RDF Portal: Enhancing Accessibility and Interoperability of Life Science RDF

Shuichi Kawashima[1], Yasunori Yamamoto[1], Hirokazu Chiba[1], Yuki Moriya[1] and Toshiaki Katayama[1]

[1]*Database Center for Life Science, 178-4-4 Wakashiba, Kashiwa-shi, Chiba 277-0871, Japan*

Abstract

The RDF Portal provides comprehensive access to high-quality RDF data in life sciences. This paper highlights recent advancements, including the automation of update processes using RDF-config, improved metadata standardization, and expanded dataset coverage. These developments address challenges such as manual dependencies and schema inconsistencies, ensuring data reliability and interoperability. By integrating public and custom RDF datasets, the portal fosters innovation and supports diverse applications. Additionally, as a foundational resource for DBCLS and external services, the RDF Portal plays a critical role in advancing life sciences research and data integration.

Keywords

RDF, SPARQL, database, metadata

1. Introduction

The RDF Portal [1], initiated in 2015 by National Bioscience Database Center (NBDC) and Database Center for Life Science (DBCLS), has established itself as a comprehensive resource for accessing RDF data in life sciences. By addressing the challenges of data accessibility and interoperability, the portal has evolved from being a mere repository to a robust platform that integrates high-quality RDF data. This paper outlines recent advancements in the RDF Portal, focusing on automated updates, improved metadata handling, and its role as a foundation for various services.

Since its inception, the RDF Portal has aimed to streamline access to RDF data for researchers and developers in life sciences. It employs rigorous guidelines and review processes to ensure data quality and interoperability. This has been made possible through collaboration with domestic RDF developers under the DB integration coordination program in Japan.

Beyond being a data repository, the RDF Portal serves as a foundational resource for other DBCLS services, including TogoDX, TogoVar, and TogoID. It also supports external services such as NITE NBRC's DBRP Stanza. By providing high-quality RDF data, the portal enables these services to deliver reliable and interoperable solutions, furthering the integration of life sciences data across platforms.

2. Automated Update Processes with RDF-config

To address inefficiencies in data update processes, the RDF Portal has adopted automation techniques using RDF-config. Currently, two key issues hinder effective updates:

Manual Dependencies: A significant portion of the update process relies on manual operations, leading to delays and increased labor.

Schema Inconsistencies: Changes in RDF data models can result in discrepancies between schema diagrams and actual data, creating barriers to seamless data integra- tion.

RDF-config, a tool developed by DBCLS, simplifies complex RDF data modeling for individual databases. It automates the generation of SPARQL queries and schema diagrams, ensuring alignment

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025
kwsm@dbcls.rois.ac.jp (S. Kawashima)
0000-0001-7883-3756 (S. Kawashima); 0000-0002-6943-6887 (Y. Yamamoto); 0000-0003-4062-8903 (H. Chiba); 0000-0001-8195-5893 (Y. Moriya); 0000-0003-2391-0384 (T. Katayama)

between models and data. The portal leverages RDF-config automating updates and minimizing manual intervention. To further enhance reliability, tools like RDF-doctor and shXer are used to validate data and models during updates. These tools identify errors and inconsistencies, enabling timely corrections and ensuring the integrity of the RDF ecosystem.

Metadata plays a pivotal role in the RDF Portal's operations. By standardizing metadata using RDF-config, the portal synchronizes RDF datasets with associated metadata, schema diagrams, and sample queries. This synchronization streamlines updates, improves user experience, and enhances the overall reliability of the platform. Automating metadata handling facilitates the rapid incorporation of new datasets and adaptations to existing data structures, ensuring the portal remains up-to-date and user-friendly.

3. Current Datasets and Accessibility

Initially launched as the NBDC RDF Portal, it transitioned to DBCLS management in 2023 and was rebranded as the RDF Portal. The portal's URL changed from https://integbio.jp/rdf to https://rdfportal.org, along with updates to endpoint URLs, reflecting its new management framework.

The RDF portal now includes an expanded range of datasets. Beyond datasets submitted by contributors, it integrates public RDF data from resources like UniProt and PubChem, as well as RDF datasets developed by DBCLS. This broadens the potential applications of RDF data, fostering innovation in data-driven research.

By incorporating public RDF data, the portal not only increases its resource base but also encourages the development of applications that leverage RDF's unique properties.

The introduction of automated processes ensures that these datasets are maintained efficiently and accurately, reducing barriers for end-users.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] S. Kawashima, T. Katayama, H. Hatanaka, T. Kushida, T. Takagi, Nbdc rdf portal: a comprehensive repository for semantic data in life sciences, Database 2018 (2018) bay123. doi:10.1093/database/bay123.