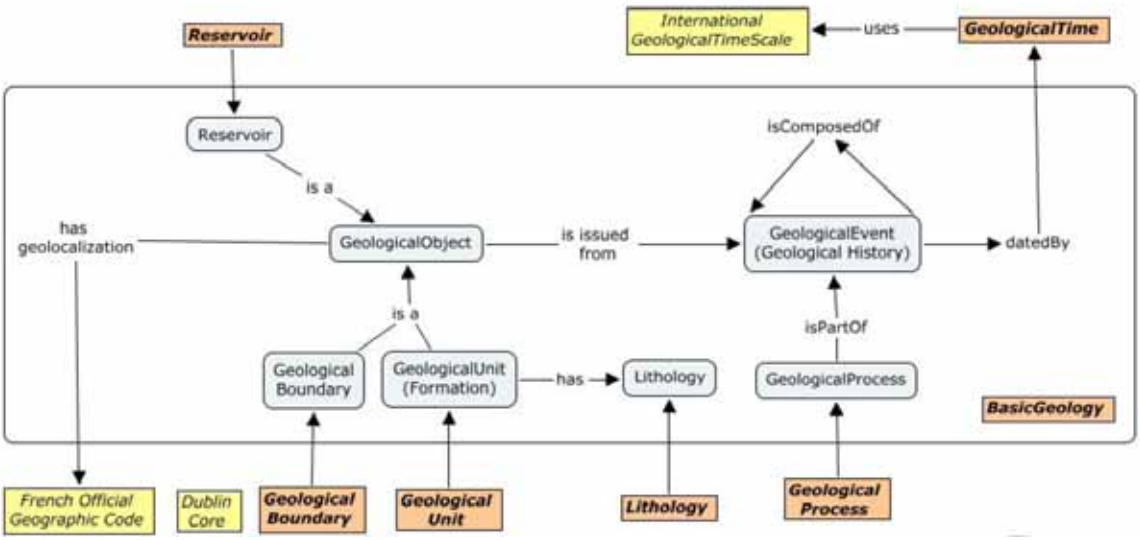
World Wide Web Consortium (W3C)
Workshop on the Semantic Web in Oil and Gas[1]
December 2008 – Houston, Texas



Example of geological ontologies[2].

The World Wide Web Consortium's (W3C) Workshop on the Semantic Web in Oil and Gas was held in Chevron's Houston offices with attendance from a good cross section of industry decision makers and semantic practitioners. Attendees hailed from BP, Oxy, Shell, Total, Halliburton, Schlumberger and Energistics *inter alia*. On the oil company side, Chevron is the most enthusiastic proponent of the semantic web although the technology still hovers between academic proof of concepts and enterprise deployment.

The presence of CEO Steve Bratt was an indication of the importance that the W3C gave to the event. Bratt described the semantic web as 'Web 3.0,' and RDF's simple 'subject, property, value' triples as set to turn the world wide web into a 'big global relational database.' Presentations from Chevron, USC, Bentley Systems, Fluor Corp. made more prosaic claims for the technology. These fall broadly into two categories; 'top down' (proper) database applications leveraging an upper ontology and 'bottom up' text-based tagging systems. The flagship oil and gas development of the ISO 15926 reference data library (RDL) falls into the first category. The above image from the E-WOK project is an example of the tagging approach. The interaction between the oil and gas domain specialists and the W3C's semantic luminaries led to some enlightening debates particularly in the context of the RDL's 'template' approach to extending RDF into a science and engineering context.

Our favorite paper was the presentation on the E-WOK project which showed how semantic tagging of an existing document library enables queries such as 'What diagenesis has affected the Bathonian of the Paris Basin?' The system 'understands' that the Bathonian is part of the Jurassic and that the Paris Basin includes the Ile de France area. Other enlightening presentations came from Fluor Corp., Bentley Systems and several more by or on behalf of Chevron. As Chevron's Frank Chum put it, RDF, the technology that underpins the semantic web has proved 'extremely useful to us in bringing together 'factoids' of unrelated information. In this context RDF is solid.' Also of note were the lively panel sessions and Q&As which we have tried capture here.

Semantic technology is a hard subject, not the least because of the tendency amongst its advocates to presuppose familiarity with concepts like 'serialization,' 'upper ontologies,' 'URIs' and a host of more abstruse jargon. While there is a hope that this technology will solve a lot of interoperability problems, exactly how this is going to come about remains unclear. Notwithstanding this, the semantic web is the most interesting thing that is happening in information management and this report has been exciting to write. The aim of the workshop was to ascertain interest in a joint industry effort to further the use of semantic technology in oil and gas under the auspices of the W3C. At the time of writing, the jury is out on this.

[1] *The official W3C report of the meeting is available at http://www.w3.org/2008/12/ogws-report.html.*

[2] *Image courtesy French Petroleum Institute and E-WOK Project - http://www-sop.inria.fr/edelweiss/projects/ewok/captureCO2/slide4.html.*

Highlights

Keynote – Jim Crompton, Chevron
Semantic technology for reservoir management
ISO 15926 in RDF/OWL
Semantic web in earth sciences – E-WOK
IBM's Upper Ontology

Contents

Subscription information - *This report was produced as a part of The Data Room's subscription-based Technology Watch reporting service. More from* www.oilit.com/tech *or email* tw@oilit.com.

TW0816_1 Welcome address – Tom Bell, Chevron

Chevron has several semantic web projects and hopes that this workshop will result in a 'critical mass' for a working group to build something substantive with this technology. Bell confessed to be 'a bit apprehensive' from a business person's perspective and issued a challenge to the audience to provide support for the technology and to communicate on technology and terminology.

TW0816_2 The semantic web a.k.a. Web 3.0[3] – Steve Bratt, W3C CEO

What is the W3C... beyond 'the html consortium?' An IBM review of standards organizations described the W3C as 'exemplary'. The W3C tries to play well with others such as ISO, ITU, IETF, UN OGF etc. www.w3.org gets 10 million hits (300,000 visitors) per day and has a Google page rank of 10 ('one of only 15 sites to have this')[4]. According to Bell, 'almost every call center in the world is based on the W3C's VoiceXML standard. The W3C is about to release a 'major new accessibility' standard. The W3C has 60 plus groups of which the Semantic Web Activity. The World Wide Web has evolved from hyperlinked documents (Web 1.0) to 'one web' of creators and consumers (Web 2.0) and now, linked data (Web 3.0) a.k.a. the Semantic Web. RDF is the core semantic web standard – using a simple 'subject, property, value' triple each with its own unique identifier (URI)[5]. This turns the web into a 'big global relational database.' RDFa allows semantic data to be embedded in XHTML. XML can be 'scraped' to produce RDF with GRIDDL[6]. Bratt showed a 'Crossing the Chasm' chart of take up stating that the semantic web was still in the early adopters end of spectrum[7]. Gartner is 'very positive' but sees full take up of semantic data by 2017, and by 2027 for a 'semantic environment.[8]'

What will be the tipping point for semantic data? Maybe something that oil and gas does. The W3C is increasingly focused on the application of its standards – helping users, and improving standards by user involvement. Domains that are 'expressing interest' include healthcare, financial services, etc. For instance the healthcare/life sciences industry is plagued by barriers to interoperability including commercial applications, external resources, a lack of APIs and data mismatches. See the October 2005 BioIT World article on 'Masters of the Semantic Web' and 'Science and the Semantic Web' in Science 2003. In healthcare the semantic web provides an interoperability platform for lots of stuff including lab information management systems (LIMS).

In 2004, healthcare domain expert John Wilbanks (of the Interoperable Informatics Infrastructure Consortium, I3C) was elected a W3C Fellow. A W3C healthcare workshop with Siemens on board has worked on BioRDF, Open Drug Data, Scientific Discourse etc. A bunch of open databases were all exposed in RDF to see what could be achieved. Data sources including Oracle, Excel, HTML have been concatenated into 500 million triples so far. SQL type query can be performed across the different data sets.

Bratt suggested a possible outcome of the present meeting would be the inception of an Energy/Oil and Gas interest group along the lines of the HCLS above. With this in mind, the workshop should discuss areas of impact, pragmatic, low hanging fruit and their target audiences.

Q&A

Fluor – You started with 35 companies in health care is this growing exponentially?

Not at all exponential – Actually we began with 2, (they came to us). They saw the W3C as a neutral venue. Since then we have seen slow growth.

Fluor – We like the idea of a workgroup for oil and gas – but it should be bigger. We want to include the process and plant industries[9].

Point taken.

---

[3] *http://www.w3.org/2008/Talks/1209-W3C-EnergyOandG-Workshop/Intro-W3C-OandG-bratt.pdf.*

[4] *At time of writing w3.org's page rank is 9.*

[5] *If there was a snappy definition of URI as opposed to URL we would love to hear it – for the meantime try http://en.wikipedia.org/wiki/URI.*

[6] *Gleaning Resource Descriptions from Dialects of Languages - http://www.w3.org/TR/grddl-primer/.*

[7] *A sluggish start for a technology announced in 1999!*

[8] *Surely far too far off for any credible forecasts!*

[9] *This is indeed an interesting point. The W3C reasonably enough has divided industry into 'Energy' with subset of 'Oil and Gas' and right away we are talking about a cross-industry activity impacting manufacturing, construction, chemicals etc.*

TW0816_3 Keynote address[10] – Jim Crompton, Chevron

Crompton is Chevron's iField program advisor and a Chevron Fellow. He also signs Chevron's WC3 invoice! He has served on several standards bodies, chairing the API PIDX for several years, while EDI was brought into the XML world. He learned a lot about procurement standards when Chevron bought a non compliant package! Another ex geophysicist, Crompton has moved 'from not knowing all the right answers to asking the right questions.' What questions do we need to answer? Oil and gas indeed does have the same problems as healthcare, we are not unique. The end goal is the fully instrumented facility – *à la* Blind Faith platform built in 6,500 ft. This cost 'only' $1 billion, some facilities have cost ten times as much.

Crompton started his career as a geophysicist and recalls the days of the 'well organized paper file room.' 'Life has never been that good.' 'We lost control in the digital world.' Companies are now trying to collaborate with contractors continents away. Also trying to update financials or anticipate well kick in near real time. Surveys show that 30-70% time is still spent finding data – a severe penalty. One person may only be able to find 10% of the data he or she created. So they 'give up looking and go with a best guess.' But as the experience of a workforce diminishes, the quality of 'best guesses' decreases.

Chevron, like other oils, exchanges a large amount of information with third parties. Maybe 70% are outside of US. Much sophisticated analysis is performed, but is it being done on good data? There is a tendency to lose understanding of the meaning of data. For instance a 'well bore' means different things to explorationists and production engineers. At a high level, information management has not changed much. It has been recognized that silos are an efficient way of managing data. But this does create access problems. Regarding search, Google is all very well, but you need to tag and you need to know if you have got everything. We can still only really search within a given environment. Chevron's intranet search is OK, but does not include email, databases etc. Search becomes siloed. Information pipelines are now fed with cheap sensor data – impacting HSE, equipment conditioning. Downhole intelligent completions provide real time data and an exponential increase in data gathering capacity. Our modeling capability is likewise now 'massive'. Between the expanding data volumes and the modeling capability there is a 'kink in the pipeline,' a yawning gap between data collection and modeling/analysis for decision support.

But we are reasonable people! How did we screw it up? The answer is in the economic cycles – when the cycle is down, there is no money for IM. When the going gets good, acquisitions and mergers kicks in. Chevron has experienced ten years of data hell following the Texaco merger! After a merger, nobody knows how the legacy systems work. The next worst thing is when someone leaves and hands over their spreadsheet. Engineers are not very good at documenting what they do. Meanwhile data sets are exploding into the gigabyte and terabyte range. Industry has worked hard at standards – religiously and with good intent – but remains poor at implementing them. We have got all the data models but we don't use them. The business impact of the current state of affairs is that engineers use month-old data for decisions. Optimization works fine at a small scale – but multiple attempts at optimization at different scales hamper the 'big picture' optimization. It is hard to react to dynamic changes such as water breakthrough or equipment failure. Today we are reactive and we need to be more proactive.

Regarding the semantic web, haven't we been here before? Yes – not exactly the semantic web but there have been other attempts to bridge IM and the business. Even IT folks shy away from this problem. Business users don't want to assure data quality. Well log users tend to check well file prints back into the system create duplication. The business is seen to be complex when you try to write business rules. You can't manage structured and non data separately. Excel is the most critical application in oil industry. Chevron has maybe a million spreadsheets and the most important function in an application is, 'export to excel,' (the second most important is import from Excel!). Do we really treat information as the asset it is? Do we even know what it is worth? Crompton believes not and suggests that the path to data sanity lies in data governance, a reference and master data architecture spanning structured and unstructured data, and data quality. Such an information architecture is a new thing for Chevron. Standardization is important but 'we need to go further and allow for regular portfolio rationalization, we need a common language.'

Crompton advocates a three tier architecture. SOA may be a way to achieve this but oil and gas has not yet got a taxonomy to support this – a potential role for the semantic web. This could be applied to legacy data by re-tagging and structuring the data. The industry has had some success with XML data protocols. Chevron came late to the WITSML party but has been more proactive with PRODML. There have been some successes here but we are still 'kicking the rock' with internal deployment. We already tried the one big data model approach and it failed miserably. Today Chevron has 600TB of data (70% technical, 25% business, 5% financial) and this is growing. There are currently 300 million office documents. The Agbami offshore Nigerian development is a floating refinery sitting on top of a complex oilfield producing 200,000

[10] *http://www.w3.org/2008/12/ogws-slides/Crompton.pdf*

bopd with gas and water injection (a no flare field). Agbami is tremendously complex. It has more instrumentation than the most sophisticated refinery – around 100,000 equipment tags (1,500 downhole). This is the digital oil field.

Q&A

Shell –I am currently researching IM for regional framework studies. In the past, Shell had unsurpassed knowledge but now we can't find the studies! I'm sure that they are in a warehouse in Schulenberg – I'm not that desperate yet! We are currently debriefing 70 year old retirees for information.

SoftQuadrant – Data quality is an unrealistic objective, the data is what it is. What is more important is quality metadata.

TW0816_4 A brief history of upstream standards – Bertrand du Castel, Schlumberger

In the 1980s companies were selling products to the upstream based on databases – but by end of the 1980s these were all different. Vendors tried to get cooperation, to share the risk and share the data. In early 1990s there was cooperation on a common data model but this did not work – it had the same fate as Esperanto. Du Castel spoke with first hand knowledge – he was head of POSC three times when it was an $8 million per year operation. In terms of data management standards nothing came out of it – an ROI of zero. So we move on to other approach to finding ways to speak together. Common exchange formats. WITSML, PRODML has taken root. We have a good conceptualization/model of what exploration is and an understanding of what we are all talking about. Now we need go to the next level and express the value of this understanding. Maybe we made a mistake in the 1990s – but now we are getting back on track.

TW0816_5 A semantic solution for reservoir management[11]– Ram Soma and Amol Bakshi, Chevron, Avanade and USC CiSOft[12]

The Chevron-backed, three year Integrated Asset Management (IAM) R&D project has been conducted at the University of California's CiSoft department. Last year saw real world deployment – in the form of a technology transfer project, with UK-based Microsoft developer Avenade. IAM is a 'comprehensive transformational approach' to integrated oilfield operations that sets out to increase integration, enable 'what if' scenarios, create a knowledge base and reduce risk. This in the face of the challenge of non interoperable data silos. IAM is said to be 'non disruptive' technology.

IAM's metacatalog is an OWL triple store of many 'dissemination models' – focusing on what data means rather than how to access it. Semantic web technology provides an expressive and rich data model suited for inference and rule based reasoning, it is also vendor-independent. USC developed a semantic SOA for IAM in 2006. In 2007/8 IAM was tested on reservoir simulation case studies. An IAM 'agent' captures results from simulations and stores them in the IAM Knowledge base which allows for metadata-based search. 'Agile development techniques (SCRUM[13] and SPRINTS iterations) proved problematical as the ontology frequently modified. Change management techniques were developed to handle this. UC is aware of performance issues and is benchmarking development to mitigate the risk of the OWL-based solution.

The IAM ontology is a three level design – with a domain-independent upper ontology containing common concepts like time. Next comes a domain-level model of asset elements. Then application/workflow specific ontologies. These are stacked so lower ontologies use the upper levels. OWL data was stored in a relational database. The system was implemented in Protégé with plugins for Jena, Pellet SPARQL. Lessons learned – keep OWL ontology small and modular, use OWL imports, track performance, design for change with SPARQL querying. It is key to separate the knowledge base from business logic and the user interface.

TW0816_6 Semantics are for People too![14] – Lee Feigenbaum, Cambridge Semantics.

Cambridge Semantics is a 'semantic vendor[15].' Often the semantic web is about 'machines doing stuff,' making information on a web machine readable, allowing inference and data integration – a 'machine centric' approach. But there is more to semantics than machines and that can impact everyday work. This is what Cambridge Semantics does – building 'models we can understand.' Semantic web technologies are an

---

[11] *http://www.w3.org/2008/11/ogws-papers/soma.pdf*

[12] *Center for Interactive Smart Oilfield Technologies a joint venture between the University of Southern California and Chevron.*

[13] *One interesting facet of SCRUM is, 'less focus on formal documentation' – and that's supposed to be a good thing!*

[14] *http://www.w3.org/2008/12/ogws-slides/cambridge-semantics.pdf.*

[15] *From Chum's presentation, at least some of this work was done for Chevron although this wasn't clear from Feigenbaum's talk.*

expressive way of modeling domains and expertise. Models align better than the relational database (RDB). OWL/RDF is a pure form of modeling. Other approaches, RDB, XML schema are often done with other things in mind (like performance for the RDB)[16]. Why does this matter? Because we can't automate everything. Data exists in legacy stores, in silos, in Excel. Today's answer is 'people power,' spending months with different silos, reformatting for data analysis and business intelligence. There is an impedance mismatch between people and how software works. That's why we love spreadsheets! But this approach creates the 'shadow IT' world – of data that is not discoverable or searchable, but is nonetheless a 'very human way of working.' But how do you get information into a machine centric format for work with business intelligence. This is how Cambridge Semantics views the world.

A drilling platform produces daily reports, stuff for joint venture partners, etc. Data in PROD/WITSML. Moving from Excel to PRODML is to put a square peg in round hole. PRODML has lots of stuff that you would never put in a spreadsheet. So that's a non starter. Cambridge Semantics' solution is 'Anzo for Excel.' This takes data from a spreadsheet and makes it searchable, discoverable and 're-castable' into other formats like PRODML. A daily production report for BHP Billiton in Excel is linked through the Anzo server via small ontology. Click on an Excel cell and say what it means. Anzo does matrices/ranges that can be identifed as 'well production data' or whatever. Linkage used a template and data changes from being human readable to become the hub of a machine centric data system. Cambridge Semantics 'Exposé' allows visualization, summary and roll-up of the above information. Exposé includes charts, data mining with 'facets' on various parameters. There is also a neat timeline display for comments. You can 'easily generate' PRODML or mash-up data with Google Earth. More from www.cambridgesemantics.com and a video of the project on http://www.cambridgesemantics.com/videos/industry/oilgas/NOJV_Reporting/.

Q&A

Schlumberger *– The problem with these tools is that nobody wants to use them.*

Yes, that's why we try to automate as much as possible. People are doing this already – solving the 'shadow IT' problem.

Oil IT Journal *– Can you use this technology to make spreadsheets better – with named ranges etc?*

No, though this is a noble goal which we are working on.

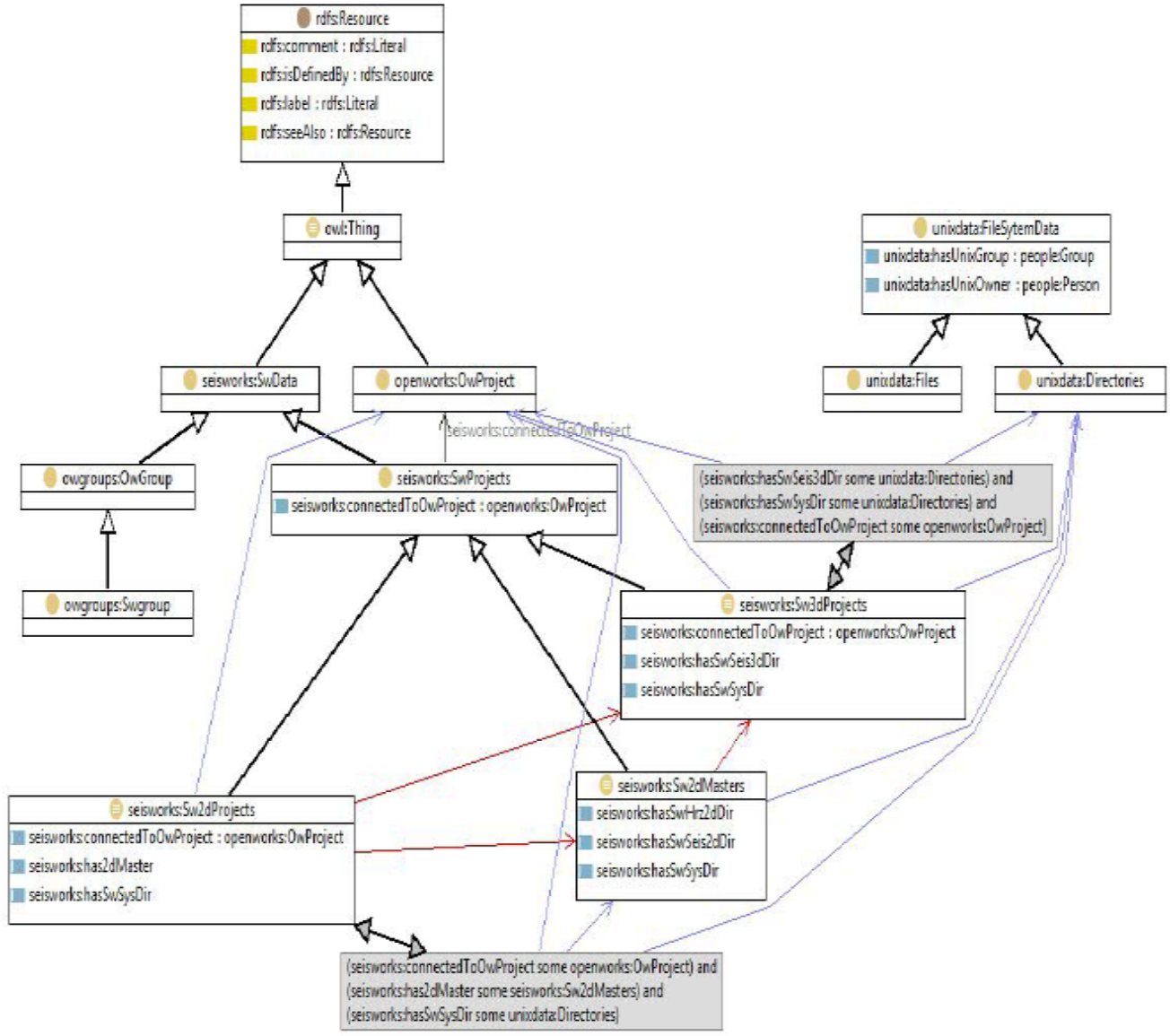
TW0816_7 Chevron Position Paper[17] – Frank Chum et al., Chevron

Chevron has been scanning and assessing the W3C's semantic web activity since 2000 and joined the W3C in 2005. Since then Chevron has been an observer on the healthcare and life sciences special interest group. In 2006, Chevron joined the semantic web education and outreach special interest group. Chevron has published ontologies for information integration in oil and gas on the W3C site (see http://www.w3.org/2001/sw/sweo/public/UseCases/slides/Slides.odp and ready yourself with a copy of Open Office!).

Chevron's problem areas include the difficulty of 'semantic reconciliation' of enterprise metadata, the standardization of information and integration across WITSML, PRODML, ISO15926, PPDM etc. The 'N° 1 role of the semantic web is data integration across applications' (2001). To achieve this you have to manage and share and ontology. Activities in Chevron include IAM (see above), drilling and production knowledge management and data exchange (Lee Feigenbaum's presentation), Major Capital Projects Operational Systems (MCPOS which leverages ISO15926 – in collaboration with Bentley and Fiatech) and a project involving Unix file system metadata – the ETC Exploratory Pilot. Here Mario Casetta posed the question, 'Is semantic web technology mature enough for use in Chevron?' It promises the holy grail of enterprise search – linking technical data and documentation. Chevron has developed a metadata store for technical data – 'in way that brought value and did things that were not possible before.' The Exploratory Pilot mustered metadata from Unix-based SeisWorks and GoCad projects, and building an ontology and RDF data store. RDF proved 'extremely useful to us in bringing together factoids of unrelated information.' In this context, 'RDF is solid.'

[16] *This is reminiscent of the design of a 'pure' model in a data modeling language like Express with relational projections. But even in the RDB, there are 'pure' designs that are subsequently 'de normalized' for performance. This distinction is not unique to OWL/RDF.*

[17] *http://www.w3.org/2008/12/ogws-slides/Chevron_Position_Paper.pdf.*



Chevron's Seisworks.owl schematic[18].

SeisWorks.owl brings the different data types together with a modular approach. Domain knowledge and OWL experience are required to do this well. Technical metadata is stored in an Oracle 11g Semantic Store. URIs are critical, they need to be unique and repeatable – embedding semantics in crawling scripts. Ontology development needed a consultant's help and several semantic web tools. The proof of concept study worked for SeisWorks/Gocad and is being extended to lifecycle management of technical data – tying in archive systems, financial legal and reserves data. Chevron sees the next step as forming partnerships to harvest these 'low hanging fruits...'

Domain ontologies are complex and require commitment from SMEs to build by hand. Automation can help. Ontology management can be hard as the knowledge-base grows. Chum would like an ontology data base management system.

Mention was made of the Open Oilfield Ontology Repository[19], an SPE initiative run by Bertrand du Castel. This has come up against IP issues and the problem of promoting an information sharing mindset. It would be nice to roll in the Schlumberger Oilfield Glossary[20] to enable context-based mining and automatic extraction.

Chum concluded by asking 'What does the semantic web bring to the table that cannot be solved by existing technology?' Chevron found that the semantic web was like a 'souped-up business intelligence system.' How does automated inference help solve business problems? It does require a lot of effort to 'corral'

[18] *Image courtesy Chevron.*

[19] *This is essentially empty and has been since its inception in September 2007 (Oil IT Journal).*

[20] *This looks eminently scrapeable – IP considerations apart...*

metadata – in fact some 70% of the effort was spent on getting the data together. The semantic web is important but real progress is going to require an industry-specific interest group to give critical mass. This raises questions such as the interest group's charter and alternatives such as possible roles to be played by Energistics and/or POSC/Caesar.

Q&A

Bechtel – All this represents a long term investment. But how can you deliver annual value?

By addressing current issues.

TW0816_8 The Cogito semantic platform in ENI – Brooke Aker, Expert Systems

Expert Systems' 'Cogito' semantic processing platform has been adopted widely by ENI. This provides fast search and analysis of text – i.e. 80% of the world's information. Aker compares semantic search to a mosaic – it's when you step back that you see the big picture, how documents interrelate. Web 3.0 turns the web 1.0 metaphor on its head; from 'a few writing for many' to web 2.0 where 'everyone has a voice' (frustrating key word technology) to web 3.0 with 'mass production and pinpoint consumption.' Knowledge is about connection, not collection. Semantic methods include morphological, grammatical analysis etc. For instance there are two meanings to gas in, 'I step on the gas and my car uses gas.' These can be identified with a semantic network, linguistic query engine and development studio. Text processed in this way results in over 90% recall precision. Semantic network contains 350,000 words and 2.8 million relationships including geographical relationships, subject, verb, objects etc. A search for 'China's nuclear energy strategy in 2020' winnowed 25 relevant documents from millions in a collection. The technique is good at capturing weak signals. Sentences in the form of subject action object are stored in a triple store. You can go forward and back – from problem to solution and back to problem. But avoid wasting time on poor quality content.

Q&A

ENI – Are the results good enough out of the box?

A linguist is needed to listen to the query and write the code.

TW0816_9 OWL integration and analysis hub for Oil and Gas – Kendall Clark, Clark Parsia

Healthcare and life sciences (HCLS) is a heavy user of ontologies including OWL. At least one novel drug discovery has been made with OWL. The financial services industry sees the future as bright for OWL as do policy makers and regulators. In defense, semantics are used in product lifecycle data management and in decision support for unmanned aerial vehicle (UAV/drones) flight planning[21]. Enterprise IT is a laggard, but is coming around slowly. The EU is ahead with the technology and is pulling away. HCLS is world leader – led by life sciences. Oracle's OWL product is crucial. China and Japan are also doing well. Heavy industry lags but there are signs of change – particularly with POSC Caesar/ISO 15926. Product modeling (PMO from SWOP) and this workshop are other examples.

The semantic web is good for information integration – it is independent of data sources, it integrates models rather than data sources at a high level with a guarantee of accuracy and logical consistency, including business rules. The semantic web offers decision support with high level concepts, using existing tools and custom code to build analysis applications. The semantic web can be seen as a 'hub' for integration and analysis, an 'empowering feedback loop.'

NASA used the technology for workforce analysis to decide who was the best person for which project. All the information needed existed inside the agency. NASA built a semantic tool last year to leverage this. This led to more integration and more analysis – and a career trajectory planning system. Is this relevant for oil and gas? What new stuff in OWL2 is important to oil and gas?

Q&A

I could make a skill set database in a spreadsheet!

Actually this included a skill database, policy manual, etc. In all, 8-10 disparate data sources. It was a bit more complicated than it looked!

Were you encouraged to use other people's ontologies – or did you start from scratch?

A bit of both. People data extended FOAF[22] and used an existing XML competency schema.

---

[21] *See for instance Project Ilium/MilInfo (Stanford) at http://protege.stanford.edu/conference/2005/slides/2.4_valents_holmes_alvidrez_ProtegeServiceOntology6.pdf - but it is unclear if this was a one-off.*

[22] *Friend of a friend - http://www.foaf-project.org/.*

TW0816_10 Mashing EPSG and WITSML – Jeremy Carroll, TopQuadrant

With all the hype we forget the basics. There are key roles for the subject matter expert (SME), the ontologist, the semantic web specialist, the user interface team and the end user. The goal is to encapsulate the SME's knowledge for reuse by an end user. Until now, technology has not been economically viable. Now there is a 'web style' of cross industry collaboration. TopBraid Ensemble is a tool for doing this. In a demo, Ralph Hodgson introduced the EPSG Geodetic Parameter Dataset of 15 to 20 tables of detailed concepts for transformations. The WITSML standard talks about similar concepts relating to coordinate reference frameworks. TopQuadrant downloaded the EPSG database into TopBraid, as well as PRODML and WITSML and mapped all in OWL. Owl was used to tie the EPSG data to WITSML's codes for ellipsoid, datum etc. The EPSG and WITSML share the same concepts of coordinate reference system, datum and ellipsoid but have a different class structure. A geodetic model in WITSML with EPSG information was generated as a proof of concept. A tool builds a query to access EPSG data for a particular WITSML ellipsoid code.

Q&A

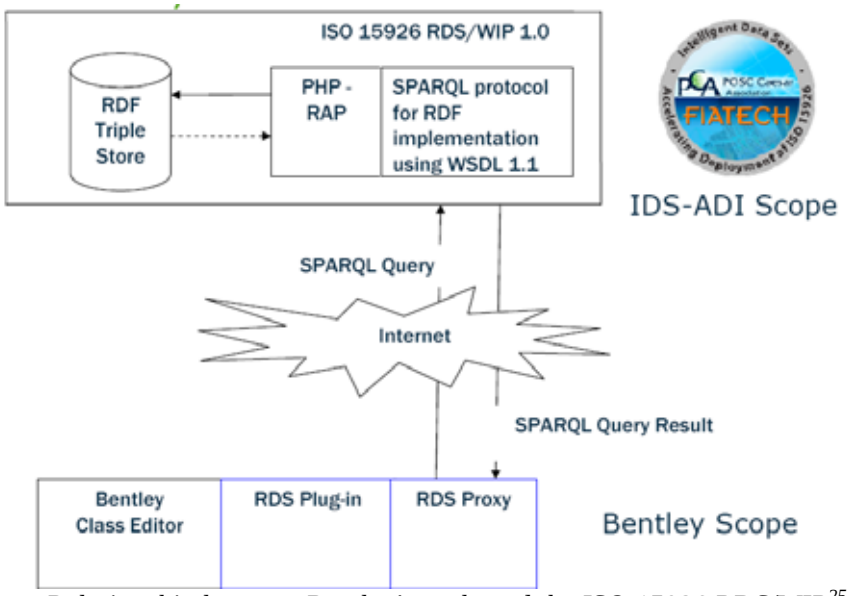
Energistics – WITSML was designed with EPSG in mind.

Schlumberger – Could you have taken the same data and done an ontology from scratch?

This would have had a richer class structure but it was much cheaper to do it this way.

TW0816_11 The Fiatech/POSC/Caesar IDS-ADI – Robin Benjamins[23] (Bechtel) and Manoj Dharwadkar, (Bentley Systems)

The ISO 15926 POSC/CAESAR IDS and Fiatech ADI (15926) projects have now merged into IDS-ADI. Bentley as a vendor of engineering software has been exposed to the semantic web in the last couple of years and has developed a method for accelerating ISO 15926 implementation. The RDS/WIP[24] 1.0 has been established, leveraging 'rapidly evolving' semantic web technologies. Now, developers need to implement this in their in products. For this to happen, we need to have wide adoption of semantic web technologies in oil and gas, for instance to model lifecycle information, and to extract information from existing formats. We tried to represent the ISO data model in XML Schema. This didn't work so now we use OWL, plus Templates, plus ISO 15926 Part2 plus Express. Bentley's product portfolio – Microstation, ProjectWise and OpenPlant require data interoperability with industry standards.



Relationship between Bentley's tools and the ISO 15926 RDS/WIP[25].

The RDS/WIP contains engineering data in an RDF triple store with access provided through RAP - the RDF API for PHP and SPARQL. There is a need for major vendors to support RDF/OWL. Support from Microsoft was described as relatively poor [laughter and cries of 'none at all[26]'].

---

[23] *http://www.w3.org/2008/12/ogws-slides/Bentley.pdf.*

[24] *Reference Data Store and Work In Progress implementation of ISO 15926.*

[25] *Image courtesy Bentley Systems.*

[26] *http://msdn.microsoft.com/en-us/library/aa303446.aspx and http://msdn.microsoft.com/en-us/library/aa303722.aspx.*

Dharwadkar introduced Bentley Components 'iRing' – described as a vision of a grid across suppliers – built on the RDS/WIP and RDF Part 8 as above and on Bentley's OpenPlant ECXML and lots of other stuff.

Q&A

Schlumberger – Why use a proprietary visualization tool for open data?

You need to check out www.rdlfacade.org for open ISO 15926-7.

W3C– What do you gain from SW technologies here? My impression is that RDF/OWL was used to represent something without actually leveraging the semantic web. OWL is used as a syntax for vocabulary – which is fine but ....

Bechtel – This is foundation technology for future work.

IFP – You start with an Express data model – then map to RDF/OWL – its easy in this direction, but not so in the opposite direction because of restrictions in Express.

Fluor – No problem. We use templates – information blocks – and a rule engine to make a two way transfer.

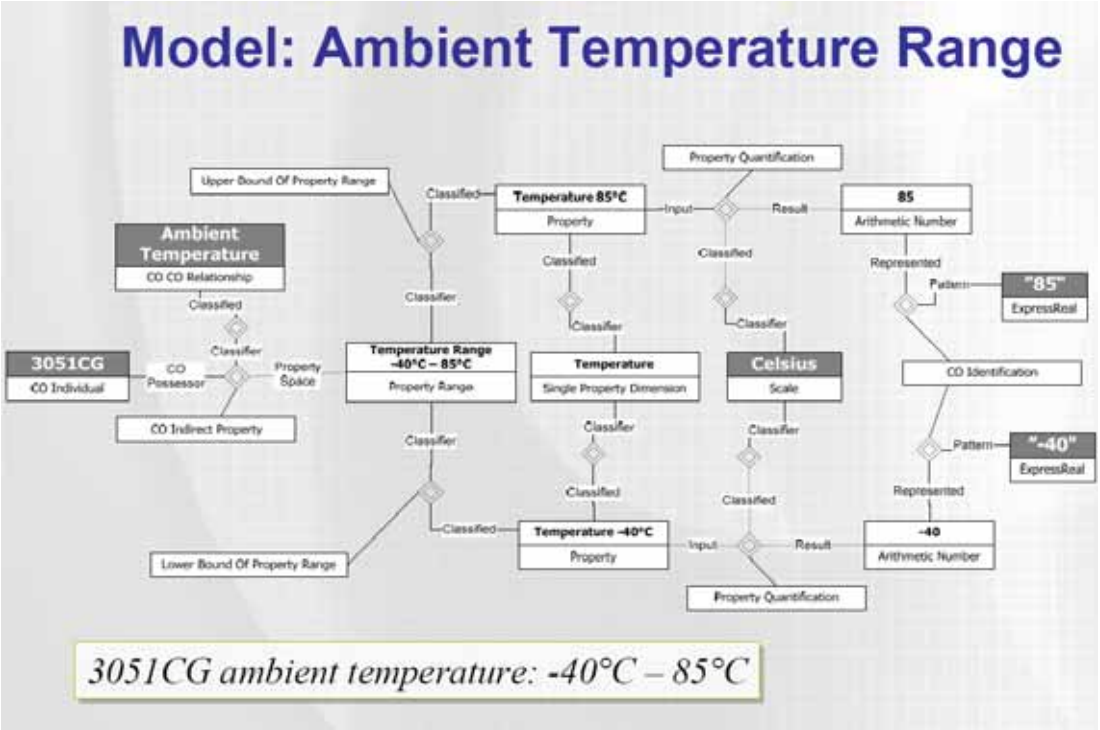
TW0816_12 ISO15926 – data modeling with RDF/OWL – Onno Paap, Fluor and DNV[27]

ISO15926 (see http://ids-adi.org) is an interoperability standard, a neutral layer for data integration. You use you own native data model(s). A few years ago, ISO started making data models for process and plant. This proved unworkable – so ISO made a generic data model – to be populated with 'meaning,' consolidating other ISO standards. ISO 15926 was chosen by FIATECH as THE interoperability standard of choice. Some 50 man years of effort have gone into developing the standard. PCA and Fiatech are collectively spending $3.5 million (most in hours) on some 20 projects. An interoperability workshop in May 2008 was attended by owner operators, BP, Chevron et al and many EPC[28]s. Today's engineers are data mappers – they spend all their time 'yellow lining' documents, checking PDF documents against the database. Paap reckons that the true amount of time spent on managing data is more like 100%! With better data management, 'one engineer could do the work of three!'

Vendors map their systems to a common data model using 'Facades,' i.e. triple stores with an API (including security, 'we don't want our competitors to see our prices!'). Facades are grouped in a 'confederation of participating facades.' A laptop with a facade browser sends SPARQL queries to the 'confederation.' The EPISTLE Reference Data Library (ERDL) Facade also ran. Fluor uses 10,000 equipment vendors on 800 current projects – around 90% are 'mom and pop' shops which will never do facades. They need a simple, compliant interface. Excel is out, and most ontology editors show too much 'assembly code.' Triples are too limited in data description. Need more than just 'subject,' 'relation,' 'object.' So use templates - 'a pattern for facts.' Paap offered a simple example of template use. A car has at least 3 wheels – so it has a **`Parts-at-least(Car, Wheel, 3)`** template.

[27] *http://www.w3.org/2008/12/ogws-slides/Fluor.pdf*

[28] *Engineering and procurement contractors.*



The data model (graph) of a temperature sensor template[29].

More seriously, a Rosemount 3051CG pressure transmitter has an operating range of -40 to +85°C and is represented by the template above. These get very complex very quickly and are the 'realm of rules, not engineers'. A similar approach is adopted for units of measure – see for instance http://rdl.rdlfacade.org/data#DegreesCelcius (actually this is not yet populated).

Q&A

W3C– The template approach is creating problems. If I add more RDF statements that conflict they will distort the whole thing. This is not an open environment. Templates use RDF syntax but there is nothing there to defend against someone who wants to break it[30].

The debate moved on to security. Bertand du Castel mentioned the SAML security standard. Someone else remarked that this was, 'a major faux pas to mention OASIS standard at a W3C meeting.' [laughter].

TW0816_13 What is a W3C interest group? – Ivan Herman, W3C

The W3C has three types of collaborative unit. At the top level is the Working Group (eg XML, HTML) that actually creates a standard. At the other end of spectrum is the Incubator Group that explores a certain area with a short term charter. There is no standards development, just a report that is published on the www.W3.org website. In the middle is the Interest Group. Again there is no standard produced, just Notes, best practices etc. Interest Groups include URIs and RDBS to RDF which is currently underway with Oracle as chair. Roger Cutler added that you do need to join the W3C to work on these groups. For a large company, one of the size required to influence the W3C, this costs $60,000 per year with a three year minimum commitment.

TW0816_14 Knowledge system for integrated operations – David Norheim, Computas

Computas specializes in expert systems and artificial intelligence – and developed the Active Knowledge System for Integrated Operations (AKSIO). This was designed to avoid Norwegian operators repeating errors and also to help with the 'big crew change.' AKSIO is an 'active socio-technical system for experience transfer in drilling,' sponsored by StatoilHydro. AKSIO regards a drill rig crew as a FOAF network. Experience transfer within the crew is OK, but weak across multiple drilling operations. Computas developed 'experience reports' including semantics to screen and annotate knowledge and to build a

[29] *Image courtesy Fluor Corp. and DNV.*

[30] *In conversation Paap told us that the templates concept was based on RDF best practices for n-ary relations developed by Stanford's Natasha Noy – see http://www.w3.org/TR/swbp-n-aryRelations/. But this debate is revealing of what could be a deficiency in semantic web technology as there seems to be just too much interpretation involved in mapping even relatively simple concepts.*

searchable, knowledge base. This can be filtered on discipline, operation, equipment state, etc. The AKSIO drilling ontology (in OWL-DL) was created by subject matter experts (SMEs) and knowledge engineers using 'question driven query scripting.' Incident reports are routed to SMEs for screening and annotating with the domain ontology. The result is an increased rate of knowledge reuse, good take up of best practices, avoiding repeating one's mistakes and process improvement. Computas is also involved in the 'Agile Process for Integrated Operations' (APRIO) project. This sets out to improve work processes with OWL-DL for BPMN[31] (RDF)/BPEL[32]. Computas has developed a business process definition metamodel and ontology. RDF, Jena, Pellet reasoner and Java GUI also ran.

Q&A

Are you going to sell or share this ontology?

Yes – absolutely.

Chevron – Send it to us, we'd love to see it!

What is the url?

I'll put it out.

Chevron – The Open Oilfield Ontology would be a suitable home.

Schlumberger – No it should be kept and maintained by the owners.

We have tried not to compete with ISO15926, but we did add processes that did not exist in ISO1596.

This kind of system requires extra work from employees which can be hard. What is your experience?

Yes, it is very hard even though it's their role. Some don't want to share or publish best practices. There is a large 'human' component in the knowledge management process. Capturing these reports is a mandatory activity when there is an exception to planned operations.

Does the ontology hold words or definitions?

It does not hold definitions in the OWL sense, but there are some explanations.

TW0816_15 Ontology building – Jan Rogier, EPSIS

EPSIS has combined two approaches to ontology building, developing a plug in to Protégé to show alignment of two ontologies – gaps and overlaps. The top down approach as exemplified by ISO 15926 contrasts with a bottom up approach developed by EPSIS in a CAD/AI project. The two approaches have been combined in EPSIS 'ERA Decide' tool. ERA Decide is being trialed in a real-time decision support system for reservoir and production engineering. This is a component of the Norwegian Integrated Operations in the High North project where 'the intention is to demonstrate use of semantic web technologies in oil and gas.'

TW0816_16 The challenge of data overload – Jennifer Sampson, EPSIS

We have terabytes of data, multiple tools and resources but we are not seeing a proportional increase in decision quality. The solution is 'better informed' decision making, being able to answer questions like 'where is the data coming from'; 'didn't we have this problem last year?' There are two possible approaches. The first involves an 'upper ontology' of knowledge representation – which should be open and non proprietary. The second is an 'experience and knowledge-based ontology' that captures experience close to the application. This is 'dynamic, less structured, competitive and proprietary.'

The Environment Web project was created by adding new terms generated by subject matter experts to ISO15926. Sampson showed the example of an acute oil discharge to sea – this is a class of compound in ISO15926 and leads to a horrendously complex data model for the top down approach. The other way, bottom up is exemplified by the 'BRAINS' system, a crime investigation tool developed for the Dutch police. This leverages text mining and 'semantic pattern recognition.'

Q&A

Chevron – We see little value coming from the upper ontology in our Integrated Asset Management trials.

---

[31] *Business Process Modeling Notation – an OMG spec - http://www.bpmn.org/ and http://en.wikipedia.org/wiki/BPMN - unclear as to relationship with RDF and semantic technology.*

[32] *Business Process Execution Language.*

You can use of parts of the upper ontology but it is hard to show concrete benefits.

Chevron – Actually we use both in our enterprise architecture. We get data sharing from the upper ontology.

You do get value but this comes at the expense of extra effort.

ASME/ASCII codes are ontologies.

You have to promote the ontology of common language before domain-specific ontologies.

TW0816_17 Panel Discussion – Raphaele Henri-Bally (Total), Danny DuCharme (BP), Roger Cutler (Chevron), Richard Sears (Shell), chair Ivan Herman (W3C)

Shell – As a humble geophysicist, I know how to spell html, witsml, prodml but have not yet shared the 'existential ontological experience.' Technical papers from the business side tend to be indecipherable and you can say the same about the IT side. A key question is how to 'marry' the jargons.

Total – We have to cater to the new generation and share knowledge – help the next generation catch up.

Chevron – There are a couple of areas where we could cooperate. We have a prima facie case that this technology has use in industry. In Chevron we have taken some baby steps (except IAM which is quite a big step!). We have solved some problems and would like to talk to others about their baby steps. If major capital projects expect 15926 then we had better agree on it. Only Wall Mart can force its vendors to align with its own specifications.

W3C – The problem of a large company talking to many suppliers is not specific to oil and gas – what is really specific?

BP – We started down the path of SOA and don't seem to be getting huge benefits. So here we are checking out the next wave.

Schlumberger– Deepwater and harsh environments mean that many new upstream ventures are costly in terms of personnel and require new technology that goes beyond human capabilities. All of which points to a long term move to automation. Our challenge is to articulate our E&P in such a way that they are amenable to automation.

Chevron – Our initial wins are very prosaic and infrastructure-based. The 'secret sauce' of IAM is that it can grow without getting top heavy. But we are a long way from reasoning through text to find biological pathways.

Shell– There is a difference between exploration and production. Exploration thrives on story telling and subtle inference from geological models. Here the semantic web will help. The problem is currently solved in Shell by moving people around the word all the time.

Chevron – It's like telling stories around the campfire.

Shell – That's what the semantic web is about.

Fluor – From an engineering perspective it is clear that we are losing money from the lack of interoperability. In operations and maintenance, there may be many changes in software over time – but the data model should remain the same. We have achieved this with ISO15926 and are also getting interoperability between engineering partners – for wells, pipe, coating etc. All now have UIDs that stay the same through manufacturing, construction and operations – in a 50 year plus lifecycle. We spotted the semantic web as appropriate. It will become 'mainstream' if we use it and there will be tools available.

Total – Engineering has proved the semantic web's ROI. For G&G, we compete initially but beyond that phase there is money to be had through cooperation and sharing.

W3C – Is ontology free of charge? Of course it's free. In the semantic web world lots of information and data is shared and reused. This is part of the culture. This is what happened in HCLS, although people were reticent at first. Is this a problem for this community?

BP – You touched on some of the limits to sharing. We are OK for sharing ontologies and how we manage our data – but not how we do our business.

Energistics– This is a continuum. We can share methodology or build ontologies which when defined will create an incentive for sharing. But as we get into the data itself, there will be no sharing.

Schlumberger– We do share a lot of stuff through publications like the Leading Edge, the SPE etc. There is a good argument to be made for publishing this material in a machine readable/understandable form.

Chevron – Granular access control is very important to us – and a problem with the semantic web. Why?

W3C – Yes, HCLS had this problem but they also have huge public databases available.

Sem web is not 'bad' for access control.

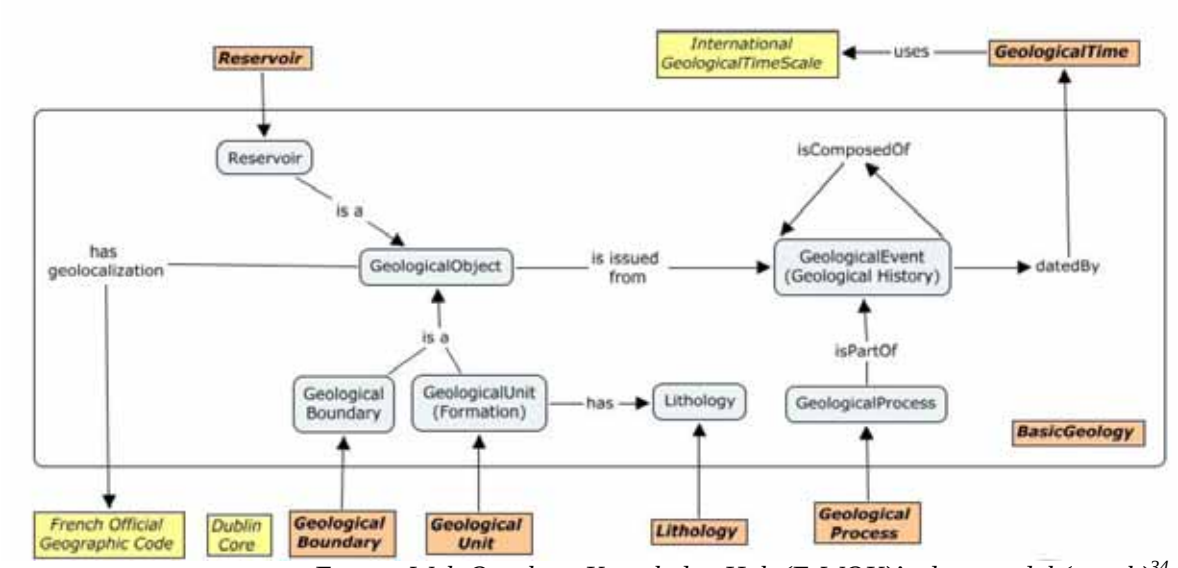
W3C – The web culture at large favors data sharing. Today there are ontologies and billions of triples in the public domain. It is natural for applications to go after this. A great example is DBPedia[33].

W3C – Intel is very interested in the semantic web and they don't share data. Semantic web could provide even better atomic level protection. Others – IBM, Oracle, Disney, Pfizer and Merck are not in it from academic interest.

Energistics– I'm impressed with today's semantic web which can clarify subtle differences from myriads of opportunities. Subtle semantic differences get in the way of automation.

Total – The semantic web can help you find information that you didn't know existed.

TW0816_18 The semantic web in earth sciences – Jean François Rainaud, IFP



Energy-Web Ontology Knowledge Hub (E-WOK)'s data model (graph)[34].

Rainaud's first use case is the Energy Web Ontology Knowledge Hub (E-WOK). This uses semantic web technology to perform intelligent document search in the context of a C02 storage project. E-WOK is a joint venture between INRIA, ENSPM, BRGM, EADS, IFP and CRITT-Informatique. The idea is to annotate (tag) documents with semantics. This allows for accurate retrieval of a document fragment. The second use case involves earth modeling for geological site qualification. This is similar to the above, adding interpretation management for reservoir characterization. The technology stack includes INRIA's Liferay portal on top of PERLS, Orchestra, JorA, RDFS/OWL, CORESE and SPARQL. A services-oriented architecture builds on ESB (Petals), Liferay (portlet), SAWSDL and persistence (storage) with OntoDB and OntoQL with Sparql. The method starts with a domain ontology. Documents are then annotated semantically, describing the meaning of the document. This works fine for text, the team is currently working on database and technical data and on data management. Queries run against annotations only.

The tagging process uses FASTR/ACABIT to extract significant words which are conceptualized in a domain ontology[35]. This re-used Dublin Core and the IGN's geo-ontology (itself inheriting Geon, NADM, and GeoSciML.) Interaction with a subject matter expert is required to define concepts. Concept Maps and ECCO, an INRIA tool for term extraction from text including natural language processing, also ran. Other facets of the process include grammatical analysis, semantic annotation, training systems etc. Lots of manual cleanup is required. The system uses http://www.owl-ontologies.com/geoTime.owl#Cenozoic (endpoint not yet populated?). OWL is used to describe data modeling of geological formations including formation boundaries, faults and fractures.

The second project is a test bed for interpretation workflow modeling. Different models and workflows need to be semantically annotated for integration. Interpretation objects result from interpretations include – line, horizon, surface. These are related by predicates between objects – 'older than,' 'member of,' fault F3 'cuts' fault F4. The project worked across seismic, well and structural geology ontologies.

---

[33] *This is an RDF database of the contents of Wikipedia. DBPedia contains 274 million pieces of information (RDF triples).*

[34] *Image courtesy IFP and the E-WOK project.*

[35] *In this context a reference to LUCENE is also appropriate - http://en.wikipedia.org/wiki/Lucene.*

The idea is to be able to answer questions such as 'What diagenesis has affected the Bathonian in Paris Basin?' How is this done? Diagenesis is defined in E-WOK's reference set. The reference set also contains the facts that Bathonian is inside 'Jurassic,' 'Secondary' and 'Mezozoic.' The system also knows about Jurassic stages like 'Comblanchian.' A similar approach is used for geographic entities; the 'Paris Basin' is defined by a polygon, along with other entities like 'Ile de France,' 'Départment du Loiret' etc. Set logic is used to find the intersection of the different reference sets. Rainaud concluded that 'E-WOK is a possible solution for the identification of CO2 storage sites.' Unlike regular search it offers 'goal-oriented' search. The team is now working to 'humanize' SPARQL and to finalize the demonstrator.

Q&A

Chevron – We are faced with similar issues. What is your experience of upper ontologies?

We are finalizing the mapping across the different time tables.

Schlumberger – What is CORESE?

The COnceptual REsource Search Engine, a graph oriented tool developed by INRIA.

W3C – Also a rule engine for SPARQL.

TW0816_19 Ontology-driven EOR system – Emilio Nunez[36], University of Texas at Austin

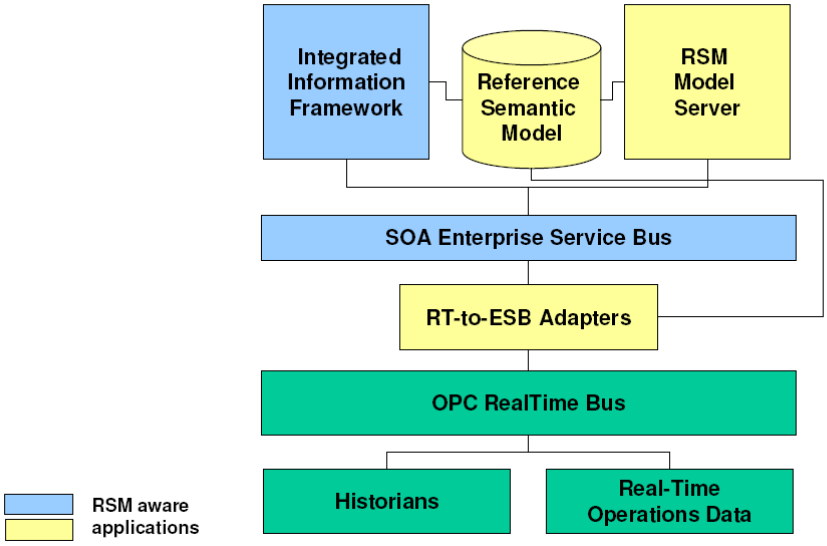
This Schlumberger-backed project is described as 'Towards an ontology-driven enhanced oil recovery (EOR) decision support system.' The University of Texas at Austin (UT) has a lot of EOR expertise and a 20 company strong EOR joint industry project. We know that industry needs help in EOR decision making. An ontology is often just the beginning – which is exactly where we are. There are no applications as yet. The EOR screening process was derived from four summary slides showing EOR methodology vs. depth and other factors. A subset of the Toris database[37] of reservoir properties was imported into Protégé. SWRL rules, the Expert System shell (all inside Protégé) were populated with 'hasEORMethod' property according to depth, porosity, permeability etc. The result was a simple ontology. For surfactant screening, Nunez interrogated the best guy in the chemical department and captured the results in C-Maps. This included lab process, core flood experiments etc. Many elements were not readily translatable to Protégé. Workflow concepts were added with El Paso's Workflow Driven Ontologies – 'isInputOf,' 'HasSuccessor' etc. These were extended to EOR with chemicals. This was a complex process that helped organize concepts. The thing is starting to look like a workflow-based ontology. Still a cork in progress. An example involved the simplified EOR recovery calculation (from 5 pages in Professor Lee's book). This diagram was quite hard to read. The calculations were translated into C-Maps and this process revealed errors and insights. The information that was previously available only to students is now open to software agents. Another workflow involved scaling up uncertainty, a complex process involving averaging porosity etc. with geostatistics. SQWRL came along in Protégé just in time. Some results of this work were presented in the SPE 109628 paper on 'Risk-based decision making and assessing the value of oilfield sensors in EOR.' Nunez concluded with another plug for Protégé and C-Maps.

TW0816_20 Upper Ontology for process industry – Udo Pletat and Vishwanath Narayan, IBM

This un-presented paper outlines IBM's research into mashing-up various chemicals and process standards like ISA88, ISA95, MIMOSA and ISO 15926. The resulting Reference Semantic Model (RSM) leverages an 'upper ontology' that forms to basis of an 'operations control and event processing system for oil and chemicals.'

[36] *'A year ago I did not know how to spell ontology'.*

[37] *The US DoE's Total Oil Recovery Information System http://www.netl.doe.gov/technologies/oil-gas/publications/brochures/TORIS.pdf.*



IBM's Reference Semantic Model for oil and gas production operations[38].

The above schematic includes the RSM Model Server, a UML application based on SISCO Inc.'s Utility Integration Bus (UIB) and IBM's Integrated Information Framework (IIF) – a Rational application for 'graphical modeling and monitoring of oil and gas equipment and assets.' The RSM is also available as an OWL application.

TW0816_21 Oil and Gas Semantic Web – Total's position paper

Total's position (not presented) noted potential application of semantic web technology in upstream, downstream, new energies and petrochemicals. Total's interest is in cross discipline corporate metadata and intelligent search. The expectation is that 'semantic browsers (RkbExplorer, Simile) will allow us to discover unexpected relationships between data.' Other areas with potential application include a common vocabulary, workflow and modeling although the impact of W3C technologies on the latter presupposes that 'innovators within companies must convince the operations of the challenge and include them in this loop.'

TW0816_22 Panel Session

Chevron – We met our expectation of learning about this space. It has been a successful meeting. What's next? Is the semantic web on the verge of valuable application? Yes. Are there real opportunities? Yes. There are opportunities to collaborate. This has been achieved before and this approach is non competitive. We have demonstrated an interest in the technology. But still there is the question, so what? Everything has to be in triples! Here is a question for Ivan, you shared the example of HCLS but what other SIGs in W3C are relevant to us? What of other industry efforts?

W3C– There is not much on oil and gas in the W3C right now. We started working on XBRL but a separate organization is now working on this. We established contact with them to use semantic web technology. Another e-Government group is looking at how technology can work in government. There has been informal contact with lawyers with the idea of making legislation understandable. So HCLS is currently the most mature group.

Bechtel – We are definitely a semantic web implementer. The challenge now is to collaborate. The next step should be the formation of an oil and gas semantic web group. Other users of W3C technology may have solutions to things that we are trying to do.

W3C– Feedback is important, but the HCLS group showed that W3C's traditional circles are big IT companies. Smaller companies will reap the benefits – half of the exciting new products are made by small start ups.

Shell – I have enjoyed the conference – thanks to Chevron. Vendors are not here today so we don't know where they are going. Where do we go today? The oil patch is traditionally outside of the W3C even if we gather all the converted, we won't have the whole of the upstream in the W3C. A fully-fledged ontology won't be done in the W3C. However we can do something like an ontology restricted to a single domain. Shell has some semantic web activity in the Netherlands.

[38] *Image borrowed from http://www.w3.org/2008/11/ogws-papers/pletat.pdf.*

Energistics – It is clear that there is a lot going on. Projects are at different stages, going in directions, old stuff is adapting to the semantic web. As Ivan said, we need information exchange, we need to identify areas in the W3C space[39] for education and to focus on an achievable target. What players need to be skilled and trained – what is achievable and amenable to pragmatic early delivery.

Chevron – We see this as game changing technology, it definitely has value. See the HCLS page on W3C as poster child – http://www.w3.org/2001/sw/hcls/.

A wide ranging debate ensued on future directions for the semantic web in the oil and gas vertical. One possibility is a W3C Fellow to build up a charter. This would require several oil companies or major service providers joining the W3C. An alternative suggestion that Energistics should join the W3C was well received while falling short of the endorsement required for a Fellow.

TW0816_23 Additional notes

For insight into the notion of an Upper Ontology see the excellent article on Wikipedia – a must read. And for a contrarian view of the semantic web and an account of the 'religious wars,' see http://www.nodalpoint.org/2007/11/30/burn_semantic_web_burn. Also of note, the semantic-mediawiki.org, and dbpedia.org. Finally for more on the HCLS see http://www.w3.org/2001/sw/hcls/, http://www.slideshare.net/fbelleau/bio2rdf-presentation-at-www2007-hcls-workshop, snpedia.com and on Parkinson's disease research on http://bio2rdf.org.

TW0816_24 The Data Room – Technology Watch subscription information

The Data Room
7 rue des Verrieres
F-92310 Sevres France

Tel (USA)	281 968 0752
Tel (UK)	020 7193 1489
Tel (France)	+33 1 4623 9596
Fax	+33 1 4623 0652

Technology Watch Home Page info@oilit.com

[39] *This interesting comment begs the question as to why the W3C conference was ring fenced to the semantic web. Other technologies (SOAP, XML etc.) in the W3C's armory are also relevant to the upstream.*