Releasing PanglaoDB human cell-type markers to Wikidata

This manuscript ([permalink](#)) was automatically generated from [lubianat/semantic_web_course_report@73f41d9](#) on November 30, 2020.

Authors

- **Tiago Lubiana**

  ⓘD [0000-0003-2473-2313](#) · ◯ [lubianat](#)

  Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil

Abstract

Introduction

PanglaoDB [1] [2] is a public database that contains data and metadata on different types of cells. The types of cells are associated with marker genes, which are used to identify the classes that best fit cells observed in biomedical experiments.

Despite its usefulness for the community, the database is on a 3-star category for Linked Open Data [3] as it does not use open standards from W3C (RDF and SPARQL). To make it 5-star, it needs to be also linked to external data via common identifiers.

The OBO Foundry provides a rich collection of linked biological identifiers [4]. However, reconciliation to OBO is challenging, as there are many ontologies, each with slightly different contribution guidelines. For that reason, we decided to reconcile PanglaoDB to Wikidata, which allows simple creation of new terms, provided they follow Wikidata`s notability criteria[5].

Data source

- Panglao

Property creation on Wikidata

- Process of creating a property on Wikidata

Class creation on Wikidata

- Process of creating species-specific cell types on Wikidata.

RDF preparation

- Use of rdflib

Integration to Wikidata

- Use of Wikidata Integrator

SPARQL queries in Wikidata endpoint

Results of the integration

SPARQL query for cell types related to neurogenesis

- Linking with Wikidata allows out-of-the-box integrative SPARQL queries (thanks to Gene Wiki and the Gene Ontology integration)

- Dataset on Wikidata can be updated by the community and improved through time

Acknowledgements

References

1. **PanglaoDB - A Single Cell Sequencing Resource For Gene Expression Data**
   https://panglaodb.se/index.html

2. **PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data**
   Oscar Franzén, Li-Ming Gan, Johan LM Björkegren
   Database (2019-01-01) https://doi.org/ggkzxr
   DOI: 10.1093/database/baz046 · PMID: 30951143 · PMCID: PMC6450036

3. **Linked Data - Design Issues** https://www.w3.org/DesignIssues/LinkedData.html

4. **The OBO Foundry** http://www.obofoundry.org/

5. **Wikidata:Notability - Wikidata** https://www.wikidata.org/wiki/Wikidata:Notability