November 09

Version 11.16A

iPG2P Data Integration:

Semantic Web Services Ground Assessment

Damian Gessler, Ph.D.
Semantic Web Architect

We assess the state of semantic web services, with special emphasis on the availability of services for genomic sequences and ortholog identification. This report is a request from the iPG2P Data Integration Working Group, November 12, 2009.

We report that deployments of semantic web services, initially restricted as web services that adhere to one of the two W3C recommendations of either SAWSDL (Semantic Annotations for WSDL [Web Service Definition Language]) or POWDER (Protocol for Web Description Resources) are essentially non-existent. While numerous research-grade implementations of SAWSDL exist (perhaps a few dozen to a few hundred), neither W3C recommended technology has been adopted by any broad community.

Recognizing that the broader informatic community has not yet standardized on a semantic web service technology, we note:

1. Web services—as traditional web service implementations using SOAP, WSDL, and/or REST interfaces, but explicitly without semantic support—are numerous. Community examples include NCBI web services, EBI web services, and Gramene web services.
2. Semantic web services using the NSF-funded technology SSWAP (Simple Semantic Web Architecture and Protocol) are deployed at Gramene, Soybase, and the Legume Information System. These services cover various genomic, transcriptomic, genetic and physical mapping, and QTL (Quantitative Trait Locus) services. Ortholog support currently exists only for discovery.

Purpose of this document To assess the current state of semantic web services relevant to iPG2P goals, with specific attention to genomic sequences and ortholog identification.

Audience iPG2P Data Integration Working Group

Synopsis W3C semantic web service technology recommendations (SAWSDL and POWDER) have yet to receive broad implementation or adoption. Non-W3C technologies exist; of these the author's work in SSWAP (Simple Semantic Web Architecture and Protocol) currently has the largest deployment in biology as restricted specifically to semantic web services.

Introduction **Semantic web services are an emerging technology**. The need for enabling computers to assess information based on context (in contradistinction to more simplistic lexical or syntactical equivalency) is widely recognized. If computers could do this, they could, for example, discover, assess, retrieve, aggregate, assimilate, and even aid in integrating data and services from a broad and disparate range of sources. Today, context determination can be done in a low-throughput, one-off manner (*i.e.*, with some component of directed, human intervention at one or more points in the data flow), but no system is yet widely deployed such that this can be done in a high-throughput manner across the web. By "high-throughput" we mean an automated, scalable process that can be applied to both existing and new entrants.

Specifically, we note that widely deployed web service installations have non-existent to poor explicit semantics amenable to high-throughput machine reasoning. (Non-semantic) web services, such as those at NCBI, EBI, Amazon, Google, and others, are high-throughput in terms of syntactical, but not semantic, processing. This means that web services—while enabling high-throughput data transfer once engaged—remain low-throughput in the areas of discovery and automated determination of suitability-for-purpose. Web services thus remain low-throughput (non-scalable) as data integration technologies. For example, there is virtually no support for machines to determine if one web service's use of the token *Gene* is semantically equivalent to another web service's use of *Gene*. Blind aggregation across semantically dissimilar tokens will result in erroneous service engagement, information flow, and data assimilation.

Semantic web technologies excel in empowering high-throughput semantic determination, but lack sufficient web service protocols and supporting informatics. Semantic web technologies such as OWL (the W3C recommended Web Ontology Language) are theoretically applicable to web service applications, but to date largely lack the specifics to allow for off-the-shelf implementation.

Semantic web services aims to address the limitations of separate pure-play web service or semantic web technologies; *i.e.*, they aim to bring either semantics to web services, or web service capabilities to semantics, but regardless of the approach, they aim to enable semantically aware, high-throughput resource discovery and engagement.

Methods To assess the state of semantic web services we:

1. Identified W3C recommended technologies for semantic web services. There are two, somewhat competing, technologies reflecting the influence, contributions, and legacy of two schools of thought:
    a. *Bring semantics to web services* The first is a response by the web services community to bring semantics to web services. This is formalized in a W3C recommended technology called **SAWSDL** (Semantic Annotations for WSDL)[1]. WSDL (Web Service Description Language) is a well recognized, non-semantic technology for web service description.
    b. *Bring web services to semantics* The second is a response by the semantic web community to bring web service support to the semantic web. This is formalized in a W3C recommended technology called **POWDER** (Protocol for Web Description Resources)[2]. POWDER is instantiated in a semantic POWDER-S document, which grounds POWDER in the well recognized semantic web language of OWL (Web Ontology Language)[3].

2. To assess semantic web service adoption, we searched for published reports of SAWSDL and POWDER implementations across:
    a. Two hundred and five BMC Central journals, notably including *BMC Bioinformatics*;
    b. Ninety Oxford Journals notably including *Bioinformatics*, *Briefings in Bioinformatics*, and *Nucleic Acids Research*;
    c. Two thousand Elsevier journals notably including the *Journal of Web Semantics*;
    d. Over 2000 Springer journals notably including *LNCS* (*Lecture Notes in Computer Science*);

3. Notable systems (not all true semantic web services) that do not use SAWSDL or POWDER include:
    a. BioMoby
    b. SADI
    c. MyGrid
    d. SSWAP

Results SAWSDL and POWDER show virtually no implementation in our search across BMC Central, Oxford Journals, and Elsevier journals. In LNCS there are over 140 papers referencing SAWSDL to some extent, though few implementations. Of the implementations reported, virtually all are of research-grade trials and prototypes. POWDER had substantially fewer reports than SAWSDL. These results

---

[1] http://www.w3.org/2002/ws/sawsdl
[2] http://www.w3.org/2007/powder
[3] http://www.w3.org/2007/OWL

support an expert opinion reflected in various points in the literature and on the web that there exists no *de facto* standard for semantic web services.

The relative lack of W3C semantic web service implementations reflects the early state of this technology development. Similarly, there are relatively few non-W3C models. We briefly note four:

1. **BioMoby** (also known as MOBY Services) has no support for formal semantics (*e.g.*, it does not require OWL compliance or compliance with the framework of any formal logic system), and thus is not a semantic web service technology in the sense of refs. [4, 5, 6], but it does support limited semantics in its use of controlled vocabularies and subclass relations. BioMoby has hundreds of implementations. BioMoby's use of controlled vocabularies in subsumption hierarchies supports an *ad hoc* application of subsumption logic. BioMoby's successor (pre-release MOBY 2) is SADI (Semantic Automated Discovery and Integration).

2. **SADI** (Semantic Automated Discovery and Integration) (MOBY 2) is a substantial technological change from BioMoby. Unlike BioMoby, SADI is OWL-compliant but does not use POWDER. SADI is not a new technology *per se,* but proposes a model for how to use OWL to achieve some of the goals of semantic web services. Its use is focused on allowing the distribution of SPARQL queries across OWL-compliant sites. SADI is newly released in 2009; more information is available at http://sadiframework.org.

3. The European eScience program and specifically the **myGrid** project had a notable and early history in ontologies and semantic web service models (*viz.*, the myGrid Ontology, Feta; see also www.semanticgrid.org). In more recent years, myGrid appears to have de-emphasized semantic web services and has focused more heavily in maturing workflow tools (Taverna), social networking for workflows (myExperiment), web service discovery (BioCatalogue).

4. **SSWAP**[7] (Simple Semantic Web Architecture and Protocol; http://sswap.info) is an NSF-funded project of the author. SSWAP is a full-featured semantic web service architecture and protocol supporting resource description, publication, discovery, querying, invocation, and data return. SSWAP supports a model of third-party, extensible ontologies under the formal logic of OWL DL. SSWAP is implemented at Gramene (10 services), Soybase (13 services), and the Legume Information System (3 services). Plans are underway to implement four services at the Plant Ontology. The LIS sequence retrieval service returns over 2,000,000 DNA and RNA sequences keyed on accession IDs. Gramene, Soybase, and LIS services have full ontological markup on both input and output data types. From a knowledge base of over 2400 resources, SSWAP discovers 57 phylogeny reconstruction services (relevant for

---

[4] McIlraith et al. Semantic Web services. *Intelligent Systems*, *IEEE* (2001) 16(2):46- 53
[5] Bruijn et al. *Semantic Web Services*. Modeling Semantic Web Services (2008) Springer Berlin Heidelberg. pp. 9-21
[6] Haniewicz et al. *Semantic Web Services Applications–a Reality Check*. Wirtschaftsinformatik (2008) 50(1): 39-46
[7] Gessler et al. *SSWAP: A Simple Semantic Web Architecture and Protocol for semantic web services*. BMC Bioinformatics (2009) 10:309.

ortholog identification or retrieval) as classified by the NAR (Nucleic Acid Research) categories. These have markup only on the service classification, not on the input and output data types. SSWAP is OWL-compliant but does not use POWDER.

Conclusion It is hopeful that a W3C recommendation will lead to standardization for semantic web services. Yet historical examples of technologies such as CORBA, ebXML, UDDI, and many others shows that industry support and official "standards" are no guarantee for wide-spread adoption. Indeed, early use of "standards" may carry far greater risk in terms of a lower ROI (return on investment) than an early adoption of simply "something that works." Conversely, use of "something that works" may drive and create a *de facto* standard that becomes a recommendation exactly because of its real-world problem solving ability (*e.g.*,. XML). In light of this, at the time of this writing we should neither adopt (nor dismiss) SAWSDL or POWDER *because* they are W3C recommendations: we should assess their use in iPlant with heavy emphasis on how they solve our real-world problems, while being cognizant of changes in their trend for broader adoption.