

RDF-based data sharing of bio-resource related information

Hiroshi Masuya^{1,2}, Terue Takatsuki¹, Mikako Saito¹, Eiki Takayama¹,
Kazuya Ohshima¹, Nozomu Ohshiro¹, Kai Lenz², Nobuhiko Tanaka¹,
Hiroshi Mori³, Shuichi Kawashima⁴ and Norio Kobayashi^{2,1}

¹RIKEN BioResource Center, 3-1-1 Kouyadai, Tsukuba, Japan

²Advanced Center for Computing and Communication, RIKEN,
2-1, Hirosawa, Wako, Japan

³Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan

⁴Database Center for Life Science, 178-4-4 Wakashiba, Kashiwa, Japan
{hmasuya, takatter, mikakosaito, etakayama, kazuya22, nobtanak}@brc.riken.jp
{ kai.lenz, norio.kobayashi}@riken.jp
hmori@bio.titech.ac.jp, kwsmb@dbcls.rois.ac.jp

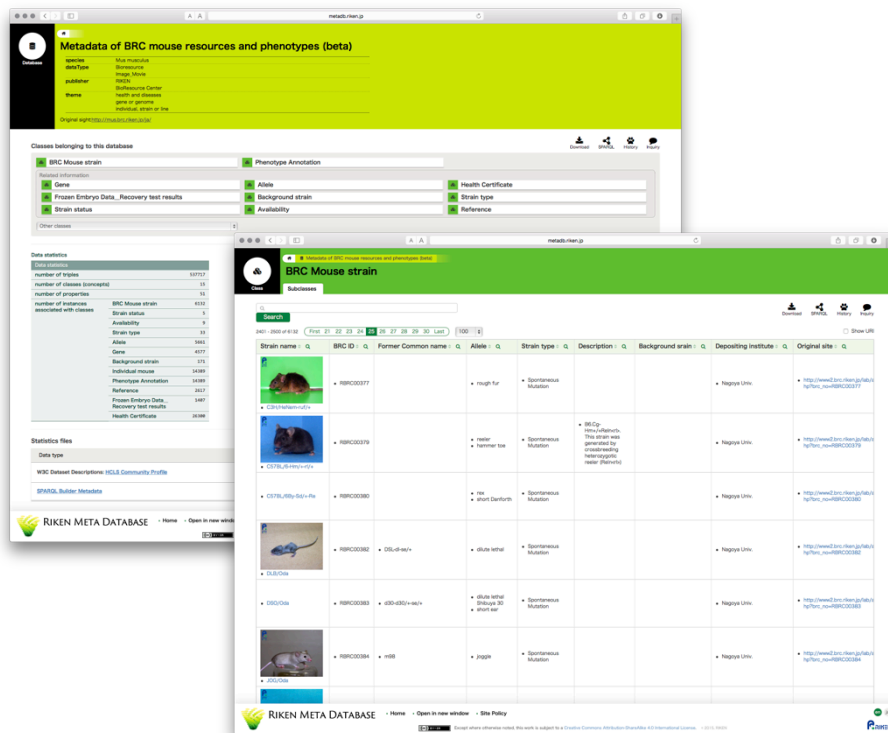
Abstract

“Bio-resources”, commonly used biological materials for experimental studies such as mouse strains, cell lines and microbe culture collections are crucial fundamentals to provide reproducibility and reliability of data in life science. To provide advanced infrastructure of life science, wider-dissemination, quality control and standardization of bio-resources are required. In this sense, data of bio-resources and related information also should be broadly “shared” in life science community. Standardized methodology of data handling across databases and software applications, which helps to maximize utility and re-use of released data is also important issues.

Resource Description Framework (RDF) and Semantic Web technologies provide suitable infrastructure for wide dissemination and re-use of bio-resource related information. Therefore, we have worked out construction of RDF data of bio-resources (mouse strains, rat strains, medaka strains, cell lines and microbe strains) collected from multiple resource centers in Japan. We adopted community-developed data schemas for cell lines (Cell Line Ontology: CLO) and microbes (Microbial Culture Collection Vocabulary (MCCV)). For descriptions of phenotypic properties of bio-resources, we designed common schema links to ontologies of phenotype (e.g. Mammalian Phenotype Ontology and Zebrafish Phenotype Ontology), body parts (e.g. Adult Mouse Anatomy, and Zebrafish Anatomy) and Phenotypic Quality (PATO). Constructed RDF data are available from RIKEN Meta Database (<http://metadb.riken.jp>), which provides web-based interfaces of relational-database like data viewer with table and card interfaces, bulk data download function and SPARQL endpoint. Each database projects in RIKEN Meta Database is accessible from the portal site J-phenome (<http://jphenome.info>).

RDF-version datasets of bio-resources help coordination across multiple databases. Common data schema of bio-resource related datasets easily enables cross-dataset

search of resources showing related phenotypes classified as a specific category. In addition, we are planning to collaborate with MicrobeDB.jp, which is an integrated database of microbial metagenomes, for sharing the latest RDF data of microbial strains in RIKEN BioResource Center. We expect that RDF-based data coordination will contribute to global sharing and improvement of utilities of bio-resources.



Screenshots of the metadata of BRC mouse resources and phenotypes (http://metadb.riken.jp/metadb/db/rikenbrc_mouse)