FORVM: a Knowledge Graph to decipher associations between metabolites and diseases

M. Delmas ,O. Filangi, N. Paulhe, F. Vinson, C. Duperier, W. Garrier, P.-E. Saunier, Y. Pitarch, F. Jourdan, F. Giacomoni and C. Frainay

Séminaire Résidentiel INRAE Semantic Linked Data 2023

MetaboHUB is a French national infrastructure dedicated to metabolomics and fluxomics research. Its primary objective is to establish an integrated platform for studying metabolic pathways and networks. This initiative collaborates with a diverse range of academic and industrial partners, including experts in analytical chemistry and bioinformatics. Their collective effort aims to advance cutting-edge technologies and methodologies in the realm of metabolomics research. Within this framework, the infrastructure's focus lies in effectively structuring data and metadata using semantic web technologies. As a result of this endeavor, FORUM[1], a comprehensive open Knowledge Graph (KG), has been developed. FORUM is constructed through a large-scale federation of life science databases and repositories of scientific literature.

While metabolic profiling plays a pivotal role in human phenotyping studies, connecting specific metabolites to potential health impacts remains a challenge. This often necessitates extensive searches within the voluminous scientific literature. By the end of 2021, PubMed alone contained over 31 million citations. FORUM KG leverages automated ontological reasoning to establish implicit connections between higher-level concepts, representing groups of chemicals or disorders. This is achieved through mining the literature. The platform offers multiple levels of abstraction, facilitating hypothesis generation for interpreting metabolic signatures. To realize this, significant relations were computed based on co-mention frequency. These relations were instantiated as a new layer within the KG, thereby supporting result interpretation and fostering further investigations.

With over 9 billion factual statements and more than 9 million inferred relations between compounds and biomedical concepts, FORUM emerges as a valuable resource, adhering to the principles of Findable, Accessible, Interoperable, and Reusable (FAIR) data. The utility of FORUM is demonstrated by comparing findings with previously published test cases and by exploring intricate pathways of relations to propose novel hypotheses. Researchers can access extracted relations through a user-friendly web interface[2] for browsing and downloading. Additionally, a SPARQL endpoint[3] is available for direct querying of the FORUM KG.

In essence, FORUM stands as a pivotal achievement within the MetaboHUB infrastructure's mission to advance metabolomics and fluxomics research. Leveraging semantic web technologies, FORUM provides a robust and FAIR-compliant resource, empowering researchers with a comprehensive web interface for in-depth and exhaustive analysis of metabolomics data, thus significantly bolstering research in the field.

---

[1] Delmas, M. et al. FORUM: building a Knowledge Graph from public databases and scientific literature to extract associations between chemicals and diseases. Bioinformatics 37, 3896–3904 (2021)

[2] https://forum-webapp.semantic-metabolomics.fr/

[3] https://forum.semantic-metabolomics.fr/