

RIKEN MetaDatabase

This database portal site is used to provide information on RIKEN's various life science databases to order to help researchers around the world make full use of RIKEN's research results.



Databases Ontologies About RIKEN Meta Database

<http://metadb.riken.jp>

生命科学とLOD—理研メタデータベース を利用したデータ利活用の実際

小林 紀郎

情報基盤センター 計算工学応用開発ユニット

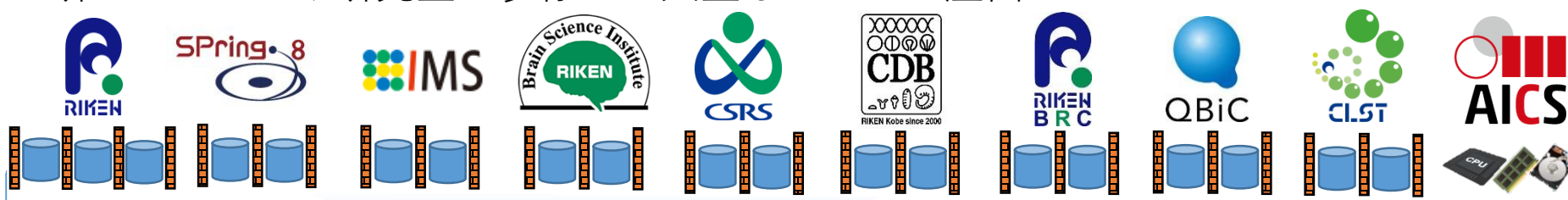


ライフサイエンスはデータ駆動科学 へテロなデータをどう活用していくか？

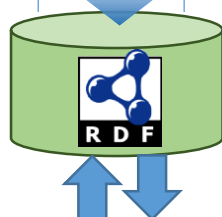
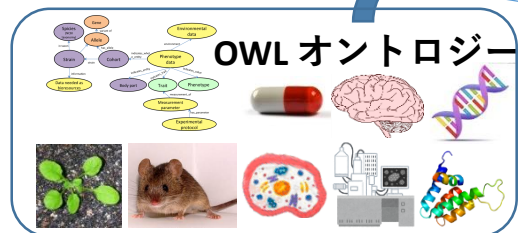
ライフサイエンスは多様化、深化している



理研のセンターや研究室は多様かつ大量なデータを産出



オントロジーを使って
RDF (標準) 化



RIKEN META DATABASE

生命科学Linked Open Data (LOD)の小宇宙

特化した異種研究分野のデータをグローバル規模で統合、活用



メタデータ

超越した、後続の、抽象度の高い、高次の



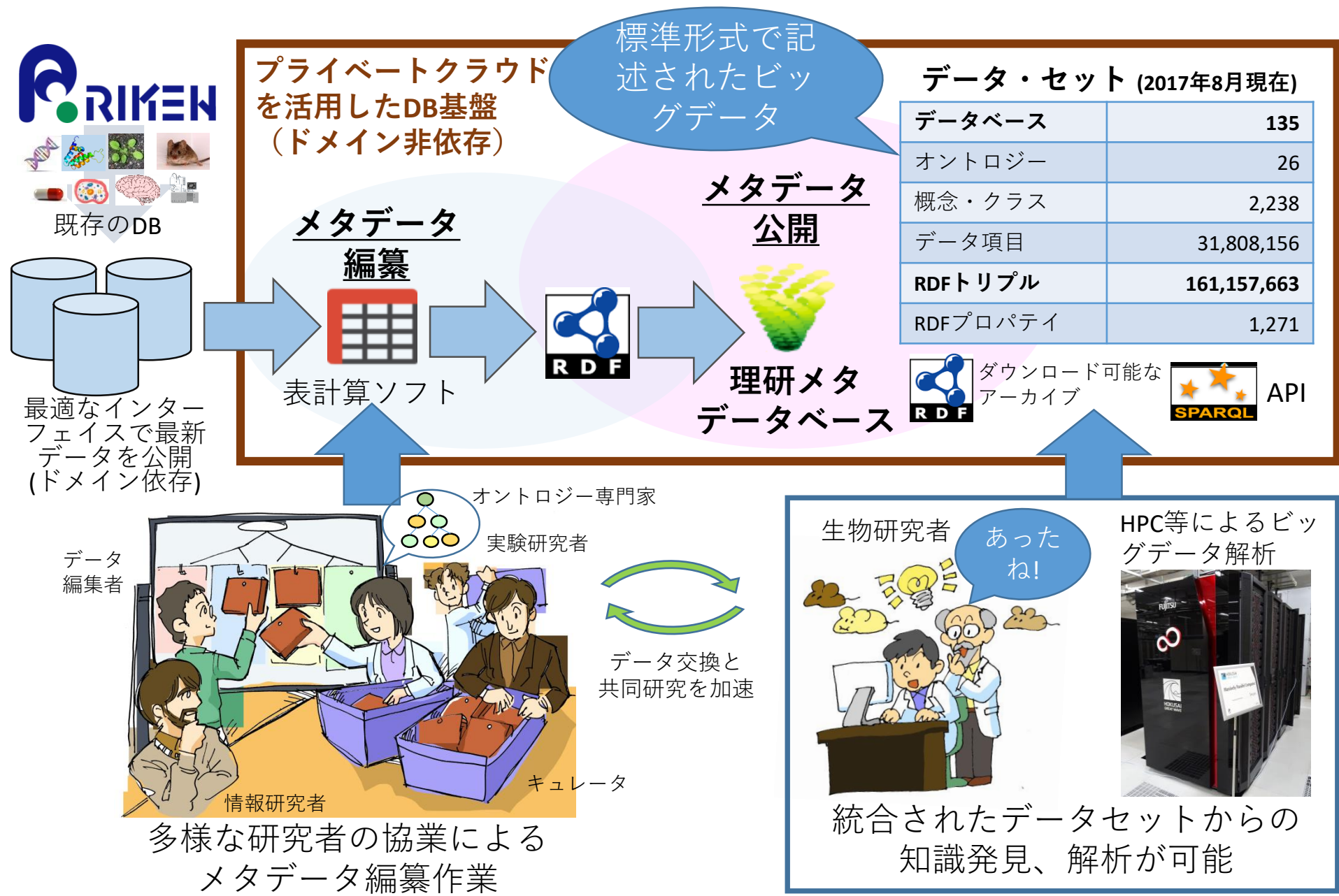
meta data

- データを抽象度の高い次元で書いたもの



- データについてのデータ
- LOD (RDF) として書ける
- 言葉の標準化が大事 （オントロジー標準化）

研究者の協業によるデータ公開作業フロー

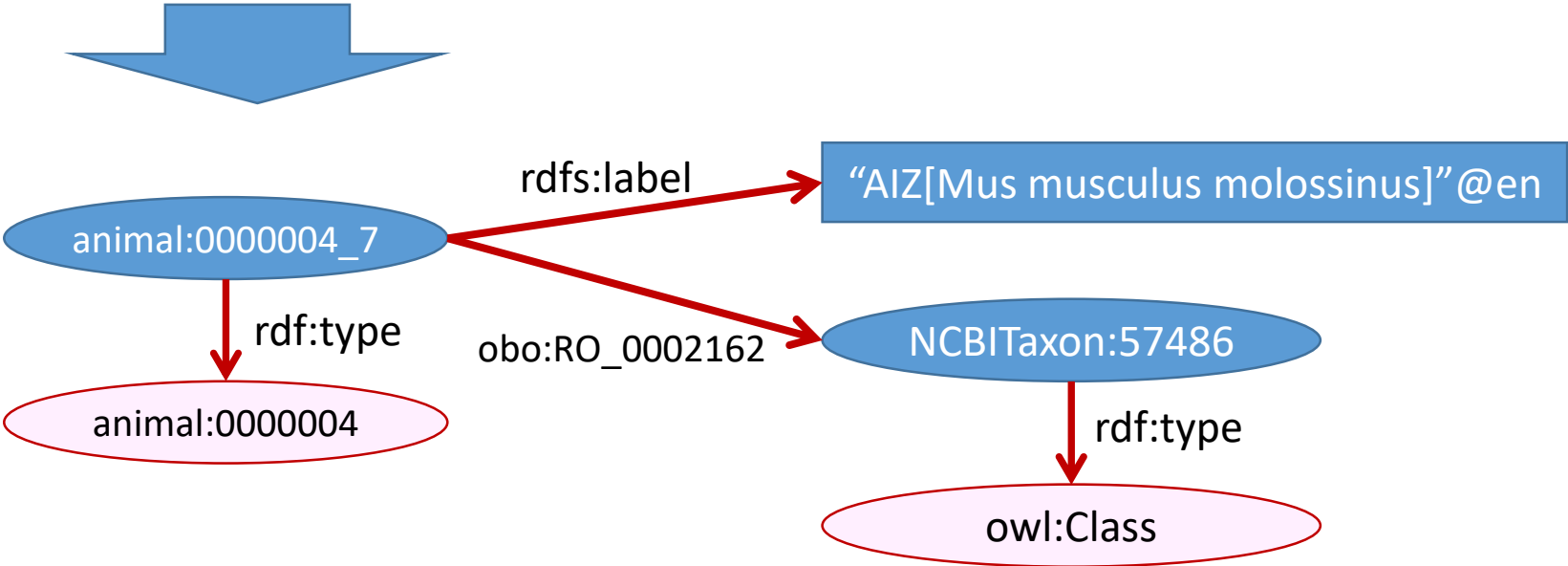


RDFグラフ記述のための表形式データモデル

RDFデータの準備: スプレッドシート (表)

RDF専門家によるデータスキーマ定義
このスキーマは生物学者、オントロジ専門家、情報学者の協業で定義する。

	1	2	3	4
1	English Attribution	Background strain	name	taxon
2	日本語属性	背景系統	名称	生物種
3	Property URI		<code>rdfs:label</code> p	<code>obo:RO_0002162</code> p
4	Data type	<code>animal:0000004</code>	<code>rdf:langString</code>	<code>owl:Class</code>
5		<code>animal:0000004_7</code> s	<code>"AIZ [Mus musculus molossinus]"@en</code> o	<code>NCBITaxon:57486</code> o
6		<code>animal:0000004_10</code>	<code>"AKT [Mus musculus musculus]"@en</code>	<code>NCBITaxon:39442</code>
7		<code>animal:0000004_12</code>	<code>"AST [Mus musculus musculus (wagneri)]"@en</code>	<code>NCBITaxon:39442</code>
8		<code>animal:000000</code>	データ産出者（生物学者）が記述するデータエリア	



RDFグラフの表形式表示

ここでデータ
統合が実現さ
れている

1 - 20 of 11837

1 2 3 4 5 6 7 8 9 10 Last

20

Show URI

Habitat	Environmental feature	Sample
http://www.w3.org/2000/01/rdf-schema#label	http://purl.jp/bio/01/mccv#MCCV_000071	http://purl.jp/bio/01/mccv#MCCV_000072
<ul style="list-style-type: none"> Habitat of JCM 10002 http://purl.jp/bio/01/mccv#MCCV_000007_1 Habitat of JCM 10124 http://purl.jp/bio/01/mccv#MCCV_000007_100 Habitat of JCM 11306 http://purl.jp/bio/01/mccv#MCCV_000007_1000 Habitat of JCM 8723 http://purl.jp/bio/01/mccv#MCCV_000007_10000 	<ul style="list-style-type: none"> plant associated http://purl.jp/bio/11/meo/MEO_0000419 Lathyrus odoratus http://purl.obolibrary.org/obo/NCBITaxon_3859 soil http://purl.jp/bio/11/meo/MEO_0000007 soil http://purl.jp/bio/11/meo/MEO_0000007 plant material http://purl.jp/bio/11/meo/MEO_0000385 	<ul style="list-style-type: none"> Sample from Fasciation of sweet peas (Lathyrus odoratus) http://purl.jp/bio/01/mccv#MCCV_000006_9006645 Sample from Soil, southern area of Vietnam http://purl.jp/bio/01/mccv#MCCV_000006_9006748 Sample from Nematode-suppressive soil, Costa Rica http://purl.jp/bio/01/mccv#MCCV_000006_9007678 Sample from Plant material http://purl.jp/bio/01/mccv#MCCV_000006_9005598

複数の目的語

逆リンク

- 例：マウス表現型アノテーションの検索とソート

Phenotype Annotation	Mammalian Phenotype ontology	Indicates entity	Indicates trait	Indicates value	Annotaiton type	Mouse strain
		tooth x	morphology x			ASC x
<ul style="list-style-type: none"> abnormal tooth morphology in RBRC02414 	<ul style="list-style-type: none"> abnormal tooth morphology 	<ul style="list-style-type: none"> tooth 	<ul style="list-style-type: none"> morphology 	<ul style="list-style-type: none"> abnormal 	<ul style="list-style-type: none"> Annotation imported from MGI (automatical) 	<ul style="list-style-type: none"> A/J.Cg-Nfe2l2<tm1Mym>
<ul style="list-style-type: none"> abnormal tooth morphology in RBRC01390 	<ul style="list-style-type: none"> abnormal tooth morphology 	<ul style="list-style-type: none"> tooth 	<ul style="list-style-type: none"> morphology 	<ul style="list-style-type: none"> abnormal 	<ul style="list-style-type: none"> Annotation imported from MGI (automatical) 	<ul style="list-style-type: none"> B6.129P2-Nfe2l2<tm1Mym>/MymRbr
<ul style="list-style-type: none"> abnormal tooth morphology in RBRC06217 	<ul style="list-style-type: none"> abnormal tooth morphology 	<ul style="list-style-type: none"> tooth 	<ul style="list-style-type: none"> morphology 	<ul style="list-style-type: none"> abnormal 	<ul style="list-style-type: none"> Annotation imported from MGI (automatical) 	<ul style="list-style-type: none"> B6;129-Slc39a13<tm1Thir>

カードビューによる詳細閲覧

cohort_33938_female

Experimental cohort

S

Sex

female

Genotype

Cdh23<v>/Cdh23<v>

Allele

Cdh23<v>

p

Cdh23<v>

Allele analyzed in IMPC

O = S'

Allele analyzed in IMPC

Cdh23<v>

Gene

Linked
Experimental cohort

cohort_33937_female
cohort_33937_male
cohort_33938_female
cohort_33938_male

Linked
Genotype analyzed in IMPC

Cdh23<v>/Cdh23<+>
Cdh23<v>/Cdh23<v>

External
Allele Name

waltzer

External
MGI marker

MGI:1856228

External
MGI marker

Linked
External
class

B6.LIII-Sgk3<fz> Mlph<In> Cdh23<v> Hps1<ep>/Rbrc

Linked
External
BRC mouse strain

B6.LIII-Sgk3<fz> Mlph<In> Cdh23<v> Hps1<ep>/Rbrc

p'

B6.LIII-Sgk3<fz> Mlph<In> Cdh23<v>

Hps1<ep>/Rbrc

BRC Mouse strain

O' = S''

Strain name

B6.LIII-Sgk3<fz> Mlph<In>
Cdh23<v> Hps1<ep>/Rbrc

BRC ID

RBRC00333

Former Common name

LIII

Strain type

Spontaneous Mutation Congenic

データベース: IMPC

クラス: Experimental cohort

インスタンス: cohort_33938_female

カードビュー
でデータ統合
を実現

データベース内で
リンクしている
データ項(逆リンク)

外部データベース
(MGI) へのリンク

外部データベース
がリンクしている
マウス系統
(逆リンク)

その他のビュー

リストビュー (データベース)

1 2 3 4 5 6 All

RIKEN Database Directory

Classes List of databases > RIKEN organizations or centers > Topics of databases >

Databases

FANTOM5 SSTAR

http://metadb.riken.jp/db/SciNetS_ria365i

Classes Human Sample > Transcription factors, relative expression of human sample > CAGE peaks of human sample > Mouse Sample > Transcription factors, relative expression of mouse sample > CAGE peaks of mouse sample >

FANTOM4

http://metadb.riken.jp/db/SciNetS_ria187i

Classes Array_Design > Experiment > GEO > Cell line analyzed in FANTOM 4 > Antibody > Barcode_Sequence > Genome_Assemblies > Tag_Sequence > Vector > Sequencer > Tissue_Type > Mouse Strain analyzed in FANTOM 4 > Se Collaborator > Developmental_Stage > Knockdown TF genes in FANTOM4 > Prim Source > Protocol > Detailed_Protocol >

Parameter Series > Lot >

Fantom_Data Analyzed_Data > Primary_Data > Mapping_Data >

FANTOM eeDB

http://metadb.riken.jp/db/SciNetS_ria281i

Classes Edge Express DB >

ツリー
ビュー

リストビュー (オントロジー)

ChEBI <http://metadb.riken.jp/metadb/ontology/ChEBI>
Chemical Entities of Biological Interest (ChEBI)
(Original site: <http://www.ebi.ac.uk/chebi/>)

CL <http://metadb.riken.jp/metadb/ontology/CL>
Cell Ontology
(Original site: <http://cellontology.org>)

CLO <http://metadb.riken.jp/metadb/ontology/CLO>
Cell Line Ontology (CLO)
(Original site: <http://www.clo-ontology.org>)

GO <http://metadb.riken.jp/metadb/ontology/GO>
Gene Ontology (GO)
(Original site: <http://geneontology.org>)

LU <http://metadb.riken.jp/metadb/ontology/LU>
Left Unit Ontology
(Original site:)

- mammalian phenotype 0
 - adipose tissue phenotype 0
 - behavior/neurological phenotype 0
 - cardiovascular system phenotype 3590
 - cellular phenotype 0
 - abnormal cell morphology 167
 - abnormal cell physiology 421**
 - abnormal extracellular matrix morphology 7
 - genetic imprinting 19
 - craniofacial phenotype 4
 - digestive/alimentary phenotype 0
 - embryogenesis phenotype 0
 - endocrine/exocrine gland phenotype 0
 - growth/size/body phenotype 0
 - hearing/vestibular/ear phenotype 0
 - hematopoietic system phenotype 0
 - homeostasis/metabolism phenotype 0
 - immune system phenotype 2221
 - integument phenotype 0

URI
http://purl.obolibrary.org/obo/MP_0005621

Label
abnormal cell physiology

SubClassOf
[cellular phenotype](#)

definition
any functional anomaly of the vital processes of the cell

has_obo_namespace
MPPheno.ontology

http://www.geneontology.org/formats/oboInOwl#id
MP:0005621

Linked from
[Phenotype annotation with MP \(RDF of MGI data \(beta version\)\)](#)
[Phenotype Annotation \(Metadata of BRC mouse resources and phenotypes \(beta\)\)](#)

その他のビュー

SPARQL実行ビュー



Output Format

HTML

SPARQL Query

```
1 SELECT DISTINCT ?Concept
2 FROM <http://metadb.riken.jp/db/rikenbrc_mouse>
3 WHERE [
4   [] a ?Concept
5 ] LIMIT 100
```

Run Query

Concept

<http://www.w3.org/2002/07/owl#Class>

<http://www.w3.org/2000/01/rdf-schema#Class>

<http://rdf.ncbi.nlm.nih.gov/pubmed>














ダウンロードビュー



Metadata of BRC mouse resources and phenotypes

Metadata of BRC mouse resources and phenotypes database download page

Download RDF data files

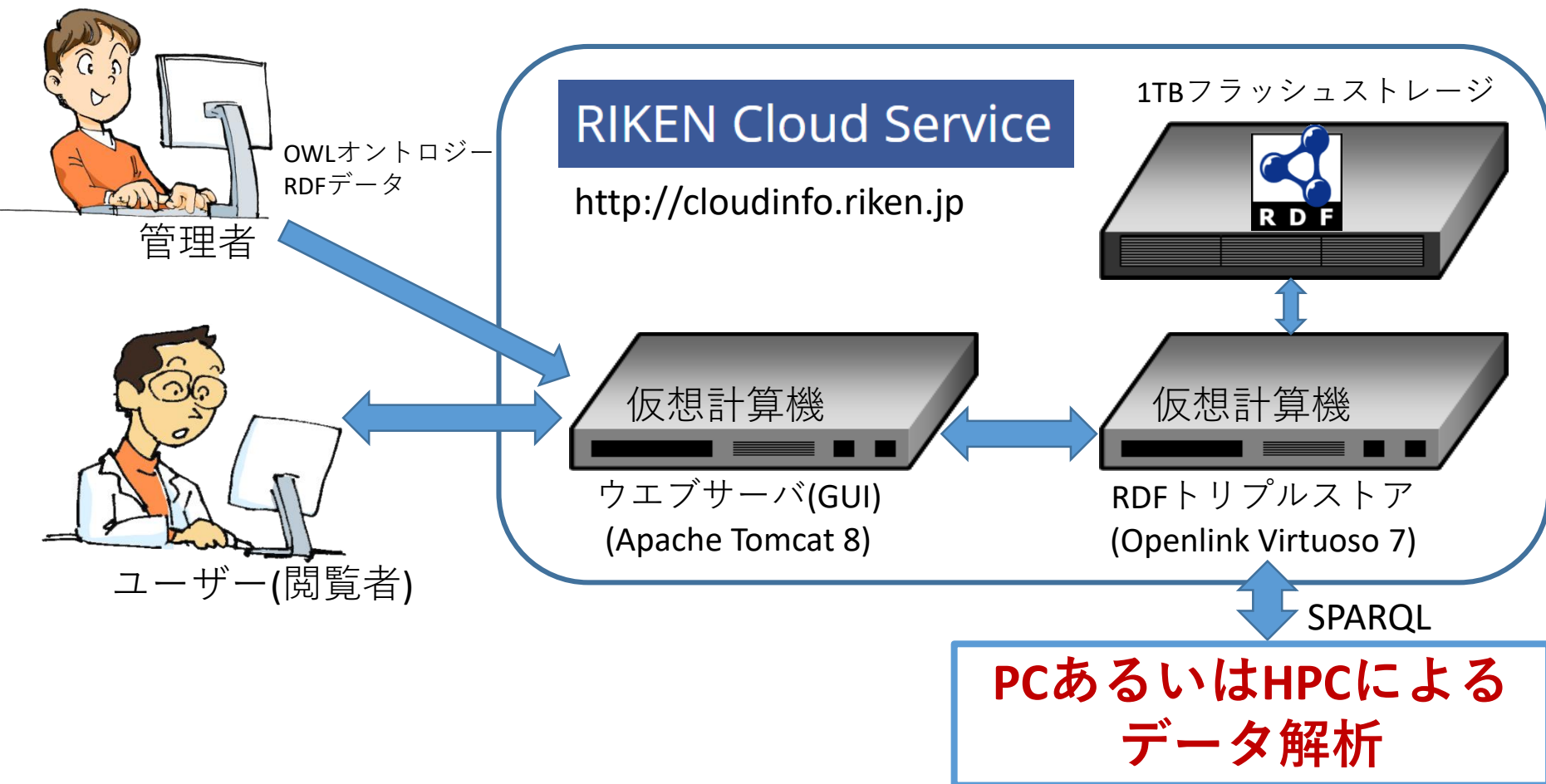
Class name	RDF in Turtle format	RDF in RDF/XML format
 Database archives View Data	 rikenbrc_mouse Turtle.zip 4.1MB	 rikenbrc_mouse RDF XML.zip 3.6MB
BRC Mouse strain	 rikenbrc_mouse-01-BRC Mouse strain Turtle.zip 1.8MB	 rikenbrc_mouse-01-BRC Mouse strain RDF XML.zip 1.6MB
Allele	 rikenbrc_mouse-02-Allele Turtle.zip 150.6KB	 rikenbrc_mouse-02-Allele RDF XML.zip 130.0KB
Gene	 rikenbrc_mouse-03-Gene Turtle.zip 47.5KB	 rikenbrc_mouse-03-Gene RDF XML.zip 40.4KB
Phenotype Annotation	 rikenbrc_mouse-04-Phenotype Annotation Turtle.zip 1.0MB	 rikenbrc_mouse-04-Phenotype Annotation RDF XML.zip 1.0MB
Health Certificate	 rikenbrc_mouse-05-Health Certificate Turtle.zip 928.6KB	 rikenbrc_mouse-05-Health Certificate RDF XML.zip 679.7KB

SPARQL: プログラムからアクセス可能な、HTTPプロトコルを介した標準化されたAPI

概念クラス毎にまとめられたデータアーカイブ

実装と運用

単純・軽量なアーキテクチャ



- プライベートクラウド「理研クラウドサービス」上に配備
- 高々2台の仮想計算機で実行可能 (軽量).

データベース

- [RIKEN Database Directory](#)

Mammalian

- [FANTOM5 SSTAR](#)
- [FANTOM4](#)
- [FANTOM eeDB](#)
- [Metadata of BRC mouse resources and phenotypes](#)
- [Metadata of BRC cell resources](#)
- [The RIKEN integrated database of mammals](#)
- [NBRP Medaka Phenotype Metadata](#)
- [Metadata of NBRP Rat](#)
- [NIG Mouse Phenotype Database Metadata](#)
- [IMPC_RDF](#)
- [Metadata of Functional Glycomics with KO mice database \(beta\)](#)
- [RDF of MGI data](#)
- [NIG Zebrafish\(beta\)](#)



Plant

- [FOX Hunting](#)
- [RIKEN Arabidopsis Activation tagging line](#)
- [The Rice Gene From RAP-DB](#)
- [RIKEN Plant gene family](#)
- [RIKEN Phenome Integration of Arabidopsis mutants](#)
- [RIKEN RARGE Promoter](#)
- [Gramene O.sativa GO Annotation \[Data\]](#)
- [RIKEN RARGE Transposon Mutant](#)
- [RIKEN RAFL cDNA](#)
- [RIKEN Arabidopsis Phenome Information Database](#)
- [RIKEN RARGE Alternative Splicing](#)
- [RIKEN Arabidopsis Transcription Factor database](#)
- [ARTADE2 predicted gene model database](#)



21 public ontologies refer to 113 databases:



62 databases and external 51 databases,
(total 140 databases)

Protein

- [RIKEN Protein Database](#)
- [RIKEN Harima Heavy-atom Database \[Data\]](#)
- [Bacpedia \(Riken Harima SPring8-Center\)](#)
- [Bacpedia data download project](#)
- [Pfam](#)
- [Protein Data Bank](#)



Others

- [Metadata of JCM resources](#)
- [Metadata of quantitative data and datasets of microscopy images provided from SSBD database](#)
- [SSBC \(Systems and Structural Biology Center\)](#)
- [RAPID: Resource of Primary Immunodeficiency Diseases](#)

Ontology

- [Metagenome and Microbes Environmental Ontology](#)
- [Chemical Entities of Biological Interest \(ChEBI\)](#)
- [Cell Ontology](#)
- [Gene Ontology \(GO\)](#)
- [Mouse Adult Gross Anatomy \(MA\)](#)
- [Microbial Culture Collection Vocabulary \(MCCV\)](#)
- [Medaka fish anatomy and development](#)
- [Mammalian Phenotype \(MP\)](#)
- [NCBI Organismal Classification \(NCBITaxon\)](#)
- [Phenotypic Quality \(PATO\)](#)
- [OBO-relation ontology \(RO\)](#)
- [Semanticscience Integrated Ontology](#)
- [Uber Anatomy Ontology \(UBERON \)](#)
- [Units of Measurement \(UO\)](#)
- [Zebrafish Anatomy and Development \(ZFA\)](#)

理研データベース電話帳

RIKEN Database Directory

List of databases

This is a catalog of databases operated in the RIKEN institute. You can filter this table by clicking icons of magnifying glass at columns.

Instance

1 - 20 of 123 1 2 3 4 5 6 7 20

Download SPARQL History Inquiry Show URI

List of databases	title	alternative	homepage	Metadata link	publisher	description	language	species	theme	dataType	reference
				ASC x					ASC x		
FANTOM4	Functional Annotation of the Mammalian Genome 4	FANTOM4	http://fantom.gsc.riken.jp/4/	http://metadb.riken.jp/metadb/db/SciNetS_ria187	Center for Life Science Technologies Chief Scientist Laboratories RIKEN	In FANTOM4 the focus has changed to understanding how these components work together	eng jpn	Homo sapiens	DNA clones cell experimental procedures gene or	GeneExpression Image_Movie Interaction_Pathway Sequence	19377474 19377475 19377478 20211142
Systems and Structural Biology Center	Systems and Structural Biology Center	SSBC	http://metadb.riken.jp/metadb/db/SciNetS_ria461	http://metadb.riken.jp/metadb/db/SciNetS_ria461	Advanced Center for Computing and Communication RIKEN	Imported from SciNetS	eng		DNA clones experimental procedures protein		
Bacpedia	Bacpedia beta in SPring8-Center	Bacpedia	http://bacpedia.harima.riken.jp/bacteria/View/Frm/Contents/topview.aspx	http://metadb.riken.jp/metadb/db/SciNetS_rib2201	Advanced Center for Computing and Communication RIKEN	Available methods for the studies of protein 3-D structures include X-ray crystallography,	eng		experimental procedures protein		23633602
RIKEN BRC Cell Bank	RIKEN BRC Cell Bank	Cell search system	http://www2.brc.riken.jp/lab/cell/search.php	http://metadb.riken.jp/metadb/db/rikenbrc_cell	BioResource Center RIKEN		eng jpn	Homo sapiens Mus musculus	cell health and diseases phenotype or phenome	Bioresource	
RIKEN BRC Mouse Catalog	RIKEN BRC Mouse Catalog	Animal Search System	http://www2.brc.riken.jp/lab/animal/search.php	http://metadb.riken.jp/metadb/db/rikenbrc_mouse	BioResource Center RIKEN		eng jpn	Mus musculus	gene or genome health and diseases individual, strain or line	Bioresource Image_Movie	19448331

- 理研が公開しているデータベースのリスト
- オールジャパンのライフサイエンスデータベースカタログ“Integbio”や、“W3C Database Description: HCLS Community Profile”の語彙に準拠

活用事例：J-phenomeプロジェクト (<http://jphenome.info>)

Phenotype: 生物に現れた性質のこと

目的: RDFによる実験動物の表現型データの統合



国際連携、日本で産出された
表現型データ・セット

RDF化された統合データを
理研メタデータベースから2次公開

(1)

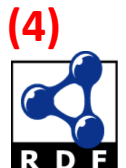
マウス

ラット

魚類

細胞

EXPERIMENTAL ANIMAL DIVISION BRC



CARD J-CRED

JCGGDB

IMPC

NBRP strains



Medaka strains



Zebrafish Gene trap

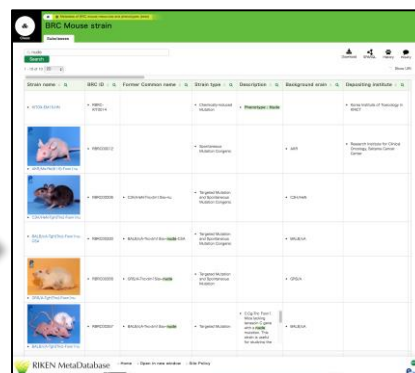
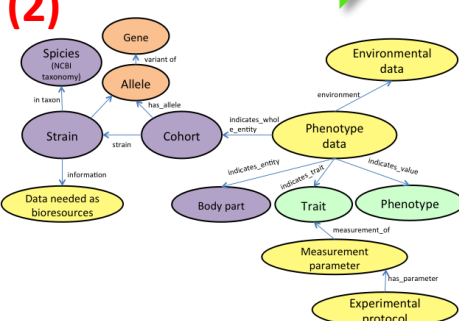


RIKEN Cell bank



共通スキーマ
RDF化

(2)

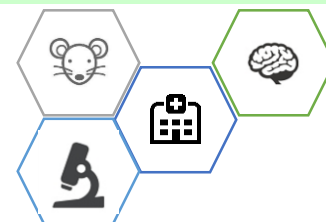


Reactome

UniProt

EMBL-EBI
RDF Platform

SPARQLエンド
ポイントを活用
した生物種横断
データ統合



RDFに詳しくな
い研究者にも
データ提供

横断検索による分子データ等と表現型データの統合：
新しいデータの価値を与える

International Mouse Phenotyping Consortium

国際マウス表現型コンソーシアム

<https://www.mousephenotype.org>



The goal of the International Mouse Phenotyping Consortium (IMPC) is to discover functional insight for every gene by generating and systematically phenotyping 20,000 knockout mouse strains.

NATURE | ARTICLE

nature

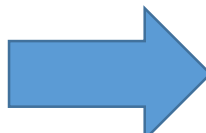


High-throughput discovery of novel developmental phenotypes

Nature (2016) | doi:10.1038/nature19356

Received 12 November 2015 | Accepted 10 August 2016 | Published online 14 September 2016

- 標準化された実験プロトコルによる遺伝子網羅的解析
- 既に2,500遺伝子の表現型データが公開されている



理研で
RDF化



世界初の表現型ビッグデータ

- 約26,000の表現型アノテーション
 - 1,150,000の解析結果 (P-値)
- 総計: **5000万RDFトリプル**