Une approche ontologique pour l'exploitation de données cliniques

Ariane Assele Kama, Giovanni Mels, Rémy Choquet, Jean Charlet, Marie-Christine Jaulent

Une approche ontologique pour l'exploitation de données cliniques

Ariane Assélé Kama[1], Giovanni Mels[2], Rémy Choquet[1], Jean Charlet[1] et Marie-Christine Jaulent[1]

[1] INSERM UMRS_872 EQ.20, Université Pierre et Marie Curie, 75006 Paris
{ariane.assele, remy.choquet, jean.charlet, marie-christine.jaulent}@crc.jussieu.fr
[2]AGFA HealthCare, Kortrijksesteenweg 157,
9830 Sint-Martens-Latem, Belgium
http://www.agfa.com/france/fr/he/index.jsp

Résumé : L'utilisation des outils et techniques du web sémantique pour intégrer, modéliser et interroger des bases de données relationnelles est un moyen de prendre en compte la dimension sémantique des données et d'en faciliter l'exploitation. Les données issues des systèmes d'information hospitaliers (SIH) peuvent êtres inutilisables dans leur format d'origine en raison de l'absence d'informations ou de connaissances. Elles nécessitent donc d'être réorganisées et liées à des ressources externes de connaissances telles que des ontologies ou terminologies. Dans le cadre de cette étude, nous utilisons des données opérationnelles du SIH de l'hôpital européen Georges Pompidou, une ontologie comme source de connaissances et des outils du web sémantique pour interroger notre base de données opérationnelle et répondre ainsi à des questions médicales dans le domaine de la résistance des bactéries aux antibiotiques. Le champ d'application de contrôle des infections des parties prenantes de cette étude est délimité par le biais de plusieurs requêtes, par exemple « la liste des patients ayant une infection spécifique (ex : infection urinaire) par un agent pathogène donné (ex : E. coli) résistante à un antibiotique donné (ex : amoxicilline) ».

Mots-clés : Sparql Endpoint, Entrepôts de données, RDF, Ontologie

1 Introduction

Les systèmes d'information hospitaliers (SIH) doivent avoir la capacité de centraliser et restituer des informations, et d'en tirer des connaissances médicales. Ils sont ensuite utilisés pour évaluer et améliorer la qualité et la sécurité des soins, et favoriser la recherche et l'épidémiologie. Les données issues des systèmes hospitaliers proviennent de sources hétérogènes et variées et peuvent être stockées dans des « entrepôts de données sémantiques » pour faciliter leur analyse (Xie et al. 2007, Skoutas et al. 2006, Sell et al. 2005), mais aussi leur exploitation (Franco et al. 1997). Dans cette phase d'intégration de données, la problématique majeure est liée aux problèmes d'hétérogénéité sémantique et structurelle des données cliniques. Les sources de données conçues indépendamment les unes des autres définissent souvent

les mêmes concepts de différentes manières selon les usages de chaque source. En médecine, l'utilisation d'une « sémantique partagée » via les ontologies, terminologies ou thésaurus facilite l'intégration de sources multiples, lorsque celles-ci sont associées à un même concept. Le web sémantique a récemment normalisé un ensemble de méthodes et d'outils permettant d'intégrer des données en les associant à des concepts d'une ontologie, suivant différentes approches : les bases de données à base ontologique (BDBO, Pierra et al. 2005) ou les SPARQL Endpoint (Prud'hommeaux et al., 2005).

Utiliser une ressource sémantique telle qu'une ontologie pourrait être un moyen d'enrichir les données cliniques en vue de répondre plus précisément à des questions d'ordre médical complexes. Couplée avec une base de données clinique, une ontologie est utilisée comme supplément d'information et devient alors un élément indispensable à l'exploitation des données dans la recherche médicale. Le travail présenté a été réalisé dans le cadre du projet européen DebugIT (Detecting and Eliminating Bacteria UsinG Information Technology). Ce projet vise à regrouper les données cliniques, stockées dans différents hôpitaux, dans un système unifié dédié à la lutte contre les maladies infectieuses et des résistances aux antimicrobiens et en particulier l'analyse de la relation potentielle entre l'émergence de résistance bactérienne et les antibiotiques prescrits (Lovis C et al. 2008). Nous proposons ici une approche visant à utiliser et exploiter les connaissances médicales (ontologie), pour répondre à des questions cliniques à partir d'un entrepôt de données. Le but de ce travail est de valider l'utilisation de méthodes et outils du web sémantique pour l'exploitation des données cliniques.

2 Les entrepôts sémantiques

Lier une ressource sémantique à une base de données est aujourd'hui abordé dans la littérature suivant trois approches : l'entrepôt RDF (les données sont stockées sous forme de triplets RDF : Resource Description Framework, « *sujet – prédicat – objet : Escherichia coli est une entérobactérie* »), les bases de données à base ontologique (BDBO, données et ontologies sont stockées au même endroit), et le SPARQL endpoint (alignement du modèle d'information à l'ontologie). Nous présenterons ici les BDBO et le SPARQL Endpoint.

2.1 Les bases de données à base ontologique

Une base de données à base ontologique est une source de données dans laquelle ontologie et données sont toutes deux stockées dans la base de données et font l'objet de mêmes traitements (insertion, mise à jour, requêtes, etc.). La cohabitation se fait de telle sorte qu'à chacune des données présentes dans la base, on associe le concept de l'ontologie qui la définit. Le modèle d'architecture d'une BDBO, telle qu'il est défini dans l'architecture OntoDB (Pierra et al. 2005), est constitué des deux parties traditionnelles des bases de données (données et méta-base), d'une partie ontologie, et

d'une partie méta-schéma contenant le méta-modèle du modèle d'ontologie utilisé, permettant de rendre générique tout traitement sur les ontologies. L'architecture OntoDB est constituée d'un Système de Gestion de base de données PostgreSQL et du modèle d'ontologie PLIB. EXPRESS est le langage utilisé pour l'interrogation de la base de données.

2.2 Le Sparql Endpoint

Cette approche consiste à considérer une base de données relationnelle comme un graphe « virtuel » RDF (Bizer et Seaborne, 2004), à travers lequel on pourra accéder au contenu de la base, en utilisant des API telles que Jena et Sesame[1]. L'entrepôt de données sémantique est dit « virtuel », puisqu'il fournit une « vue sémantique » de la base de données. Le langage de requête SPARQL (Simple Protocol and RDF Query language) est utilisé pour consulter la source de données (De la Borda & Conrad, 2006). Des projets tels que Bio2RDF (Atlas du web sémantique de la post-génomique des connaissances sur l'homme et la souris), DailyMed[2] (Fournit des informations sur les médicaments prescrits sous ordonnance, approuvée par la FDA) utilisent respectivement des outils comme OpenLinkVirtuoso[3] et D2RQ[4], pour mettre en œuvre des « SPARQL Endpoint ».

3 Les outils et standards du web sémantique

Le Web sémantique désigne un ensemble de technologies visant à rendre le contenu des ressources du World Wide Web (W3C) accessible et utilisable par les programmes et agents logiciels, grâce à un système de métadonnées formelles, utilisant notamment des langages, technologies, outils et standards développés par le W3C. Nous présentons ici ceux que nous avons utilisés dans le cadre de cette étude.

Le langage SPARQL

SPARQL[5] est à la fois un langage et un protocole de requête. Le protocole va permettre à un client Web de consulter, en exécutant une requête SPARQL, un service ou point d'accès SPARQL (endpoint) qui traitera la requête pour retourner la réponse sous différents format (HTML, XML, RDF/XML, N3, JSON etc.). Le langage permet d'interroger des descriptions RDF en utilisant des clauses (similaires dans certains cas à celles du langage SQL) telles que **PREFIX** (spécifie l'adresse exploitée dans la construction de la requête), **SELECT** … [**FROM**] … **WHERE** (requête interrogative), **CONSTRUCT** (requête constructive), **UNION**, **OPTIONNAL** (jointures, conditions optionnelles), **FILTER** (conditions obligatoires) et

[1] http://jena.sourceforge.net/index.html - http://www.openrdf.org/doc/sesame/users/ch07.html
[2] http://bio2rdf.org/ - http://dailymed.nlm.nih.gov/dailymed/about.cfm
[3] http://esw.w3.org/topic/VirtuosoUniversalServer
[4] http://www4.wiwiss.fu-berlin.de/bizer/d2rq/spec/#jena-assembler
[5] http://www.w3.org/TR/rdf-sparql-query/

DESCRIBE, **ASK** (description d'une ressource, évaluation d'une requête). Le projet caBIG (cancer du Biomedical Informatics Grid) utilise le langage SPARQL pour formuler des requêtes, au cours de la représentation sémantique des services de données liés au cancer (Shironoshita et al. 2008).

La Notation 3 (N3)
Tout comme les langages OWL et RDF, la Notation 3[6] est un langage de description des connaissances. Développé par Tim Berners-Lee et des membres de la communauté du web sémantique, la Notation 3 est un langage plus compact et plus lisible que la notation RDF/XML[7]. Un peu comme en XML, des abréviations d'URI (Uniform Ressource Identifier) utilisant les préfixes (@prefix) liés à un espace de noms (namespace en anglais) sont utilisées.

D2R Server
D2R Server[8]est un outil de publication de bases de données relationnelles sur le web sémantique, développé par Chris Rizer de l'université de Berlin. Il est accessible depuis un navigateur RDF ou HTML et permet de créer un graphe virtuel RDF de la base de données relationnelle. Le langage de mapping D2RQ permet d'aligner le schéma de la base de données à une ou plusieurs ontologies (Barrasa et al. 2003). Ces correspondances, définies dans un "fichier de mapping" (D2RQ mapping file) sont exprimées en Notation 3 (N3).

Joseki
Joseki[9] est un moteur http qui soutient le protocole et langage SPARQL, pour l'édition des graphes RDF sur le Web. A chacun des graphes RDF est associé une URL, et ces derniers peuvent être accessibles par la méthode GET du protocole HTTP. L'intérêt de cet outil est de pouvoir accéder, au même titre que D2R Server, à des données RDF depuis des fichiers et des bases de données. L'utilisateur dispose d'une interface d'interrogation, similaire à SNORQL (D2R server), pour éditer et exécuter des requêtes SPARQL. Les résultats obtenus sont disponibles sous divers formats.

4 Matériel

Dans le cadre de cette étude nous utilisons deux ressources : une base de données clinique multidimensionnelle construite à partir des données issues de l' **HEGP** (Hôpital Européen Georges Pompidou) et une ontologie construite dans le cadre du projet européen **DebugIT** (Detecting and Eliminating Bacteria UsinG Information Technology). Le projet européen DebugIT vise à optimiser la qualité des soins et la sécurité des patients, en proposant des solutions de lutte contre la résistance des bactéries aux antibiotiques.

---

[6] http://www.w3.org/DesignIssues/Notation3
[7] http://www.w3.org/TR/REC-rdf-syntax/
[8] http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/
[9] http://www.joseki.org/ Joseki – A SPARQL Server for Jena

4.1 La base de données clinique

La base de données source mise à notre disposition (13 tables 357,5 Mo), nous permet d'obtenir des informations relatives aux prescriptions médicamenteuses, aux résultats d'examens biologiques, aux données du patient (données anonymisées) et aux médicaments prescrits. Nous nous sommes intéressés uniquement aux résultats d'antibiogrammes (bactéries identifiées, antibiotiques testés) pour créer notre base de données, en utilisant des procédures ETL (*Extract – Transform – Load*) à l'aide de l'outil Talend (7 tables – 20 Mo). Les tables **BACTERIA** (ID_BACTERIA, NAME) **ANTIBIOTIC** (ID_ANTIBIOTIC, ANTIBIOTIC), contiennent respectivement la liste distincte des bactéries identifiées et antibiotiques testés dans la base source. La table **BILAN** (ID_BILAN, ID_PATIENT) contient l'identifiant des bilans d'examens biologiques d'un **PATIENT** (ID_PATIENT, SEX, DATE_NAISS). Les résultats d'antibiogrammes sont stockés dans les tables **ANTIBIOGRAM** (ID_ANTIBIOGRAM), **TEST** (ID_TEST, ID_ANTIBIOTIC) et **EVENTRESULT** (ID_BILAN, ID_ANTIBIOGRAM, ID_TEST, RESULT, VALEUR). Enfin, la table **EVENTBACTERIA** (ID_ANTIBIOGRAM, ID_BACTERIA, NB_TEST, FREQUENCY) permet d'obtenir la fréquence d'apparition d'une bactérie et le nombre de tests effectués sur celle-ci.

4.2 La ressource ontologique

Dans le cadre du projet, une ontologie a été conçue, DCO (DebugIT Core Ontology), pour décrire le domaine des maladies infectieuses (agents pathogènes, prescriptions médicamenteuses, diagnostics …) et permettre l'intégration sémantique de sources de données hétérogènes. Des correspondances à des terminologies externes telles que SNOMED CT, BioTop et ATC[10] sont effectuées en utilisant des propriétés annotées. Cette ressource ontologique est constamment mise à jour et contient actuellement 894 classes.

5 Méthode

L'exploitation classique des bases de données cliniques ne permet pas de répondre à des questions médicales nécessitant la connaissance d'un domaine. Les technologies du Web Sémantique doivent donc permettre à des utilisateurs d'obtenir des résultats incluant cette connaissance en utilisant des ontologies comme support d'interrogation, mais surtout comme source de données. La **figure 1** présente l'architecture générale du système. L'utilisateur exécute une requête SPARQL depuis une interface dédiée (SNORQL ou Joseki), accessible via un navigateur HTML ou RDF. En se basant sur les correspondances définies dans le fichier de mapping et une ressource ontologique, le processeur va exécuter la requête pour extraire les informations de la base de données clinique, renvoyées sous différents formats.

[10] http://www.ihtsdo.org/snomed-ct/ - http://purl.org/biotop/1.0/biotop.owl - http://www.whocc.no/atcddd/

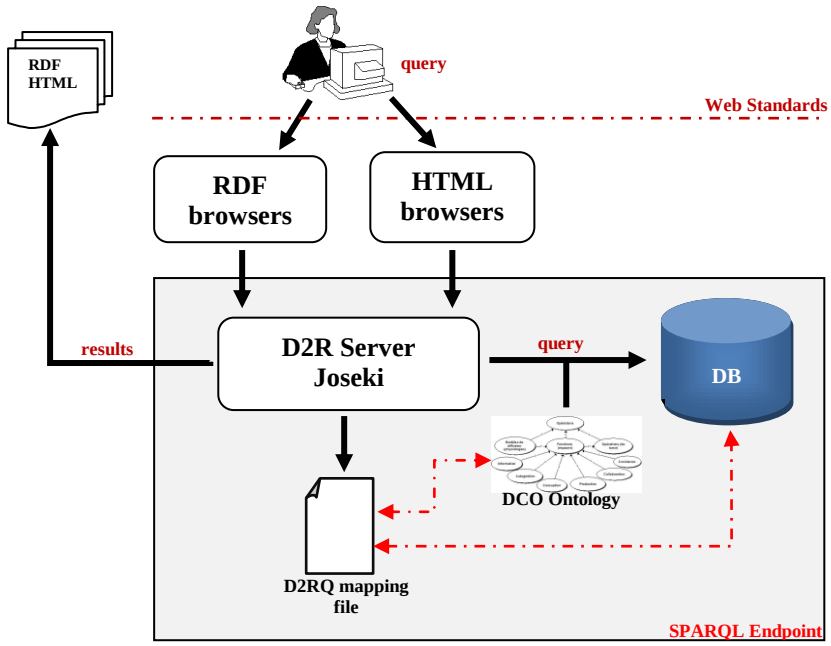

Fig.1- Architecture générale du système

La **figure 2** illustre la requête suivante : « *Résultats d'antibiothérapie d'une infection d'E. coli, résistante aux fluoroquinolones ?* ». La table « Antibiogram » comporte 4 colonnes, l'identifiant de l'antibiogramme (ID), le nom de la bactérie, le nom de l'antibiotique testé et une mesure de la résistance (S pour « sensible » et R pour « résistant »).

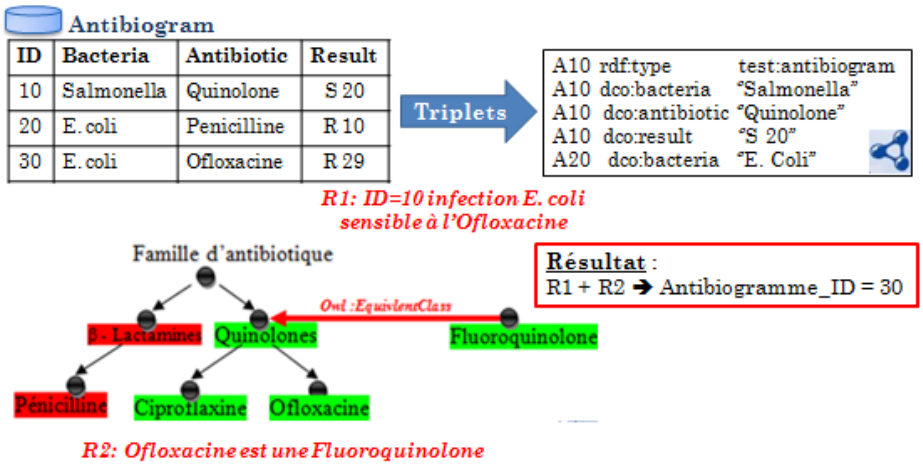

Fig.2 - Illustration d'un processus d'association de données cliniques à une connaissance ontologique. La base de données (R1) nous donne le résultat du test de sensibilité de la bactérie à un antibiotique, et la ressource ontologique (R2), les caractéristiques de cet antibiotique (ex. famille).

5.1 Utilisation de D2R Server

D2R Server nous permet de créer le graphe virtuel RDF de la base de données, généré automatiquement à partir d'un processeur. Ce fichier respecte la notation 3 et les propriétés du langage D2RQ pour mettre en correspondance les différentes tables/colonnes et les concepts/instances de nos ressources. La **figure 3** présente un exemple d'alignement entre la table « antibiotic » (map :antibiotic) et le concept « Antibiotic » (dco :Antibiotic). Les données de cette table auront un URI (défini à l'aide de la propriété « d2rq:uriPattern »), et seront accessibles via le SPARQL Endpoint. Pour effectuer nos requêtes sur les deux ressources, il est nécessaire de créer un deuxième type d'alignement entre les données de la base telles que définies dans le premier fichier de mapping (ex. URI du nom d'un antibiotique) et le concept/instance de l'ontologie correspondant, en notation 3. (Cf. figure 4)

```
# Table Antibiotic
map:antibiotic  a  d2rq:ClassMap;
d2rq: dataStorage   map:dasic;
d2rq:class    dco:Antibiotic;
d2rq:uriPattern "antibiotic/@@antibiotic.ID_ANTIBIOTIC@@";
.
```

Fig. 3 –Correspondance entre la table « antibiotic » et le concept « Antibiotic » effectuée dans le fichier « inserm-d2r.n3 »

```
@prefix rdfs:          <http://www.w3.org/2000/01/rdf-schema#>.
@prefix biotop:        <http://purl.org/biotop/1.0/biotop.owl#>.
@prefix dco:           <http://www.debugit.eu/ontology/1.0/dco.owl#>.

<http://debugit1.spim.jussieu.fr/resource/bacteria/60>     a biotop:SpeciesEscherichiaColiRegion.
<http://debugit1.spim.jussieu.fr/resource/antibiotic/4>    a dco:Penicillin.
<http://debugit1.spim.jussieu.fr/resource/antibiotic/21>   a dco:Ciprofloxacin.
<http://debugit1.spim.jussieu.fr/resource/antibiotic/2>    a dco:Quinolone.
```

Fig. 4 –Alignement des données dans le fichier « inserm-map.n3 »

5.2 Utilisation de joseki

En nous basant sur les fichiers de mapping précédemment définis, nous allons interroger la ressource ontologique et le graphe RDF de la base clinique. Pour cela, il est nécessaire de configurer le serveur joseki, en spécifiant les différents fichiers de mapping, l'URI de la base de données et de l'ontologie utilisée. En utilisant les propriétés de Jena[11], les différents graphes vont être assemblés pour permettre ainsi d'exploiter les relations directes de subsumption existants entre les concepts et les

[11] Jena est un framework Java, permettant de construire des applications du Web Sémantique. Un environnement de programmation RDF(S), OWL, SPARQL est disponible. http://jena.sourceforge.net/

instances de l'ontologie, et rattacher par exemple un antibiotique prescrit dans la base de données à sa famille définie dans l'ontologie.

6 Résultats

La base de données contient **238.623** résultats d'antibiogrammes effectués sur 61 (12 pénicillines, 13 céphalosporines soit 21 béta-lactamines) antibiotiques et 165 bactéries distinctes. Pour obtenir les résultats de nos requêtes, nous avons dans un premier temps utilisé l'interface dédiée SNORQL de D2R Server pour la consultation et l'interrogation la base de données. L'outil **D2R Server** dans sa version actuelle, ne nous permet pas de spécifier un graphe autre que celui de la base de données. Il nous est donc impossible, depuis cette interface, d'exploiter notre ressource ontologique. La requête SPARQL de la **figure 5** permet d'obtenir *« les résultats d'antibiogrammes des tests de sensibilité de l'E. Coli à l'Amoxicilline (de la famille des pénicillines)* ». Les résultats obtenus sont accessibles sous différents formats (HTML, XML, XML+XSLT, JSON[12]). La base de données contient **1 971** résultats d'antibiogrammes (E. Coli – Amoxicilline), dont **1 081** sont résistants.

Fig. 5 –Requête SPARQL pour obtenir la liste des résultats d'antibiogrammes (ici limitée à 10) d'E. Coli sur l'Amoxicilline.

[12]JavaScript Objet Notation - http://www.json.org/

En utilisant l'interface **joseki**, nous interrogeons les deux ressources pour obtenir « *les résultats d'antibiogrammes d'E. Coli, résistantes aux Béta-lactamines* ». (Cf. **figure 6**). On obtient **4 001** résultats de résistance de l'E. Coli aux Béta-lactamines. (**25 673** sensibles, **2 716** Intermédiaires). Le fichier de mapping nous permet de récupérer tous les URI des « pénicillines et céphalosporines » et celui de l'E. coli, pour récupérer les informations dans la base de données. Les résultats sont disponibles sous différents formats, mais aussi sous forme de graphe RDF, en utilisant la clause **CONTRUCT**.

```
PREFIX biotop: <http://purl.org/biotop/1.0/biotop.owl#>
PREFIX dco:    <http://www.debugit.eu/ontology/1.0/dco.owl#>
PREFIX inserm: <http://debugit1.spim.jussieu.fr/resource/vocab/>

SELECT DISTINCT *  WHERE {
GRAPH <http://debugit.eu/inserm-map.n3> {
            ?antibiotic1 a dco:BetalactamAntibiotic.
            ?bacteria     a biotop:SpeciesEscherichiaColiRegion.
            ?r1             a dco:Resistant. }

GRAPH <http://debugit1.spim.jussieu.fr/resource> {
            ?susceptibility1 a inserm:ResultAntibiogram;
                                    inserm:antibiogram_ID ?antibiogram;
                                    inserm:antibiotic_tested ?antibiotic1;
                                    inserm:antibiotic_RESULT ?r1.

            ?antibiogram     a inserm:antibiogram;
                                    inserm:bacteria_analyzed ?bacteria.  }
}
```

Fig. 7 –Requête SPARQL pour obtenir les résultats d'antibiogrammes d'E. Coli, résistants aux bêta-lactamines. On se base sur les concepts de l'ontologie pour regrouper les données par famille

7 Discussion

Les perspectives envisageables consistent à exploiter de façon plus complète la connaissance disponible via l'ontologie de manière à construire des requêtes SPARQL plus spécifiques (ex. *« Quelle est l'entérobactérie la plus souvent résistante aux fluoroquinolones, en cas d'infection urinaire »*). On pourrait par exemple utiliser des « règles » en logique N3, pour bénéficier de formules, d'implications logiques etc. et enrichir l'écriture de nos requêtes SPARQL. Le SPARQL Endpoint mis en œuvre dans le cadre de ce projet nous a permis d'expérimenter l'utilisation d'une ontologie, qui fournit un vocabulaire structuré et une connaissance formalisée, comme « support » à l'expression des requêtes mais aussi comme « source de connaissance ».

Une limite de ce travail est liée d'une part à l'aspect technique des outils utilisés dans leur version actuelle. Avec l'outil D2R Server, nous ne pouvons pas regrouper les données par type dans une table (ex : la liste distincte des bactéries), en utilisant la clause GROUP BY pour effectuer un alignement cardinalité « 1 – n » (correspondance entre une table et un concept). Compte tenu de cette contrainte, nous avons choisi une architecture multidimensionnelle pour faciliter le mapping entre les différentes ressources. On pourrait aussi envisager de créer des vues matérialisées des tables de la base de données relationnelle et les utiliser dans le fichier de mapping.

D'autre part, la mise en œuvre de cette approche nécessite de lier les instances de la base de données à celles de l'ontologie. Dans cette expérimentation, cette annotation a été réalisée « manuellement ». Cette approche reste assez fragile, fastidieuse, et peu juste car sujette à des erreurs de saisie. Nous avons listé les URI des instances (antibiotiques, bactéries) à lier aux concepts correspondant en notation 3, dans un fichier de mapping. On pourrait envisager de mettre en œuvre des outils « semi-automatiques » pour effectuer la mise en correspondance entre les données, ou encore d'utiliser des outils de transformation de modèles pour créer une ontologie à partir du schéma de la base de données relationnelle (Krivine S. et al. 2009). Cette ontologie sera ensuite aligner à l'ontologie de domaine en utilisant des techniques d'alignement d'ontologies (Mazuel et al. 2009).

Analyser des données cliniques en prenant en compte la connaissance d'un domaine permet d'effectuer une analyse complète et riche du problème étudié. Les résultats obtenus au cours de ce travail ne pourront être considérés comme « bons » qu'en supposant que les données d'origine soient riches et de qualité (Choquet R. et al. 2010).

8 Conclusion

Les approches actuelles préconisent l'utilisation des technologies et outils du web sémantique pour concevoir des entrepôts sémantiques et faciliter l'intégration sémantique des données, à travers leurs métadonnées. Nous avons constaté au début de ce projet que les bases de données cliniques ne peuvent pas être directement exploitables, en l'état actuel des outils disponibles, et nécessitent donc d'être réadaptées, voire réorganisées, en adoptant une modélisation multidimensionnelle par exemple. La base de connaissance nous a permis d'obtenir des informations de classification des antibiotiques que nous avons associées aux résultats d'antibiogrammes issus de notre base de données cliniques. La mise en œuvre de ces approches permettra, dans le cadre du projet européen DebugIT, d'interroger des données hétérogènes pour en extraire une connaissance susceptible d'alimenter des systèmes experts d'aide à la prescription d'antibiotiques.

Références

BARRASA J., CORCHO O. AND GOMEZ-PEREZ (2003), *Fund Finder Wrapper: A Case Study of Database-to-ontology Mapping*, in International Workshop on Semantic Integration, pp9-15.

BIZER C., SEABORNE A. (2004). D2RQ – Treating Non-RDF Database as virtual RDF Graphs. In *the 3rd International Semantic Web Conference* (ISWC 2004).

CHOQUET R, QOUIYD S, OUAGNE D, PASCHE E, DANIEL C, BOUSSAID O, JAULENT M-C (2010). The Information Quality Triangle: a methodology to assess Clinical Information quality. In *Medical and Health Informatics*.

FRANCO J-M (1997). Le Data Warehouse (Le Data Mining). Editions Eyrolles, Paris.

KRIVINE S, NOBÉCOURT J, SOUALMIA LF, CERBAH F, DUCLOS C. (2009). Construction automatique d'ontologies à partir d'une base de données relationnelles : application au médicament dans le domaine de la pharmacovigilance.

LOVIS C, COLAERT D, STROETMANN VN. (2008) DebugIT for patient safety - improving the treatment with antibiotics through multimedia data mining of heterogeneous clinical data. *Stud Health Technol Inform*. 136, 641-6.

MAZUEL L, CHARLET JEAN (2009). Alignement entre des ontologies de domaine et la snomed: trois études de cas. In *20ème conférence sur l'Ingénierie des Connaissances - IC2009*, Hammamett : Tunisia (2009).

MINVIELLE E, DE POURVOURILLE G. (2001). La mesure de la qualité des soins : un enjeu et un défi de santé publique. *Revue Epidémiologie Sante Publique*. 49 : 113-5.

PIERRA G., DEHAINSALA H., AIT AMEUR Y., BELLATRECHE L., CHONON J., EL-HADJ MIMOUNE M. (2005). Base de données à base ontologique : principe et mise en œuvre. *Ingénierie des systèmes d'information*, Hermès 10(2): 91–116.

PRUD'HOMMEAUX E. AND SEABORNE A. (eds.) (2005): SPARQL Query Language for RDF. *W3C Working Draft, November*, available at http://www.w3.org/TR/rdf-sparql-query/

SELL D., CABRAL L., MOTTA E., ET AL. (2005) A Semantic Web based Architecture for Analytical Tools, In *Proceeding of the Seventh IEEE International Conference on E-Commerce Technology* (CEC'05) pp. 347-354.

SHIRONOSHITA E. YVES R. JEAN-MARIE, M. RAY BRADLEY, MANSUR R. KABUKA (2008). semCDI: A Query Formulation for Semantic Data Integration in caBIG, *In Journal of the American Medical Informatics Association.* 2008: 15(4).

SKOUTAS D., SIMITSIS A. (2006) Designing ETL processes using semantic web technologies, In *Proceeding of the 9th ACM international workshop on Data-warehousing and OLAP* pp. 67-74.

XIE GT., SHENG PING LIU, YANG YANG, ZHAO MING QIU, YUE PAN, XIONG ZHI ZHOU. (2007) EIAW: Towards a Business-friendly Data Warehouse Using Semantic Web Technologies, ISWC 2007, Busan, Korea, November 13rd, 2007