Web interface defines new paradigm for life science data-sharing
31 May 2011

The SciNetS cloud service provides virtual laboratories that undertake advanced research activities by collaboration among scientists on the Web, achieving systematic sharing of life-science data resources obtained utilizing the latest bioinformatics technologies. Credit: RIKEN

A new lightweight web service interface for accessing massive amounts of life science research data across multiple public and private domains has been developed by researchers at RIKEN, Japan’s flagship research institute. Through the powerful RIKEN Scientists’ Networking System (SciNetS), the service provides a secure, flexible and light weight interface to millions of data records and their network of semantic relationships, ushering in a new era of collaboration, analysis and information-sharing for life science research and applied innovation.

Gene annotation, protein structure analysis, plant ontologies, transcriptomes - dramatic increases in the size, variety and complexity of data resources in the life sciences have accentuated the challenges of data analysis in the information age. Adding to these challenges, much of the data handled at each step of the research process is private, making integration with public data more difficult and hindering collaboration. Overcoming these challenges requires systems for securely integrating data resources and making their information widely available through a flexible interface.

The RIKEN Bioinformatics And Systems Engineering (BASE) division, Japan’s leading research institute focusing on the integration and publication of life science research data, has now developed such an interface. Referred to as Semantic-JSON, the interface accesses a "virtual laboratory cloud centre" also developed at BASE named the Scientists’ Networking System (SciNetS), which brings together, as of May 2011, a total of 192 public database projects both internal and external to RIKEN. SciNetS creates common ground for sharing life science data resources by linking these resources together in a network of semantic relationships based on standardized Semantic Web techniques.

Pink circles represent individual "virtual laboratory" projects. Yellow squares and green circles denote respectively organizational reality of centers at RIKEN and organizations outside RIKEN. Blue lines show the number of links between data in proportion to thickness. Red lines show relationships between organizations that produced the data, and green lines show comprehensive collaboration within RIKEN. Credit: RIKEN
Semantic-JSON provides a flexible interface to SciNetS on the web, enabling bioinformaticians to access specific data from across the SciNetS network using the programming languages and information tools they normally use in their research. The interface does so by defining a set of simple but relevant commands for accessing and searching SciNetS data and their semantic relationships, delivering results in the widely-used JavaScript Object Notation (JSON) format.

Semantic-JSON extends the concept of short URL services to Semantic Web. It also provides functions of data access control, data search and inference and access to biomedical raw data such as DNA sequences. Credit: RIKEN

Already, RIKEN has successfully applied Semantic-JSON to a number of projects, including international data collaborations on mouse phenotypes, domestic integrated database projects, and the GenoCon International Rational Genome Design Contest. Looking ahead, RIKEN plans to use the interface to distribute life science data across its research centres and with international collaborators via the SciNetS project, broadening the life-sciences Semantic Web data universe and promising to achieve not just comprehensive understanding of various life phenomena, but also collaborative breakthroughs for medicine, industry and the environment.

This research result will appear in the online version of the British scientific journal *Nucleic Acids Research* on June 1.