

queryMed : enrichissement des données de recherche en pharmaco-épidémiologie

Y. Rivault^{a,b}, O. Dameron^b, and N. Le Meur^a

^a Univ Rennes, EHESP, REPERES (Recherche en Pharmaco-épidémiologie et Recours aux Soins) - EA 7449, F-35000 Rennes, France
{yann.rivault ; nolwenn.lemeur}@ehesp.fr

^b Univ Rennes, IRISA, Dyliss, F-35000 Rennes, France
olivier.dameron@univ-rennes1.fr

Mots clefs : pharmaco-épidémiologie, Web des données, Web sémantique, ontologies

La réutilisation des données médico-administratives présente un intérêt pour la recherche en santé publique, notamment pour la pharmaco-épidémiologie[1]. En France, le Système National des Données de Santé (SNDS) recueille des informations sur toutes les consommations de soins remboursées par l'assurance maladie (consultations, hospitalisations et prescriptions médicamenteuses). Leur analyse vise par exemple à évaluer les risques et bénéfices de consommations de soins, les effets secondaires ou indésirables des médicaments. Les enjeux de l'exploitation de ces données résident dans la volumétrie et l'hétérogénéité des sources (diversités des nomenclatures).

Les standards de représentation structurée des données des Linked Open Data et du Web sémantique fournissent les solutions techniques permettant l'intégration de données distribuées, leur interrogation et leur interprétation. Ces technologies permettent également le raisonnement grâce à des connaissances, notamment sous la forme d'ontologies, de classifications ou de thésaurus médicaux. Des connaissances essentielles en pharmacologie, les interactions, indications et contre-indications médicamenteuses, ont ainsi pu être prises en compte lors d'analyses de données cliniques[2]. Cependant, si des travaux comme la *Drug Indication Database* (DID) et la *Drug Interaction Knowledge Base* (DIKB) ont regroupé différentes sources de connaissances médicales du Web des données[3, 4], l'utilisation et le chaînage de celles-ci restent encore peu aisés compte tenu de la multitude de classifications médicales et de sources de connaissances.

Nous proposons un package R, queryMed¹, pour fournir aux pharmaco-épidémiologistes des outils permettant d'exploiter les différentes sources de représentations de connaissances médicales issues du Web des données et de les lier aux principales classifications médicales, pour l'enrichissement de l'analyse des données cliniques et médico-administratives.

Pour interroger le Web des données, queryMed propose des fonctions de requêtage des principaux SPARQL *endpoints* spécifiques à la médecine et à la pharmacologie[5]. Ces serveurs sur lesquels des ontologies médicales sont publiées aux standards du Web Sémantique sont ainsi interrogés depuis R, grâce à SPARQL, un langage de requête dédié à de telles données. D'autres services, les REST API, rendent également possible l'extraction d'informations issues d'ontologies médicales. Grâce à l'intégration des bases de données DID et DIKB[3, 4], le package se concentre particulièrement sur les indications des médicaments,

¹ <https://github.com/yannrivault/queryMed>

leurs interactions et contre-indications avec d'autres médicaments. L'interrogation de SPARQL *endpoints* et de REST API ajoute à cela des connaissances sur les contre-indications entre médicaments et diagnostics. D'autres requêtes permettent de chaîner toutes ces connaissances à différentes nomenclatures et classifications médicales internationales et d'extraire les structures hiérarchiques qui leur sont propres.

Les applications de ce package sont notamment nombreuses en pharmaco-épidémiologie, allant de l'enrichissement des études, à la formulation d'hypothèses pharmacologiques. Par exemple, le clopidogrel, agent antiagrégant plaquettaire, possède 26 interactions avec d'autres médicaments selon DIKB, accessibles via queryMed (Fig1A). Lorsque l'on considère des classes plus générales de la nomenclature de l'ATC (Fig1B, Fig1C), les interactions concernent respectivement 15 puis 11 classes. Certaines classes se caractérisent par un nombre élevé d'interactions et de sources les citant : les anti-thrombotiques et gastro-oesophagiens. On pourrait alors formuler des hypothèses sur l'intérêt d'élargir l'étude des interactions entre le clopidogrel et d'autres médicaments à celles entre le clopidogrel et ces familles plus générales de médicaments.

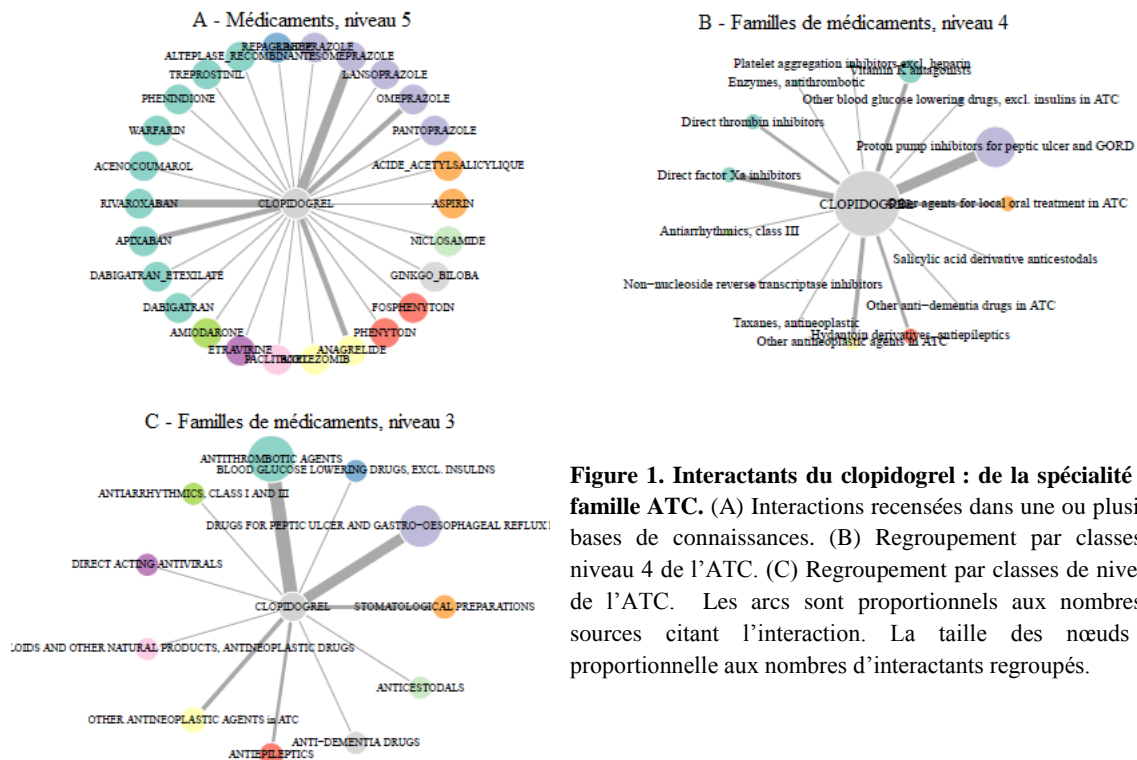


Figure 1. Interactants du clopidogrel : de la spécialité à la famille ATC. (A) Interactions recensées dans une ou plusieurs bases de connaissances. (B) Regroupement par classes de niveau 4 de l'ATC. (C) Regroupement par classes de niveau 3 de l'ATC. Les arcs sont proportionnels aux nombres de sources citant l'interaction. La taille des nœuds est proportionnelle aux nombres d'interactants regroupés.

Références

- [1] Weill, A et al (2017). Utilisation des données de l'Assurance maladie française pour étudier la qualité des pratiques et la sécurité des soins dans le domaine du médicament : revue de 204 articles publiés entre 2007 et 2016. *Revue d'Épidémiologie et de Santé Publique*. 65, S22.
- [2] Pathak, J et al (2013). Using Linked Data for Mining Drug-Drug Interactions in Electronic Health Records. *Stud Health Technol Inform*. 192, 682–686.
- [3] Ayvaz, S. et al (2015). Toward a complete dataset of drug-drug interaction information from publicly available sources. *J Biomed Inform*. 55, 206–217.
- [4] Sharp, M.E. (2017). Toward a comprehensive drug ontology: extraction of drug-indication relations from diverse information sources. *Journal of Biomedical Semantics*. 8, 2.
- [5] Salvadores, M. et al (2012). Using SPARQL to Query BioPortal Ontologies and Metadata. In: *The Semantic Web – ISWC 2012*. pp. 180–195. Springer, Berlin, Heidelberg.