Building and Mining Knowledge Graphs
Course Code 2223-KEN4256
Academic Year: 2022-2023

Maastricht University

Written project report
NTDs2RDF: A heterogeneous and integrated knowledge graph for the exploration of neglected tropical diseases.

Sebastian Ayala Ruano i6314501

1. Abstract

Neglected tropical diseases (NTDs) affect millions of people in developing countries and have been historically overlooked in the global health agenda. Although progress has been made in controlling and eradicating these diseases, there are still open questions regarding key biological mechanisms and potential drug therapies. Biological and biomedical databases can facilitate the identification of novel drugs and targets for treating diseases like NTDs, but they present issues of standardization and integration. The semantic web, with its set of technologies and standards, can provide an interoperable and unified framework to study the NTDs. The NTDs2RDF project aims to create a knowledge graph (KG) of genes, proteins, metabolic pathways, gene ontologies, single nucleotide variants, drugs, and other relevant data for three NTDs (Chagas disease, Leishmaniasis, and African trypanosomiasis), integrating all the information in a single data structure that can be explored through a query interface implemented with a *Streamlit* web application. This software provides a user-friendly platform to extract information from the KG using *SPARQL* queries. The project represents an initial step towards the creation of a heterogeneous database for different NTDs with several potential applications in advancing the understanding of NTDs biology and providing insights that cannot be obtained through alternative resources. All underlying data and code are accessible through GitHub (https://github.com/sayalaruano/NTDs2RDF) under the MIT and CC0 licenses and archived on Zenodo (https://doi.org/10.5281/zenodo.7772555). Figure SI1 shows a graphical abstract for this project.

2. Significance

Neglected tropical diseases (NTDs) are a heterogeneous group of 20 bacterial, viral, parasitic, and fungal conditions (Figure SI2) that generally occur in developing countries in the Americas, Africa, and Asia[1]. NTDs mainly affect poor populations that do not have access to safe water, sanitation, and high-quality healthcare. Because of the severe effects of NTDs (i.e., long-lasting disabilities), they reinforce the cycle of poverty in vulnerable communities [2,3]. According to some surveys, around 20% of the world's population is susceptible to the harmful effects of NTDs[4]. Nonetheless, these diseases have been historically omitted from the global health agenda, leading to inadequate public health strategies[2,5].

There have been international projects to control and eradicate the NTDs, including the World Health Organization (WHO) NTD roadmaps 2012–2020 and 2021–2030[1], the London Declaration on NTDs[6], among other initiatives. According to data from the WHO (2023), the number of people requiring interventions against NTDs in the last decade has reduced by 25%. Despite the substantial progress made in NTDs research, there are still open questions about the key biological mechanisms and potential drug therapies for treating and preventing these complex diseases. Several studies have demonstrated that biological databases facilitate the identification of novel drugs, drug targets, side effects, and other information for treating diseases[7,8]. However, these resources present issues regarding standardization practices, creating difficulties in linking the data across databases[9]. Currently, there are several independent databases that contain biological knowledge of NTDs, but no integrated resource with all the information. This unified database could enable the systematic exploration of all the components of the NTDs, contributing to the research of potential therapies for these diseases.

Therefore, the NTDs2RDF project aims to create a knowledge graph (KG) of genes, proteins, metabolic pathways, gene ontologies, single nucleotide variants (SNVs), and drugs for three NTDs (Chagas disease, Leishmaniasis, and African trypanosomiasis), integrating everything in a single resource that can be queried through a user-friendly web application.

3. Related work

Life science research communities have shown increasing interest in semantic technologies and standards (e.g., RDF, OWL). These tools have contributed to organizing and linking data from diverse databases and provided an interoperable and unified framework to foster biomedical research[10,11]. There have been several projects to convert conventional biological databases into RDF stores. *Bio2RDF*[12] and *EBI-RDF*[13] are examples of projects for converting large structured biological databases into RDF graphs, which integrate knowledge from heterogeneous sources to facilitate the exploration and analysis of such complex biological information. Considering NTDs, a study created a semantic problem-solving environment for finding information about *Trypanosoma cruzi* using an RDF store, in which they integrated information proteomics, gene expression, and metabolic pathways[14]. Another research developed an RDF graph and a visual querying tool for the genomics data of T. *cruzi*[15]. Furthermore, *Bioportal*[16] and other similar projects have developed comprehensive repositories of biological ontologies and terminologies to standardize the data required to create RDF stores.

Despite the significant progress and efforts to develop biological databases for NTDs, there is still a lack of integrated data sources for many of these diseases. For instance, *EuPathDB*[17] contains genomic data for eukaryotic pathogens that cause some NTDs, *GNTD*[18] provides epidemiological information, and so on. To the best of our knowledge, no database has integrated biological data of NTDs from multiple sources in a standardized format.

4. Goal and specific objectives

The main goal of NTDs2RDF is to build a KG of genes, proteins, metabolic pathways, gene ontologies, SNVs, and drugs for three NTDs - Chagas disease, Leishmaniasis, and African trypanosomiasis. By creating the KG, researchers would have access to an integrated database that can be easily queried for information related to the three studied NTDs. This resource would benefit researchers and clinicians, who could use the insights gained from the data to develop more effective treatments, which ultimately could help NTD patients. Hence, this KG could improve the understanding of the molecular basis of NTDs, identifying novel drug targets and treatments, and eventually leading to better health outcomes for patients.

The specific objectives of the NTDs2RDF project are:

1. Collect relevant biological data for three NTDs from publicly available sources and identify relevant ontologies from standard repositories to build the KG.
2. Construct a KG with the collected data and ontologies.
3. Develop a query interface using a web application that allows users to retrieve information from the KG. Also, include examples of *SPARQL* queries to show the potential of the KG to advance research efforts for NTDs.

5. Methodology

This project was divided into various modules for data collection, data cleaning, programmatic creation of the KG, and the development of a web application for querying the RDF graph. An overview diagram of all the parts of the NTDs2RDF project is presented in **Figure 1**. Further details of all the elements of this workflow are presented below.

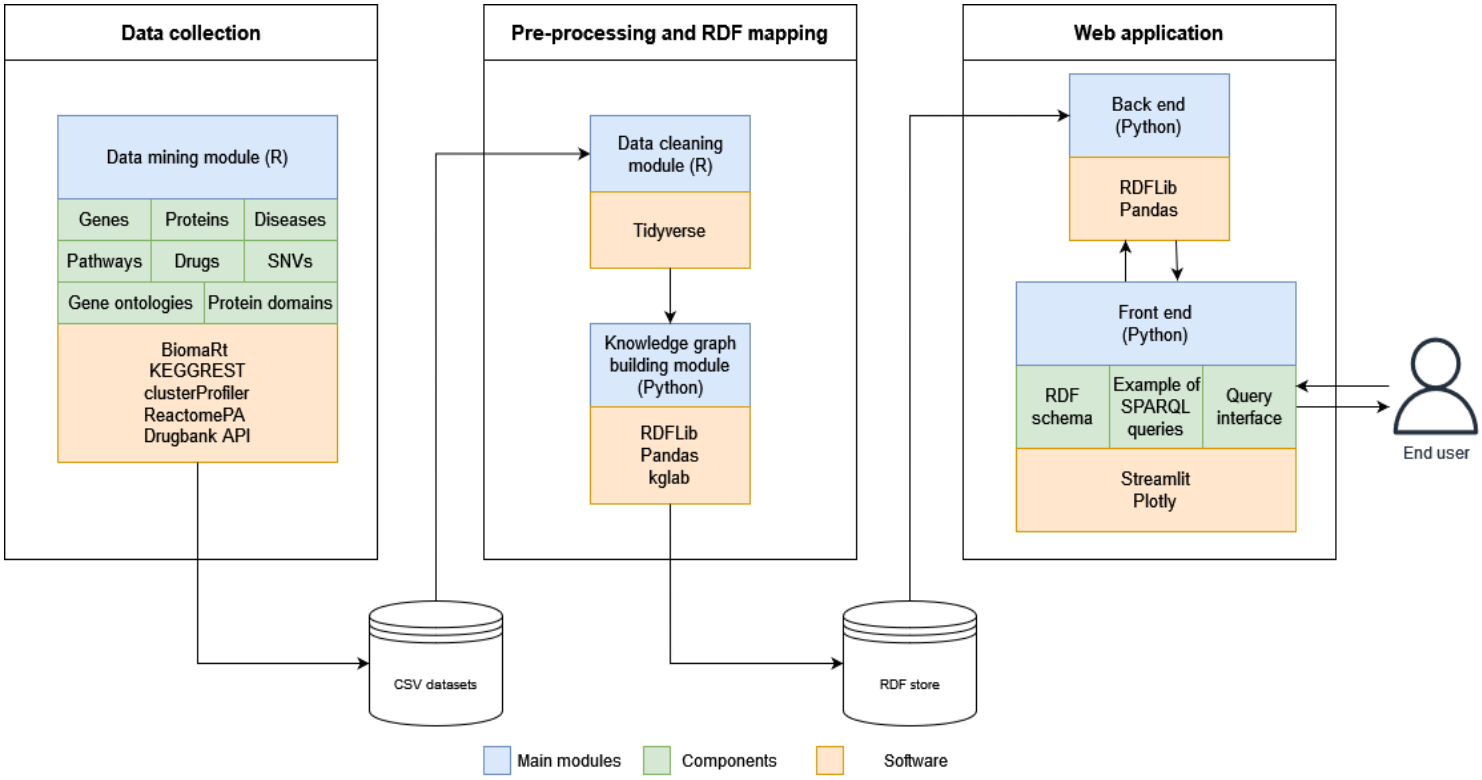

Figure 1.-Software architecture of the NTDs2RDF project. The workflow was divided into three main stages with some modules (blue rectangles) and components (green rectangles). The software libraries are shown as orange rectangles.

5.1. Data collection

The data collection for this project was developed through a data mining approach, using several R packages, APIs, and a systematic literature search (**Figure 1**). First, the human genes involved in the infection process of the three NTDs were obtained from the *KEGG*[19] database using the *KEGGREST* package[20]. The information about proteins, gene ontologies, protein domains, and some identifiers for external databases were retrieved with the *BiomaRt* package[21]. The *clusterProfiler*[22] and *ReactomePA*[23] packages were used to obtain the pathways associated to the genes for each disease, applying gene set enrichment analysis. The SNVs and their relationships with the genes and drugs were obtained from the *PharmGKB*[24] database, as well as the relationships between the drugs and proteins. Some of the relationships between the drugs and the diseases were obtained using the *Drugbank*[25] database API. Finally, additional external identifiers were mapped using the *TogoID*[26] web application. All the data was enriched with information from scientific articles and domain-specific databases through a systematic literature search. The scripts for this part can be found on this folder of the GitHub repository or the /Data_collection_processing folder in the provided zip file.

5.2. Preprocessing and RDF mapping

After retrieving all the datasets from the different sources as *CSV* files, a data cleaning and preprocessing step was applied to assure the correct format, datatype, among other details using some packages from *Tidyverse*[27]. Then, I looked for the proper terms to define the classes, predicates, and properties of the KG, which were mainly obtained from the *Biolink Model*[28] (v3.2.5) and some general terms from the *rdfs* and *xds* namespaces. Furthermore, the *Bioregistry* [29] metaregistry (v0.7.1) was used to assign standardized identifiers for all the biological entities of the KG. This information was used to create our KG with the proper subjects, objects, and predicates. Finally, the KG was constructed using *pandas*[30], *kglab*[31] and *rdflib*[32], exporting the result with the *n-triples* and *turtle* syntaxes. The script for this part can be found on the jupyter notebook of the GitHub repository or the /RDF_graph_building.ipynb file in the provided zip file.

5.3. Web application

To guarantee the accessibility and reusability of the KG, a multi-page web application was created. This software was divided into three pages: Meta-graph, examples of *SPARQL* queries, and query interface. The back end of the software was developed with *pandas*[30] and *rdflib*[32], while the front end was created with *Streamlit*[33] and *Plotly*[34] (**Figure 1**). The app was deployed with the *Streamlit*'s Community Cloud. All the files required to create and deploy the web application are available in the GitHub repository of this project or the /Webapp folder in the provided zip file. Furthermore, it is possible to run the web application locally by creating a Python virtual environment with *pipenv*[35] or *conda*[36] and installing the software dependencies. The instructions on how to run the web application locally are available in the README files of the GitHub repository and the zip file.

6. Results

6.1. Knowledge graph

The NTDs2RDF KG is available at Zenodo, the GitHub repository, or the /KG_building/Data/RDF_graphs folder of the zip file with *n-triples* and *turtle* syntaxes. The KG has ten classes and 33.892 triplets. **Figure 2** shows the meta-graph of the NTDs2RDF KG. The *Biolink Model*[28] was the main source to define classes and predicates of the RDF graph, which gave several advantages over other resources. *Biolink* has many human-readable and domain-specific predicates for Biology, which captures the details of the biological knowledge (**Figure 2**). For instance, the predicates "biolink:is_sequence_variant_of", "biolink:gene_associated_with_condition", and "biolink:has_sequence_location" are explicit and easy to understand, which do not have counterparts in other ontologies. Before working with *Biolink*, I was using the terms suggested by *BioPortal*[37], which were mainly from the *National Cancer Institute Thesaurus (NCIT)*[38] and the *Open Biomedical Ontologies (OBO)*[39]. The classes and predicates of the *NCIT* and *OBO* ontologies were not human-readable because they were associated with codes (e.g., C16612, NCIT_C25281), and most of the URIs did not work (e.g., http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C16612).

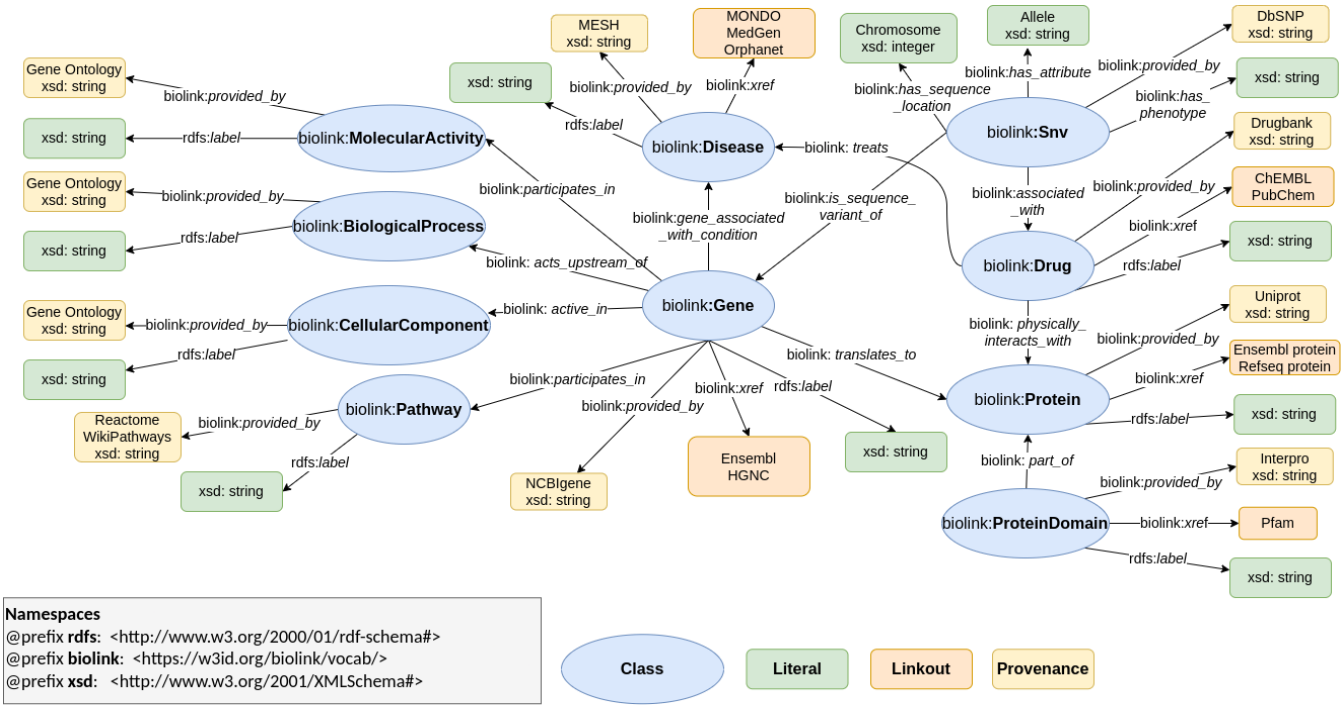

Figure 2.-Meta-graph of the NTDs2RDF KG. The classes are represented as blue ovals, the literals as green rounded rectangles, the references to external databases as orange rounded rectangles, and the provenance for the classes as yellow rounded rectangles.

Besides being human-readable, the entities from *Biolink* follow a machine-readable format and they are integrated with other ontologies using semantic mappings, which makes the KGs interoperable. Other ontologies like *NCIT* and *OBO* are machine readable, but they lack interoperability Finally, the *Biolink* data model provides class properties to link external identifiers ("biolink:x_ref" predicate) and to track the provenance of the data ("biolink:provided_by" predicate).

The standardized identification of entities is another important aspect to consider when creating KGs. Biological databases tend to create their own unique identifiers and they are not always cross-referenced with external entries, hindering the integration and interoperability of biological knowledge[40]. To assure that all the entities of the NTDs2RDF KG have standardized and stable identifiers, the *Bioregistry*[29] metaregistry was used to assign the URIs and CURIEs.

6.2. Web application

To guarantee the accessibility and findability of the NTDs2RDF KG, a multi-page *Streamlit* web application was developed. The home page of the web app has general information about the project and loads the RDF graph and the *SPARQL* queries objects (**Figure SI3**). The next page shows the meta-graph of the KG with all the classes and predicates from the *Biolink Model* and other standard ontologies (**Figure 2**). The third page has examples of *SPARQL* queries to extract relevant information from the RDF graph and demonstrate its usefulness in gaining new insights into NTDs. The loading time for this page typically takes approximately two minutes due to the intricacy of some queries. Also, the query examples are available on a txt file of the GitHub repository or at the /sparql_queries_NTDs_RDF_examples.txt file of the zip file. The last page of the web application consists of a user-friendly *SPARQL* query interface, which provides a simple method for extracting information from the KG (**Figure SI4**).

Figure 3 presents screenshots of two *SPARQL* queries from the web application and their visualization plots. The query of **Figure 3A** reveals that Toll-interleukin receptor domains are among the top 20 protein domains of genes associated with *African Trypanosomiasis*. Moreover, the query featured in **Figure 3B** shows that the Toll-like receptor signaling pathways have the highest number of genes involved in the three NTDs. Other *SPARQL* queries support these results (refer to the corresponding page of the web application). The Toll-like receptors have been identified as key components of the host immune response against several parasitic diseases[41], so maybe these proteins can also play a role in the NTDs infection. These findings could serve as the starting point for novel research hypotheses that could advance the understanding of NTDs biology and provide insights that cannot be obtained through alternative resources.

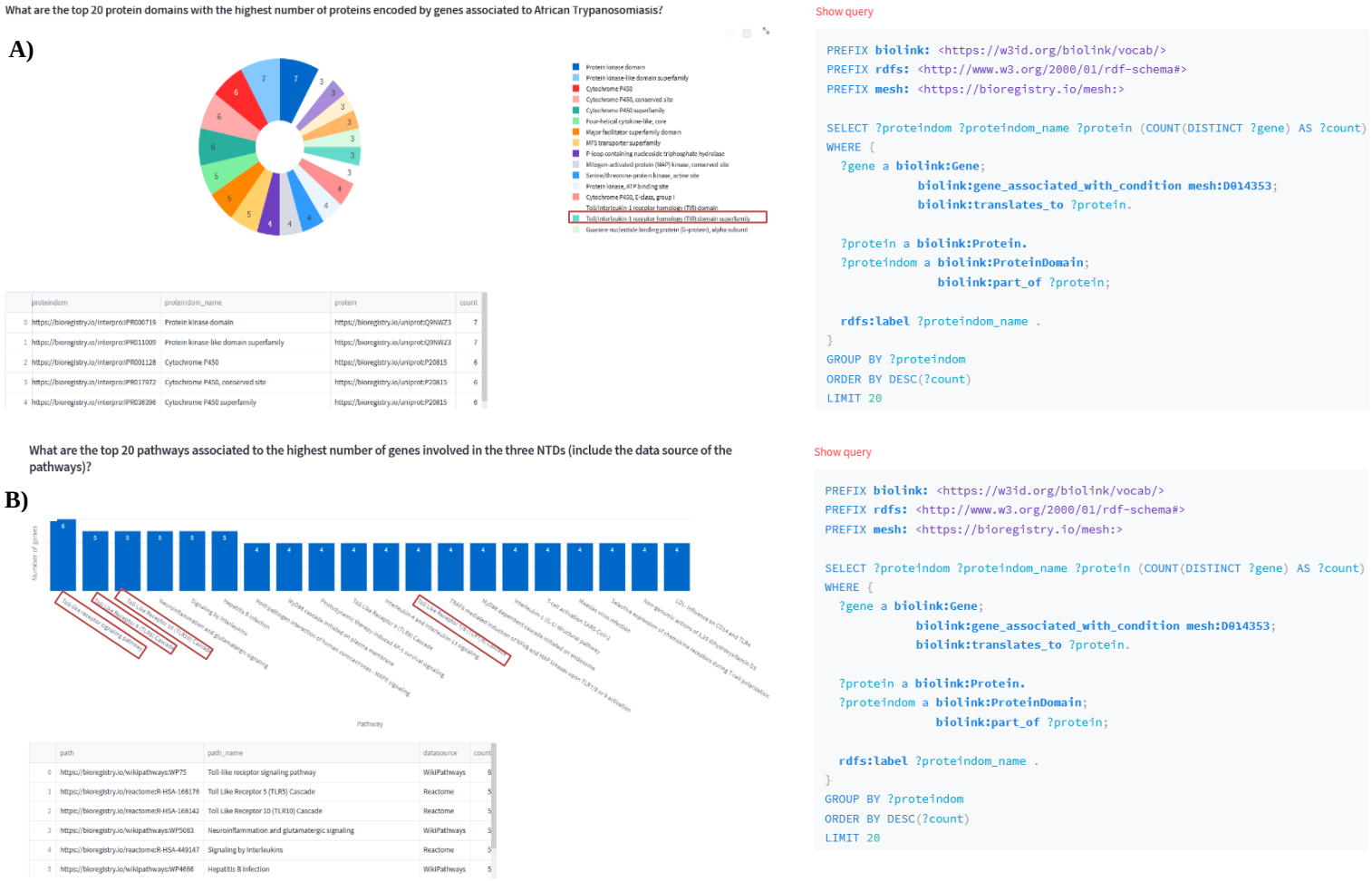

Figure 3.-Two examples of SPARQL queries of the NTDs2RDF web application. Each query is linked with a research question and shows the results in a table and different plots. The protein domains and pathways related to the Toll-interleukin receptors are enclosed by red rectangles.

7. Discussion

As mentioned in the *Significance section*, several projects have converted huge biological databases into RDF stores (e.g., *Bio2RDF* and *EBI-RDF*). However, these resources are broad and do not have specific information to understand diseases like NTDs because most of the knowledge about them is in scientific articles and specialized repositories that are not included in big databases. Therefore, the systematic literature search step was crucial on the creation of the NTDs2RDF KG, making this resource more suited to contribute to the study of NTDs than general RDF stores. Considering the application of semantic web technologies to understand NTDs, previous studies have developed RDF graphs for individual NTDs and specific biological knowledge (refer to *section 3* for the specific examples). Hence, the NTDs2RDF KG is an initial step in the creation of a heterogeneous database for different NTDs with several potential applications.

Another important feature of the NTDs2RDF KG is the use of the *Biolink Model* to provide human and machine-readable names for the classes and predicates, and the *Bioregistry* metaregistry to assign standard identifiers. As explained in the *section 4.1* of the Results, traditional ontologies such as *NCIT* and *OBO* present several limitations in terms of human-readability, broken links, and interoperability, so it is better to use the terms from the *Biolink Model*. By using these integrative and open-source standards, the KG represents a useful and unique resource that can be queried for information related to the three studied NTDs. One potential disadvantage of using resources like *Biolink* and *Bioregistry* is that they are in continuous development, meaning that some of the information derived from them will be outdated at some point. As a result, it is important to specify the version of the resources and try to maintain the KG with the upcoming updates.

To make a KG useful for the research community, it should adhere to the FAIR principles[42]. To this end, I archived the KG in Zenodo and included information about its license and other details on how to use it. Moreover, I created a web

application to retrieve information from the KG through *SPARQL* queries. Considering the lack of programming expertise of some researchers in the life sciences, creating a user-friendly query interface with examples of use-case *SPARQL* queries was an important contribution of this project. By doing so, the information of KGs can be available to a broad range of users, including those lacking extensive technical expertise.

Nonetheless, it is noteworthy that the NTDs2RDF KG has several limitations. First, it only contains data from 3 out of the 20 NTDs, which prevents finding biological commonalities or similar therapeutic strategies among most NTDs. Furthermore, most *SPARQL* queries retrieve broad information that provide general facts for the diseases, which are not necessarily novel knowledge. Creating more complex queries and expanding the range of data sources to build the KG may facilitate the generation of customized *SPARQL* queries, thereby obtaining novel insights about the NTDs. In addition, the literature review was done by a non-expert of the field, so it is likely that crucial information about the NTDs was omitted. The final point can be improved by asking the advice of domain experts or using machine learning-based strategies (e.g., natural language processing algorithms) that find patterns in large amounts of articles and domain-specific databases.

Despite these shortcomings, the NTDs2RDF project is a starting point to create an heterogenous and integrated KG for NTDs. This resource can be expanded for other NTDs, representing a significant contribution to the research community working in this field. In the long term, the NTDs2RDF can benefit researchers and clinicians, who could use the insights gained from the data to develop more effective treatments, which ultimately could help NTD patients.

8. Conclusions and further work

In conclusion, the NTDs2RDF project has successfully transformed heterogeneous and dispersed data of three NTDs (Chagas disease, Leishmaniasis, and African trypanosomiasis) into a unified KG, which can be queried and explored through a user-friendly web application. By using the *Biolink* and *Bioregistry* standards, the KG had human/machine-readable and domain-specific entities, standardized identifiers, and interoperability with other ontologies. Moreover, I endorsed the application of the *FAIR* principles in this project by publishing the results in open repositories (Zenodo and GitHub), providing proper licenses and metadata, and by creating a web application with a query interface and some examples of *SPARQL* queries to demonstrate use-cases of the KG. Overall, the NTDs2RDF project represents a significant contribution to the research of NTDs, which could improve the understanding of the molecular basis of these diseases, and eventually leading to better health outcomes for patients.

Looking forward, the NTDs2RDF project has the potential for future development. One possibility is to include more NTDs in the KG, which could provide a broader understanding of the molecular mechanisms underlying the diseases. Another avenue for improvement is to automate the literature searching process using Natural Language Processing (NLP) algorithms, which could extract more specific data about the diseases and improve the quality of the data in the KG. This, in turn, would enhance the effectiveness of *SPARQL* queries. In addition, the application of rule mining or other machine learning strategies could be used to predict links between entities, which could uncover new relationships between the molecular components of the diseases. Finally, graph analytics could be used to identify hub nodes, shortest pathways, communities, and other relevant insights, which could provide a deeper understanding of the NTDs and help to identify new targets for drug discovery.

9. Supplementary information

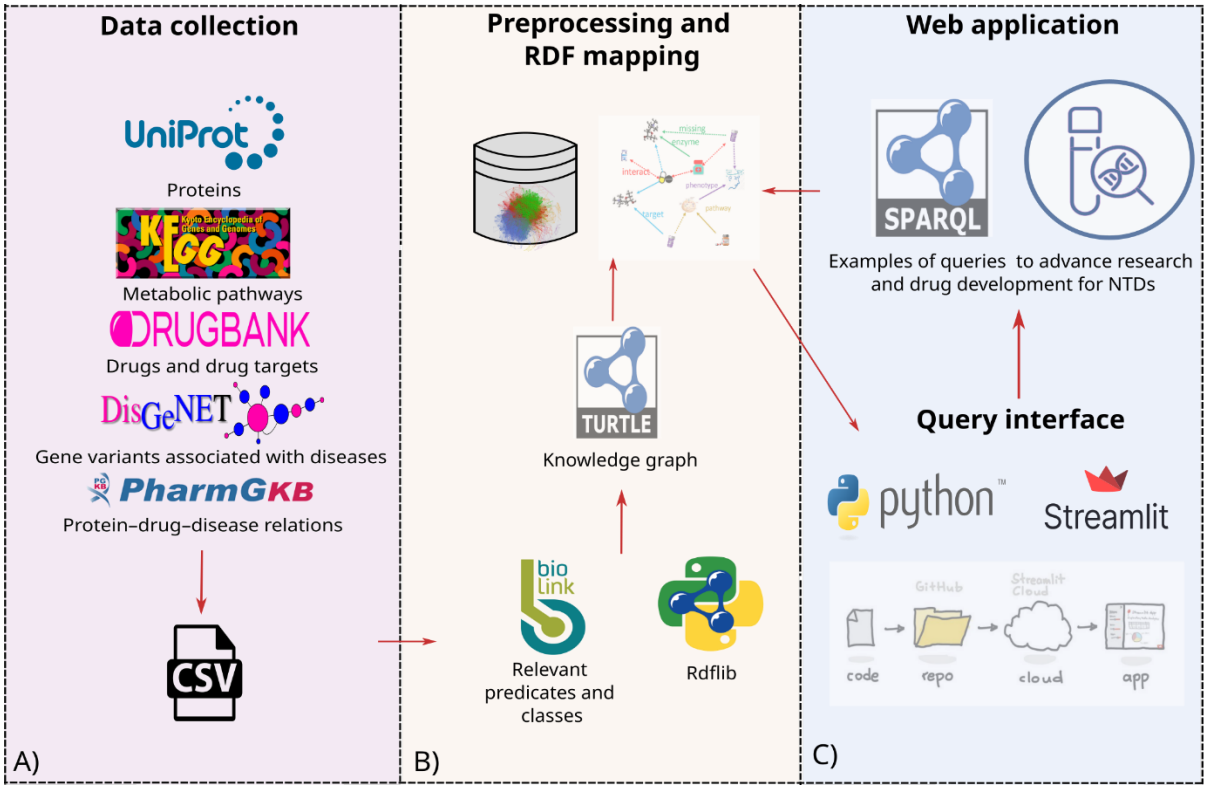

Figure SI1. Overview diagram of the NTDs2RDF project.

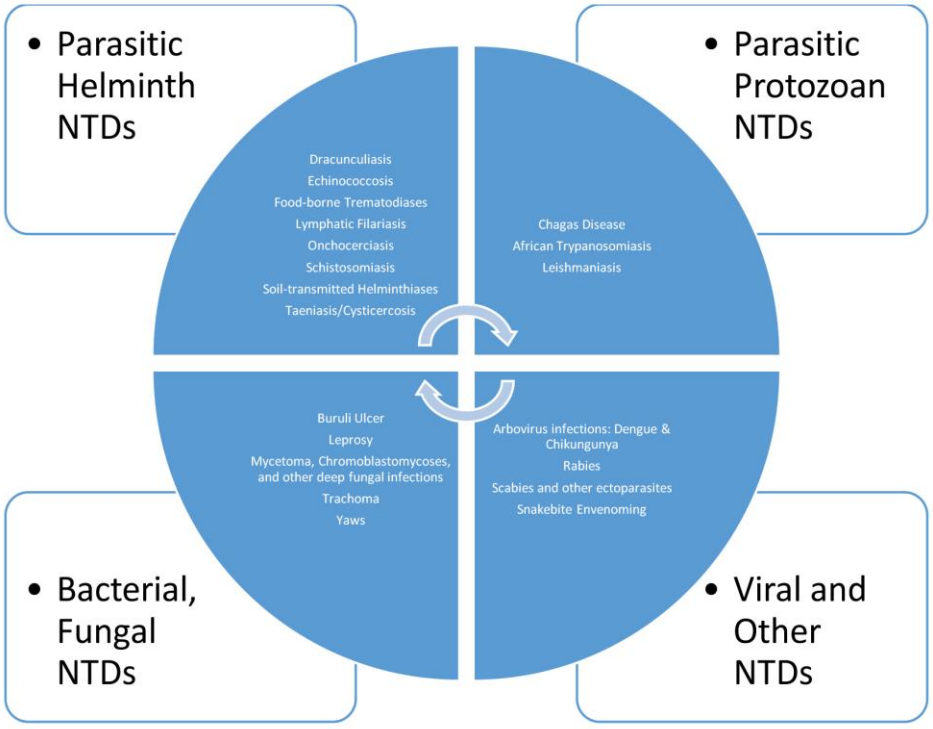

Figure SI2. Classification of NTDs according to the WHO. Retrieved from [43]

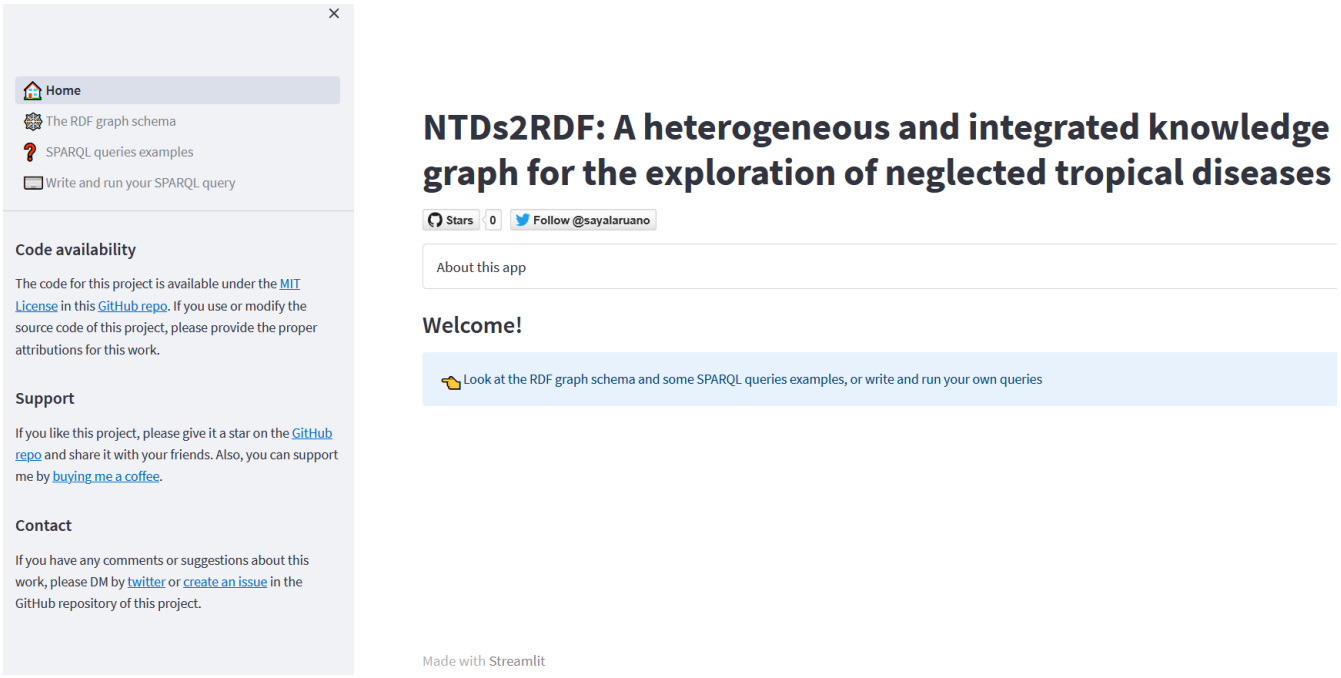

Figure SI3. Homepage of the NTDs2RDF web application.

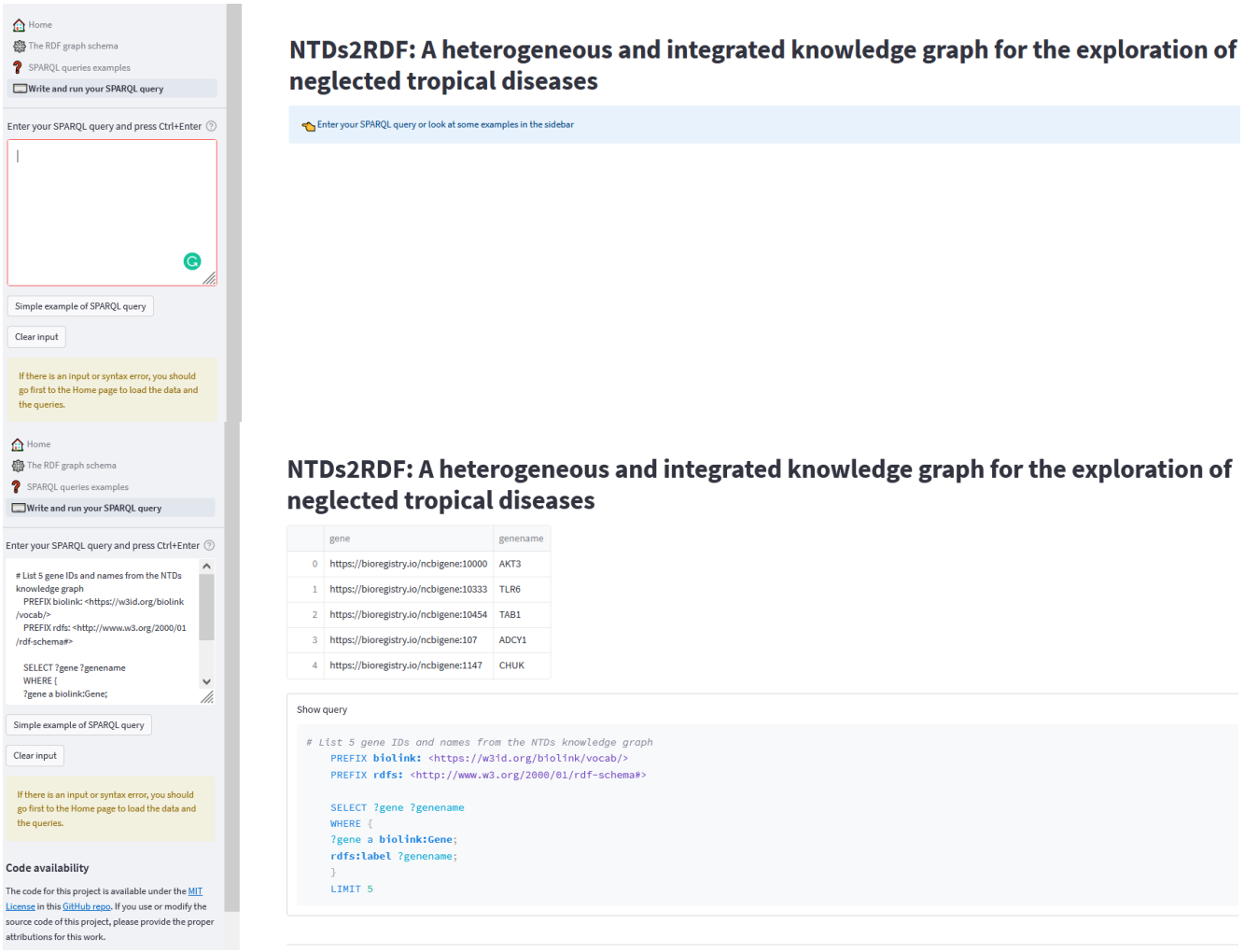

Figure SI4. Query interface page of the NTDs2RDF web application.

10. References

1. World Health Organization. *Global report on neglected tropical diseases 2023*. (World Health Organization, 2023).
2. Lin, Y., Fang, K., Zheng, Y., Wang, H. & Wu, J. Global burden and trends of neglected tropical diseases from 1990 to 2019. *Journal of Travel Medicine* **29**, taac031 (2022).
3. Hotez, P. J., Fenwick, A., Savioli, L. & Molyneux, D. H. Rescuing the bottom billion through control of neglected tropical diseases. *The Lancet* **373**, 1570–1575 (2009).
4. The Lancet. Neglected tropical diseases: ending the neglect of populations. *The Lancet* **399**, 411 (2022).
5. Engels, D. & Zhou, X.-N. Neglected tropical diseases: an effective global response to local poverty-related disease priorities. *Infectious Diseases of Poverty* **9**, 10 (2020).
6. Molyneux, D. H. The London Declaration on Neglected Tropical Diseases: 5 years on. *Transactions of The Royal Society of Tropical Medicine and Hygiene* **110**, 623–625 (2016).
7. Hassani-Pak, K. & Rawlings, C. Knowledge Discovery in Biological Databases for Revealing Candidate Genes Linked to Complex Phenotypes. *Journal of Integrative Bioinformatics* **14**, (2017).
8. Baxevanis, A. D. & Bateman, A. The Importance of Biological Databases in Biological Discovery. *Current Protocols in Bioinformatics* **50**, 1.1.1-1.1.8 (2015).
9. Wilson, S. L. *et al.* Sharing biological data: why, when, and how. *FEBS Letters* **595**, 847–863 (2021).
10. Cheung, K.-H. *et al.* A journey to Semantic Web query federation in the life sciences. *BMC Bioinformatics* **10**, S10 (2009).
11. Kamdar, M. R., Fernández, J. D., Polleres, A., Tudorache, T. & Musen, M. A. Enabling Web-scale data integration in biomedicine through Linked Open Data. *npj Digit. Med.* **2**, 1–14 (2019).
12. Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P. & Morissette, J. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics* **41**, 706–716 (2008).
13. Jupp, S. *et al.* The EBI RDF platform: linked open data for the life sciences. *Bioinformatics* **30**, 1338–1339 (2014).
14. Parikh, P. P. *et al.* A Semantic Problem Solving Environment for Integrative Parasite Research: Identification of Intervention Targets for Trypanosoma cruzi. *PLOS Neglected Tropical Diseases* **6**, e1458 (2012).
15. Mendes, P. N., McKnight, B., Sheth, A. P. & Kissinger, J. C. TcruziKB: Enabling Complex Queries for Genomic Data Exploration. in *2008 IEEE International Conference on Semantic Computing* 432–439 (2008). doi:10.1109/ICSC.2008.93.
16. Martínez-Romero, M. *et al.* NCBO Ontology Recommender 2.0: an enhanced approach for biomedical ontology recommendation. *Journal of Biomedical Semantics* **8**, 21 (2017).
17. Aurrecoechea, C. *et al.* EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Research* **38**, D415–D419 (2010).
18. Hürlimann, E. *et al.* Toward an Open-Access Global Database for Mapping, Control, and Surveillance of Neglected Tropical Diseases. *PLOS Neglected Tropical Diseases* **5**, e1404 (2011).
19. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* **42**, D199–D205 (2014).
20. Tenenbaum, D. KEGGREST: Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG). (2022).
21. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184–1191 (2009).
22. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).
23. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. BioSyst.* **12**, 477–479 (2016).
24. Barbarino, J. M., Whirl-Carrillo, M., Altman, R. B. & Klein, T. E. PharmGKB: A worldwide resource for pharmacogenomic information. *WIREs Systems Biology and Medicine* **10**, e1417 (2018).
25. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* **46**, D1074–D1082 (2018).
26. Ikeda, S. *et al.* TogoID: an exploratory ID converter to bridge biological datasets. *Bioinformatics* **38**, 4194–4199 (2022).
27. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).
28. Unni, D. R. *et al.* Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. *Clinical and Translational Science* **15**, 1848–1855 (2022).
29. Hoyt, C. T. *et al.* Unifying the identification of biomedical entities with the Bioregistry. *Sci Data* **9**, 714 (2022).
30. The pandas development team. pandas-dev/pandas: Pandas. (2023) doi:10.5281/zenodo.7741580.
31. Nathan, P. *et al.* DerwenAI/kglab: v0.5.2 release on PyPi. (2022) doi:10.5281/zenodo.6360664.

32. Grimnes, G. A. *et al.* RDFLib/rdflib: RDFlib 6.3.1. (2023) doi:10.5281/zenodo.7748890.
33. Streamlit. Streamlit - The fastest way to build and share data apps. https://streamlit.io/undefined/ (2021).
34. Plotly Technologies Inc. Plotly: Low-Code Data App Development. *Plotly: Low-Code Data App Development* https://plotly.com/.
35. The pipenv development team. pipenv: Python Development Workflow for Humans.
36. Anaconda. Anaconda - The World's Most Popular Data Science Platform. *Anaconda* https://www.anaconda.com/.
37. Noy, N. F. *et al.* BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research* **37**, W170–W173 (2009).
38. de Coronado, S. *et al.* The NCI Thesaurus quality assurance life cycle. *Journal of Biomedical Informatics* **42**, 530–539 (2009).
39. Smith, B. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* **25**, 1251–1255 (2007).
40. Lapatas, V., Stefanidakis, M., Jimenez, R. C., Via, A. & Schneider, M. V. Data integration in biological research: an overview. *Journal of Biological Research-Thessaloniki* **22**, 9 (2015).
41. Aguirre-García, M. M. *et al. TLR-Mediated Host Immune Response to Parasitic Infectious Diseases*. *Toll-like Receptors* (IntechOpen, 2019). doi:10.5772/intechopen.84679.
42. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
43. Hotez, P. J., Aksoy, S., Brindley, P. J. & Kamhawi, S. World neglected tropical diseases day. *PLOS Neglected Tropical Diseases* **14**, e0007999 (2020).