

GlaxoSmithKline Position Paper for Workshop of Ontologies in Agents Systems

Robin McEntire

GlaxoSmithKline Pharmaceuticals
709 Swedeland Road, PO Box 1539
King of Prussia, PA 19406-0939
610.270.6527

Robin_A_McEntire@gsk.com

ABSTRACT

This paper describes the position of GlaxoSmithKline with respect to the development of ontologies and intelligent, agent-based systems.

Keywords

ontologies, agents, life sciences.

1. Motivation

It is clear that the number of data sources available and necessary to move forward in the study of the life sciences is quite large and diverse and is growing at a remarkable rate. It is also clear that there is tremendous growth in the field in computational methods and services for manipulating this data, often manipulating multiple, disparate, distributed and heterogeneous sources of data. The proliferation of data formats and schemas for life sciences objects and the many methods for accessing or computationally manipulating those objects and services is a serious impediment to the efficient, effective and scalable use of these data and services. The problem will become only more severe as the field continues to grow. However, the recent developments in web technology, including web languages, and object-oriented and knowledge-representation technology, are at a level of maturity that they can be leveraged to help solve this problem.

As outlined by the workshop organisers a progressive framework for the life sciences has a number of components, which are;

1. the development of common, shareable ontologies for objects within the life sciences domain
2. the development, or, more likely, the adoption of one or a few common languages for the exchange of these common ontologies, and
3. the definition of services that provide well-characterised, uniform (or, at least, consistent) access to data/information and services

Common, shareable ontologies provide the field with an understandable semantics for discussion and for the programmatic computation of life sciences objects. Exchange languages provide

a common syntax, which enables the semantic representations to be moved from one site to another without undo, or inaccurate, translation of the definitions of the underlying objects. And well-defined services allow easy programmatic access to life sciences objects and to services that manipulate those objects (and produce other life sciences objects).

2. Commercial Opportunities

More importantly, over the long run, this combination of well-defined services and well-understood, easily exchangeable objects will create a marketplace in which vendors may compete to provide services and new data sources. As we have seen in other technology areas the greatest driver for the adoption and use of a technology is a strong marketplace. GSK, and other pharmas, are interested in services that provide scientific information that can be easily geared to the specific objectives of its scientific investigators. So, for example, there is currently work being done to provide more targeted access to scientific journal articles for the research scientist by many of the electronic journal vendors. This process could be significantly aided by categories and ontologies that are shared across the industry. The workshop organisers have already suggested that one of our goals should be finding applications that provide the greatest internal business impact or "the most bang for the buck", which is a laudable goal. However, we should also consider the markets that our work might target and what opportunities there might be for us to leverage our work into new or existing COTS products. By focusing on the greatest external business impact this work will yield the greatest impact on our industry as a whole.

3. Additional Technologies

In addition, GSK is very interested in intelligent integration of multiple, heterogeneous data sources. Well-defined services and common object definitions will make the job of mediated access to multiple data sources much easier. It will also make the display and visualization of this information more intelligent and geared more to the particular interests of each scientist. There are already some vendors in the market and others emerging. Our efforts could provide substantial assistance to these vendors. The same could be said of text mining software and products. We are seeing a number of vendors emerge in this area. Ontologies and related services could be a significant aid here as well.

4. Relationships and Synergies

We also encourage the organisers of the workshop to consider what other organisations might provide in terms of valuable synergies for our work. For example, the OMG has been in existence for a considerable time and has been very successful in

defining common services within the life sciences research domain. Another significant strength of the OMG is that its membership includes a large number of IT and Life Sciences organisations, who now have a history of working together on a number of problems.

5. References

- Cattell, R.G.G., et al, 2000. *The Object Data Standard: ODMG 3.0*, Morgan Kaufmann Publishers
- Fikes, R., Kehler, T. 1985. The Role of Frame-Based Representation in Reasoning *Communications of the Association for Computing Machinery*, 28(9):904-920.
- Gruber, T.R. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199-220.
- Karp, P., Riley, M., Paley, S., Pellegrini-Toole, A., Krummenacker, M. 1999. EcoCyc: Electronic Encyclopedia of *E. coli* Genes and Metabolism. *Nuc. Acids Res.*, Vol 27 No 1, pp 55-58.
- Karp, P.D., Chaudhri, V, *XOL Specification*, <http://www.ai.sri.com/pkarp/xol/>.
- McEntire, R, et al, An Evaluation of Ontology Exchange Languages for Bioinformatics, In *Proceedings of ISMB 2000*.
- Object Management Group, 1997. *UML Semantics, Version 1.1*, <ftp://ftp.omg.org/pub/docs/ad/97-08-04.pdf>.
- Riley, M., 1993. Functions of the gene products of *Escherichia coli*, *Microbiological Reviews*, 57:862-952.
- Rumbaugh, J., Jacobson, I., Booch, G., 1998. *The Unified Modeling Language Reference Manual*, Addison-Wesley.