INTEROPERABILITY IN THE PETROLEUM INDUSTRY

Jon Atle Gulla
Department of Computer and Information Science
Norwegian University of Science and Technology, Norway

Keywords: Interoperability, data integration, ontology engineering, enterprise integration.

Abstract: The petroleum industry is a technically challenging business with highly specialized companies and complex operational structures. Several terminological standards have been introduced over the last few years, though they address particular disciplines and cannot help people collaborate efficiently across disciplines and organizational borders. This paper discusses the results from the industrally driven Integrated Information Platform project, which has developed and formalized an extensive OWL ontology for the Norwegian petroleum business. The ontology is now used in production reports, and the ontology is considered vital to semantic interoperability and the concept of integrated operations on the Norwegian continental shelf.

1 INTRODUCTION

The petroleum industry on the Norwegian continental shelf (NCS) is technically challenging with challenging subsea installations and difficult climatic conditions. It is a fragmented business, in the sense that there is little collaboration between phases and disciplines in large petroleum projects. There are many specialized companies involved, though their databases and applications tend not to be well integrated with each other. Research done by the Norwegian Oil Industry Association (OLF) shows that there is a need for more collaboration and integration across phases, disciplines and companies to maintain the industry's profitability (OLF, 2005b). The existing standards do not provide the necessary support for this, and the result is costly and risky projects and decisions based on wrong or outdated data.

This paper presents the vision and some main results of the Integration Information Platform (IIP) project. The idea of the IIP project was to extend and formalize an existing terminology standard for the petroleum industry, ISO 15926. Using Semantic Web technologies, we have turned this standard into a real ontology that provides a consistent unambiguous terminology for selected areas in the oil and gas industry. The results of the project so far are promising, and the ontology developed by IIP is now being adopted by industry and is used in production reporting to the government.

The work in IIP is the first step towards the concept of *integrated operations* in the petroleum sector. In this long-term vision semantic standards and tools enable companies to work seamlessly together across geographical and organizational borders, and people from different disciplines or phases can cooperate without terminological confusion and misunderstandings.

The paper is organized as follows. In Section 2 we go through the structures and challenges in the subsea petroleum industry, explaining the status of current standards and the vision of future integrated operations. Section 3 briefly presents the parts of the Semantic Web initiative relevant to this project. Whereas the ontological work in the IIP project is introduced in Section 4, we discuss the issue of introducing semantic standards in the petroleum business in Section 5. Conclusions are found in Section 6.

2 THE SUBSEA PETROLEUM INDUSTRY

The Norwegian subsea petroleum industry is characterized by sophisticated technologies and highly competent and specialized companies. Many disciplines and competences need to come together in oil and gas projects, and their success is highly affected by the way people and systems collaborate and coordinate their work. On the Norwegian

Atle Gulla J. (2008).
INTEROPERABILITY IN THE PETROLEUM INDUSTRY.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - DISI*, pages 33-40
DOI: 10.5220/0001694100330040

Continental Shelf (NCS) there are traditional oil companies like Statoil, Norsk Hydro and ElfTotalFina, but also specialized service companies like Schlumberger, Haliburton, Baker Hughes, Aker Kværner, FMC KongsbergSub, and smaller ICT service companies.

Both the projects and the subsequent production systems are information-intensive. When a well is put into operation, the production has to be monitored closely to detect any deviation or problems. The next generation subsea systems will include numerous sensors that measure the status of the systems and send real-time production data back to onshore operation centers. For these centers to be effective, they need tools that allow them to understand and harmonize data, relate it to other relevant information, and help them deal with the situation at hand. There is a challenge in dealing with the sheer size of this information, but also in interpreting information that is deeply rooted in very technical terminologies.

The Norwegian petroleum industry is now facing a number of challenges (OLF, 2005a): Firstly, as most of the resources are in the decline phase, we now produce 2-3 times more oil than what is added through the development of new fields. Secondly, the costs on all the bigger fields are increasing significantly as we enter the decline phase. Thirdly, we see a development from traditional big oil fields of 300-400 million Sm3 (standard cubic meters, equal to 6.29 barrels) to fields of only 3-5 million Sm3, which also implies that many small and specialized companies enter the market. Lastly, the exploration in the north is environmentally very sensitive and requires new approaches to deal with climatic and geographical issues.

All these trends pose a challenge to the profitability of existing and future petroleum fields on NCS. While the costs of old large fields are increasing, the new ones are financially less attractive due to scalability problems. The multitude of companies involved, with their own applications and databases, makes coordination and collaboration more important than in the past. For the industry as a whole, this severely hampers the integration of applications and organizations as well as the decision making processes in general:

- **Integration.** Even though there is some cooperation between companies in the petroleum sector, this cooperation tends to be set up on an ad-hoc basis for a particular purpose and supported by specifically designed mappings between applications and databases. There is little collaboration across disciplines and phases, as they usually have separate databases structured according to different goals, processes and terminologies. It is of course possible to map data from one database to another, but with the complexity of data and the multitude of companies and applications in the business this is not a viable approach for the industry as a whole.
- **Decision Making.** A current problem is the lack of relevant high-quality information in decision making processes. Some data is available too late or not at all because of lack of integration of databases. In other cases relevant data is not found due to differences in terminology or format. And even when information is available, it is often difficult to interpret its real content and understand its limitations and premises. This is for example the case when companies report production figures to the government using slightly different terminologies and structures, making it very hard to compare figures from one company to another.

XML is already used extensively in the petroleum industry as a syntactic format for exchanging data. Over the last few years, there have been several initiatives for defining semantic standards to support information sharing in the business, but they have typically been limited to particular disciplines, companies or activities.

2.1 ISO 15926 Integration of Life-Cycle Data

ISO 15926 is a standard for integrating life-cycle data across phases (e.g. concept, design, construction, operation, decommissioning) and across disciplines (e.g. geology, reservoir, process, automation). It consists of 7 parts, of which part 1, 2 and 4 are the most relevant to this work. Whereas part 1 gives a general introduction to the principles and purpose of the standard, part 2 specifies the representation language for defining application-specific terminologies. Part 2 comes in the form of a data model and includes 201 entities that are related in a specialization hierarchy of types and sub-types. It is intended to provide the basic types necessary for defining any kind of industrial data. Being specified in EXPRESS (International Standards Association, 2007), it has a formal definition based on set theory and first order logic.

Part 4 of ISO 15926 is comprised of application or discipline-specific terminologies, and is usually referred to as the Reference Data Library (RDL). These terminologies, described as RDL classes, are instances of the data types from part 2, are related to each other in a specialization hierarchy of classes and sub-classes as well as through memberships and relationships. If part 2 defines the language for describing standardized terminologies, part 4 describes the semantics of these terminologies. Part 4 today contains approximately 50.000 general concepts like motor, turbine, pump, pipes and valves.

ISO 15926 is still under development, and only Part 1 and 2 have so far become ISO standards. In addition to adding more RDL classes for new applications and disciplines in Part 4, there is also a discussion about standards for geometry and topology (Part 3), procedures for adding and maintaining reference data (Part 5 and 6), and methods for integrating distributed systems (Part 7). Neither ISO 15926 nor other standards have the scope and formality to enable proper integration of data across phases and disciplines in the petroleum industry.

2.2 Integrated Operations

The Norwegian Oil Industry Association launched the Integrated Operations program in 2004. The fundamental idea is to integrate processes and people onshore and offshore using new information and communication technologies. Facilities to improve onshore's abilities to support offshore operationally are considered vital in the first phase of this program. Personnel onshore and offshore should have access to the same information in real-time and their work processes should be redefined to allow more collaboration and be less constrained by time and space. OLF has estimated that the implementation of integrated operations on NCS can increase oil recovery by 3-4%, accelerate production by 5-10% and lower operational costs by 20-30% (OLF, 2005b).

Central in the program is the semantic and uniform manipulation of heterogeneous data that can be shared by all relevant parties. Decisions often depend on real-time production data, visualization data, and background documents and policies, and the data range from highly structured database tables to unstructured textual documents. This necessitates intelligent facilities for capturing, tracking, retrieving and reasoning about data.

The first generation of OLF's integrated operations includes the definition of common terminologies that enable the automatic transfer of data between applications in the same discipline or inside the same company. Onshore operation centers for monitoring and controlling subsea oil installations are also part of this generation. The second generation requires complete formal ontologies that cover multiple domains and disciplines and support reasoning and inference of data using real-time data and rules. This will allow operators and vendors to integrate their operation centers, and subsea installations can to some extent control themselves using smart sensors and rule-based control systems that make use of semantic standards to integrate and interpret data from highly heterogeneous sources. Figure 1 shows how a comprehensive oil and gas ontology based on ISO 15926 is intended to support integration across disciplines and phases.

3 SEMANTIC WEB AND INTEROPERABILITY

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to

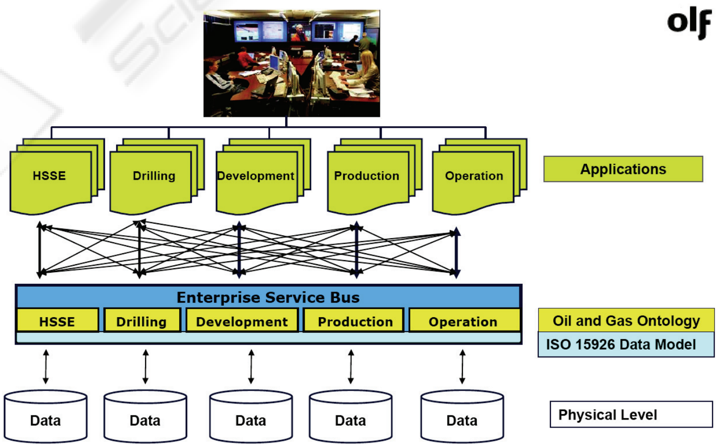

Figure 1: An oil and gas ontology allows cooperation across companies and disciplines (adapted from OLF).

work in cooperation" (Berners-Lee *et al.*, 2001). The Semantic Web is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. The general idea is to annotate data and services with machine-processable semantic descriptions. These descriptions must be specified according to a certain grammar and with reference to a standardized domain vocabulary. The domain vocabulary is referred to as an ontology and is meant to represent a common conceptualization of some domain. The grammar is a semantic markup language, as for example the OWL web ontology language recommended by W3C. With these semantic annotations in place, intelligent applications can retrieve and combine documents and services at a semantic level, they can share, understand and reason about each other's data, and they can operate more independently and adapt to a changing environment by consulting a shared ontology (Sheth *et al.*, 2002; Zhong *et al.*, 2002).

Interoperability can be defined as a state in which two application entities can accept and understand data from the other and perform a given task in a satisfactory manner without human intervention. We often distinguish between syntactic, structural and semantic interoperability (Aguilar, 2005; Dublin Core, 2004):

- *Syntactic interoperability* denotes the ability of two or more systems to exchange and share information by marking up data in a similar fashion (e.g. using XML).
- *Structural interoperability* means that the systems share semantic schemas (data models) that enable them to exchange and structure information (e.g. using RDF).
- *Semantic interoperability* is the ability of systems to share and understand information at the level of formally defined and mutually accepted domain concepts, enabling machine-processable interpretation and reasoning.

For the Semantic Web technology to enable semantic interoperability in the petroleum industry, it needs to tackle the problem of *semantic conflicts,* also called *semantic heterogeneity*. Since the databases are developed by different companies and for different phases and/or disciplines, it is often difficult to relate information that is found in different applications. Even if they represent the same type of information, they may use formats or structures that prevent the computers from detecting the correspondence between data. For example, the tables *ORG_NAME* and *COMPNY* in two different applications may in fact contain the same information about organizations. Similarly, while a time period may be modeled with the variables *"StartTime"* and *"Endtime"* in one database, the same information may be represented with *"StartTime"* and *"Duration"* in another (see for example (Pollock & Hodgson, 2004)). Even for concepts that are well understood and subjected to international conventions, the definitions may be slightly different from one source to another. The descriptions of 'mean time between failure' in Figure 2, which are extracted from various sources used in the petroleum industry, are almost identical, but it turns out that the differences are large enough to cause problems when data about mean times are transferred between applications.

	Mean time between failure
1	"A period of time which is the mean period of time interval between failures"
2	"The time duration between two consecutive failures of a repaired item" (International Electrotechnical Vocabulary online database)
3	"The expectation of the time between failures" (International Electrotechnical Vocabulary online database)
4	"The expectation of the operating time between failures" (MIL-HDBK-29612-4)
5	"Total time duration of operating time between two consecutive failures of a repaired item" (International Electrotechnical Vocabulary online database)
6	"Predicts the average number of hours that an item, assembly, or piece part will operate before it fails" (Jones, J. V. Integrated Logistics Support Handbook, McGraw Hill Inc, 1987)
7	"For a particular interval, the total functional life of a population of an item divided by the total number of failures within the population during the measurement interval. The definition hoolds for time, rounds, miles, events, or other measure of life units". (MIL-PRF-49506, 1996, Performance Specification Logistics Management Information)
8	"The average length of time a system or component works without failure" (MIL-HDBK-29612-4)

Figure 2: Different definitions of 'mean time between failure'.

The Semantic Web's approach to these problems is the construction of shared formal ontologies of all important domain concepts. These may be specified in OWL, which is a semantic markup language based on Description Logic. It has an XML syntax, is built on top of RDF(S)'s property statements and class hierarchies, and adds constraints for class membership, equivalence, consistency and classification (Antoniou *et al.*, 2005; W3C, 2006).

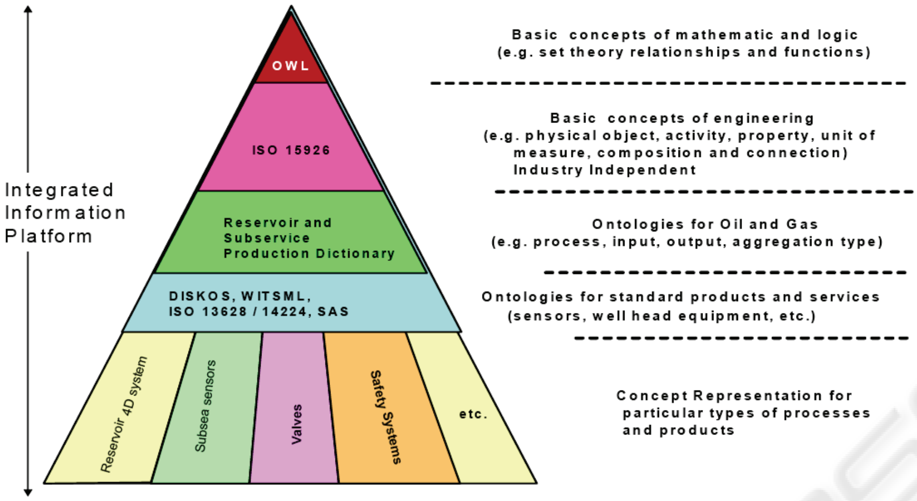

Figure 3: The standardization approach in IIP.

4 DEVELOPING OIL AND GAS ONTOLOGIES

The Integrated Information Platform (IIP) project was a collaboration project between companies active on NCS and academic institutions, supported by the Norwegian Research Council (Sandsmark & Mehta, 2004). Its long-term target was to increase petroleum production from subsea systems by making high quality real-time information for decision support accessible to onshore operation centers. The IIP project started in June 2004 and terminated at the end of June 2007 with a total budget of 26 million NOK (about 3.25 million Euro). The participants included Det Norske Veritas, Statoil, Norsk Hydro, Cap Gemini, Poseidon, OLF, FMC Technologies, National Oilwell Varco, OilCamp, POSC, IBM and NTNU.

The project addressed the need for a common understanding of terms and structures in the subsea petroleum industry. The objective was to ease the integration of data and processes across phases and disciplines by providing a comprehensive unambiguous and well accepted terminology standard that lends itself to machine-processable interpretation and reasoning. This should reduce risks and costs in petroleum projects and indirectly lead to faster, better and cheaper decisions

The project has identified a representative set of real-time data from reservoirs, wells and subsea production facilities. The OWL web ontology language was chosen as the markup language for describing these terms semantically in an ontology. The entire standard is thus rooted in the formal properties of OWL, which has a model-theoretic interpretation and to some extent support formal reasoning. A major part of the project was to convert and formalize the terms already defined in ISO 15926 Part 2 (Data Model) and Part 4 (Reference Data Library). Since the ISO standard addresses rather generic concepts, though, the ontology also includes more specialized terminologies for the oil and gas segment. Detailed terminologies for standard products and services were included from other dictionaries and initiatives (DISKOS,WITSML, ISO 13628/14224, SAS), and the project also opened for the inclusion of terms from particular processes and products at the bottom level. In sum, the ontology built in IIP has a structure as shown in Figure 3.

The ontology engineering approach in IIP was a combination of converting formal ISO 15926 definitions to manual modeling and verification of ontological structures. Due to the formality of ISO 15926's EXPRESS notation most of the ISO concepts could be automatically converted into legal OWL constructs. The manual modeling part was led by Det Norske Veritas and was handled by multi-disciplinary teams with years of experiences from standardization work and modeling projects.

This conversion of ISO 15926-2/4 from EXPRESS gave us an OWL hierarchy that has formed the backbone of the new oil and gas ontology. Additional terms were gradually and manually added to this hierarchy to reflect the larger scope of the new standard. In these initial stages it was considered important to concentrate on hierarchical relationships between concepts. Relationships and constraints of classes and relationships, which are needed for more

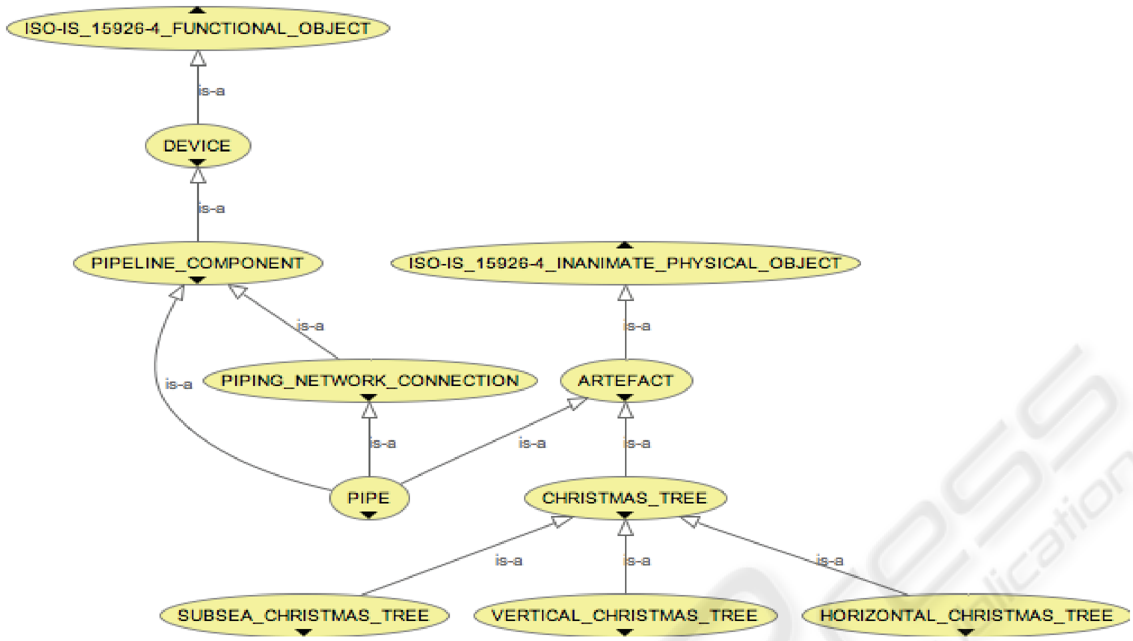

Figure 4: Christmas tree OWL hierarchy.

sophisticated reasoning with rules, are assumed to be added over time as the ontlogy matures.

Take for example the concept *Christmas tree,* which is an assembly of parts that is connected to the top of a wellhead to control the flow out of the well. Its OWL definition (without relationships and constraints) is:

```
<owl:Class rdf:about="#CHRISTMAS_TREE">
        …
        <dc:description
        rdf:datatype="http://www.w3.org/2001/XMLSchema#strin
        g">
                An artefact that is an assembly of pipes and
                piping parts, with valves and associated
                control equipment that is connected to the top
                of a wellhead and is intended for control of
                fluid from a well.
        </dc:description>
        <dc:title
        rdf:datatype="http://www.w3.org/2001/XMLSchema#strin
        g">
                CHRISTMAS TREE
        </dc:title>
        …
        <rdfs:subClassOf rdf:resource="#ARTEFACT"/>
</owl:Class>
```

These statements give us an informal definition of Christmas trees and reveal that they are subclasses of artefact. Looking at the excerpt of the class hierarchy in Figure 4, we see that there are at least three types of Christmas tree (subsea, vertical, and horizontal). It is a specialization of Artefact, which in turn is an Inanimate physical object that is made or given a shape by man. The Pipe class is also a specialization of Artefact, but it is also a specialization of two other classes. This is quite natural, as the pipe both has a physical (artefact) and a functional dimension (pipeline or network connection). More details about the construction of the ontology can be found in (Christiansen *et al.*, 2005).

The IIP project has now converted the ISO 15926 Part 2 (210 elements) and Part 4 (about 50.000) elements into OWL class hierarchies. In addition, we have incorporated additional terms from the following disciplines:

- Geometry and topology: ca. 400 terms
- Drilling and logging: ca. 2.700 terms
- Production: ca. 2.000 terms
- Safety and automation: ca. 150 terms
- Subsea equipment: ca. 1.000 terms
- Reservoir characterization
- Reliability and maintenance

The Tyrihans oil field, operated by Statoil, was used as a case in the IIP project. This means that the initial terms included in the ontology were based on the Tyrihans specifications, though they had been generalized and verified against other specifications as well, like ISO 13628 “Petroleum and natural gas industries – Design and operation of subsea production systems”. The ontology is the basis for developing new semantically interoperable applications, and IIP has already started experimenting with integrated visualization and information retrieval environments.

5 INDUSTRIAL ADOPTION OF SEMANTIC STANDARDS

In recent years a number of powerful new ontologies have been constructed and applied in domains like medicine and biology, where Semantic Web technologies and web mining have been exploited in new intelligent applications (Aguilar, 2005; Gene Ontology Consortium, 2000; Pisanelli, 2004). However, these disciplines are heavily influenced by government support and are not as commercially fragmented as the petroleum industry. Creating an industry-wide standard in a fragmented industry is a huge undertaking that should not be underestimated. In this particular case, we have been able to build on an existing standard, ISO 15926. This has ensured sufficient support from companies and public institutions. There is still an open question, though, what the coverage of such an ontology should be. There are other smaller standards out there, and many companies use their own internal terminologies for particular areas. The scope of this standard has been discussed throughout the project as the ontology grew and new companies signalled their interest. For any standard of this complexity, it is important also to decide where the ontology stops and to what extent hierarchical or complementing ontologies are to be encouraged. Techniques for handling ontology hierarchies and ontology alignment and enrichment must be considered in a broader perspective.

As far as the construction of the ontology is concerned, there was a need for both domain experts and ontology engineers. Since both the syntax and the semantics of OWL are non-trivial, it cannot be assumed that domain experts do the modeling themselves. To handle the complexity, the IIP project decided to model only the hierarchical relations in the first round, delaying relationships and constraints until the hierarchies were stable. For later update and quality assessment, it may be useful to use text mining techniques for automatic term extraction (Gulla *et al.*, 2004; Maedche, 2002).

The quality of ontologies is a delicate topic. It is important to choose an appropriate level of granularity. In this project we have been fortunate to have an existing standard to start with. What was considered satisfactory in ISO 15926 may however not be optimal for the ontology-driven applications that will make use of the future ontology. Ultimately, we need to consider how the ontology will be used in these applications and the nature of the source data to be annotated with ontological descriptions.

Since the Semantic Web is still a rather immature technology, there are still open issues that need to be addressed in the future. One problem in the IIP project is that we needed the full expressive power of OWL (OWL Full) to represent the structures of ISO 15926-2/4. Reasoning with OWL specifications is then incomplete. The lack of industrial SW applications is another issue worth taking into consideration. There may be performance and maintenance complexities that are still unclear with such an untested technology. However, there is now a large community promoting SW technologies and developing innovative applications, and the first commercial products have also emerged. Additionally, the tool development in IIP indicates that the technology can form the semantic foundation for a new generation of intelligent, interoperable information services.

The success of the new ontology, and standardization work in general, depends on the users' willingness to commit to the standard and devote the necessary resources. If people do not find it worthwhile to take the effort to follow the new terminology, it will be difficult to build up the necessary support. This means that it is important to provide environments and tools that simplify the use and maintenance of the ontology. Intelligent ontology-driven applications must demonstrate the benefits of the new technology and convince the users that the additional sophistication pays off. A positive sign is that daily production reports and daily drilling reports are now standardized across companies with the help of our ontology, and the major oil companies on NCS as well as IBM are now working on a similar semantic standarization of monthly production reports. The industry has received the standard with enthusiasm and are already planning new projects for further expansion of the standard and the development of appropriate semantic applications.

6 CONCLUSIONS

The Integrated Information Platform project is one of the first attempts at applying state-of-the-art Semantic Web technologies in an industrial context. Existing standards have been converted and extended into a comprehensive OWL ontology for reservoir and subsea production systems. The intention is that this ontology will later be approved as an ISO standard and form a basis for developing interoperable applications in the industry.

With the new ontology at hand, the industry will have taken the first step towards integrated operations on the Norwegian Continental Shelf. Data can be related across phases and disciplines, helping people collaborate and reducing costs and risks. However, there are costs associated with building and maintaining such an ambitious ontology. It remains to be seen if the industry is able to take advantage of the additional expressive power and formality of the new ontology. The work in IIP indicates that both information retrieval systems and sensor monitoring systems can benefit from having access to an underlying ontology for analyzing data and interpreting user needs.

As the class hierarchies in the ontology are completed, the emphasis of the IIP project will be put on adding more relationships and constraints to the ontology. This also includes specifying rules that will be used to analyze anomalies in the real-time data from subsea sensors. At that point we can start exploiting the logical properties of OWL and start experimenting with the next generation rule-based notification systems. We can also use agents to simplify the coordination of work and improve cooperation along the entire value chain. We will then see if a strong semantic foundation makes it easier for us to handle and interpret the vast amount of data that are so typical to the petroleum industry.

ACKNOWLEDGEMENTS

This research is funded by the Integrated Information Platform for reservoir and subsea production systems project under the Petromax research program.

REFERENCES

Aguilar, A. (2005). Semantic interoperability in the context of e-health: CDH Seminar. http://m3pe.org/seminar/aguilar.pdf.

Antoniou, G., Franconi, E., & van Harmelen, F. (2005). Introduction to semantic web ontology languages. In N. Eisinger & J. Maluszynski (Eds.), *Reasoning web, first international summer school*. Malta: Springer.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American, 284*(5), 34-43.

Christiansen, T., Jensen, M., & Valen-Sendstad, M. (2005). Defining iso 15926-4 reference data library classes in owl: The Norwegian Oil Industry Association. http://www.olf.no/io/kunnskapsind/?28140.pdf.

Dublin Core. (2004). Dublin core metadata glossary: http://library.csun.edu/mwoodley/dublincoreglossary.html.

Gene Ontology Consortium. (2000). Gene ontology: Tool for the unification of biology. *Nature Genet, 25*, 25-29.

Gulla, J. A., Brasethvik, T., & Kaada, H. (2004, June). *A flexible workbench for document analysis and text mining*. Paper presented at the 9th International Conference on Applications of Natural Language to Information Systems (NLDB'04), Salford.

International Standards Association. (2007). Industrial automation systems and integration - product data representation and exchange. Par 11: Description methods: The express language reference manual.: http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=18348.

Maedche, A. (2002). *Ontology learning for the semantic web*: Kluwer Academic Publishers.

OLF. (2005a). Digital infrastructure offshore - common network operation management for digital infrastructure offshore on the norwegian continental shelf: The Norwegian Oil Industry Association.

OLF. (2005b). Integrated work processes: Future work processes on the norwegian continental shelf: The Norwegian Oil Industry Association.

Pisanelli, D. M. (Ed.). (2004). *Ontologies in medicine. Volume 102 studies in health technology and informatics*: IOS Press.

Pollock, J. T., & Hodgson, R. (2004). *Adaptive information: Improving business through semantic interoperability, grid computing, and enterprise integration*: Wiley Publishers.

Sandsmark, N., & Mehta, S. (2004). Integrated information platform for reservoir and subsea production systems, *Proceedings of the 13th Product Data Technology Europe Symposium (PDT 2004)*. Stockholm.

Sheth, A., Bertram, C., Avant, D., & Hammond, B. (2002). Managing semantic content for the web. *IEEE Internet Computing, July/August*, 80-87.

W3C. (2006). Owl web ontology language overview. Http://www.W3c.Org/tr/owl-features/.

Zhong, N., Liu, J., & Yao, Y. (2002). In search of the wisdom web. *Computer*, 27-31.