Interoperability with BioMoby - Past, Present, and Future

Ben Vander Valk, Luke McCarthy, Mark Wilkinson
Heart + Lung Research Institute, St. Paul's Hospital,
Center for High Throughput Biology, University of British Columbia,
Vancouver, British Columbia, Canada
markw@illuminae.com

In 2001 the term "Deep Web" was coined to describe the information on the Web that was not accessible to search engines – for example, data from databases or analytical tools that were hidden behind Web Forms; it was estimated at the time that the Deep Web contained upwards of 100-fold more data than the Web itself. In that same year, the BioMoby project was founded with the goal of establishing a framework for syntactic and semantic interoperability between Deep Web-style bioinformatics resources. By mid-2002, a prototype interoperability architecture and supporting codebase had been developed that had four main components: an ontology of bioinformatics data-types ("Objects"), an ontology of bioinformatics identifier-types ("Namespaces"), an ontology describing various computational operations ("Services"), and a novel Web Service Registry API ("Moby Central") that was aware of these three ontologies and was capable of semantic brokering between Web Services providers and the clients who required these Services.

BioMoby has been moderately successful in providing an interoperability platform within discreet communities of users in the absence of competing standards; however it failed to be widely adopted throughout the life sciences community. In the interim, novel technologies have emerged from the World Wide Web consortium's Semantic Web activity that achieve many of the same goals as the BioMoby framework does, but have the broader endorsement of the W3C itself. Nevertheless, the "niche" that BioMoby fills – the intersection of semantically and syntactically-standardized data-types and the ability to discover and invoke appropriate Web Services based on the type and content of these data elements – remains unfilled by current Semantic Web activities within the Health Care and Life Science domains. As such, it is now prudent for the BioMoby project to examine its past successes and failures, learn from these, and re-invent itself. The next generation of Moby will use these emergent Semantic Web standards to represent the Deep Web, and thereby parallel their utilization by the wider life sciences community to represent traditional Web data.

To this end, I introduce here the architecture of the Moby 2.0 interoperability proposal, and its first prototype implementation in the CardioSHARE (Cardiovascular Semantic Health and Research Environment) project. Moby 2.0 is a proposed extension to the SPARQL Protocol and RDF Query Language in which predicates in a SPARQL query are mapped onto Web Service categories (effectively the BioMoby Service Ontology), and Web Service registry queries and executions are achieved in-line with SPARQL query resolution. Thus data is generated dynamically on an as-needed basis in response to specific queries, prior to the resolution of those queries. This is further extended by the CardioSHARE architecture such that concepts (subjects/objects) in the SPARQL query can refer to concepts defined in OWL ontologies, where the necessary and sufficient conditions defining those concepts are evaluated through a similar process of Web Service discovery and execution. In this way, complex data-types referred to in the query, but not present in the initial data-store, can be assembled by a combination of DL-reasoning and Web Service invocation prior to query resolution. If successful, this will simplify the process of discovery and query-answering, and will significantly change the way we interact with data both in our local databases, as well as on the Web.

Websites: http://biomoby.org; http://cardioshare.icapture.ubc.ca