



Semantics are for people too!

Lee Feigenbaum, Cambridge Semantics
W3C Workshop on Semantic Web in Energy Industries
09 December 2008

About This Presentation

- This presentation is extremely visual. Please use the included speaker's notes for the context of the imagery within.
- Questions to: lee@cambridgesemantics.com

Why are we here?



December 9, 2008



3

- + Why are we at this workshop?
- + What are we hoping to get from it?
- + What are we hoping to contribute to it?

Why are we here?



December 9, 2008



4

Most important reason (with homage/apologies to Vanilla Ice)

- + Vendor - SemWeb expertise in applications in enterprise software
 - + no significant O&G / Energy exposure
 - + probably similar to other vendors here
- + Listen
 - + O&G challenges
 - + Specific use cases, experiences
- + Collaborate – do things for real
 - + With industry partners
 - + proofs of concept, pilots, production deployments to use these technologies to solve problems
 - + With other vendors
 - + prove out the point that open standards enable cross-vendor solutions
 - + take advantage of multiple vendors particular expertise focus in this technology hegemony

Photo credit: <http://flickr.com/photos/wonderferret/2900631165/>

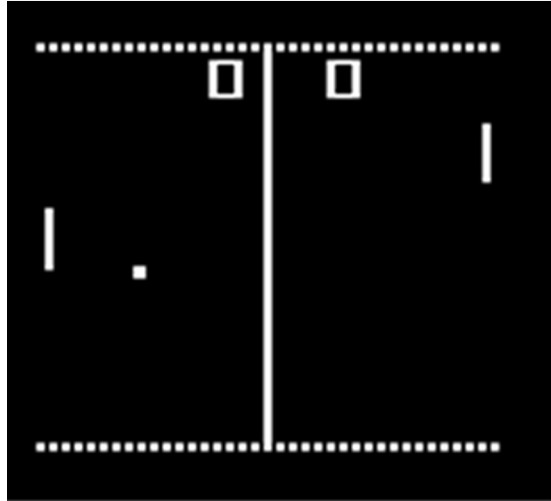
Why are we here?



2nd reason - talk about Cambridge Semantics's position

- + what's the underlying world view that unites the 9 Cambridge Semantics's employees?
- + how might that view be applicable to O&G
- + why do we care about semantics (web technologies) in the first place?

Why are we here?



December 9, 2008



6

Final reason –

+ To demonstrate some novel & interesting software in the context of an Oil & Gas scenario.

Man or Machine?



December 9, 2008



7

Semantics (semantic web technologies) are often characterized in terms of what they enable for machines.

“make information machine-readable”
“infer new relationships in a knowledgebase”
“enable (automated) data integration”

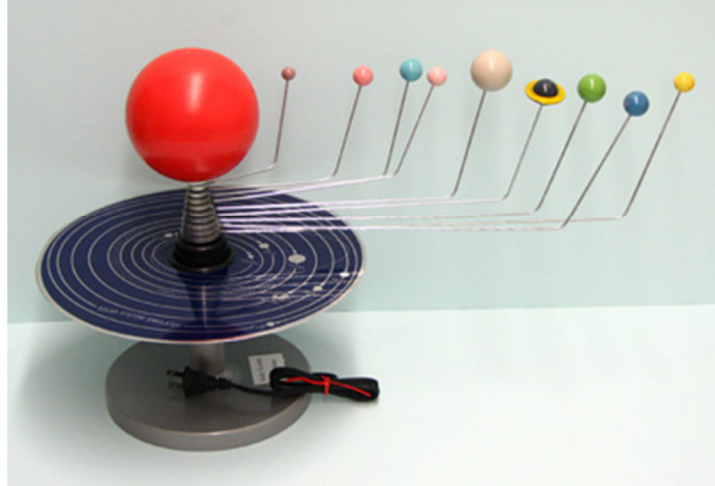
- + these end up benefitting people (of course)
- + making use of automated agent/analytics software
- + finding otherwise unknown answers

...but at their core, these are capabilities of semantics that rely on some degree of *machine automation*.

And of course there are other “machine-centric” things to do with data (might not be semantic):

- + optimization algorithms
- + search / query (fast! → relational database)
- + ...and more, we’ll look at one other example a bit later

Models we understand



December 9, 2008



8

But there are “human-centric” reasons to like semantic web techs as well!

+ Modeling a domain using semantic technologies – and then using software that relies on that model (ontology) – allows us to create software that speaks to people (SMEs) in the language and with the mental pictures that are familiar to them. There are two main reasons for this

+ Technical reason:

- + deal with “information” rather than “data”
- + the flexibility of the graph model
- + expressivity of RDFS and OWL
- + ...models can often be closer to reality – closer to how people think about the domain
 - + (compared to relational DDL or XML Schema, for example)

+ Social reason:

- + building an ontology is a “purer” form of modeling
 - + building a DB model is about modeling the domain AND defining storage structure
 - + building an XML model is about modeling the domain AND defining a wire serialization
- + concerns about the latter often trump concerns about the former
- + also: enable both top-down and bottom-up modeling

We can't automate everything.



December 9, 2008



9

Why does this matter? This matters because we can't automate everything.

- + Real world
- + Real data
- + Legacy software
- + Opaque document formats
- + ... silos!
- + The solution today is person-power
 - + hours, days, weeks spent getting information into the right machine-centric places & tools & forms
 - + RDBMS for storage and query
 - + DW/DM/OLAP cubes for BI analysis
 - + XML documents for interchange
 - + tedious, time consuming, increase cycle time in decision making

All this largely because of the impedance mismatch between how an expert views his/her domain and how software (machines) does.

That's why we love spreadsheets



December 9, 2008



10

- + Excel is easygoing & human-centric
- + It lets us put whatever we want into it
- + We can shape the info however we want
- + Labels, colors, formulas, etc.

Excel is popular for data analysis, but it's **really** popular for communicating data to other people (sometimes to ourselves).

Via:

- + email
- + doc servers
- + portals
- + etc.

Of course, this often results in a complete Shadow IT system – information that works for the people that have access to it, but is ungoverned, not discoverable, can't be used for interchange or inference or query, etc.

Where does this leave us?



- People use spreadsheets because they're easygoing.
- Information needs to end up in strict formats for technical reasons (XML for interchange, relational for storage or query, ...)
- We work well with semantic models that operate at or near the same conceptual level that we operate at

Which—to us—begs the question[1] of how can we bring semantics into Excel in such a way that it's easy to do the things we want to do with data.

That's what Cambridge Semantics has been after with Anzo for Excel.

[1] That's not what "begs the question" actually means, but it's how it's always used and as much as I'd like it to be, there isn't room in a 30-minute presentation for a linguistics diatribe.

Enough with the boring stuff



December 9, 2008



12

Enough of build-up. What I've said so far is our position at Cambridge Semantics, and while it's a position that is true across many industries, it's particularly applicable in an industry like O&G that:

- + is very dependent on raw data
- + benefits from diminished cycle times in fixing problems and optimizing production
- + engaged in many cross-company / cross-organizational partnerships O&G.

So we've been working with our friends at Chevron (Frank, Roger, David) to attempt to demonstrate our position in the context of some O&G scenarios.

<http://flickr.com/photos/joshme17/1557627176/>

The Scenario



December 9, 2008



13

- + drilling platform out in the ocean – likely a joint venture operated by various organizations
- + ownership stakes in the platform by various (different) organizations as well
- + daily production reports—Excel spreadsheets—are sent from the platform to the overseeing companies
 - + we want to easily share this data (with people)
 - + we want to get this data in a form that's easily analyzed (e.g. monthly roll-ups, but also more sophisticated tasks such as rules- or reasoning-based tasks for detecting potential production problem states, optimization)
 - + we want to easily share this data (with other software)
 - + we want to address the general problems of Excel as Shadow IT – governance, query, management, discoverability, accuracy (single version of truth)

From <http://www.reuters.com/article/pressRelease/idUS122850+25-Sep-2008+BW20080925>

- + Excel is the single most-popular production data management application – but all that data ends up being unmanaged, scattered on different people's hard drives, ...

PRODML & WITSML

```
<productVolume>
  <name>PRODML Flow 1</name>
  <installation kind="field">FLOW-1</installation>
  <kind>PRODML optimization</kind>
  <periodKind>report start</periodKind>
  <dateStart>2005-10-26</dateStart>
  <productFlowModel uidRef="PFM-1">Flow #1 Network
  Model</productFlowModel>
  <facility>
    <name kind="compressor">Compressor Bank A</name>
    <facilityParent1 kind="field">FLOW-1</facilityParent1>
    <unit uidRef="A">A</unit>
    <flow>
      <kind>gas lift</kind>
      <port uidRef="port-2">2</port>
      <direction>outlet</direction>
    ...
```

December 9, 2008



14

One thing in particular we wanted to look at was PRODML. Industry standards for interchange of well & production data. From Energistics.

This is one of the machine-centric destinations of our human-centric Excel data. Another might be a database tuned for optimization algorithms. But we also wanted to show the sort of flexible, ad-hoc views we get “for free” (i.e. can be put together on-demand in the matter of days or sometimes even minutes & hours).

“”“”

The objective of PRODML is to be a low cost, low risk, and highly innovative environment for the configuration and running of advanced optimization processes.

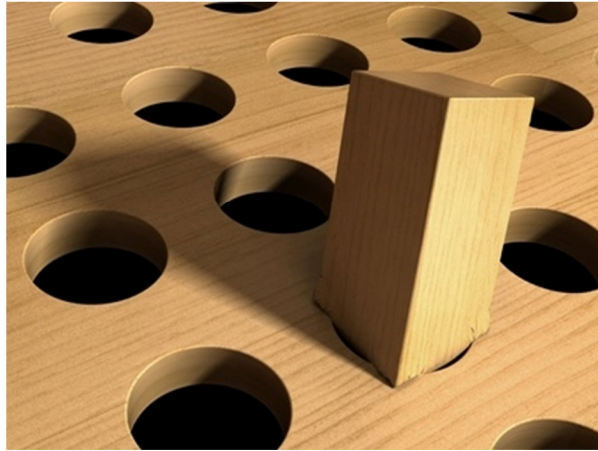
...

In August 2005, a group of energy companies, software and service providers, and an industry standards organization launched an initiative focused on helping producers independently optimize their oil and gas production by improving data exchange and work process efficiency.

“”“”

Prodml.org

From Excel direct to PRODML?



December 9, 2008



15

What would we have to do to map directly to PRODML / WITSML?

This is an industry-standard for data interchange, and more and more software will emerge that is built on top of it. Bad idea not to embrace it into any solution that deals with production data. That said, this goes hand-in-hand with what we saw earlier:

- + Databases are optimized for storage & analytics (machine-analytics-centric)
- + PRODML/WITSML are optimized for interchange between software (machine-interchange-centric)
- + Ontologies are (can be) optimized for conceptual representation of the domain – human-centric

""""

In other words PRODML Version 1.0 leans toward general functionality rather than performance or ease-of-use. It is hoped that this initial version has struck a balance appropriate for a foundation layer of an industry standard.

""""

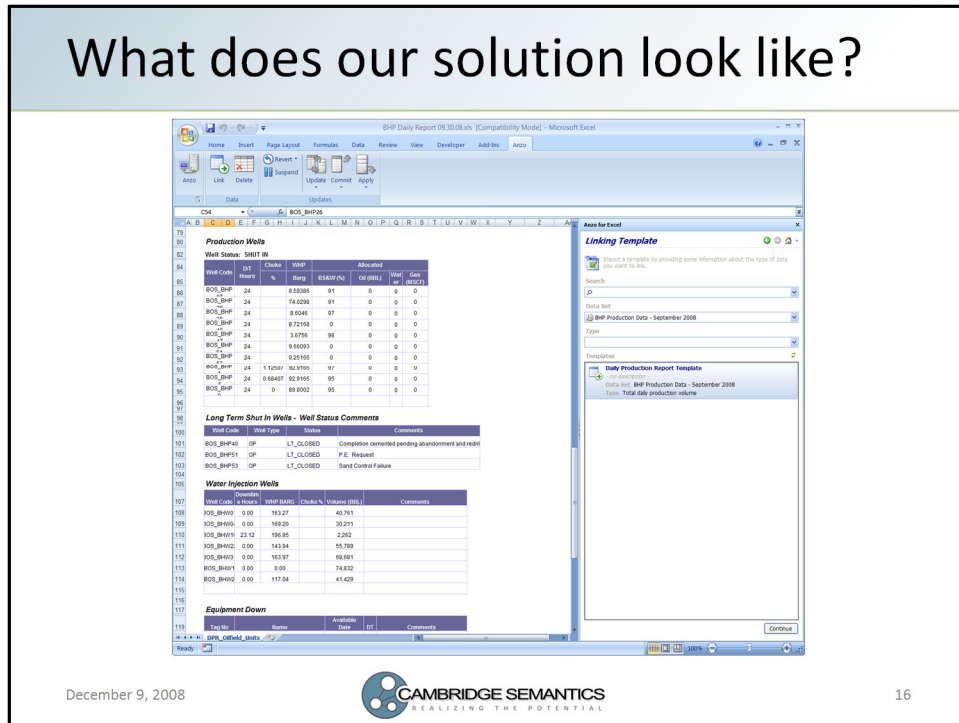
(from
<http://www.prodml.org/prodml/NewsBot.asp?MODE=VIEW&ID=666&SnID=662191862>)

(Mismatches: id refs, coding schemes, facility1, facility2, ...)

(Example of mismatch between human-oriented daily production report spreadsheets and expected PRODML XML serialization.)

- + round peg in square hole

What does our solution look like?



December 9, 2008



16

Demo (please contact lee@cambridgesemantics.com to see this demo live):

- + Spreadsheet comes in
 - + We're using Excel with the Anzo for Excel plug-in
 - + First time, need to link the spreadsheet data to its semantics. [We've created a small ontology for this demonstration, but we could also pull in ontologies from the various other projects that we're hearing about today & tomorrow]
 - + This is how we visually link our spreadsheet data to the semantic concepts. Note that our ontology thinks about the domain in much the same way that our spreadsheet did – "human-centric" if you will.
 - + This just has to be done the first time, so like a good cooking show, we'll skip the rest of the labor (~5 minutes work), and show the common case
- + Spreadsheet comes in
 - + Choose a linking template to be used to link the spreadsheet data to its semantics
 - + Now that we've done that, we've pulled out all of the information from this spreadsheet into a format that can be reused all over the place (in our case, into an RDF store, but this could also back onto an existing relational database)
- + To the Web
 - + Anzo Exposé lets us slice & dice this information on the Web
 - + Here we can see a table of our daily reports over the past year
 - + This is also a faceted browsing view, allowing us to filter our data along any criteria
 - + Point out the data from the spreadsheet that was just linked
 - + All of this runs on the Anzo Server, we've implemented a domain-specific module that can publish PRODML from our semantic model - point out that our Web view contains a link to the generated PRODML from our spreadsheet
 - + other flexible, extensible views:
 - + maps, timelines, details, charts
 - + Exposé allows end users to create & extend these views at will – e.g. I don't seem to have a date filter, but we can easily add one
 - + Add filter, show in action

What did we just see?



December 9, 2008



17

Excel spreadsheets – human-centric

Semantic model – human- & machine-centric

PRODML – machine-centric

Faceted browsing – human-centric UI driven by machine-centric data storage

Hub & spoke model of work makes it easy to do these things – weeks instead of months, minutes instead of days

Thanks & Questions

- Thanks: Frank Chum, Roger Cutler, and David Shipley
- Questions: lee@cambridgesemantics.com

December 9, 2008



18

Thanks to Frank Chum, Roger Cutler, and David Shipley for helping develop this particular use case. Happy to take questions. Please contact lee@cambridgesemantics.com for more details.