

Integrated bioinformatics gateways portal and interface

September 22 2011



Credit: AI-generated image ([disclaimer](#))

An easy-to-use bioinformatics interface has been developed by a research group led by Tetsuro Toyoda called the RIKEN Bioinformatics And Systems Engineering division (BASE), Yokohama. The web-service-based tool, called Semantic-JSON, and the portal, BioLOD, integrate access to information contained within genomics,

proteomics, and other ‘omics’-based data repositories.

“Advances in life sciences increasingly depend upon cross-analysis and integration of diverse information from multiple large databases maintained on remote computer servers,” explains Toyoda. “The challenge is to facilitate data retrieval, integration and collaboration while maintaining database security.”

As a first step, various research organizations worldwide, including RIKEN, recently published 192 public and 190 private mammalian, plant and protein databases. The data are integrated by SciNetS.org, the Scientists’ Networking System. These databases contain more than 8.2 million individual data records.

Pioneering a new global trend, BASE provides ‘structured linked open data’ and private data via the newly developed BioLOD.org portal connecting with the World Wide Web Consortium (W3C) Linking Open Data initiative. These self-described data are interlinked using standard web technologies allowing automatic reading by computers, thereby making them more useful to researchers. The system facilitates information sharing and collaboration between researchers, but brings new challenges.

“The sheer amount of data contained in our biological data cloud outstripped the capacity of existing [bioinformatics](#) interfaces to cope with the complexity of researcher queries, motivating us to develop Semantic-JSON,” explains Toyoda.

Semantic-JSON has two major components. The secured, unified data repository integrates data meaningfully—or ‘semantically’ in computer parlance—from numerous sources. The web-based [interface](#) allows researchers to retrieve linked data seamlessly and securely using established bioinformatics programming languages and processing.

Bioinformatics researchers can then use their specialized computational tools to analyze raw biological data (Fig. 1).

Databases already available through Semantic-JSON and BioLOD.org include the RIKEN Integrated Database of Mammals with 79 human and mouse omics databases, the RIKEN Integrated Database of Plants incorporating 30 similar databases for the plant species *Arabidopsis thaliana*, and the RIKEN Integrated Protein Database containing 18 databases.

Since December 2009, international researchers have successfully used the system to identify 28 million data relationships, generating some 4.5 terabytes of associated files. Around 134,000 programs from non-RIKEN researchers have accessed the server as of March 2011. Biological applications include genome design, DNA sequence processing, and the inference of phenotype biological characteristics from genomic information.

“Our next goal is to develop and improve the system to increase its functionality and the usefulness of its linked open data to the worldwide biological community,” says Toyoda.

More information: Kobayashi, N., et al. Semantic-JSON: a lightweight web service interface for Semantic Web contents integrating multiple life science databases. [Nucleic Acids Research](#) published online 1 June, 2011 ([doi: 10.1093/nar/gkr353](https://doi.org/10.1093/nar/gkr353)).

Provided by RIKEN

Citation: Integrated bioinformatics gateways portal and interface (2011, September 22) retrieved 4 July 2024 from <https://phys.org/news/2011-09-bioinformatics-gateways-portal-interface.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.