

The Interplay of Wikidata and the Cell Ontology

Tiago Lubiana¹

¹ University of São Paulo, R. Cidade Universitária, 374 - Butantã, São Paulo, SP, Brazil

Abstract

This poster presentation centers on the potential of Wikidata, a comprehensive and collaborative knowledge base, to advance biomedical ontologies, particularly the Open Biological and Biomedical Ontologies (OBO) Foundry ontologies. The spotlight is on the integration of Wikidata with the ontology landscape, showcasing its integration with the Cell Ontology (CL). The CL, established around 2005, has been expanding to currently include over 2,600 cell classes, with technical enhancements and the addition of novel cell types over the years. Despite these advancements, contributing to it can be daunting, and its coverage of cell types is yet far from complete. Ontologies, though highly effective, require substantial technical expertise and specific software installations, restricting potential contributors with limited access or skills. In contrast, Wikidata promotes communal contribution and is welcoming for users of varying technical proficiency. Its strength stems from a vast community of active contributors, with broad coverage of diverse concepts. The speed, breadth, and accessibility of Wikidata, along with its CC0 public domain license, render it an inviting platform for widespread use and contribution. Our work exemplifies the advantages arising from the integration of Wikidata with OBO Foundry ontologies. For instance, Wikidata's multilingual capability promotes inclusivity in stark contrast to the predominantly anglocentric ontologies and, its direct linkages to Wikipedia facilitate access to textual descriptions often lacking in ontology development. Finally, Wikidata's connections to Wikipedia and other controlled vocabularies can support ontologies in validating information. To map concepts in Wikidata to the Cell Ontology, we developed a streamlined workflow combining ROBOT and a custom Python package for real-time natural language term curation in Wikidata, resulting in over 2,600 cross-references from Wikidata to the Cell Ontology. These curation efforts have revealed several cell types listed on Wikipedia but absent from CL, providing a unique opportunity for expansion and enhancement of CL's scope, as well as opening a gateway for multilingualism in applications that use CL identifiers. To sum up, the integration of Wikidata and ontologies can significantly benefit the ontology ecosystem, enriching the value of ontologies and fostering broader collaboration. Our work with the Cell Ontology suggests its connection other ontologies is a promising work direction for the future.

Keywords

Wikidata, Cell Ontology, cell types, knowledge graph.

1. Acknowledgements

This work was funded by FAPESP grant #19/26284-1.

Proceedings of the International Conference on Biomedical Ontologies 2023, August 28th-September 1st, 2023, Brasilia, Brazil

EMAIL: tiago.lubiana.alves@usp.br (A. 1)

ORCID: : 0000-0003-2473-2313 (A. 1)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)