

Integration of Life Science Databases through Knowledge Graphs



APBJC
2024

Database Center for Life Science (DBCLS) & BioData Science Initiative (BSI), Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Japan

At the Database Center for Life Science (DBCLS), we have been developing technologies for the integrated use of life science databases.

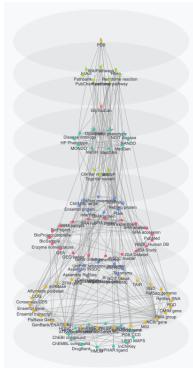
In the [RDF Portal](#), numerous databases are integrated by converting them into knowledge graphs, and [TogolD](#) aggregates relationships between database IDs. Based on these developments, [TogolD/Human](#) has realized data integration and visualization to explore diverse information related to humans.

For the integrated use of data, standardization and improved interoperability across databases are essential. To address this, DBCLS has been organizing the international developer conference, [BioHackathon](#), since 2008, fostering international collaboration with database researchers and developers worldwide.

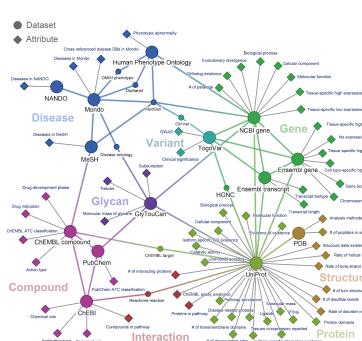
Compilation of Knowledge Graphs in the RDF Portal

The screenshot shows the RDF Portal's main dashboard. On the left, there is a sidebar with navigation links like Home, About, Help, and Contact. The central area features a large table with columns for ID, Name, Type, Status, Last Update, Version, Size, and Progress. Below the table is a detailed view of a specific database entry, showing its schema and data. At the bottom, there is a search bar and a footer with copyright information.

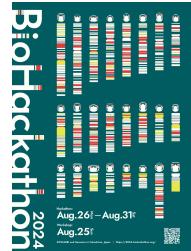
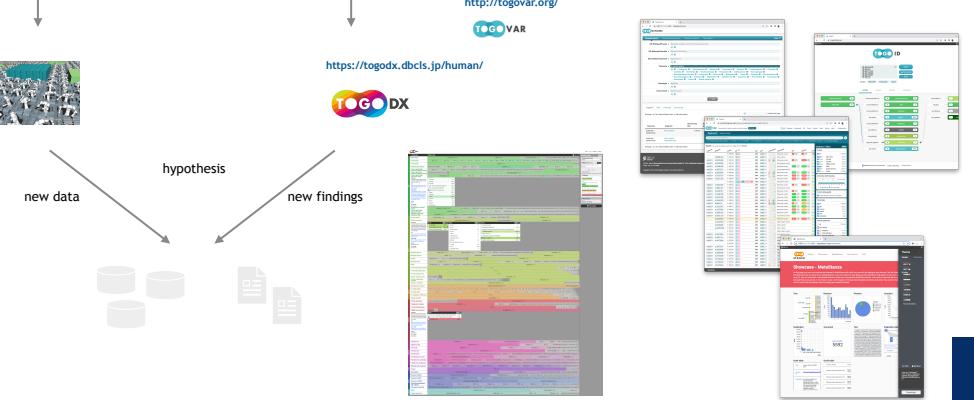
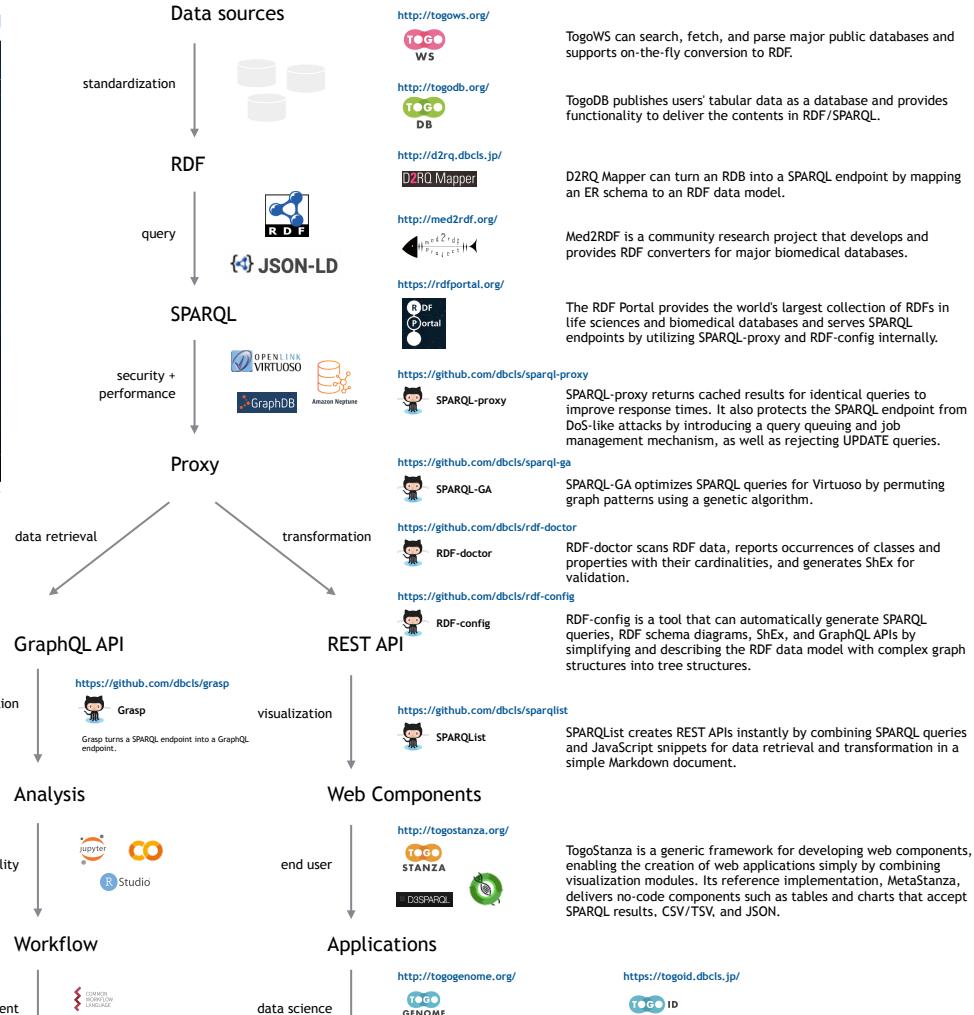
Semantic-based integration of ID relations across DBs with TogolD



Consolidation of Human Data in TogolD



Full-Stack Solutions for Knowledge Graph Development and Deployment



DIVE: Data Integration Visual Exploration

Database Center for Life Science (DBCLS) & BioData Science Initiative (BSI), Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Japan

Pursuing an Integrated Database Designed for Everyone to Explore and Comprehend

Due to the complex graph structure of integrated data, it is not easy to browse and intuitively understand diverse information. Addressing this challenge is the project Data Integration Visual Exploration (DIVE), which aims to explore various exhibit applications that enable the interactive visualization of integrated data. The acronym "DIVE" also embodies the meaning of "exploration" into the depths of databases.

For our first project, we took on the challenge of visualizing the relationships between human genes, organs, and evolutionary connections. The human body model uses Bodyparts 3D, developed by DBCLS, while the data showing the relationships between human genes and organs, as well as the evolutionary novelty of each gene, is based on data organized through TogoDX.

Visual Showcase of Related Data with TogoStanza and LLM Summarization

Detailed information about each gene can also be displayed on the web using TogoStanza/MetaStanza. Expression data is retrieved from the RefEx, GTEx, and ProteinAtlas databases, protein function from the UniProt database, 3D structures from the PDB database, variant information and clinical significance from the TogoVar and ClinVar databases, and reference information is obtained from the PubMed database via TogoGenome. Additionally, gene descriptions are generated using Large Language Models.

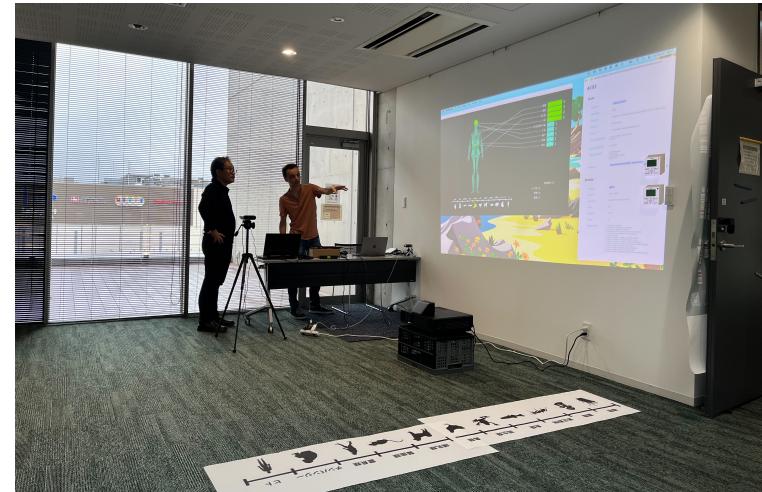
Future Prospects

This time, we proposed an interactive interface that allows for exploratory analysis of gene set expression across various organs and their evolutionary medical significance, while also being highly intuitive for the general public with an interest in science. Although deriving meaningful hypotheses from the reference data contained in databases is not always easy, we will continue to explore how data can be represented in a way that enables intuitive and efficient data exploration.

About DIVE

DIVE started as a voluntary group at the BSI office in DMM.make AKIBA, with the goal of creating new scientific exhibits through hands-on craftsmanship.

For example, when considering the future of genome medicine, it is essential to improve the public's literacy regarding genomes, genes, and variants. However, we feel that there is still much room for improvement in genome exhibits at science museums both in Japan and abroad. At DIVE, we aim to create exhibits that allow people to fully experience cutting-edge scientific data.



GBAT

Overview

The GBAT gene encodes the enzyme glucosidase, which is crucial for the metabolism of glycopolymers. Mutations in this gene are associated with several diseases, including Gaucher disease.

Function

- Enzyme Role: GBAT produces glucosidase, which breaks down glucosidase into glucose and neutral glycosaminoglycans.
- Cellular Role: This process is vital for metabolism and cellular homeostasis.

Associated Diseases

- **Gaucher Disease:** Type 1 (non-neuronopathic), Type 2 (acute neuronopathic), Type 3 (chronic neuronopathic).
- **Parkinson's Disease:** Link: Some studies suggest a correlation between GBAT mutations and an increased risk of developing Parkinson's disease.
- **Other Conditions:** Previous associations with other neurodegenerative disorders and certain cancers.

Genetic Variants

Mutations in the GBAT gene have been identified in the GBAT gene, with the most common being the N370S and L44P mutations.

Diagnosis and Future Directions

- Enzyme Replacement Therapy (ERT): Administering recombinant glucosidase to manage Gaucher disease symptoms.
- Sphingomyelin Reductase Therapy (SRT): Reducing the production of glucosidase to alleviate the burden on the enzyme.

Research and Future Directions

Ongoing research is focused on understanding the full spectrum of GBAT-related diseases, potential gene therapies, and therapeutic interventions.

Conclusion

The GBAT gene plays a critical role in lipid metabolism and is linked to significant health conditions, making it a key focus in genetic research and therapeutic development.

LRRK2

Overview

The LRRK2 gene encodes Protein kinase 2, which is located on chromosome 12 and encodes a protein that is involved in various cellular processes, including neuronal function, membrane vesicle transport, and autophagy.

Function

- Protein Kinase Activity: LRRK2 has kinase activity, meaning it can add phosphate groups to other proteins, which is crucial for regulating various cellular functions.
- Autophagy: LRRK2 is involved in the regulation of autophagy, a process that affects cell survival and differentiation.
- Neural Health: LRRK2 is particularly important in neurons, where it helps maintain cellular homeostasis and function.

Association with Parkinson's Disease

- Genetic Mutations: Mutations in the LRRK2 gene are one of the most common genetic causes of familial and sporadic Parkinson's disease (PD).
- Parkinson's Disease: The most studied mutation, associated with increased risk of early-onset and progressive PD.

Clinical Implications

- Diagnoses: Genetic testing for LRRK2 mutations can aid in the diagnosis of Parkinson's disease, especially in families.
- Therapeutic Targets: LRRK2 is a target for drug development, with efforts focused on inhibiting its kinase activity to slow disease progression.

Research Directions

- Mechanism of Action: Ongoing research aims to elucidate the precise mechanisms by which LRRK2 mutations lead to the development of PD.
- Biomarkers: Identifying biomarkers associated with LRRK2 activity could improve diagnosis and treatment monitoring.

Conclusion

The LRRK2 gene is a critical player in both normal cellular function and the pathogenesis of Parkinson's disease, making it a vital focal point of genetic and therapeutic research.

PARK7

Overview

The PARK7 gene, also known as DJ-1, is located on chromosome 12 and encodes a protein that is involved in cell cycle regulation and maintaining mitochondrial function.

Function

- Disease Resistance Response: PARK7 is involved in the defense against oxidative stress and maintaining mitochondrial function.
- Mitochondrial Protection: The protein is involved in mitochondrial protection, maintaining membrane integrity and preventing apoptosis.
- Neuronal Health: PARK7 is particularly important in neurons, where it helps maintain cellular homeostasis and function.

Mutations and Disease

- Parkinson Disease: Mutations in the PARK7 gene are one of the most common genetic causes of familial and sporadic Parkinson's disease (PD).
- Parkinson's Disease: The most studied mutation, associated with increased risk of early-onset and progressive PD.

Pathophysiology

- Protein Aggregation: Mutations can lead to protein aggregation, forming Lewy bodies.
- Mitochondrial Dysfunction: Mutations can disrupt mitochondrial function, leading to energy deficits.
- Types of Mutations: Common mutations include missense, nonsense, and frameshift mutations.

Research and Therapeutics

- Parkinson Disease: Initiatives are underway to develop treatments for PD, including gene therapy and drug development.
- Therapeutic Targets: Understanding the role of PARK7 in PD can help identify new therapeutic targets.

Conclusion

The PARK7 gene is critical for neuronal health, and its ongoing research aims to elucidate its mechanisms and potential therapeutic applications.

