OLSP Ontology working group

Hiroshi Masuya
RIKEN BioResource Center (BRC)

OLSP Ontology Working Group (2020~)

Promoting the interoperability of life science data at RIKEN through the development of a "Shared Vocabulary"

- Conduct a survey on the usage and demand for shared vocabularies, such as ontologies and public IDs.
- Examine and propose actionable strategies to accelerate data interoperability for cross-domain integration.

BDR	Itoga, H., Wang, FF., Onami, S.
CSRS	Takahashi, M., Arita, M.
CBS	Okumura, Y., Imaizumi, I.
BRC	Kushida, T., Masuya, H.*
R-IH	Kato, M., Yamagata, Y., Kobayashi, N.**
RAP	Yokota, H.
IMS	Kasukawa, T.**

* Chair
** Observer

A survey of research topics covered by RIKEN's database (2021)

■ As main topic　　■ As metadata of main topic

Databases operated by each center

(Rows labeled: BDR, BRC, IMS, RAP, CBS)

Highlighted column group headers:
- Biological entity interested (omics level)
- Biological process interested
- Biological property interested
- Experimental procedure for data production

Summary of the survey

- Overall, it is rich in variety, but the topics are relatively concentrated within each center

- Many datasets contains genes, but they are not interconnected due to variations such as species differences

- To connect datasets from RIKEN, it is essential to introduce public data for the life sciences

- Ortholog data is considered useful for linking RIKEN data and demonstrating its utility in disease research and other applications

- OMA (https://omabrowser.org/oma/) developed in SIB, which is available as RDF, is a strong candidate of the ortholog data

- (We collaborate with SIB through RIKEN hackathons)

An example in RIKEN BRC (using RDF)

(An example of mouse resource)

bioresource

RIKEN BRC mouse---> MGI gene

RIKEN BRC Mouse RDF Graph

mouse gene

NCBI Mouse gene <- MGI gene

MGI RDF Graph

mouse gene

NCBI mouse gene <---- Orthologs Cluster --> OMA human protein ->

OMA human gene -> NCBI/Ensembl human gene

OMA RDF Graph

human gene

disease

NCBI/Ensembl human gene <- Gene-Disease Association -> UMLS disease

human gene

DisGeNET RDF Graph
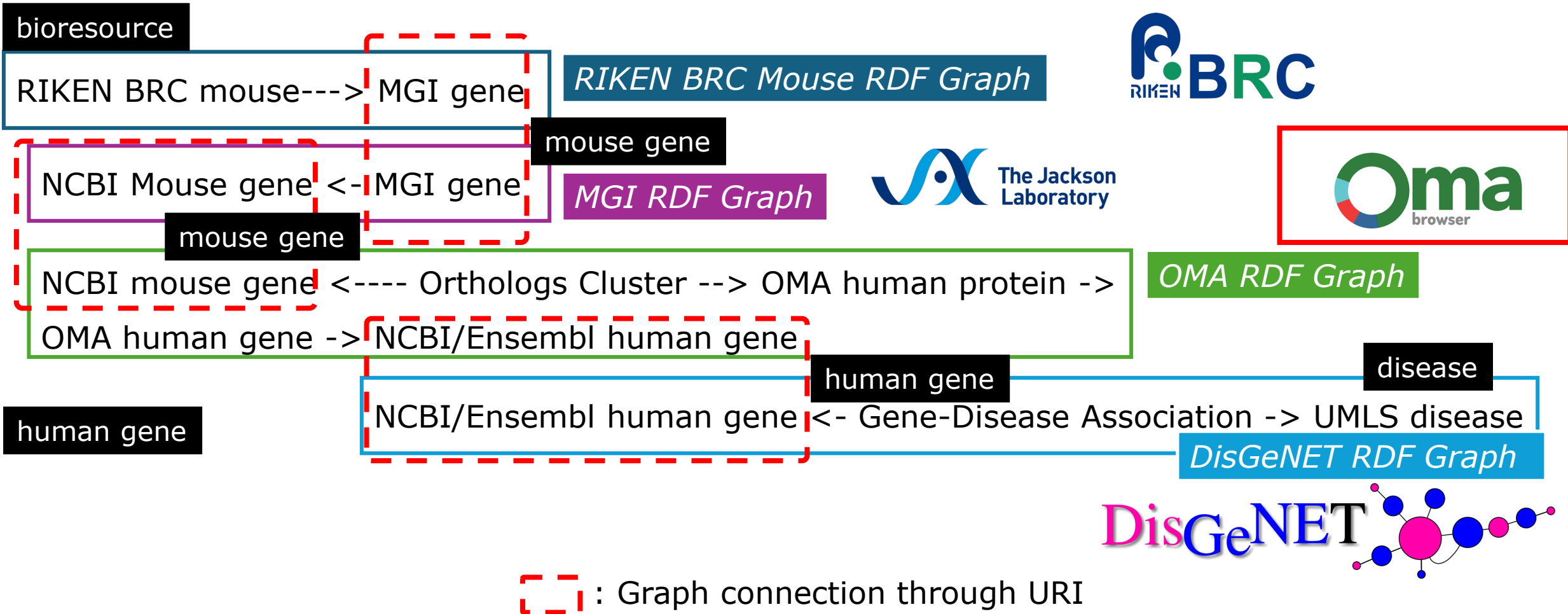
┌─ ─ ─┐
└─ ─ ─┘ : Graph connection through URI

RIKEN has a data integration platform using RDF.

RIKEN Metadatabase (2016-) https://metadb.riken.jp/

Report: How should we address URI inconsistencies in the RIKEN MetaDB?

Masuya, H., Yamamot, Y., Kushida, T., Kato, M (2024)

https://docs.google.com/document/d/1oMXUrNCK2oc5GuuJNbU_93jiOF-X0Ao87hOyLUt0tB8/edit?usp=sharing

A key issue in RDF-based data integration is the failure to connect data due to URI inconsistencies. This report proposes practical solutions for addressing URI inconsistencies in RIKEN MetaDB operations for the RIKEN OLSP.

Conclusion:
- Using the data registration system of the RIKEN MetaDB, correcting URIs during data registraiton is efficient considering cost and performance after the change.
- However, it will require rewriting existing data.
- The modification targets should be limited (e.g., genes, species, proteins, diseases).
- Overview of the modification algorithm: A conversion table for the modification targets will be prepared. After correcting typos with tools like RDFDoctor, the conversion table will be referenced to correct any inconsistencies.

Development of a system on the OLSP website to offer introductions on the RIKEN's databases (2023)
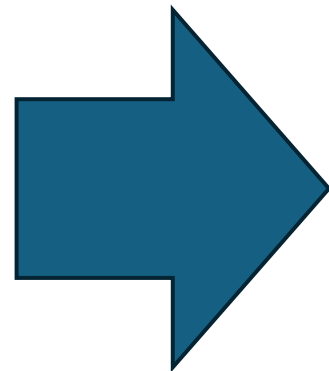
Development of a system on the OLSP website to offer introductions on the RIKEN's databases (2023-2025)
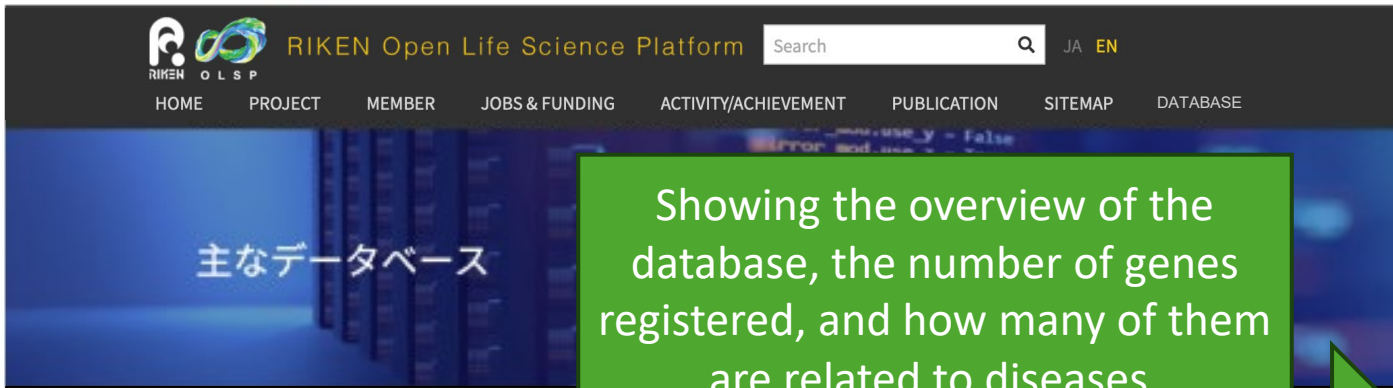
Showing the overview of the database, the number of genes registered, and how many of them are related to diseases.
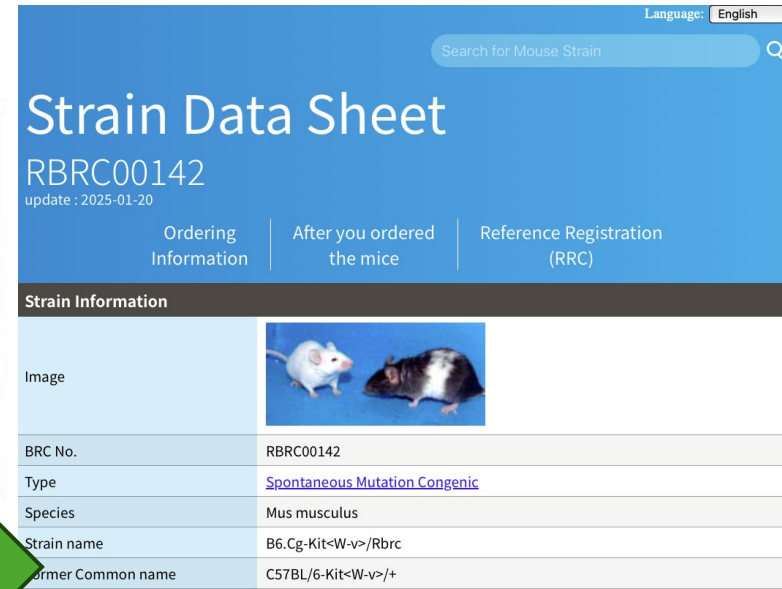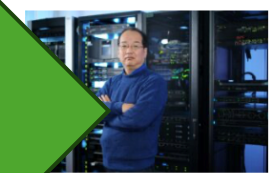
Link to original database

Links to another content explaining the database

Ongoing (2024-2025)…

- Installing the system to the server
- Preparing metadata of databases

Coming soon

Acknowledgements
OLSP Ontology Working Group (2020~)