

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.



en ja



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

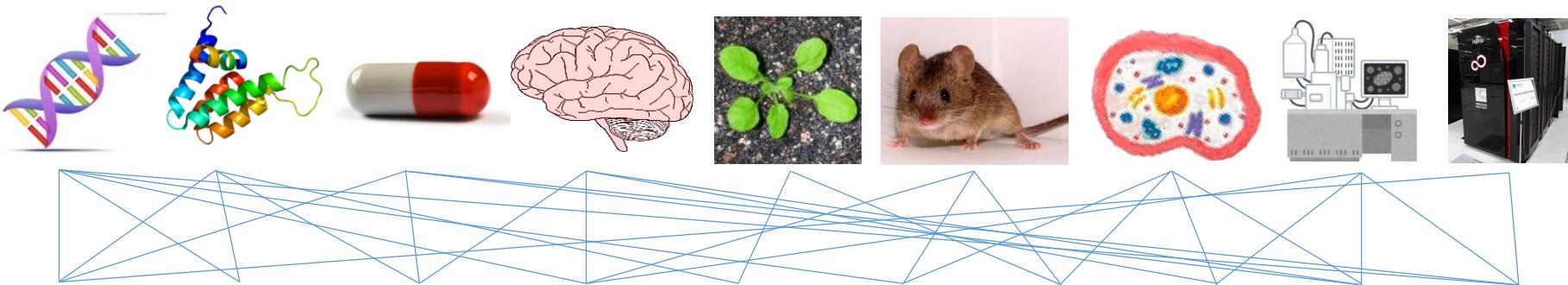
小林 紀郎

情報基盤センター 計算工学応用開発ユニット

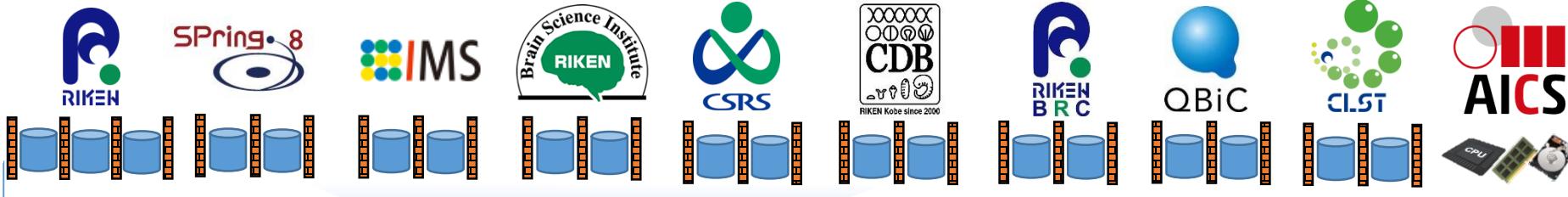


ライフサイエンスはデータ駆動科学 へテロなデータをどう活用していくか？

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



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

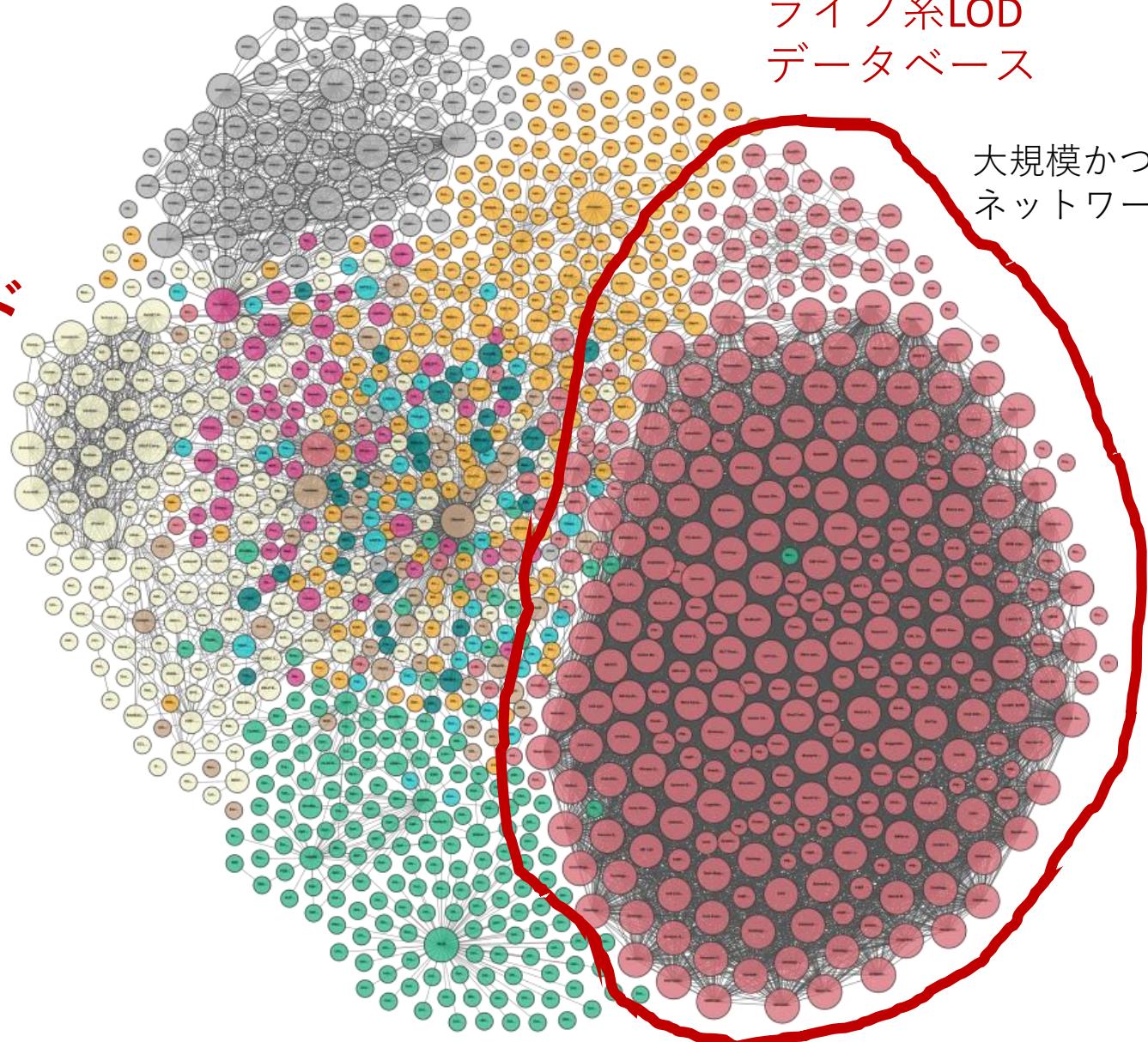


Legend	
Cross Domain	
Geography	
Government	
Life Sciences	
Linguistics	
Media	
Publications	
Social Networking	
User Generated	
Incoming Links	—
Outgoing Links	—

ライフ系LOD
データベース

大規模かつ緻密な
ネットワーク

LODクラウド



"Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>"

メタデータ

超越した、後続の、抽象度の高い、高次の

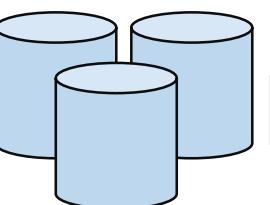
meta data

- 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



既存のDB



最適なインターフェイスで最新データを公開
(ドメイン依存)

プライベートクラウド
を活用したDB基盤
(ドメイン非依存)

メタデータ
編纂

表計算ソフト



メタデータ
公開



理研メタ
データベース

データ・セット (2017年9月現在)

データベース	135
オントロジー	26
概念・クラス	2,238
データ項目	31,808,156
RDFトリプル	161,157,663
RDFプロパティ	1,271



ダウンロード可能な
アーカイブ



API

データ
編集者



多様な研究者の協業による
メタデータ編纂作業

オントロジー専門家

実験研究者

キュレータ

データ交換と
共同研究を加速

生物研究者

あった
ね!



統合されたデータセットからの
知識発見、解析が可能

HPC等によるビッグデータ解析



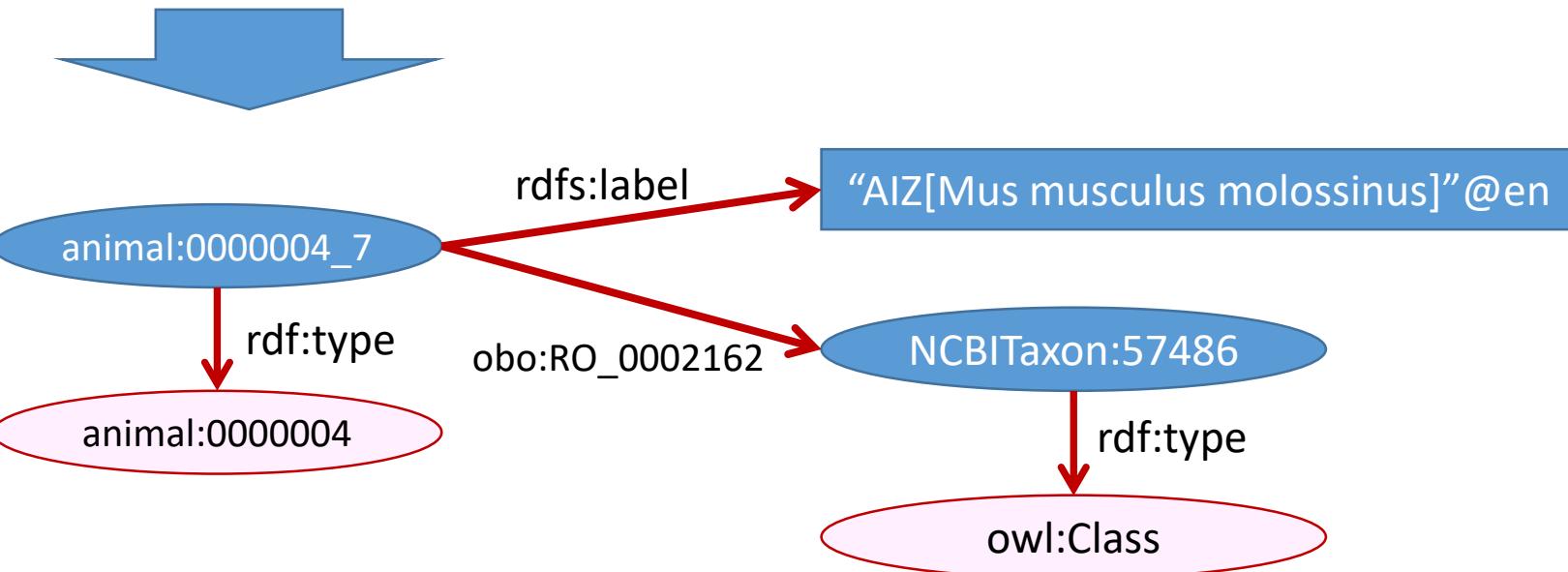
RDFグラフ記述のための表形式データモデル

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		rdfs:label p	obo:RO_0002162 p
4 Data type	animal:0000004	rdf:langString	owl:Class
5	animal:0000004_7 S	"AIZ [Mus musculus molossinus]"@en O	NCBITaxon:57486 O
6	animal:0000004_10	"AKT [Mus musculus musculus]"@en	NCBITaxon:39442
7	animal:0000004_12	"AST [Mus musculus musculus (wagneri)]"@en	NCBITaxon:39442
8	animal:0000004	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 ▾

Habitat	Environmental feature	Sample
<ul style="list-style-type: none"> Habitat of JCM 10002 http://purl.jp/bio/01/mccv#MCCV_000007_1 	<p>複数の目的語</p> <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 	<p>逆リンク</p> <ul style="list-style-type: none"> Sample from Fasciation of sweet peas (<i>Lathyrus odoratus</i>) 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
<ul style="list-style-type: none"> Habitat of JCM 10124 http://purl.jp/bio/01/mccv#MCCV_000007_100 		
<ul style="list-style-type: none"> Habitat of JCM 11306 http://purl.jp/bio/01/mccv#MCCV_000007_1000 		
<ul style="list-style-type: none"> Habitat of JCM 8723 http://purl.jp/bio/01/mccv#MCCV_000007_10000 		

- 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
• abnormal tooth morphology in RBRC02414	• abnormal tooth morphology	• tooth	• morphology	• abnormal	• Annotation imported from MGI (automatical)	• A/J.Cg-Nfe2l2<tm1Myrm>
• abnormal tooth morphology in RBRC01390	• abnormal tooth morphology	• tooth	• morphology	• abnormal	• Annotation imported from MGI (automatical)	• B6.129P2-Nfe2l2<tm1Myrm>/MyrmRbc
• abnormal tooth morphology in RBRC06217	• abnormal tooth morphology	• tooth	• morphology	• abnormal	• Annotation imported from MGI (automatical)	• 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

p

Cdh23<v> o = s'

Allele analyzed in IMPC

- Cdh23<v>

Gene



Experimenal cohort

- cohort_33937_female
- cohort_33937_male
- cohort_33938_female
- cohort_33938_male



Genotype analyzed in
IMPC

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



Allele Name

- waltzer



MGI

- MGId:1856228

External
MGI marker

Linked External
class

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

Linked External
BRC mouse strain

p'

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

B6.LIII-Sgk3<fz> Mlph<ln> Cdh23<v>
Hps1<ep>/Rbrc
BRC Mouse strain
o' = s''

Strain name
p''

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

O''

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_ria365

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_ria187

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_ria281

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

MPheno.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

RIKEN MetaDatabase
SPARQL Endpoint

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>

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

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

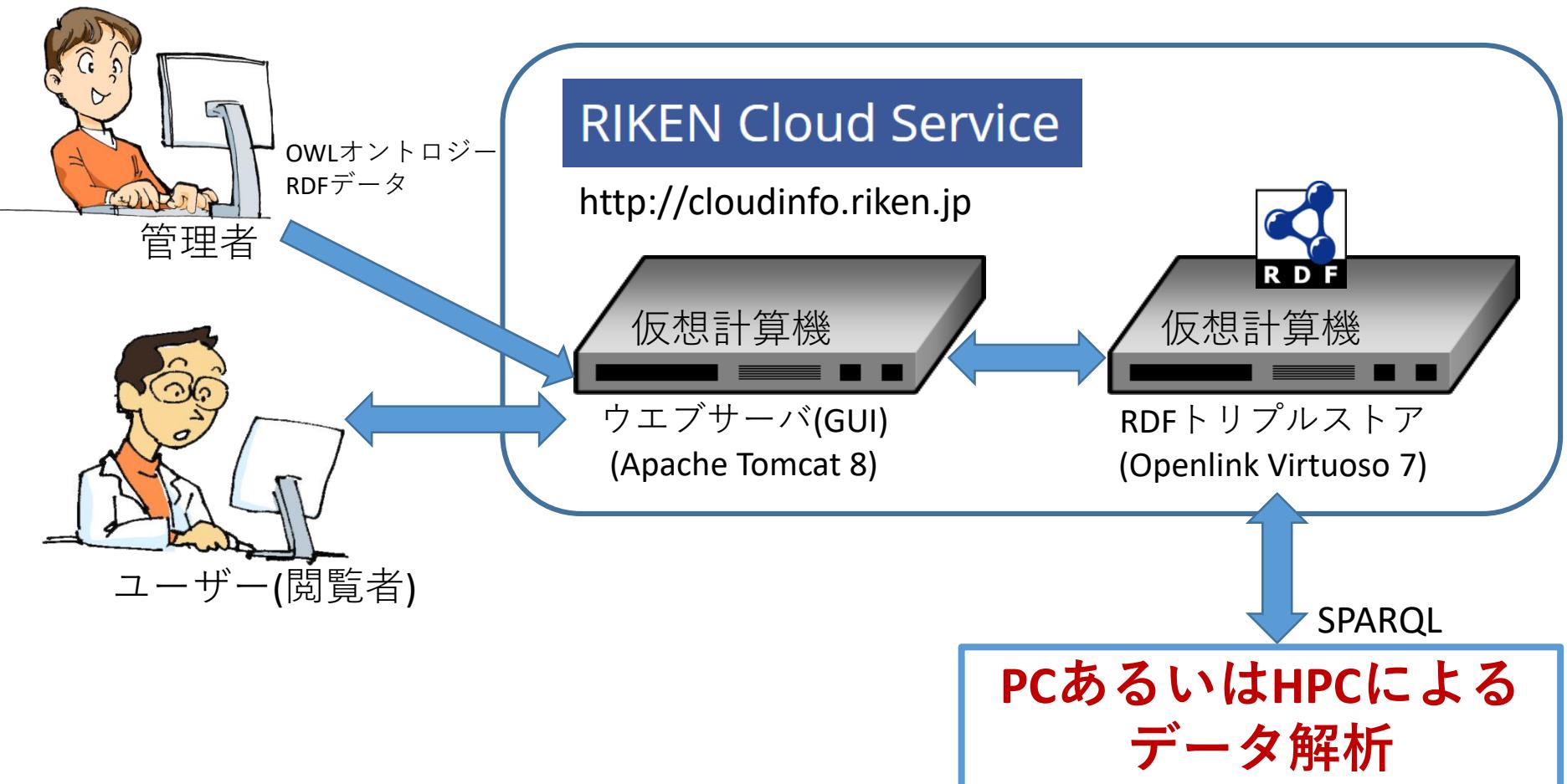
Download RDF data files

Class name	RDF in Turtle format	RDF in RDF/XML format
Database archives	rikenbrc_mouse_Turtle.zip 4.1MB View Data ▾	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

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ー



- 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)

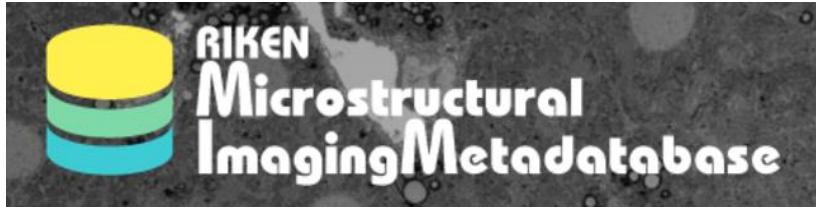
理研データベース電話帳

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.													
Class		Instance											
1 - 20 of 123													20
List of databases	title	alternative	homepage	Metadata link	publisher	description	language	species	theme	dataType	reference		
• FANTOM4	• Functional Annotation of the Mammalian Genome 4	• FANTOM4	• http://fantom.gsc.riken.jp/	• http://metadb.riken.jp/metadb/db/SciNet5_rn187	• Center for Life Science Technologies • Chief Scientist Laboratories • RIKEN	• In FANTOM4 the focus has changed to understanding how these components work together	• eng • jap	• 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/SciNet5_rie46i	• http://metadb.riken.jp/metadb/db/SciNet5_rie46i	• 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/SciNet5_rlb220i	• Advanced Center for Computing and Communication • RIKEN	• Available methods for the studies of protein 3-D structures include X-ray crystallography,	• eng					• 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 • jap	• Homo sapiens • Mus musculus	• cell • health and diseases • phenotype or pheno	• 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 • jap	• 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

電子顕微鏡画像データベース

理研CLST 久米研究員との共同研究



	Thumbnail	ID	Description	RIKEN Metadatabase URL
Species		DZI_RatLiverNCMIR002	DZI for Rat Liver NCMIR Images Obtained by SEM	DZI_NormalRatLiverNCMIR002
Rat Human		DZI_RatLiver001	DZI for Rat Liver Images Obtained by SEM	DZI_NormalRatLiver001

メタデータを取得し、
画像のリストを表示

SPARQL

メタデータ

オントロジー

RIKEN META DATABASE

Project

- Project
- Experiment
- Experimenter G...

BioSample

- Biosample
- Phenotype Data of Biosample
- derives_form (in...

- Animal Strain Used in This Database
- Sample Preparation
- Staining Method...

- EM Staining class
- Sample conteiner

Imaging Condition

- Electron Microscope Imaging
- Accellerating Voltage
- Manufacture spe...

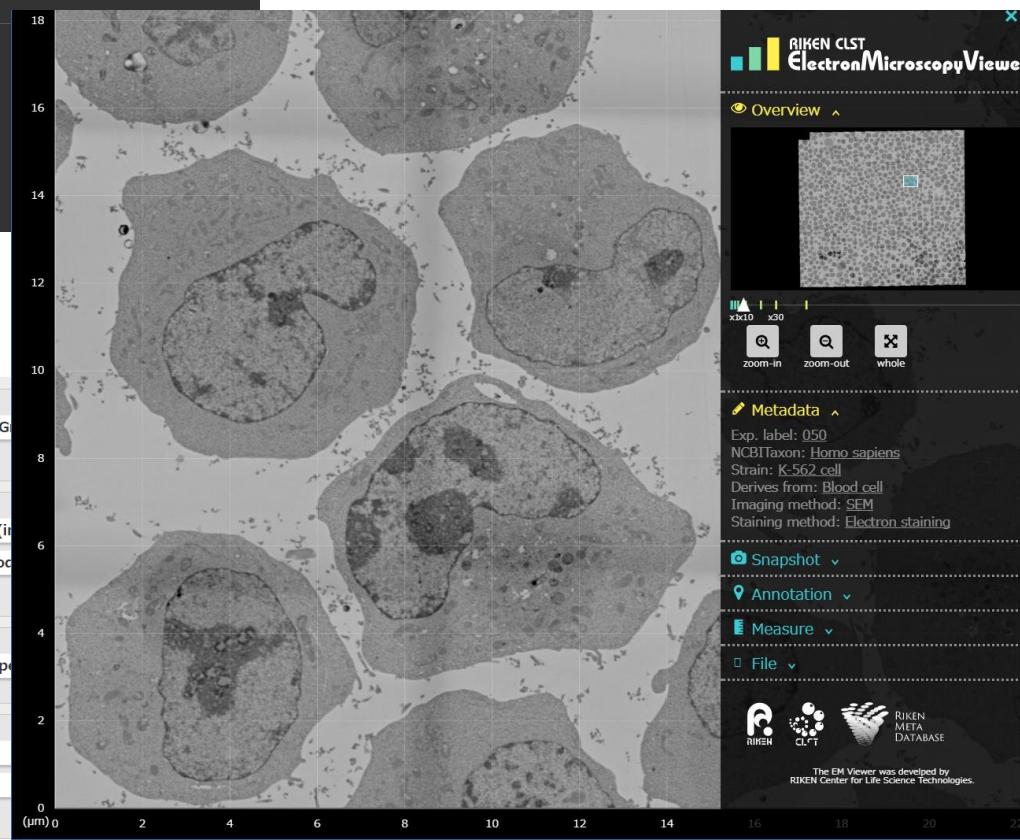
Image

- Image
- Phenotype Data of Images
- Dataset

- DZI Dataset
- Tiling Map
- Tiling Image

高性能ビューワを備えた、研究・教育現場で活用可能な画像公開基盤

メタデータと画像を表示：
画像、実験の情報がわかる



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

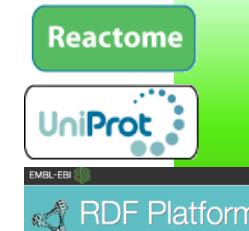
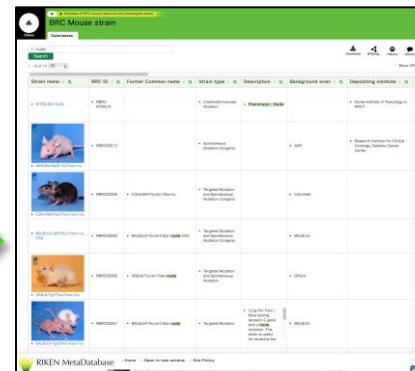
国際連携、日本で産出された表現型データ・セット



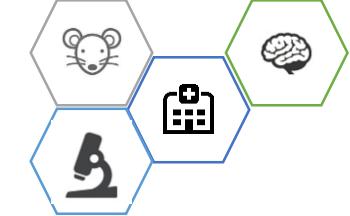
魚類
細胞



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



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

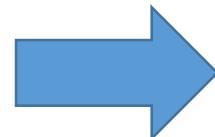


High-throughput discovery of novel developmental phenotypes

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