Ontology-driven Exploration of RIKEN Bioresources via ChEBI Roles and Gene Ontology

Tatsuya Kushida[1, *,†], Daiki Usuda[1,†], Takatomo Fujisawa[2,†], Yasunori Yamamoto[3,†], Norio Kobayashi[4,†], and Hiroshi Masuya[1, *,†]

[1] BioResource Research Center, RIKEN, Tsukuba-shi, Japan
[2] Research Organization of Information and Systems, NIG, Mishima-shi, Japan
[3] Database Center for Life Science, ROIS, Kashiwa-shi, Japan
[4] Information R&D and Strategy Headquarters, RIKEN, Tokyo, Japan

Abstract

To accelerate life science research, establishing a knowledge base that enables the rapid and accurate selection of research materials suited to specific research objectives is indispensable. In this report, we attempted to realize ontology-driven discovery of bioresources that participate in the biosynthesis or utilization of chemical entities bearing specific roles (*e.g.*, anti-inflammatory agent) and that are associated with gene functions, processes, or locations. We extended the existing knowledge graph (KG) of the RIKEN BioResource Research Center for bioresources by integrating the KG with well-curated ontologies and RDF datasets: the chemical ontology ChEBI, the Gene Ontology (GO), UniProt RDF, Rhea, and KNApSAcK. Executing SPARQL queries over the integrated KG identified 61,499 plant DNA materials linked to 163 biological/chemical roles, 5,401 plant DNA materials linked to 289 biological activities, and 108,589 human DNA materials linked to 18,273 GO biological processes. Collectively, these results show that the KG provides intuitive entry points for experimental researchers and, by reusing ChEBI roles and GO annotations, delivers a verifiable path from ontology terms to testable interventions—improving the relevance and efficiency of bioresource discovery.

Keywords

Bioresource, Knowledge Graph, ChEBI, Gene Ontology, SPARQL, REST API

1. Introduction

Contemporary life science research relies on vast and diverse datasets, making the efficient integration and exploration of this data essential. Among these efforts, supporting experimental researchers in identifying candidate biological resources that facilitate the production of useful chemical compounds and elucidation of molecular functions and biological processes is of particular importance. This study addresses the following question: How can researchers systematically identify experimental materials (*e.g.*, cDNA clones, mutant strains, or cell lines)

16th International Conference on Biological and Biomedical Ontology (ICBO 2025) — ICBO EAST, November 1, 2025, Nara, Japan.

* Corresponding author.
† These authors contributed equally.
tatsuya.kushida@riken.jp (T. Kushida); hiroshi.masuya@riken.jp (H. Masuya)
0000-0002-0784-4113 (T. Kushida); 0000-0001-8978-3344 (T. Fujisawa); 0000-0002-6943-6887 (Y. Yamamoto); 0000-0002-2962-0073 (N. Kobayashi); 0000-0002-3392-466X (H. Masuya)

that are relevant to a desired chemical role or gene function across multiple ontological resources?

As one of the world's leading biological resource repositories, the RIKEN BioResource Research Center (BRC) manages and provides diverse bioresources—such as mouse strains, plant seeds, cultured cell lines, microbial strains, and DNA/cDNA materials—under rigorous quality control. These bioresources support life science research across mechanistic studies, chemical production, breeding, and drug discovery.

For the dissemination and retrieval of bioresource data, metadata are integrated into the RIKEN BioResource MetaDatabase (MetaDB) [1], an RDF-based platform on Virtuoso, and published as the RIKEN Bioresource Knowledge Graph (KG) [2]. To enhance interoperability, the KG follows the Biological Resource Schema Ontology (BRSO) [2] and has already been linked with an Orthologous MAtrix (OMA) database [3], a gene-disease association database: DisGeNET [4], and disease ontologies (MONDO [5], DOID [6], Orphanet [7], NANDO [8]), enabling the identification of disease-relevant mice, cell lines, and genetic materials. Disease ontologies such as MONDO and DOID are relevant because they are already integrated into the existing RIKEN Bioresource KG to enable disease-based searches for mouse and human resources, forming the foundation on which the present chemical and gene-function extension is built.

In this study, we integrated the Bioresource KG with additional biomedical ontologies and RDF graph datasets—namely, the chemical compound ontology: ChEBI [9], Gene Ontology (GO) [10], the proteome dataset: UniProt RDF [11], the chemical reaction dataset: Rhea RDF [12], and a biological activity RDF: KNApSAcK [13] —to enable the exploration of bioresources that are involved in the production of useful chemical compounds and in research aimed at elucidating biological phenomena. This study is expected to contribute to the advancement of agriculture and to addressing global environmental and food issues by providing information for creating new high-yield and stress-tolerant crops, such as salt- and hyperosmolarity-tolerant cultivars. Furthermore, the established methodology can be applied to other domains, contributing to a deeper understanding of biological phenomena, drug discovery research, and diverse fields in biology. GO terms are routinely produced by GO enrichment analyses and come as stable, resolvable identifiers; our KG ingests them directly. ChEBI roles capture interpretable chemical and pharmacological functions. Taken together, these vocabularies provide practical entry points, which we operationalize into an end-to-end path from GO/ChEBI terms through Rhea/EC and UniProt/NCBI Gene to orderable bioresources.

The rest of the paper is organized as follows: Section 2 reviews related work on bioresource RDF data and the development of its retrieval systems. In Section 3, we explain the external datasets and ontologies integrated with the RIKEN Bioresource KG. Section 4 shows examples of SPARQL queries used to retrieve candidate bioresources related to a given chemical or biological role (*e.g.*, anti-inflammatory agent), biological activities (*e.g.*, induce flowering), and gene function (*e.g.*, structural constituent of cytoskeleton). Section 5 presents the performance comparison results using local datasets within BioResource MetaDB versus using remote (federated) queries. We also discuss the implementation of a keyword-search API, the advantages of using ChEBI and GO for bioresource exploration, and future work.

2. Related work

The Monarch Initiative is an international consortium working to expand the use of genome information in biology and biomedical research. It is a vast knowledge graph that semantically harmonizes and integrates a wide array of data sources, including bioresources, with the Biolink Model serving as its core schema [14].

DBRP Stanza is a retrieval system for microbial strains collected by the Biological Resource Center in the National Institute of Technology and Evaluation (NBRC) and the Japan Collection of Microorganisms (JCM) of RIKEN BRC. These data are stored and published via the DBCLS RDF Portal [15]. The retrieval results of the DBRP Stanza are obtained from the RDF Portal using SPARQL and graphically visualized using TogoStanza [16], a web framework. For example, the user can execute a keyword search to retrieve microorganisms using microbial habitat terms (*e.g.*, hot spring) as defined by the Metagenome and Microbes Environmental Ontology.

BioDB Search, developed by DBCLS, is a retrieval system covering biochemical reactions—substrates, products, and enzymes—across various organisms in addition to NBRC/JCM microbial strains, similar in spirit to DBRP Stanza. Unlike DBRP Stanza, which uses only microorganism RDF, BioDB Search integrates ChEBI, UniProt RDF, GO, and Rhea RDF. Users can search by chemical entity (*e.g.*, caffeine), protein (*e.g.*, starch synthase), biological process (*e.g.*, xanthophyll biosynthetic process), taxonomy (*e.g.*, *Micromonospora chalcea*), and EC number (*e.g.*, 1.14.15.24; beta-carotene 3-hydroxylase).

3. Constructing the Bioresource KG

3.1. Bioresource KG data schema

The Bioresource KG is based on the BRSO [2] and contains labels (*e.g.*, iso-D), identifiers (*e.g.*, psi00001), organisms (*e.g.*, NCBITaxon_3702 and *Arabidopsis thaliana*), related genes (*e.g.*, ncbigene:843873 and At1g65620), and associated phenotypes (*e.g.*, Dwarf). Users can browse KG data through a web interface, execute SPARQL queries at the BioResource MetaDB SPARQL endpoint, and download all the data from the BioResource MetaDB [1].

Table 1: External ontologies and RDF datasets integrated into the Bioresource KG.

Integrated ontologies/RDF graphs	Data format of original data or conversion methods to N-Triples	Data acquisition date	License
ChEBI (ver. 244)	OWL	2025-09-09	CC BY 4.0
Gene Ontology (ver. 2025-07-22)	OWL	2025-09-10	
UniProt-NCBI Gene (ver. 2025-06-18)	Conversion from TSV data to N-Triples	2025-07-30	
UniProt-Enzyme	Execution of the SPARQL query and acquisition of N-Triples	2025-07-30	
UniProt-GO		2025-07-30	
Rhea	RDF/XML	2024-03-10	
KNApSAcK	Turtle	2023-08-23	

3.2. Integrated biomedical ontologies and public RDF graph datasets

We link RIKEN bioresources to ChEBI—which organizes chemical entities and biological roles (*e.g.*, anti-inflammatory agent)—and to the GO, which structures the molecular functions of gene products (*e.g.*, naringenin-chalcone synthase activity), biological processes (*e.g.*, cuticle development), and cellular components (*e.g.*, chloroplast). We further integrate the Bioresource KG with UniProt (ID mapping, enzyme/EC and GO annotations), Rhea, and KNApSAcK (Table 1). Together, UniProt and Rhea provide curated bridges from genes/gene products to ChEBI chemical entities and GO terms, enabling typed traversals from ontology terms to distributable RIKEN BRC resources. These external datasets, distributed under open licenses (see Table 1), were ingested into BioResource MetaDB, which stores the Bioresource KG. All inputs were ingested as RDF graphs: OWL ontologies (e.g., ChEBI, GO) in RDF/XML or Turtle, and other datasets in RDF/XML, Turtle, or N-Triples.

3.3. Identifier strategy

We connect datasets using resolvable, community-standard URIs: enzyme classes in the UniProt namespace (http://purl.uniprot.org/enzyme/) and chemical entities via ChEBI OBO PURLs (http://purl.obolibrary.org/obo/CHEBI_). For NCBI Gene, UniProt RDF employs the http://purl.uniprot.org/geneid/ namespace. To maximize interoperability with external graphs, the Bioresource KG accepts four resolvable URI patterns for NCBI Gene: https://www.ncbi.nlm.nih.gov/gene/, http://identifiers.org/ncbigene/, https://identifiers.org/ncbigene:, and http://purl.uniprot.org/geneid/. Our query templates accept all four variants, enabling joins to UniProt and Rhea without custom mappings while preserving compatibility with third-party KGs.

4. Results

Interpretation note: Traversals illustrate ontology-level associations across curated graphs and do not imply experimentally validated activities in specific organisms or strains.

Figure 1 summarizes our integration path from ontology terms to RIKEN BRC resources across UniProt, Rhea, ChEBI, and KNApSAcK. In the integrated KG, each bioresource is linked to an NCBI Gene (via the Bioresource KG); the gene encodes a UniProt protein (via the UniProt–NCBI Gene mappings); the protein is associated with an enzyme class (EC number) using the UniProt enzyme dataset (via the UniProt–Enzyme mappings); the EC class is associated with Rhea reactions whose participants include ChEBI chemical entities as substrates or products (via the Rhea RDF graph); and the ChEBI entities are annotated with ChEBI roles (via the ChEBI Ontology graph). This typed path lets users start from an interpretable role and resolve specific, distributable BRC resources. We assessed integration completeness by counting resolvable links across graphs (*e.g.*, Gene⟶UniProt protein, EC⟶Rhea reaction, Rhea participant⟶ChEBI entity) and by reporting distinct counts and coverage (Supplemental Data 01[1]–02[2]).

---

[1] https://purl.archive.org/purl/rbrc/icbo_east2025/blob/main/supplementalData01.pdf
[2] https://purl.archive.org/purl/rbrc/icbo_east2025/blob/main/supplementalData02.pdf

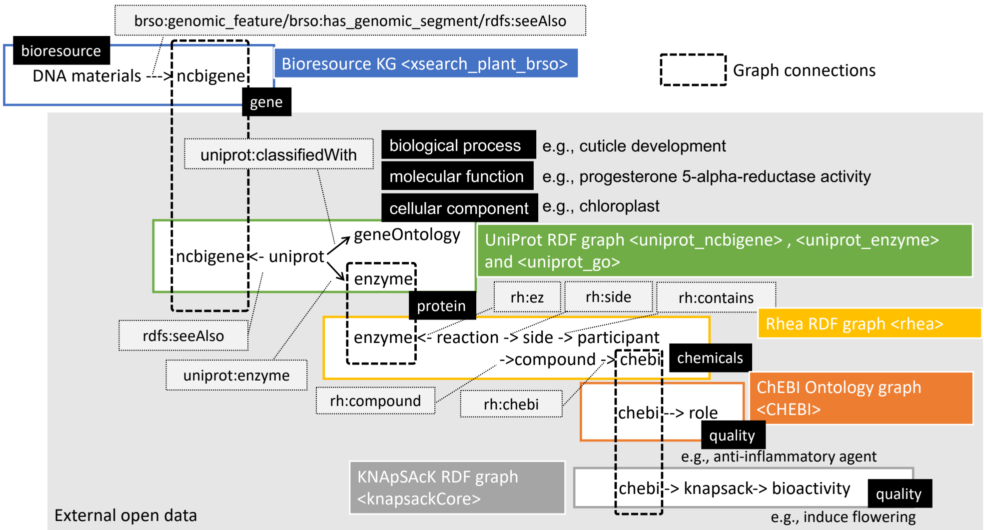

Figure 1: Semantic network of the integrated Bioresource KG showing links among RIKEN BRC bioresources, NCBI Gene, UniProt proteins, EC classes, Rhea reactions, and ChEBI chemical entities/roles; for GO-based paths, UniProt GO annotations are also used. This path enables ontology-driven traversal from ChEBI roles, KNApSAcK bioactivity terms or GO terms to distributable bioresources.

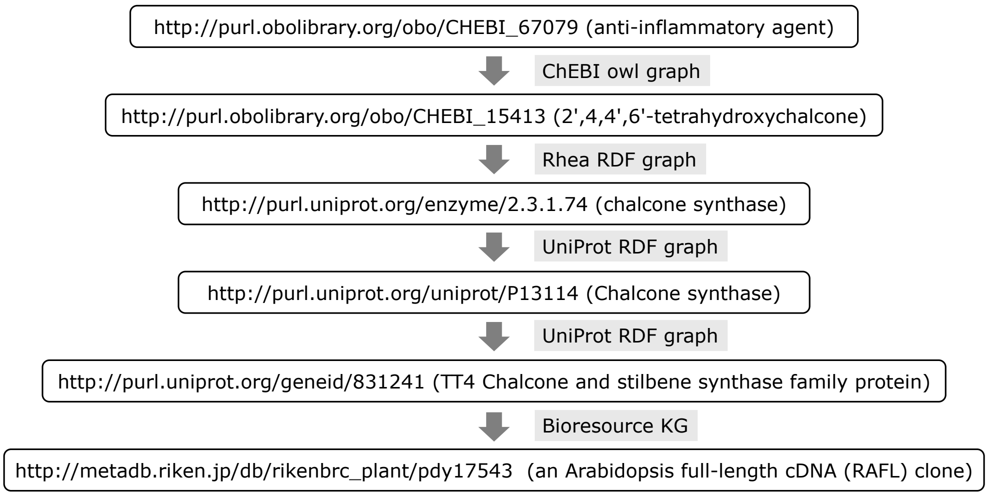

Figure 2: Example traversal from a ChEBI role to an orderable Arabidopsis DNA resource. Starting at the ChEBI role anti-inflammatory agent (CHEBI:67079), the query resolves to the compound 2′,4,4′,6′-tetrahydroxychalcone (CHEBI:15413), via Rhea reactions catalyzed by EC 2.3.1.74 (chalcone synthase), to UniProt proteins (*e.g.*, P13114), to NCBI Gene TT4 (GeneID:831241), and finally to a RIKEN BRC Arabidopsis full-length cDNA clone (pdy17543).

Figure 2 illustrates the Rhea-mediated traversal from a ChEBI role to a concrete RAFL clone [17]. We reuse the same pattern in our SPARQL query templates throughout this study.

The component datasets of this integrated KG are similar to that of the BioDB Search; however, they differ in the following ways: the former enables the exploration of all bioresources with related gene information, and the latter allows for the exploration of only microorganisms as bioresources. Moreover, the former enables the retrieval of associated bioresources by using the ChEBI role and KNApSAcK's biological activity terms.

Figure 1 also shows the data schema of integrated Bioresource KG with UniProt RDF and GO, where users can explore related RIKEN bioresources using GO biological process, molecular function, and cellular component terms. For example, executing a SPARQL query over the KG starting from the GO biological process "cuticle development" (GO:0042335) retrieves a path to relevant UniProt protein entries (Transmembrane protein 79; UniProt:Q9BSE2), to the NCBI Gene TMEM79 (GeneID:84283) in *Homo sapiens*, and finally to a RIKEN BRC Plasmid clone of human TMEM79 cDNA (*e.g.*, HGY087496) [18].

As of 15 October 2025, RIKEN BRC holds 612,129 plant DNA materials and 169,107 human DNA materials, of which 267,613 and 136,989 have at least one NCBI Gene, respectively. When we executed SPARQL queries, we retrieved 61,499 plant DNA materials associated with 163 ChEBI roles [19], 5,401 plant DNA materials associated with 289 KNApSAcK biological activities [20], 254,082 plant DNA materials associated with 6,459 GO biological processes [21], and 108,589 human DNA materials associated with 18,273 GO biological processes [22]. Connectivity analysis revealed that for DNA materials, 11.7% (19,809/169,107) of bioresources reach ≥1 ChEBI role and 64.2% (108,589/169,107) reach ≥1 GO term; for plant DNA materials, the corresponding figures are 10.0% (61,499/612,129) and 41.5% (254,082/612,129) (Supplemental Data 01). The residual isolated nodes are primarily records without gene annotations, which blocks resolution to UniProt and thus to the EC/Rhea/ChEBI chain. We will periodically reconcile these nodes with future releases of BRSO and the external graphs.

We summarize the integrated graphs used in this study—original formats, conversion to N-Triples, acquisition dates, licenses, and the numbers of distinct triples and properties—in Supplemental Data 02. All analyses in Section 4 are based on this snapshot hosted locally in MetaDB (no federation).

5. Discussion and future work

5.1. Comparison of local and federated SPARQL execution results

In this study, we expanded the Bioresource KG by collecting external open datasets into a single local environment (MetaDB). SPARQL also enables federated queries via the SERVICE keyword, allowing us to combine locally hosted graphs with official remote endpoints on the Internet (*e.g.*, the UniProt SPARQL endpoint). A clear advantage of using original endpoints is near-real-time freshness and reduced local maintenance. This is particularly relevant for very large resources—UniProt RDF is published as more than 600 files with a total size of ≳1.9 TB—for which full local mirroring and regular updates are time-consuming.

We therefore evaluated the practicality of offering interactive retrieval over federated queries by reusing the representative SPARQL patterns from Section 4 and measuring runtimes. In our environment, federated execution frequently exceeded a 10-minute transaction limit or returned 502 Proxy Errors, yielding no results [23-25], while local execution completed successfully and returned 957 bioresources [17].

These findings suggest that federated SPARQL is currently unsuitable for interactive use in our setting. As future work, we will pursue query-level optimizations (*e.g.*, subqueries), triple-store tuning, server scaling, and cost-based federated query engines to shorten retrieval times and make federation viable [26].

Hosting a local mirror within MetaDB ensures reproducible query execution and stable performance even when remote endpoints are unavailable or updated. Federated SPARQL remains conceptually valuable for freshness, but operational reproducibility and responsiveness favored local integration.

5.2. Implementation of a keyword-search API

We demonstrated that executing SPARQL queries could obtain bioresources relevant to chemical roles, biological activities, biological processes, molecular functions, and cellular components. However, it is difficult for biological experimental researchers to perform SPARQL queries. Therefore, we implemented keyword-search APIs to improve usability for experimental researchers. SPARQList is a REST API server that executes a SPARQL query [27], and the following API Examples 1-4 are RESTful APIs utilizing SPARQList. API Example 1 [28] is an API for exploring plant DNA materials and related genes using identifiers (*e.g.*, CHEBI:67079) of the ChEBI role terms (*e.g.*, Anti-inflammatory agent). These REST endpoints wrap representative SPARQL queries so that users can supply either keywords or stable identifiers (*e.g.*, Anti-inflammatory agent, CHEBI:67079) without writing SPARQL.

Similarly, API Example 2 [29] is an API for exploring plant DNA materials and related genes using biological activity terms (*e.g.*, induce flowering), and API Examples 3 [30] and 4 [31] are APIs for exploring plant and human DNA materials and related genes using identifiers (*e.g.*, GO:0042335) of GO terms (*e.g.*, cuticle development), respectively. The users can obtain the retrieval results in the JSON format by embedding keywords, such as CHEBI:67079, into the defined URLs and accessing the embedded URLs or accessing each API webpage, typing keywords in the input box, and accessing the created URLs.

The search for bioresources [32], which is currently in operation, has enabled the exploration of RIKEN bioresources using keywords, such as common disease names (in English and Japanese), phenotype terms, gene symbols, bioresource and organism names, and their IDs. This system has implemented a keyword input assist function that presents keyword suggestion lists. Although the current system exposes REST APIs through SPARQList, we plan to develop a lightweight web interface that wraps these endpoints, allowing non-expert users to search bioresources by keyword, by ontology term (*e.g.*, ChEBI role or GO term), or by pasting GO enrichment results—without writing SPARQL.

5.3. Benefits of using ChEBI roles and GO annotations

ChEBI and GO offer complementary semantics that are immediately useful for experimental researchers. ChEBI roles describe the function of a chemical (*e.g.*, anti-inflammatory agent), while GO terms describe the function of a gene product, the processes it is involved in, and its location (molecular function, biological process, and cellular component). As long-standing, community-curated, openly licensed resources (see Table 1), they are already familiar to biomedical researchers, which lowers the adoption barrier for ontology-based searches. Their active communities also keep chemical and gene-function semantics up to date, which is critical for the long-term maintenance and sustainability of the KG. In typical workflows, GO enrichment outputs stable GO identifiers, and ChEBI defines interpretable chemical/pharmacological roles; in our KG, either can serve as a starting term to resolve downstream to genes and distributable resources.

In our integrated KG, bioresources are connected to NCBI Gene and UniProt (gene products); gene products carry EC and GO annotations. EC classes connect to Rhea reactions whose participants include ChEBI entities as substrates or products, and ChEBI entities are annotated with roles. Declared domains and ranges on these predicates provide stable, typed traversals from a "desired role or function" to "candidate bioresources." Practically, this lets users start from a ChEBI role or a GO term and resolve to concrete RIKEN BRC bioresources. Collectively, these results indicate that ChEBI and GO concepts scale to extensive bioresource catalogs, including those of genetically modified mouse strains.

Our approach also has distinctive aspects when compared to more expansive, global knowledge graph initiatives like the Monarch Initiative. The Monarch Initiative, centered on a standardized data model such as the Biolink Model, integrates OBO ontologies, including ChEBI and GO, to support a wider range of applications, including phenotype-driven discovery, graph machine learning, and cross-species inference [33]. However, its lack of an official SPARQL endpoint requires users to implement their own RDF store, presenting a significant barrier for direct data exploration.

As of 9 September 2025, Monarch's public datasets do not expose the specific Rhea-driven chain illustrated in Figures 1–2 (EC ⟶ Rhea ⟶ ChEBI participants). In contrast, our Bioresource KG represents a more practical solution tailored to the specific mission of efficiently discovering RIKEN's bioresources. While approaches like the Monarch Initiative, which pursue universal interoperability, are powerful for general-purpose applications like cross-species inference, their schema may not be optimized for detailed, mission-specific queries. For example, as of 9 September 2025, Monarch's published import sources do not include, and Monarch's public APIs do not expose a reaction-level chain with ChEBI-typed participants. Consequently, our Rhea-based traversals in Figures 1–2 cannot be reproduced against Monarch's public datasets, whereas they are explicitly modeled and supported in our KG. Note that our statement that the public Monarch stack "cannot reproduce" our path is limited to the Rhea-driven chain illustrated in Figures 1–2 (EC⟶Rhea⟶ChEBI participants); other aspects may be available in Monarch's public datasets.

On the other hand, in typical transcriptomics pipelines (*e.g.*, RNA-seq, qPCR), differentially expressed genes are routinely subjected to GO enrichment analysis using widely adopted tools such as DAVID [34], and g:Profiler [35]. The outputs of these tools provide the same GO identifiers and labels that our KG accepts, allowing investigators to copy a significant GO term from their enrichment results and directly traverse to related reactions, enzymes, genes, and, ultimately, the distributable RIKEN BRC bioresources. This alignment makes GO terms a highly practical entry point for ontology-driven bioresource discovery.

6. Conclusion

We integrated the RIKEN Bioresource KG with ChEBI, GO, UniProt, Rhea, and KNApSAcK to enable ontology-driven discovery of bioresources from chemical roles and gene-centric functions. Representative SPARQL patterns were exposed as REST APIs so that users can start with familiar keywords or stable identifiers without writing SPARQL. Local (non-federated) execution returned complete results, while federated SPARQL frequently timed out, suggesting pragmatic deployment choices. Overall, reusing ChEBI roles and GO annotations within the KG improved both the relevance and efficiency of target bioresource discovery, providing a verifiable path from ontology terms to testable interventions and accelerating life science research.

Acknowledgements

This work was supported by the Cross-ministerial Strategic Innovation Promotion Program (SIP), "Technologies for Smart Bio-industry and Agriculture" (funding agency: Bio-oriented Technology Research Advancement Institution).

Declaration on Generative AI

We used a large language model (LLM) to proofread and refine the English expression of this paper. The content and core ideas of the paper were entirely developed by the authors.

References

[1] Kobayashi N, Kume S, Lenz K, Masuya H. Riken metadatabase: a database platform for health care and life sciences as a microcosm of linked open data cloud, International Journal on Semantic Web and Information Systems (IJSWIS) 14 (2018) 140-164.

[2] Kushida T, Mendes de Farias T, Sima AC, Dessimoz C, Chiba H, Bastian F, Masuya H. Federated SPARQL query performance evaluation for exploring disease model mouse: combining gene expression, orthology, and disease knowledge graphs. BMC Med Inform Decis Mak. 2025 May 16;25(Suppl 1):189. doi: 10.1186/s12911-025-03013-8. PMID: 40380154; PMCID: PMC12082848.

[3] Altenhoff AM, Train CM, Gilbert KJ, Mediratta I, Mendes de Farias T, Moi D, et al. OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. Nucleic Acids Res. 2021 Jan 8;49(D1):D373-D379. doi: 10.1093/nar/gkaa1007. PMID: 33174605; PMCID: PMC7779010.

[4] Piñero J, Saüch J, Sanz F, Furlong LI. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. Comput Struct Biotechnol J. 2021 May 11;19:2960-2967. doi: 10.1016/j.csbj.2021.05.015. PMID: 34136095; PMCID: PMC8163863.

[5] Mungall CJ, McMurry JA, Köhler S, Balhoff JP, Borromeo C, Brush M, et al. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 2017 Jan 4;45(D1):D712-D722. doi: 10.1093/nar/gkw1128. Epub 2016 Nov 29. PMID: 27899636; PMCID: PMC5210586.

[6] Schriml LM, Munro JB, Schor M, Olley D, McCracken C, Felix V, et al. The Human Disease Ontology 2022 update. Nucleic Acids Res. 2022;50(D1):D1255-D1261. doi: 10.1093/nar/gkab1063.

[7] Mohtashamian M, Abeysinghe R, Hao X, Cui L. Identifying Missing IS-A Relations in Orphanet Rare Disease Ontology. Proceedings (IEEE Int Conf Bioinformatics Biomed). 2022 Dec;2022:3274-3279. doi: 10.1109/bibm55620.2022.9995614. Epub 2023 Jan 2. PMID: 36776767; PMCID: PMC9918376.

[8] Kushida T, Usuda D, Takada T, Yamagata, Y Masuya H. Ontology Integration for Discovering Bioresources Contributing to Medical Science Research [abstract] ICBO 2022: International Conference on Biomedical Ontology. [https://icbo-conference.github.io/icbo2022/papers/ICBO-2022_paper_1944.pdf] Accessed on 9 Sept. 2025.

[9] Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res. 2016 Jan 4;44(D1):D1214-9. doi: 10.1093/nar/gkv1031. Epub 2015 Oct 13. PMID: 26467479; PMCID: PMC4702775.

[10] Gene Ontology Consortium; Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023 May 4;224(1):iyad031. doi: 10.1093/genetics/iyad031. PMID: 36866529; PMCID: PMC10158837.

[11] UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023 Jan 6;51(D1):D523-D531. doi: 10.1093/nar/gkac1052. PMID: 36408920; PMCID: PMC9825514.

[12] Bansal P, Morgat A, Axelsen KB, Muthukrishnan V, Coudert E, Aimo L, et al. Ignatchenko A, Redaschi N, Bridge A. Rhea, the reaction knowledgebase in 2022. Nucleic Acids Res. 2022 Jan 7;50(D1):D693-D700. doi: 10.1093/nar/gkab1016. PMID: 34755880; PMCID: PMC8728268.

[13] Nakamura Y, Afendi FM, Parvin AK, Ono N, Tanaka K, Hirai Morita A, et al. KNApSAcK Metabolite Activity Database for retrieving the relationships between metabolites and biological activities. Plant Cell Physiol. 2014 Jan;55(1):e7. doi: 10.1093/pcp/pct176. Epub 2013 Nov 27. PMID: 24285751.

[14] Shefchek KA, Harris NL, Gargano M, Matentzoglu N, Unni D, Brush M, et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 2020 Jan 8;48(D1):D704-D715. doi: 10.1093/nar/gkz997. PMID: 31701156; PMCID: PMC7056945.

[15] Kawashima S, Katayama T, Hatanaka H, Kushida T, Takagi T. NBDC RDF portal: a comprehensive repository for semantic data in life sciences. Database (Oxford). 2018 Jan 1;2018:bay123. doi: 10.1093/database/bay123. PMID: 30576482; PMCID: PMC6301334.

[16] Katayama T, Kawashima S, Okamoto S, Moriya Y, Chiba H, Naito Y, et al. TogoGenome/TogoStanza: modularized Semantic Web genome database, Database, Volume 2019, bay132.

[17] SPARQL Query for obtaining plant DNA materials relevant to ChEBI roles "anti-inflammatory agent (CHEBI:67079)" (non-federated search). [https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Example1.rq] Accessed on 9 Sept. 2025.

[18] SPARQL Query for obtaining human DNA materials relevant to GO biological process "cuticle development (GO:0042335)". [https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Example9.rq] Accessed on 9 Sept. 2025.

[19] SPARQL Query for obtaining plant DNA materials relevant to ChEBI roles. [https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Statistics1_plantD_role.rq] Accessed on 9 Sept. 2025.

[20] SPARQL Query for obtaining plant DNA materials relevant to KNApSAcK biological activities.
[https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Statistics2_plantD_BA.rq] Accessed on 9 Sept. 2025.
[21] SPARQL Query for obtaining plant DNA materials relevant to GO biological processes. [https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Statistics4_plantD_BP.rq] Accessed on 9 Sept. 2025.
[22] SPARQL Query for obtaining human DNA materials relevant to GO biological processes . [https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Statistics7_humanD_BP.rq] Accessed on 9 Sept. 2025.
[23] SPARQL Query for obtaining plant DNA materials relevant to ChEBI roles "anti-inflammatory agent (CHEBI:67079)" (federated search for UniProt official SPARQL endpoint)
[https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Example2.rq] Accessed on 9 Sept. 2025.
[24] SPARQL Query for obtaining plant DNA materials relevant to ChEBI roles "anti-inflammatory agent (CHEBI:67079)" (federated search for Rhea official SPARQL endpoint) [https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Example3.rq] Accessed on 9 Sept. 2025.
[25] SPARQL Query for obtaining plant DNA materials relevant to ChEBI roles "anti-inflammatory agent (CHEBI:67079)" (federated search for ChEBI's SPARQL endpoint within RDF Portal) [https://github.com/kushidat/bioresourceKG_SPARQLQueryExamples/blob/main/SPARQL_Query_Example4.rq] Accessed on 9 Sept. 2025.
[26] Qudus U, Saleem M, Ngonga Ngomo AC, Lee Young-Kooa. An Empirical Evaluation of Cost-based Federated SPARQL Query Processing Engines. 1 Jan. 2021 : 843-868. doi: 10.3233/SW-200420.
[27] Katayama T, Kawashima S. SPARQList: Markdown-Based Highly Configurable REST API Hosting Server for SPARQL. In Proceedings of the 10th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences (SWAT4LS 2017) [https://ceur-ws.org/Vol-2042/paper47.pdf].
[28] API Example 1 for obtaining RIKEN plant DNA materials with a ChEBI role identifier [https://splist.brc.riken.jp/sparqlist/api_example1_bioresourceKG] Accessed on 9 Sept. 2025.
[29] API Example 2 for obtaining RIKEN plant DNA materials with a KNApSAcK biological activity keyword [https://splist.brc.riken.jp/sparqlist/api_example2_bioresourceKG] Accessed on 9 Sept. 2025.
[30] API Example 3 for obtaining RIKEN plant DNA materials with a GO identifier [https://splist.brc.riken.jp/sparqlist/api_example3_bioresourceKG] Accessed on 9 Sept. 2025.
[31] API Example 4 for obtaining RIKEN human DNA materials with a GO identifier [https://splist.brc.riken.jp/sparqlist/api_example4_bioresourceKG] Accessed on 9 Sept. 2025.
[32] Masuya H, Usuda D, Nakata H, Yuhara N, Kurihara K, Namiki Y, et al. Establishment and application of information resource of mutant mice in RIKEN BioResource Research Center.

Lab Anim Res. 2021 Jan 18;37(1):6. doi: 10.1186/s42826-020-00068-8. PMID: 33455583; PMCID: PMC7811887.
[33] Caufield JH, Putman T, Schaper K, Unni DR, Hegde H, Callahan TJ, et al. KG-Hub-building and exchanging biological knowledge graphs. Bioinformatics. 2023 Jul 1;39(7):btad418. doi: 10.1093/bioinformatics/btad418. PMID: 37389415; PMCID: PMC10336030.
[34] Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022 Jul 5;50(W1):W216-W221. doi: 10.1093/nar/gkac194. PMID: 35325185; PMCID: PMC9252805.
[35] Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. 2023 Jul 5;51(W1):W207-W212. doi: 10.1093/nar/gkad347. PMID: 37144459; PMCID: PMC10320099.

A. Online Resources

1. BRSO webpage at https://github.com/dbcls/brso
2. BioResource MetaDatabase webpage at https://knowledge.brc.riken.jp
3. The Nanbyo Disease Ontology (NANDO) at http://nanbyodata.jp/ontology/nando
4. DBRP Stanza webpage at https://www.nite.go.jp/nbrc/dbrp/dbrpstanza/top.html
5. BioDB Search webpage at https://sip-db.dbcls.jp/bdsearch/#/
6. Bioresource KG data schema at https://github.com/kushidat/bioresourceKG_schema
7. BioResource MetaDB SPARQL endpoint at https://knowledge.brc.riken.jp/sparql
8. chebi.owl.gz can be downloaded at https://ftp.ebi.ac.uk/pub/databases/chebi/ontology/
9. go.owl can be downloaded at https://purl.obolibrary.org/obo/go.owl
10. ID mapping data of UniProt can be downloaded at https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/idmapping_selected.tab.gz
11. UniProt official SPARQL endpoint at https://sparql.uniprot.org/sparql/
12. rhea.rdf.gz can be downloaded at https://ftp.expasy.org/databases/rhea/rdf/
13. KNApSAcK GitHub webpage at https://github.com/ddbj/mb_knapsack
14. UniProt RDF data can be downloaded at https://ftp.uniprot.org/pub/databases/uniprot/current_release/rdf/
15. Arabidopsis full-length cDNA (RAFL) clone; pdy17543 webpage at https://plant.rtc.riken.jp/resource/rafl/rafl_detail.html?brcno=pdy17543
16. Plasmid clone of human TMEM79 cDNA; HGY087496 webpage at https://dnaconda.riken.jp/search/GNPhum/IRAL018/IRAL018M08.html
17. Rhea official SPARQL endpoint at https://sparql.rhea-db.org/
18. RDF Portal for EBI data SPARQL endpoint at https://rdfportal.org/ebi/sparql
19. RIKEN BRC webpage at https://web.brc.riken.jp/
20. Biolink-Model webpage at https://biolink.github.io/biolink-model/