



HAL
open science

La gestion de données médico-administratives grâce aux outils du Web sémantique

Yann Rivault, Olivier Dameron, Nolwenn Le Meur

► To cite this version:

Yann Rivault, Olivier Dameron, Nolwenn Le Meur. La gestion de données médico-administratives grâce aux outils du Web sémantique. Journées ADELFF-EMOIS “ Système d’information hospitalier et Epidémiologie ”, Mar 2016, Dijon, France. pp.S15, 10.1016/j.respe.2016.01.051 . hal-01297389

HAL Id: hal-01297389

<https://hal.science/hal-01297389v1>

Submitted on 4 Apr 2016

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.

La gestion de données médico-administratives grâce aux outils du Web Sémantique

Yann Rivault¹, Olivier Dameron², Nolwenn Le Meur^{1,3}

¹ Département METIS, EHESP, Sorbonne Paris Cite, France.
{yann.rivault, nolwenn.lemeur}@ehesp.fr

² Université Rennes 1, IRISA UMR6074, Rennes.
olivier.dameron@univ-rennes1.fr

³ EHESP, EA 7348 MOS Management des organisations en santé, France.

Mots-clés : Bases médico-administratives, Ontologie, Web Sémantique.

Introduction

Le traitement statistique de données provenant des bases médico-administratives, comme le *Programme Médicalisé des Systèmes d'information* (PMSI) et l'*Echantillon Généraliste des Bénéficiaires* de l'Assurance Maladie (ÉGB), est en plein essor. Cependant le volume, l'hétérogénéité ou la complexité des données issues de ces systèmes d'information remettent en question l'usage exclusif d'outils statistiques classiques.

Cette étude pilote vise à élaborer et évaluer l'efficacité d'une méthode basée sur les outils du Web Sémantique pour améliorer l'intégration, la gestion et l'exploration de ces données, relevant du « big data ».

Méthodes

Pour illustrer notre approche, l'étude s'est intéressée à la détection des complications post-chirurgie.

Des tables de données issues du PMSI et de l'ÉGB, sur la période 2012-2013, ont été transformées au format *Ressource Description Framework* (RDF). Le langage de requête *SPARQL Protocol And RDF Query Language* a ensuite permis de sélectionner les sous-populations d'étude. Les performances de ces requêtes ont été comparées à celles d'algorithmes équivalents basés sur le format initialement tabulaire des données, grâce au logiciel R et au package *rrdf*.

Les données au format RDF ont ensuite été enrichies d'ontologies relatives aux domaines de connaissances des médicaments et des actes chirurgicaux. Des requêtes bénéficiant de l'intégration de ces connaissances ont alors été exécutées sur un serveur de requête SPARQL, FUSEKI.

Résultats

L'étude comparative a montré que l'exécution des requêtes SPARQL est plus rapide que celle des algorithmes équivalents sous R, moins d'une seconde contre cinq à dix secondes. L'intégration d'ontologies a rendu possible la recherche de populations statistiques très spécifiques, basée sur différents critères, relevant par exemple de prescriptions médicamenteuses ou de diagnostics, selon des degrés de granularité plus ou moins fins.

Conclusion

Cette étude a permis le développement d'une nouvelle méthode de gestion des données massives issues des bases médico-administratives. Elle a également fourni des premiers éléments sur son efficacité et son intérêt.