

Abstract

The relevance of this demonstration is justified as follows:

1. World Wide Web Consortium technology recommendations are evolvable, scalable and decentralized methodologies for managing large, rapidly changing databases and metadata.
2. Science increasingly depends on managing massive quantities of machine-generated data and metadata., and the use of semantic-supporting information has been increasing in the life sciences (e.g., GeneOntology, MESH).
3. There is a need for overlap between the Web community of practice and the scientific community of practice:
 - More scientists are requesting databases to recognize semantic kinds of content and queries
 - As life science moves more towards understanding systems it needs to handle more than just data constructs, it must represent conceptual forms and logical statements, that can easily change as the model is updated
 - Scientists need to place semantically enhanced resources across the Web, so that researchers from any background domain can access any set of semantically defined information on the web and use it intelligently.

Current research has demonstrated that Knowledge can be assembled from structured and unstructured sources and represented in a machine-readable format that is web-compatible using Resource Description framework (RDF) [<http://www.w3.org/RDF/>]. RDF is a World Wide Web Consortium (W3C) standard that is used to define sets of relations between data and concepts. W3C's XML standard has proven useful in making many kinds of structured data exchangeable in a self-describing format, yet the semantic(s) a scientist projects on to that data is not formally encoded. Hence most uses of such information, though structured, have limited application flexibility and are devoid of semantic interoperability.

RDF, a new type of XML, is the cornerstone for the Semantic Web [<http://www.w3.org/2001/sw/>], which will enable anybody to clearly and commonly define the concepts and logic within any document. Any information expressed in RDF may be connected to any other information expressed in RDF, in much the same manner that any document expressed in HTML may link to any other document expressed in HTML. In this manner, the expression of discrete facts in RDF makes them available for sharing and analyses by the larger scientific community. RDF has the potential to let scientists apply all available knowledge to decision-making, including target prioritization, assessment of compound liabilities and clinical trial design.

This proposal is intended to familiarize both scientists and bioinformaticists to the practical applications and uses of web-based semantic technologies in support of research objectives: sharing findings, summarizing insights, and encoding knowledge independent of the amount of data. The related principles and implementations will be thoroughly described from a life science perspective, focusing on key challenges in the life science field. The demonstration will emphasize the relations between structured information (databases) and semantic forms (statements and annotations), and how communities can take advantage of them together. In some cases knowledge representations (ontologies) will be utilized, and in other cases it will address data aggregation, annotations and information sharing paradigms that rely on less intricate semantics.

This demonstration will introduce students and scientists to the technologies that create the Semantic Web, including RDF, the Web Ontology Language (OWL) [<http://www.w3.org/2004/OWL/>] and the Uniform Resource Indicator (URI) [<http://www.w3.org/Addressing/>], with a particular focus on how Semantic Web can impact querying across multiple distributed heterogenous databases, enterprise knowledge aggregation and elements of scientific inferencing. The demonstration will touch on such topics as:

- Content creation applications: Authors can connect metadata (subject, creator, location, language, copyright status, or any other terms) with documents, making the new enhanced documents searchable.
- Software that takes advantage of both RDF and OWL: Organizations can integrate enterprise applications, publishing and subscriptions using flexible models
- Cross-application data reuse: RDF and OWL formats are standard, not proprietary, allowing data reuse from diverse sources.