A Secure Data Repository Integrating Huge Volumes of Digital Contents as a Semantic Web

Norio Kobayashi[1], Yuko Makita[1], Eli Kaminuma[1], Shuji Kawaguchi[1], Yuko Yoshida[1], Yoshiki Mochizuki[1], Akihiro Matsushima[1], Manabu Ishii[1], Ryo Umetsu[1], Satomi Asano[1], Naohiko Heida[1], Tetsuya Sakurai[2], Takashi Kuromori[2], Kazuo Shinozaki[2], and Tetsuro Toyoda[1]

[1] Bioinformatics And Systems Engineering division (BASE), RIKEN, JAPAN
toyoda@base.riken.jp[**]
[2] Plant Sciences Center (PSC), RIKEN, JAPAN

1 Abstract

Recent omics sciences are strongly supported by the advances in various detection technologies, such as the next generations high-throughput sequencers, high-resolution imaging devices and remote-sensing technologies. The obtained data, the processed data and knowledge extracted from the data are huge, heterologous, complex and interconnected, and preventing us from sharing or depositing all of the data to the public databases. We have developed a semantic-web-based data repository by using the Semantic Web Folders (SWF for short) technology which is able to share large-scale omics data, experimental raw data, images, sequences and analysed data within a closed community and to publish all or a part of the data from the repository. SWF is designed not only for the data publishing infrastructure, but also as a collaboration tool that allows researchers to generate a database. More concretely, SWF is designed to achieve the following features: 1) implementing an ontology-based data structure including file folders which enables users to write and store various kind of omics data, 2) supporting various data input/output formats including standardized data formats, 3) implementing a data viewer/editor including a meta search engine over SWF data, and 4) providing a communication tool that allow users to collaborate with each other to edit their data. In the current implementation of SWF, we employ a 30 TB Network Attached Strage system that stores a file folder associated with a ontology term. Moreover, more than 100 database projects including public ontology data and RIKEN's original and public databases have been integrated on SWF. SWF is available at `http://database.riken.jp`.

[**] To whom correspondence should be addressed.