SmithKline Beecham Position Paper for

IEEE Workshop on

XML-Enabled Wide Area Search in Bioinformatics (XEWA)

Motivations: It is clear that the number of data sources available and necessary to move forward in the study of the life sciences field is quite large 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 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 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 authors the way forward for the life sciences has a number of components, which are;

1. the development of common, shareable ontologies for objects within the life sciences
2. the development, or, more likely, the adoption of one or a few common languages for the exchange of these common ontologies
3. the definition of services that provide well-defined, 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 over 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 allows easy programmatic access to life sciences objects and to services that manipulate those objects (and produce other life sciences objects).

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. SB, 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 authors have already suggested that one of our goals should be finding applications that provide "the most bang for the buck", which is a laudable goal. 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.

Additional Technologies: In addition, SB is very interested in intelligent intergration of multiple, heterogeous 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 on 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 services could be a significant aid here as well.

Relationships and Synergies: We also encourage the authors of the workshop to consider what other organisations might provide valuable synergies for our work. For example the OMG has been in existence for a considerable length of time and has been very successful in defining common services within the life sciences. Another significant strength of the organisation is that it's membership includes a large number of IT and Life Sciences organisations, who now have a history of working together on a number of problems. The OMG has not been as successful in bringing products to market quickly (in "internet-time") or in wide-spread adoption of these products; however, this may not be a drawback to the goals of this group.