



queryMed: Semantic Web functions for linking pharmacological and medical knowledge to data

Yann Rivault, Olivier Dameron, Nolwenn Le Meur

► To cite this version:

Yann Rivault, Olivier Dameron, Nolwenn Le Meur. queryMed: Semantic Web functions for linking pharmacological and medical knowledge to data. *Bioinformatics*, Oxford University Press (OUP), 2019, 35 (17), pp.3203-3205. 10.1093/bioinformatics/btz034 . hal-01988699

HAL Id: hal-01988699

<https://hal.archives-ouvertes.fr/hal-01988699>

Submitted on 5 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

queryMed: Semantic Web functions for linking pharmacological and medical knowledge to data

Y. Rivault^{1,2}, O. Dameron¹ and N. Le Meur²

¹ Univ Rennes, CNRS, Inria, IRISA - UMR 6074, F-35000 Rennes, France.

²Univ Rennes, EHESP, REPERES (Pharmacology and health services research) - EA 7449, F-35000 Rennes, France.

Summary: In public health research and more precisely in the reuse of electronic health data, selecting patients, identifying specific events and interpreting results typically requires biomedical knowledge. The queryMed R package aims to facilitate the integration of medical and pharmacological knowledge stored in formats compliant with the Linked Data paradigm (e.g. OWL ontologies and RDF datasets) into the R statistical programming environment. We show how it allowed us to identify all the drugs prescribed for critical limb ischemia (CLI) and also to detect one contraindicated prescription for one patient by linking a medical database of 1003 CLI patients to ontologies.

Availability: queryMed is readily usable for medical data mappings and enrichment. Sources, R vignettes and test data are available on GitHub (<https://github.com/yannrivault/queryMed>) and are archived on Zenodo (<https://doi.org/10.5281/zenodo.1323481>).

1 Introduction

Medico-administrative databases (e.g. electronic prescription databases and medical insurance databases) are useful for public health research, especially in pharmaco-epidemiology. Patient health status and care prescriptions enable identifying severe drug interactions (Nobili et al., 2009) or adverse drug reactions (Durrieu et al., 2014), and studying their determinant on a large scale population. Medical knowledge is necessary for each research step: before the analysis to select patients, data, and to identify specific events, or after the analysis to interpret the results. Typically, data in such databases are stored following various semantics and syntactic standards. These data encodings are instrumental for semantic interoperability and knowledge-based data exploration (Pathak et al., 2013). Their exploitation typically requires non-trivial ad-hoc processing.

Semantic Web technologies provide the underlying framework for both syntactic and semantic interoperability, ultimately facilitating data integration. Based on these technologies, the Linked Data paradigm promotes the publication of structured and interlinked knowledge bases. They can be queried on remote servers, with SPARQL (<https://www.w3.org/TR/sparql11-overview/>), a query language targeting RDF graphs, or alternatively on REST APIs. For example, BioPortal, one of the principal biomedical ontology repositories, provides methods and tools to access biomedical knowledge through SPARQL endpoints and REST APIs (Whetzel et al., 2011). However, the growing number of knowledge bases, the heterogeneity of their schema representation, and the lack of their conceptual description, make the reuse of knowledge bases –and especially SPARQL query design– a challenge (Jain et al., 2010). Some important works contributed to reduce these bottlenecks. In particular, in medical domain, the Concept Unique Identifier (CUI) from the Unified Medical Language System has been widely used in ontologies to provide mappings between medical terms from different nomenclatures. In addition, the Drug Indication Database (DID) (Sharp, 2017) and the Drug Interaction Knowledge Base (DIKB) (Ayvaz et al., 2015) have gathered knowledge sources of drug indications and interactions from more than a dozen of Linked Data sources. Despite these works, knowledge integration in a patient database is still a laborious process for a non-expert. (Ferreira et al., 2012) highlighted the need for tools that facilitate the use of ontologies and knowledge bases, especially in epidemiology.

Because data exploration is strongly linked with data analysis, several tools (Van Hage et al., 2013; Willighagen, 2014; Kurbatova et al., 2015) have made available Linked Data knowledge bases in the R statistical programming environment (R Core Team, 2017). However, designing SPARQL queries remains a technical barrier that none of the current R packages help to overcome. We present queryMed, an R library of functions that aims to help users overcoming this barrier in the field of health data. By providing predefined queries and functions, queryMed thus aims to facilitate the access and linkage of medical and pharmacological knowledge to patient databases and thereby knowledge-based exploration.

2 Methods

The main low-level function, `sparql(query, endpoint url)` sends queries to SPARQL endpoints. As this function requires an expertise in SPARQL query design, biomedical ontologies and appropriate SPARQL endpoints, higher level functions hide predefined SPARQL queries. `queryMed` offers embedded SPARQL queries suitable for Bio2RDF (Callahan et al., 2013), DBpedia (Lehmann et al., 2015) and Ontobee(Ong et al., 2017) servers. The provided queries focus on drugs and diseases and retrieve definitions, abstracts, labels, comments, synonyms, mappings between nomenclatures or particular annotations like contraindications between drugs and diagnosis.

Alternatively, `search(terms, ontologies)` uses the search REST APIs

from BioPortal and SIFR BioPortal (Jonquet et al., 2016). This function also requires an expertise in BioPortal repository and its ontologies but is the cornerstone for user-friendly CUI based mapping function. Moreover, `mapping_cui(codes,source,target)` enables nomenclatures linkage, and so expand the potential sources of useful knowledge (Figure 1). Additionally, queryMed provides curated versions of DID and DIKB. Using the retrieved knowledge, `find_relations(...)` function searches for complex semantic relations that involve two medical terms, like drug indications, drug interactions or drug-disease contraindications.

3 Application

We used queryMed in the context of a pharmacovigilance scenario, searching for critical limb ischemia (CLI) patients with at least one prescription contraindicated with CLI or without any indicated prescription. Drug prescriptions 15 days prior surgery to 31 days after surgery were extracted from the French national system of health data (Medico-administrative data) for 1003 patients operated in 2015. Using queryMed functions calls, we extracted information on drug indications regarding CLI from DID, and we could identify 72 CLI patients that had potentially no prescription listed as indicated for CLI during the respective period. These patients may either had a flawed care trajectory, or had a drug delivery in hospital, or DID might not be exhaustive for CLI drugs. queryMed allowed us to query Linked Data sets without having to learn SPARQL first. It thus enabled the identification of 91 drugs contraindicated with CLI, according to NDF-RT ontology. As a result, using a combination of queryMed function calls, we identified 1 patient with a prescription for a vasoconstrictor, which is contraindicated with CLI (Figure 1 and queryMed vignette for details).

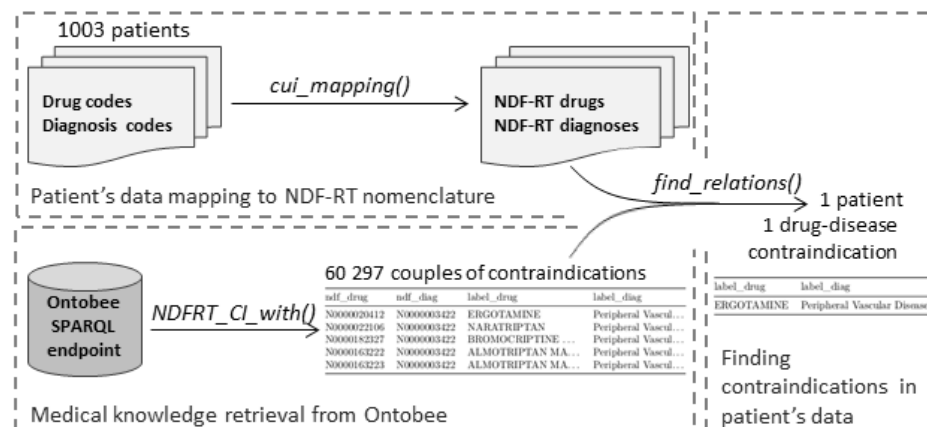


Figure 1: queryMed use in a pharmacovigilance oriented scenario.

4 Conclusion and discussion

queryMed provides R functions to facilitate linking external knowledge stored in data sources following the Linked Data paradigm, typically accessible through SPARQL interfaces, with health data. In the context of CLI disease, it enabled answering healthcare recommendations follow up questions and detecting possible drug interactions. queryMed design provides simplicity to the users while dealing with complex and various knowledge representations. The current limitation of queryMed is its small number of predefined queries. Future work will focus on expanding this set of prebuilt queries for other use cases as well as for accommodating the possible Linked Data evolutions (e.g. new knowledge sources).

Acknowledgements and Fundings

This work was funded by the French Health Products Agency. The study protocol was approved by the National Institute of Health Data review board (#201) and by the French data protection authority (#1968571). The authors declare no conflict of interest.

References

- Ayvaz, S. *et al.* Toward a complete dataset of drug-drug interaction information from publicly available sources. *Journal of Biomedical Informatics*, 55:206–217, June 2015. ISSN 15320464. doi: 10.1016/j.jbi.2015.04.006. 2
- Callahan, A. *et al.* Bio2RDF Release 2: Improved Coverage, Interoperability and Provenance of Life Science Linked Data. In *The Semantic Web: Semantics and Big Data*, Lecture Notes in Computer Science, pages 200–212. Springer, Berlin, Heidelberg, May 2013. ISBN 978-3-642-38287-1 978-3-642-38288-8. doi: 10.1007/978-3-642-38288-8_14. 2
- Durrieu, G. *et al.* Use of administrative hospital database to identify adverse drug reactions in a Pediatric University Hospital. *European Journal of Clinical Pharmacology*, 70(12):1519–1526, December 2014. ISSN 1432-1041. doi: 10.1007/s00228-014-1763-1. 1
- Ferreira, J. D. *et al.* On the usefulness of ontologies in epidemiology research and practice. *Journal of Epidemiology and Community Health*, 67(5):385–388, nov 2012. doi: 10.1136/jech-2012-201142. 2
- Jain, P. *et al.* Linked Data is Merely More Data. *Papers from the AAAI Spring Symposium: Linked Data Meets Artificial Intelligence*, pages 82–86, January 2010. 2
- Jonquet, C. *et al.* SIFR BioPortal : Un portail ouvert et gnrique d’ontologies et de terminologies biomdicales franaises au service de l’annotation smantique. In *JFIM: Journes Francophones d’Informatique Mdicale*, e-health pour tous, Genve, Switzerland, June 2016. 3
- Kurbatova, N. *et al.* *ontoCAT: Ontology traversal and search*, 2015. R package version 1.29.0. 2
- Lehmann, J. *et al.* DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6:167–195, 2015. doi: 10.3233/SW-140134. 2

- Nobili, A. *et al.* Potentially severe drug interactions in elderly outpatients: results of an observational study of an administrative prescription database. *Journal of Clinical Pharmacy and Therapeutics*, 34(4):377–386, August 2009. ISSN 1365-2710. doi: 10.1111/j.1365-2710.2009.01021.x. 1
- Ong, E. *et al.* Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic Acids Research*, 45(D1):D347–D352, January 2017. ISSN 1362-4962. doi: 10.1093/nar/gkw918. 2
- Pathak, J. *et al.* Using linked data for mining drug-drug interactions in electronic health records. *Studies in Health Technology and Informatics*, 192:682–686, 2013. ISSN 0926-9630. doi: 10.3233/978-1-61499-289-9-682. 1
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. 2
- Sharp, M. E. Toward a comprehensive drug ontology: extraction of drug-indication relations from diverse information sources. *Journal of Biomedical Semantics*, 8(1), December 2017. ISSN 2041-1480. doi: 10.1186/s13326-016-0110-0. 2
- Van Hage, W. R. *et al.* *SPARQL: SPARQL client*, 2013. R package version 1.16. 2
- Whetzel, P. L. *et al.* BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research*, 39(Web Server issue):W541–545, July 2011. ISSN 1362-4962. doi: 10.1093/nar/gkr469. 2
- Willighagen, E. Accessing biological data in R with semantic web technologies. *PeerJ PrePrints*, mar 2014. doi: 10.7287/peerj.preprints.185v3. 2