PETRIS

A Semantic GPS:

Navigating Data in the Geotechnical World

Jeffrey W Pferd, PhD, Senior VP for Product Management, Petris Technology, Inc., Houston (speaker)

and Eric P. Déliac, DSc, President of Eastern Hemisphere Operations, Petris Technology, Inc.

---

1. The Problem

Introduction

The Exploration and Production community is swamped by glossaries and standards, resulting in complex problems of data definition and the multiple classifications of energy industry information (taxonomy). Besides this, new definition efforts for transport tagging are using languages such as WITSML to record information. Existing archival standards models such as PPDM and PODS continue to be enhanced as new definitions emerge. And as interest increases in knowledge management, the vast amount of unstructured information residing in documents, emails and images must now also be considered and managed.

Emerging technology can combine and deliver information from all these disparate sources. Adding the geospatial component and a portal-based presentation to this medley results in a situation where data integration is ultimately dependent upon a sound foundation based on meaning. Navigation through this reservoir of knowledge requires the waypoints of semantics to meet the challenges of finding the right data for key business purposes – in essence, a "Semantic GPS". This paper provides a perspective of the challenges and opportunities that these new capabilities place before us.

Business Drivers and Industry Realities

The energy industry is becoming more technical and specialized everyday, while industry demographics are making this knowledge, expertise and insight harder to access and preserve. Real challenges concerning knowledge face the industry. Controlling the cost of production, reducing technical and political risk and meeting stricter regulatory requirements stand out as prime examples that rely on the collective and specialized knowledge of the enterprise. At the same time, the management of massive volumes of primary data with fewer people and shorter project cycle times demands that information technology tools be applied to do more with fewer resources.

Successfully meeting these challenges with people, processes and information technology yield real bottom line results. Increased staff and equipment productivity and increased accountability compliance to regulatory bodies are just a few of the benefits. All these increase corporate value with lower risk and have real bottom line value.

2. *Structured and Unstructured Information*

Relationships

Access to knowledge is key to the solutions envisioned. That knowledge is assembled from multiple information components that are present in structured and unstructured sources (Figure 1). How do these sources 'fit' together when we have the opportunity to produce a search result that draws from both types of repositories?

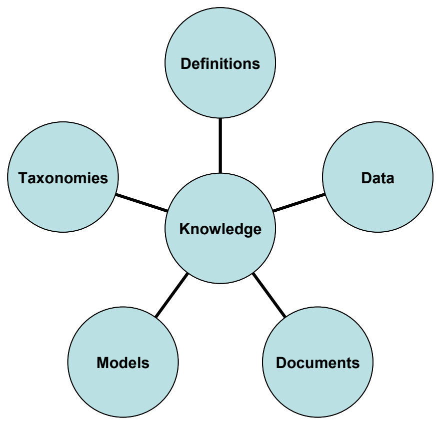

Figure 1 - Components of Knowledge Representation

Figure 2 represents the parallel paths to knowledge and meaning that must take place to join these two rich sources of information.

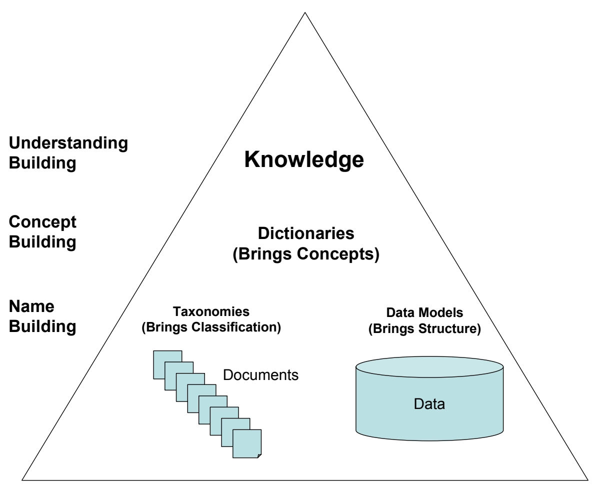

Figure 2 - Unstructured and Structured Information

Considerable effort has been spent at the base of this triangle developing the attribute and taxonomic naming conventions for each source. Much of this effort was expended well before computerization to establish definitions for the geoscientific and engineering domains. Commonly recognized online collections are the

Petroleum Industry Data Dictionary (PIDD), the PPDM Glossary and the Schlumberger Oilfield Glossary. These are useful resources and have been built upon notable physical publications such as The Concise Oxford Dictionary of Earth Sciences[i], Dictionary of Geological Terms[ii], Encyclopedic Dictionary of Exploration Geophysics[iii], A Dictionary for the Petroleum Industry[iv]and Dictionary of Petroleum Exploration, Drilling and Production.[v]

The challenge to make sense, that is, to produce knowledge and understanding from the delivery of structured and unstructured data, is dependent upon how well we can apply the dictionary concepts to the information presented to us. The real issue is that the information that we are all trying to capture and store or transport is based on an understanding of the "meaning" of a set of variables and classifications.

Value to Standards Bodies

The dictionary resources are emerging as critical companions to the development of standard storage data models, such as the PPDM data modeling efforts and the xml tagging standards that are being developed through the Special Interest Groups of POSC. In addition, the dictionaries have a key role to play in the development of taxonomic classifications that are being built for our unstructured data collections.

3. Facing the Problem – Case Study Example

Core Data Integration

A recent project with a large national oil company required that we integrate information across three applications, plus a fourth with multiple versions. A canonical model was evaluated, but it was found to leave considerable information behind and was hence undesirable. The project required the preservation of all data while enabling the exchange of as much of the data as possible and compatible between applications. A partial subset of data for core information is shown in Figure 3.

Listed are the variable names that each system uses in its structured data model. Clearly, these names are truly variable as some mean the same thing and some have the same name and mean different things. The far left column is taken from a taxonomy applied to a core sample library. Correspondence across the unstructured classification and structured data models faces the same difficulty. A dictionary reference for each and all of these attributes and classification would enable rapid and accurate integration of this data.

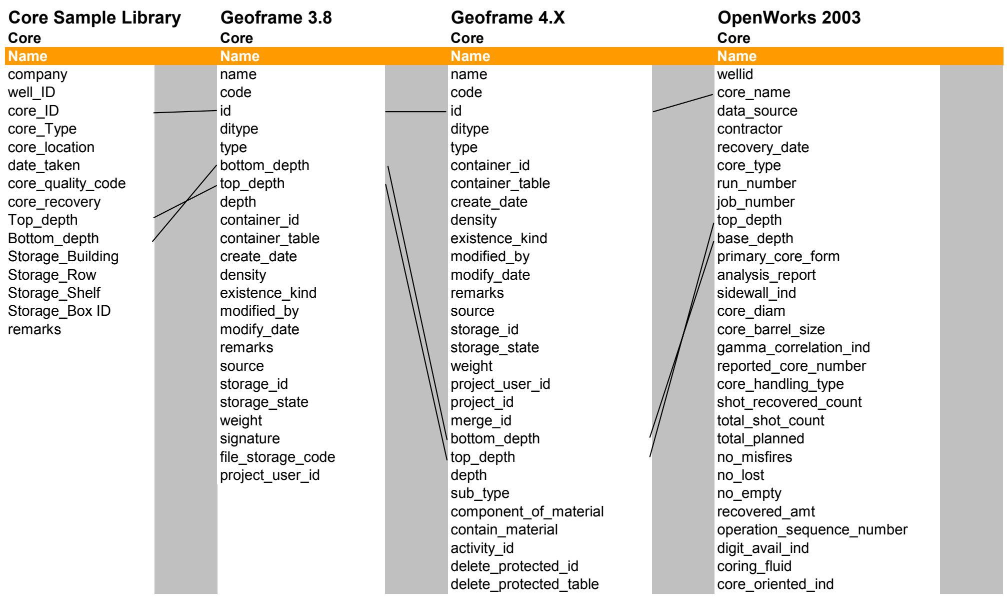

Figure 3 - Core Data Integration Challenge

Log data processing integration

In another project, we implemented a workflow in a sophisticated petrophysical department (Figure 4), which consisted of the transfer of log information across four applications. Specialists at each step processed the incoming logs until the final logs were ready for publication to production projects. The attribute names across each of these applications varied widely. Figure 5 summarizes the full scope of interchanges that were required. Using a dictionary resource in the project helps this process significantly. It also shows that if application vendors introduced a dictionary reference to their documentation, these models would be integrated correctly and completely. The bindings would be easily extended to unstructured indexes of log images in the same correct and complete manner.

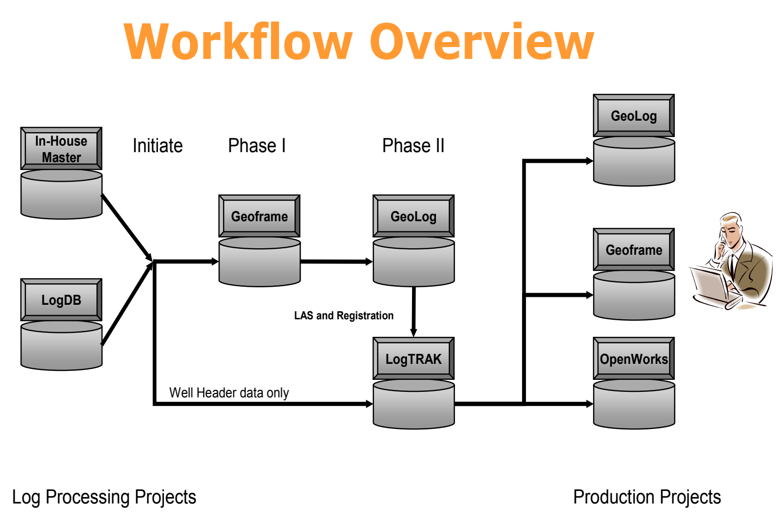

Figure 4 - Log Processing Workflow

Wells, Logs Transfer Summary

Wells, Logs Curves	Target							
Source	**Wells**	LogDB	Geoframe	Geolog	LogTRAK	OpenWorks	GoCad*	Emeraude*
	LogDB	No	Demo	Yes	No	Yes	Demo	Demo
	Geoframe	No	Yes	Demo	No	Yes	Yes	No
	GeoLog	No	Yes	Yes	No	Yes	Yes	No
	LogTRAK	No	Yes	Yes	No	Demo	Yes	No
	OpenWorks	No	Yes	Yes	No	Yes	Yes	No
	GoCad	No	No	No	No	No	No	No
	Emeraude	No	No	No	No	No	No	No

* Compatible File Format

Figure 5 - Integration Extent

4. *Emerging Tool Concept*

Integration Engines and Glossaries – The Semantic Designer Toolkit™

The Semantic Designer Toolkit™ supports integration across structured and unstructured information repositories. It uses a data mapping process to make the transition from hand-crafted XSL's to a more automated, easier to maintain process that is integrated with electronic dictionaries. The Semantic Designer Toolkit provides domain experts with the ability to modify and enhance the data definitions and correlations by exposing the integration design to them. It eliminates problem of 'black box' associations that are embedded and not visible in other integration applications.

Another function of the Toolkit is to enable web service or download access to on-line data definition libraries, such as PIDD, PPDM Glossary or the Schlumberger dictionary, for inclusion in the managed definitions where possible. When external definitions are not applicable or useful to a specific information collection, company definitions can be recorded and maintained. A 'conversion/calculation library' for semantic designers would also be included that enables the use of computation or transformations to map from one data source to the standard target.

Value Proposition

The common approach to mapping attribute relationships is to rely on an xml programmer who either has the domain expertise themselves or works directly with a domain expert. The Semantic Designer Toolkit reduces the time required for integration and improve the accuracy of the integration definitions by giving domain experts a simple interface to map data between applications. Also, the parsing, concatenation and calculations that were implemented by programmers would now be more visible to the domain experts, who can adjust when needed. This information is then presented in quick-to-comprehend GUI screens.

Web Service Dictionary

The Semantic Designer Toolkit becomes even more valuable to companies, as well as more consistent throughout the industry, if the standards organizations take on the deployment and maintenance of online glossaries. If these are deployed as web services, then application builders can easily associate their attribute information using these standard definitions. New definitions can be added and vetted using processes similar to those used to evolve the structured data models and taxonomies.

5. Conclusions

The opportunity to combine structured and unstructured data together brings with it the challenge of establishing the meaning of the information retrieved from a search. The technology exists today. We have good data models and tagging schemes. We have developed useful taxonomies to classify unstructured data. We have a number of electronically-based dictionary tools. By combining and linking these capabilities in a system that manages meaning, we can make significant progress toward the goal of managing and preserving the knowledge assets of our corporations, all through a 'Semantic GPS' that makes navigating information space as simple as using a GPS to navigate from place to place in the physical world.

---

[i] Allaby A and Allaby M: *The Concise Oxford Dictionary of Earth Sciences*. New York, New York, USA: Oxford University Press, 1991.

[ii] Bates RL and Jackson JA: *Dictionary of Geological Terms*, 3rd ed. New York, New York, USA: Anchor Books/Doubleday, 1984.

[iii] Sheriff RE: *Encyclopedic Dictionary of Exploration Geophysics*. Tulsa, Oklahoma, USA: Society of Exploration Geophysicists, 1991.

[iv] *A Dictionary for the Petroleum Industry, 1st edition.* Austin, Texas, USA: Petroleum Extension Service, 1991.

[v] Hyne NJ: *Dictionary of Petroleum Exploration, Drilling and Production.* Tulsa, Oklahoma, USA: PennWell, 1991.