Position Paper – W3C Workshop on Semantic Web in the Energy Industries; Part I: Oil & Gas

Prepared by:	Alan Doniger, Energistics CTO
Date created:	September 26, 2008

Energistics and Semantics

Energistics™ is an energy industry consortium that leads collaborative, non-commercial initiatives to improve operational efficiency within member organizations and across the industry. The tangible results of Energistics initiatives are best practice guidelines and industry standards related to data and process definition and use.

Energistics results are usually centered in the space between individual company point solutions through enterprise-wide company standards on one side and industry-independent technologies, practices, and standards on the other side. As such, Energistics results almost always consist of fundamental definitions and semantics, interaction software frameworks, and formal data specifications for exchange and/or storage/archiving that enable interfacing and integration across traditional organizational and functional discipline boundaries.

Energistics results are employed directly or indirectly in organizational policies, practices, work process definitions, manual procedures, in-house and research automated solutions, and commercial automated solutions.

- Some Energistics results are simply structured sets of definitions, such as the Units of Measure standards and Well Purpose Reference Standards.

- Most Energistics results are framed as elements of implementable computing and communications technology, such as the XML and Web Services based industry standards families: WITSML™ for drilling and completions, PRODML™ for production optimization and reporting, and RESQML™ for the characterization of earth models and reservoir models. In these cases, structured sets of definitions are embedded in the results, for instance as XML Schema enumerations.

The challenges posed to Energistics are the collective and compounded challenges of individual energy companies and their constituent component organizational units to achieve local, specific excellence and enable foreseen and even unforeseen integration excellence. As with any re-engineering opportunity, the use of Energistics results faces industry company policies that limit investments in new initiatives only to those that are absolutely necessary and those that offer a very large, short term return on investment.

In all of these ways, Energistics initiatives are all about working with *semantics*, from discovering and discerning semantic intent of existing solutions and procedures to asserting adequately complete and clear semantic definitions within Energistics best practice and industry standards. Energistics works both with challenges of defining and supporting the use of consistent units of data and process as well as the challenges of defining adequately workable transformations between separately defined solutions. The separate solutions may include those that are internal proprietary, those that are commercial product proprietary, and those that are the results of standards efforts within parts of a company, for an entire company, or across companies.

Semantic Web Technologies

The development of the Semantic Web Technologies, notably the RDF/OWL formalism for defining ontologies, expands our ability to define semantics beyond what has been available previously. This applies whether the intended use of such ontologies is for Web-based wide-area searching or for other application settings.

We have often noted gaps between the desired expressions of semantics and the ability of various formal definition mechanisms to define them. Remedies have often required supplementary, informal documentation.

Even Semantic Web Technologies are limited in the degree of precision and the ability to express subtle semantic nuances. Note the three levels of OWL and their progressive expressive abilities versus the limitations on automation presented by the additional expressive abilities.

Energistics is strongly interested in Semantic Web Technologies and is monitoring the early work by others as well as considering our own work efforts. We believe that it would be useful if the results of the various Energistics initiatives were adapted and published using Semantic Web Technologies. Clearly, this would enhance dialogue and joint work with our member organizations and with other industry and cross-industry groups also engaged with Semantic Web Technologies.

Context, Context, Context

There are many challenges facing any effort to establish and use sets of definitions that have clarity, completeness, correctness, and consistency as would be needed to achieve a business objective, whether that objective is of the nature of searching and qualifying subsets of documents (pages, instances, etc.) or integrating data from multiple sources as part of a specific business operation. These range from technical aspects of what must be denoted and expressed to organizational and social aspects of framing the scope of work and obtaining sufficient resource commitments.

The challenge that seems to be both the most prevalent and the most difficult to solve, at least for the classes of problems addressed by Energistics, is *context*. Wikipedia offers this definition: “the relevant constraints of the communicative situation that influence language use, language variation and discourse.” A corollary definition of *context analysis* is quite informative: “analysis of the environment in which a business operates”.

We view context as that which is supposed to be understood and need not be stated. When I see a traffic signal at a road intersection, I understand the colored lights in that context without a sign being posted identifying that the colored lights follow an accepted road traffic signal convention. The integration of data from multiple sources always faces the challenge of context.

We know what is documented about each data source, but do we know enough about what is not documented about each one? I presume that each data source has sufficient clarity, completeness, correctness, and consistency to operate successfully in its source environment. This does not, however, mean that a third-party attempt to integrate data from these sources will succeed. The worst outcome may be apparent success that leaves unnoticed problems in the result for a subsequent business process to find. This violates one of the fundamental rules of data management, which is to validate data as close to the source as possible.

E. Paslaru Bontas’ paper, Using Context Information to Improve Ontology Reuse[i] asserts that the context problem inhibits the reuse of ontologies: “Existing ontologies contain valuable amounts of knowledge. Nevertheless, they cannot be easily evaluated for and reused in new usage settings.”

One may believe that the solution to the context problem when developing sets of definition, including an ontology using Semantic Web Technology, is to insist on explicitly defining all aspects of context, thus, leaving nothing to the imagination. This is a good theory, but is not practically achievable. A related principle is this: “Use the widest possible explicit scope – leaving the smallest possible context.” Even more important is this principle: “Make the best possible effort to explicitly define the context.”

We came across a proposed method for explicitly defining context in aspects of the Core Components specifications of ebXML[ii]. We are favorably disposed to such efforts to formalize the specification of an otherwise unstated context.

We have tried to apply the 'widest possible explicit scope' principle in our work on the Energistics families of XML data exchange industry standards. We use a single XML Schema Namespace. We encourage the sharing of sub-components across XML schemas and across standards families.

We also recognize the need to be concerned about local context, that is, within individual elemental definitions. We require natural language definitions be composed and annotated with each XML schema element specification. We are cautious about using unqualified element names.

We believe that concerted work is needed to clarify approaches to associating context with ontologies as a whole and with elemental portions of ontologies. We also support efforts to interface multiple ontologies – short of fully integrating them – so that readers/users of the ontologies can use the interface bridges to obtain full value always recognizing the source context and the presumed interfaced context.

Ultimately, we believe that Semantic Web Technologies can contribute to fairly local solutions, even those involving automated interpretation and/or automated searches. We believe that Semantic Web Technologies can also contribute to the broader challenges of integration across internal and external boundaries through an iterative series of interfacing and integrating efforts built up from high quality source material.

Conclusion

This paper reviewed the positioning of Energistics as an organization that by its nature addresses challenges of data and relationship definitions, structures, and semantics. Energistics seeks clarity, completeness, correctness, and consistency, collectively known as industry standards, which apply across traditional boundaries both within and among industry organizations.

Energistics results currently use various formalisms for specifying data definitions, structures, and semantics, notably XML Schema for our families of data exchange and Web Services industry standards. We recognize the refinements and capabilities of the Semantic Web Technologies and look forward to opportunities to adapt and extend Energistics specifications to use these technologies. We also look forward to the opportunities to relate the resulting ontologies with similar work done by our members and by industry groups.

We approach the specifying of definitions, structures, and semantics with a great deal of respect for the challenges that stand in the way of achieving exceptionally good results. This applies to challenges inherent in any such effort and the additional challenges that result from expanding scope, integrating multiple existing specifications, etc., especially challenges from unstated context implications.

Finally, Energistics believes in driving industry initiatives (and, by extension, cross-industry initiatives) in an incremental manner in which each incremental step is strongly driven by business strategic interests with well-defined objectives and goals. We look forward to active participation in the further development of Semantic Web Technologies and their application in the energy industries.

[i] E. Paslaru Bontas, *Using Context Information to Improve Ontology Reuse*, Freie Universitat Berlin, 2004

[ii] See http://www.ebxml.org/