

# Chapter 3

## Web Services for Bioinformatics

**Abad Shah**

*University of Engineering and Technology, Pakistan*

**Zafar Singhera**

*Oracle Corporation, USA*

**Syed Ahsan**

*University of Engineering and Technology, Pakistan*

### EXECUTIVE SUMMARY

*A large number of tools are available to Bioinformaticians to analyze the rapidly growing databanks of molecular biological data. These databanks represent complex biological systems and in order to understand them, it is often necessary to link many disparate data sets and use more than one analysis tool. However, owing to the lack of standards for data sets and the interfaces of the tools this is not a trivial task. Over the past few years, web services has become a popular way of sharing the data and tools distributed over the web and used by different researchers all over the globe. In this chapter we discuss the interoperability problem of databanks and tools and how web services are being used to try to solve it. These efforts have resulted in the evolution of web services tools from HTML/web form-based tools not suited for automatic workflow generation to advances in Semantic Web and Ontologies that have revolutionized the role of semantics. Also included is a discussion on two extensively used Web Service systems for Life Sciences, myGrid and Semantic-MOBY. In the end we discuss how the state-of-art research and technological development in Semantic Web, Ontology and Database Management can help address these issues.*

### INTRODUCTION

The two major problems that biological scientists are facing are distribution and heterogeneity of the data and its analysis tools. These problems are due to autonomous, decentralized and individualistic

web based approach towards the biological research (Bodenreider & Stevens, 2006). Integration of the data and tools is a difficult task but it is vital for the integrative *insilico* experimentation and exchange of results (Lord et al., 2004). Biology has coped with this work in an effective but in ad-hoc manner. Almost all databases and tools of bioinformatics that have been made available on

DOI: 10.4018/978-1-60960-015-0.ch003

the web and the data integration techniques have been applied to the bioinformatics domain have met limited success because the data and information are made available in a non-standardized way (Lord et al., 2004; Post et al., 2007). However, unlike other domains, the bioinformatics domain on the Web has embraced the standards, such as XML and web services, and there exists a large number of bioinformatics data sources that are either accessible as web services or provide data using XML (Thakar, Ambite & Knoblock, 2005). A web service is a program/software that can be executed on a remote machine owing to the industry efforts to standardize web service description, discovery and invocation. These efforts have led to standards such as WSDL (Christenson et al., 2001), UDDI (UDDI2002), and SOAP (SOAP 2000) (Thakar, Ambite & Knoblock, 2005).

The integration of such services and their interoperability is now feasible by using web services technologies and the researchers can easily construct bioinformatics workflows and pipelines by combining two or more web services to solve their complex biological tasks such as protein function prediction, genome annotation, micro array analysis, etc (Cannta N., et al., 2008). However, these standards, in their current form, suffer from the lack of semantic representation leaving the promise of automatic integration of applications written to web services standards unfulfilled (Labarga et al., 2007).

More recently efforts have been made to populate web services with semantic metadata and semantic descriptions to enhance data exchange and integration (Lord et al., 2004; Thakar et al., 2005; Post et al., 2007). A semantic web approach provides standardized formats (such as RDF, RDF Schema (RDFS) and OWL) to achieve a formalized computational environment. The objective of Semantic Web is to bring meaning to the raw data content by defining relationships between distinct concepts using ontologies (Cabrall L. et al., 2004). The existing life sciences databanks can be built with better retrieval performance using ontological

abstractions. Fortunately, the life sciences community has realized that the semantic modeling is a necessity for the biological knowledge bases (Ruttenberg et al., 2007) and many biological ontology initiatives exist (<http://obo.sourceforge.net>), with Gene Ontology (GO) and it is the most widely adopted ontology ((Bodenreider O. & Stevens R., 2006; Ashburner et al., 2000).

However, a complete and seamless semantic integration of data and information sources and tools is a challenging objective that we are facing, amongst others. Problems related to the shared definitions of knowledge domains, i.e., ontologies, association of biological concepts to the existing data, semantic descriptions of services/requirements and automatic workflow generation (Bodenreider O. & Stevens R., 2006).

We feel that for a complete understanding and appreciation of the problems faced by the integrative biology researchers, the evolution of web services tools from HTML/web form-based tools not suited for automatic workflow generation to advances in Semantic Web and Ontologies that have revolutionized the role of semantics must be traced. Also, we must examine how the state-of-art research and technological development in Semantic Web, Ontology and Database Management can help address these issues. Remainder of the chapter is organized as follows. In Section 2, we describe the special nature of the life sciences research which has led to adoption of e-science to the life science community as a necessity. In Section 3, the past efforts and solutions to cope with the problems of distribution and heterogeneity of life science resources are described. Section 4 discusses the evolution of web services for the life sciences domain. The web services architecture and its limitations are also included in this section. The role of Semantic Web for realization of semantic web services and its promising potential for the life sciences research in Section 5 and Section 6. In Section 6, we also analyze two widely used life sciences web service systems, Semantic-MOBY and myGRID.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/web-services-bioinformatics/49213](http://www.igi-global.com/chapter/web-services-bioinformatics/49213)

## Related Content

---

### Mining Repetitive Patterns in Multimedia Data

Junsong Yuan (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1287-1291).

[www.irma-international.org/chapter/mining-repetitive-patterns-multimedia-data/10988](http://www.irma-international.org/chapter/mining-repetitive-patterns-multimedia-data/10988)

### Learning Kernels for Semi-Supervised Clustering

Bojun Yan (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1142-1145).

[www.irma-international.org/chapter/learning-kernels-semi-supervised-clustering/10965](http://www.irma-international.org/chapter/learning-kernels-semi-supervised-clustering/10965)

### A Data Mining Methodology for Product Family Design

Seung Ki Moon (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 497-505).

[www.irma-international.org/chapter/data-mining-methodology-product-family/10866](http://www.irma-international.org/chapter/data-mining-methodology-product-family/10866)

### Mining Email Data

Steffen Bickel (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1262-1267).

[www.irma-international.org/chapter/mining-email-data/10984](http://www.irma-international.org/chapter/mining-email-data/10984)

### Count Models for Software Quality Estimation

Kehan Gao and Taghi M. Khoshgoftaar (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 346-352).

[www.irma-international.org/chapter/count-models-software-quality-estimation/10843](http://www.irma-international.org/chapter/count-models-software-quality-estimation/10843)