## **RDF-based data sharing** of bio-resource related information

Hiroshi Masuya<sup>1,2</sup>, Terue Takatsuki<sup>1</sup>, Mikako Saito<sup>1</sup>, Eiki Takayama<sup>1</sup>, Kazuya Ohshima<sup>1</sup>, Nozomu Ohshiro<sup>1</sup>, Kai Lenz<sup>2</sup>, Nobuhiko Tanaka<sup>1</sup>, Hiroshi Mori<sup>3</sup>, Shuichi Kawashima<sup>4</sup> and Norio Kobayashi<sup>2,1</sup>

 <sup>1</sup> RIKEN BioResource Center, 3-1-1 Kouyadai, Tsukuba, Japan
 <sup>2</sup>Advanced Center for Computing and Communication, RIKEN, 2-1, Hirosawa, Wako, Japan
 <sup>3</sup>Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan
 <sup>4</sup>Database 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, kwsm@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 bioresources (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 <u>MicrobeDB.jp</u>, 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.

		A A		metadb.riken.jp				0	00+		
•											
Metadat	a of BRC mous	e resourc	es and phenotypes	(beta)							
species dataType	Mus musculus Bioresource										
publisher	image_Movie RIKEN										
Perse	BoResource Center health and diseases										
	gene or genome										
Original sight hour	Introduce, strain or line										
or gran spreight.	and the second states of the s										
Classes belonging to this	database							Cowritions SPARIAL Has	ory heaty		
BRC Mouse strain			Phenotype Annotation								
Related information											
Gene		Aleie			Health Certificate						
Frozen Embryo Data_Recovery test results		Background strain			Strain type						
Strain status			Availability			Reference					
Other classes											
			000()	_	A A		metado.rii	en.jp	Ċ		0.0
Data statistics		_	A B Meson								
number of triples		\$37717	& BRC Mc	ouse strai	n						
number of classes (concept	ta)	15	Ann Subvision								
number of properties		58	0.004949								
number of instances associated with classes	BRC Mouse strain	6532	9.							-	* 4 *
	Availability		Search							Dow	serveral matery in
	Strain type	33	2401 - 2500 of 6132 First	21 22 23 24 2	5 25 27 28 29 30 Last	a 00					Show
	Aleie	5663	Strain name = Q	BRC ID : Q	Former Common name : Q	Allele : Q	Strain type : Q	Description : Q	Background srain : Q	Depositing institute : Q	Original site : Q
	Gene Backpround strain	4577									
	Individual mouse	14389									
	Phenotype Annotation	14309		• RBRC00377		<ul> <li>rough fur</li> </ul>	Spontaneous     Mutation			<ul> <li>Negoye Univ.</li> </ul>	<ul> <li>http://www2.brc.riken.jp/</li> <li>heifter: novi8880</li> </ul>
	Reference	2617	- 25								
	Recovery test results	2487	<ul> <li>C3P(PleNem-ruf/+</li> </ul>								
	Health Certificate	26300	8					. 8600			
Read and an and a second						. min	· Sportsperus	Hm+/+Rein-rb. This strain was			• http://www.it.box.clean.in.it.
Hattatica mes			and the second s	<ul> <li>RBRC00379</li> </ul>		<ul> <li>hammer toe</li> </ul>	Mutation	generated by crossbreeding		<ul> <li>Nagoya Univ.</li> </ul>	hp?brc_no=RBRC00379
Data type			- CS28 #Junited/a					heterozygotic realiar (Reinvetv)			
W3C Dataset Descriptions	HCLS Community Profile		• Gradaud+40+								
SPAROL Builder Metadata						• rex	Spontaneous			· Name Inte	<ul> <li>http://www2.brc.riken.jp/f</li> </ul>
			<ul> <li>Us/sU/oxy-od/+Re</li> </ul>	<ul> <li>HBHU00380</li> </ul>		<ul> <li>short Denforth</li> </ul>	Mutation			<ul> <li>Hagoya Umic.</li> </ul>	hp?brc_no=RBRC00380
RIKEN META	DATABASE · Hum	e - Open in new v	vinde								
/		(D)····	-								
			and a	<ul> <li>RIRC00382</li> </ul>	• DSL-d-m/+	<ul> <li>diute intrai</li> </ul>	<ul> <li>Spontaneous</li> </ul>			<ul> <li>Nagova Univ.</li> </ul>	<ul> <li>http://www2.brc.riken.jp/</li> </ul>
					·	. own wid	Mutation			gete over	npRes_no=RBRC00382
			· DLB/Oda								
			<ul> <li>DSO/Oda</li> </ul>	<ul> <li>RBRC00383</li> </ul>	<ul> <li>d30-d30/+-se/+</li> </ul>	<ul> <li>dilute lethal Shibuye 30</li> </ul>	Spontaneous     Mutation			<ul> <li>Nagoya Univ.</li> </ul>	<ul> <li>http://www2.brc.riken.jp.</li> <li>he/her. nov.000000000000000000000000000000000000</li></ul>
						<ul> <li>short ear</li> </ul>					
			the second se								
			600	<ul> <li>RBRC00384</li> </ul>	• m98	<ul> <li>joggie</li> </ul>	<ul> <li>Spontaneous Mutation</li> </ul>			<ul> <li>Nagoya Univ.</li> </ul>	<ul> <li>http://www2.brc.riken.jp/ hp?brc.no=RBRC00384</li> </ul>
			<ul> <li>J00/06a</li> </ul>								
			Provide State								
			🥳 RIKEN META	DATABAS	E · Home · Open in new wind	ow · Site Policy					
			<b>V</b>		(( <b>0</b> ) +	igt where otherwise noted	this work is subject to a Ore	athe Commons Attribution Die	arsAlia 4.0 International Science. + 2	115. FIREN	

Screenshots of the metadata of BRC mouse resources and phenotypes (<u>http://metadb.riken.jp/metadb/db/rikenbrc\_mouse</u>)