knowledge for reuse, and oil and gas domain experts who understand the specific knowledge required, and are aware of both open and closed sources of information to be included in the solution. As well as Semantic Web specific validation steps such as checking an ontology for consistency, and concept-satisfiability, there are rather more trivial, but onerous, data laundry tasks involved with any information integration project, and creation of missing information. Thus the tool suite being used should also provide a solution for semi-skilled data review, and data entry.

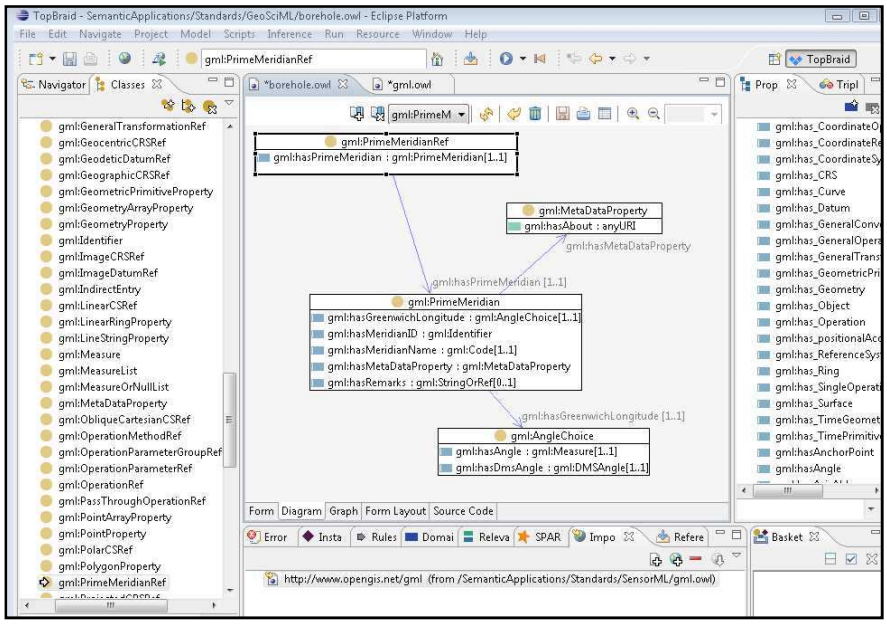

Fig. 1: Using TopBraid Composer for Ontology Design (GeoSciML Ontology)[1]

[1] This paper's illustrations are a shameless plug for the TopBraid Suite. To be less partisan, there are already several providers for high-level Semantic Web tools, with a high degree of interoperability, since the tools all conform with the W3C recommendations. It is advisable when planning a Semantic Web deployment to work with one of these providers, who often also provide ontology development and training.

Ontology design and provision of some data sources is a potential fruitful area for pre-competitive cooperation. To the extent that industry participants are prepared to share either or both of their domain knowledge, or their meta-knowledge about how to structure the domain knowledge, the costs associated with these steps can be amortized over the industry as a whole.
With NASA's permission, TopQuadrant can share foundational engineering ontologies developed in the areas of units, dimension, datatypes, etc. which we believe greatly assist in making shared frameworks for semantic applications.

Out of scope

In this paper various important topics are not addressed, these include:

- Examples of use cases for the workshop
- Specific commercial benefits of using Semantic Web technology in the Oil and Gas industries

This reflects TopQuadrant's expertise in the provision of semantic solutions provider, rather than in oil and gas. We trust that other participants will bring example use cases, and the workshop discussion will highlight the commercial benefits.

Semantic Solutions

Before describing the various roles, we sketch a few solution areas in which a specific project might be working.

Information Integration/Mash-ups

Exploratory aspects of the industry require the ability to integrate information in an essentially ad hoc fashion, and present that information in a fashion intelligible to appropriate experts, (fig. 2).

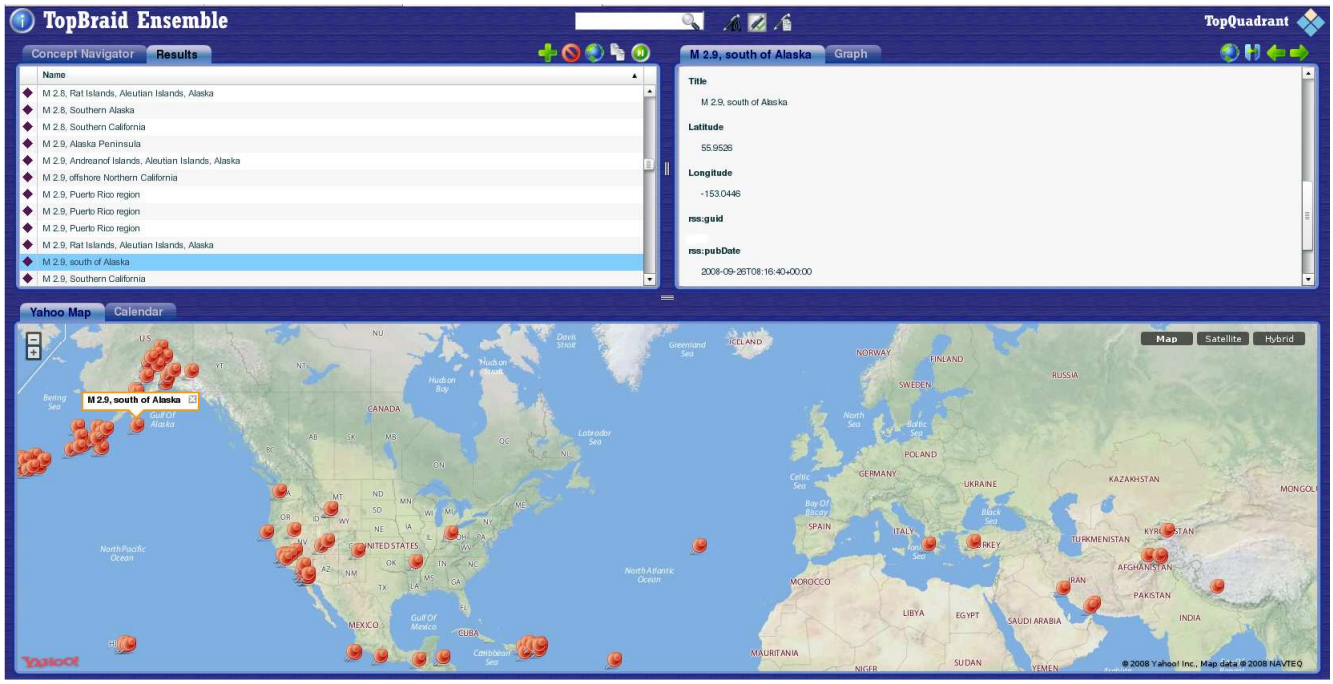

Fig. 2: Exploring an earthquake RSS feed in TopBraid Ensemble

Linked Data Discovery with On-Demand Search and Analysis

Semantic relationships between data enable powerful queries, leveraging expert knowledge. Non-programmers can connect diverse data sources, search and analyze data without the help of the IT department.

Enterprise Architecture

Many potential semantic web solutions for the O&G industry will be internal to some of the largest companies. Developing enterprise architectures with strong Semantic Web components allows better information reuse across the enterprise.

Enterprise Metadata Management

A related application area is in managing metadata across an enterprise to avoid information silos.

Vocabulary Management

Integrating heterogenous supervisory, control, and data acquisition systems requires a common shared vocabulary. Careful, methodical use of namespaces and naming conventions can address these needs, with URIs providing truly global identifiers.

Key roles

Domain Expert:

The role of domain expert is crucial. Such people are both a key target audience for an oil and gas solution, and also central in the development of the capabilities in the first place.
Responsibilities include:

- Identifying the domain concepts needed both for specific solutions and more broadly
- Identifying data sources for such information
- Interpreting the results from any information integration, and collaborating with the businesses for exploiting the new knowledge (fig. 2)

Ontologist:

The ontologist acts as a bridge between the domain expert and the Semantic Web technologies. They help in the correct formalization of the domain knowledge, critique the presentation of that knowledge based on various best practices, and explore the applicability of key concepts such as the open world assumption, and a lack of unique naming. In cases where closed world assumptions, or unique naming are more appropriate they configure Semantic Web tools appropriately.
Responsibilities include:

- Constructing ontologies based on the input of the domain expert, (fig. 1).
- Ensuring ontologies conform with the expectations of the Semantic Web recommendations
- Ensuring ontologies are fit for reuse in different applications
- Information integration (fig. 3)

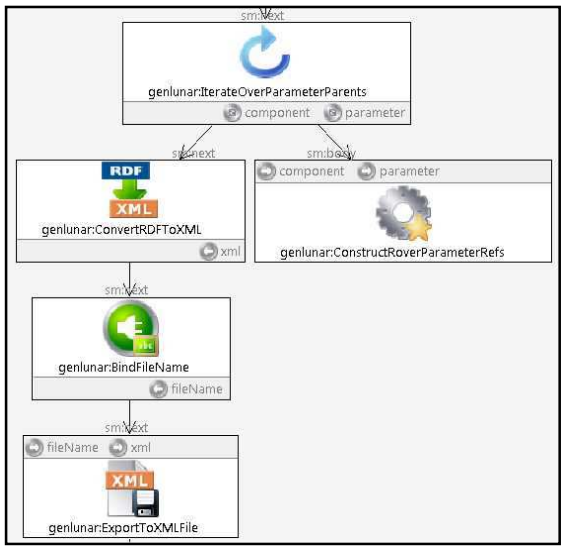

Fig. 3: Transforming a Lunar Rover ontology with TopBraid's SPARQL Motion

Semantic Web Specialist:

Correct production and use of RDF, OWL, SPARQL and GRDDL is itself an expert task. While everyone involved in a Semantic Web application will need some grounding in these technologies several subtleties should be left to a specialist.
Responsibilities include:

- Integration with low level Semantic Web toolkits and frameworks (such as Jena)
- Integration with OWL reasoners (such as Pellet)

- Importing both Semantic Web and non-Semantic Web data (fig. 4) and ensuring correct representation in terms of the ontologies developed by the ontologist (fig. 3)
- Appropriate presentation of integrated information for the domain expert
- Providing appropriate tools for semi-skilled assistance.

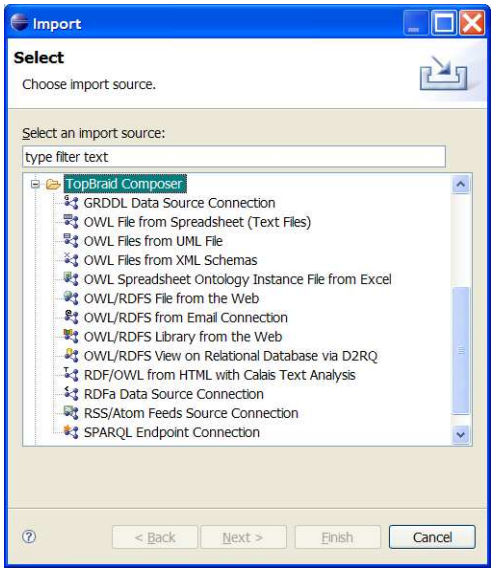

Fig. 4: Importing data with TopBraid

Data entry, data laundry:

In any application involving large quantities of information, there are issues both of data quality and missing data. While some automated support can be provided, often human intervention is required on an item by item basis. This partly depends on the application, where some tasks may be able to make do with dirtier data than others. To the extent that human intervention is needed, the project design for a semantic solution needs to keep the costs down in this step, which can be very extensive.

Knowledge Reuse and Amortizing Costs

In some sense, the basic ingredients of a semantic solution have been around for years: knowledge representation was a main plank of artificial intelligence in the 70s and 80s; these included reasoners, modeling languages etc. However, except in certain specific domains, typically characterized by a relatively small quantity of bounded knowledge, such solutions have not been economically viable.
Increasing machine size and performance addresses part of the problem, but a more significant aspect is simply the cost of generating enough information of sufficiently high quality to be useful.

Some of the key values that a Semantic Web solution offers over and above older solutions are to do with reuse on a web scale, enabling cost reductions.

- Reuse of knowledge frameworks: instead of each vendor having their own framework, a single standardized framework from the W3C enables reuse of tools, training and expertise across many industries. The lack of vendor lock in that open standards enable, also ensure that these cost reductions are realized by the end customer, rather than by the vendors.
- Reuse of information. By providing a single framework integrated with Web technology, the Semantic Web enables the reuse of information between many applications, in many areas, by many companies and individuals.
- Reuse of ontologies and schema. As well as the instance data, a key component of any information is how to organize it. Semantic Web technologies allow for sharing these even when the lower level information itself cannot be shared (for example, because it is viewed as a proprietary asset)

A useful outcome of this workshop would be a list of potential information sources available for reuse, and a list of area for collaborative development of ontologies and schemas. However, to the extent that the information required for oil and gas applications is expensive to collect it is unlikely to be freely available (e.g. RigLogix); and without the active participation of the information owners, development of ontologies for such information is difficult, and quite possibly futile.
A further issue is that developing ontologies for such proprietary data is likely to be useful, but presupposes that the application users have paid for licenses to the data. This is in conflict with the normal W3C royalty free commitments, and the consortium is unlikely to be the best forum for such work.
One area where such proprietary concerns are likely to be less prominent is in foundational engineering ontologies, for example, concerning units, dimensions, properties, datatypes and information structures. We have worked with NASA developing such ontologies (fig.5) for the Constellation Program for the return to the Moon and missions to Mars. Establishing agreement on such basics is vital in the development of higher level ontologies. With permission from NASA, this work can be shared under some governed licensing arrangement.

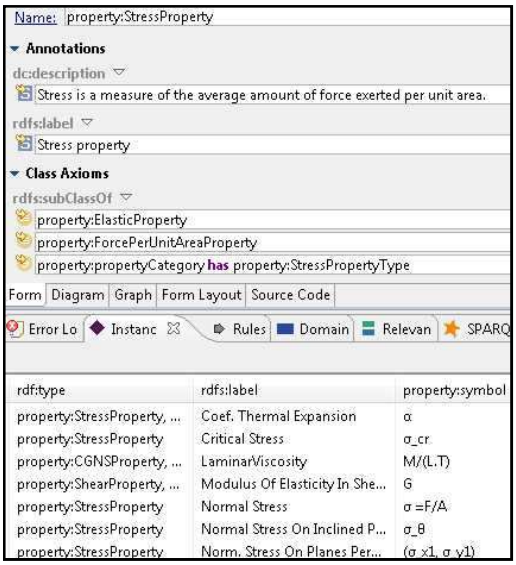

Fig. 5: StressProperty for NASA

TopBraid Suite

TopQuadrant's interest in this workshop is primarily to promote the TopBraid Suite of user-friendly Semantic Web tools to take the role of Semantic Web specialist identified above.
The three components are:

- *TopBraid Composer*, a specialist tool for ontologists to edit and manage ontologies.
- *TopBraid Live*, a server used in conjunction with TopBraid Composer to make both ontologies and their instance data available to a variety of applications.
- *TopBraid Ensemble*, a generic client to TopBraid Live that can be customized for example, to provide information browsing capabilities (with, for example a map viewer and a graph viewer) or to provide ontology driven data entry capabilities.

They enable integrating information from a wide range of formats, including native RDF and OWL formats, spreadsheets, databases, and RSS feeds, and provide a SPARQL based framework for manipulating the information to conform with a desired ontology.
The various illustrations show the TopBraid Suite being used for the different activities described. These could have been replaced with illustrations from other vendors' tools suites: the take home message is use a tool suite rather than building your own!

Conclusion

Deployment of semantic solutions in the oil and gas industry is more cost-effective when:

- There is pre-competitive cooperation on foundational ontologies and datasources, produced collaboratively between domain experts and ontologists.
- Tool suites are used to encapsulate the knowledge of semantic web experts.