

IDOSCHISTO : une extension de l'ontologie noyau des maladies infectieuses (IDO-Core) pour la schistosomiase

Gaoussou Camara, Sylvie Despres, Moussa Lo

► **To cite this version:**

Gaoussou Camara, Sylvie Despres, Moussa Lo. IDOSCHISTO : une extension de l'ontologie noyau des maladies infectieuses (IDO-Core) pour la schistosomiase. Catherine Faron-Zucker. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. pp.39-50, 2014. <hal-01015104>

HAL Id: hal-01015104

<https://hal.inria.fr/hal-01015104>

Submitted on 25 Jun 2014

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.

IDOSCHISTO : une extension de l'ontologie noyau des maladies infectieuses (IDO-Core) pour la schistosomiase

Gaoussou Camara^{1,2,3}, Sylvie Despres¹, Moussa Lo²

¹INSERM, U1142, LIMICS, F-75006, Paris, France.
Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), F-93430, Villetaneuse, France.
sylvie.despres@univ-paris13.fr

²LANI, Université Gaston Berger, B.P. 234 Saint-Louis, Sénégal.
moussa.lo@ugb.edu.sn

³Université Alioune Diop de Bambey, B.P. 30, Bambey, Sénégal.
gaoussou.camara@uadb.edu.sn

Résumé : Cet article décrit la construction de l'ontologie de la schistosomiase (IDOSCHISTO). Elle est conçue comme une extension de l'ontologie noyau des maladies infectieuses (IDO) afin d'assurer son interopérabilité avec les ontologies de domaine des maladies infectieuses existantes. IDOSCHISTO importe intégralement ou partiellement des ontologies du domaine médical conçues pour des besoins spécifiques propres aux maladies infectieuses. Il s'agit par exemple des ontologies concernant la transmission des agents pathogènes, le diagnostic, les études cliniques et biologiques, etc. IDOSCHISTO est une ontologie modulaire de la schistosomiase intégrant les points de vue épidémiologique, clinique et biologique de la maladie. Elle a pour finalité de devenir une référence dans le domaine des maladies infectieuses en facilitant la communication et l'interopérabilité entre les différents acteurs impliqués et en supportant le raisonnement dans des contextes d'usage variés tels que la veille sanitaire, l'aide au diagnostic et à la prescription médicamenteuse, la gestion du dossier patient, les annotations biologiques, etc.

Mots-clés : ontologie de domaine, ontologie noyau, ontologie modulaire, maladie infectieuse, schistosomiase.

1. Introduction

La dynamique des populations humaines et leurs interactions avec l'environnement favorisent la propagation rapide des maladies infectieuses. La maîtrise du mode de contamination de la maladie et de ses facteurs de risque est indispensable pour en prévenir l'émergence ou en contrôler l'évolution. Dans le cas de la schistosomiase, une maladie parasitaire dont le parasite vit dans les eaux douces, les interactions des populations avec les points d'eau augmentent le risque de propagation. Ainsi, il devient indispensable de mettre en place un système de veille épidémiologique pour la prévention et le contrôle de ces types de phénomènes. L'analyse des risques et la prise de décision, phases incontournables du processus de veille (Camara et al., 2012a), nécessitent l'implication de plusieurs acteurs aux profils différents et ayant chacun une vision partielle du phénomène étudié. Leur appartenance à des communautés différentes fait qu'ils n'utilisent pas les mêmes vocabulaires pour désigner les mêmes concepts du domaine. Par exemple, un médecin parlera de *patient* alors que l'épidémiologiste parlera d'*hôte infecté*. Pour la veille d'une maladie spécifique telle que la schistosomiase, il est par conséquent nécessaire de disposer d'une ontologie jouant le rôle de médiateur entre les différents acteurs mais aussi d'un modèle formel des connaissances du domaine pour automatiser les raisonnements dans un contexte de veille épidémiologique.

Des ontologies relatives à différentes maladies ont été construites pour améliorer les études biologiques, les traitements cliniques des patients, etc., comme par exemple dans (Lin et al., 2011). Néanmoins, ces ontologies ne sont pas conçues pour la veille et la prévention en santé publique d'une maladie infectieuse. En outre, il n'existe aucune ontologie décrivant la schistosomiase.

Dans cet article, nous proposons la construction d'une ontologie de domaine de la schistosomiase (IDOSCHISTO¹). IDOSCHISTO est conçue en tenant compte des différentes perspectives sur la maladie (biologique, clinique et épidémiologique) dont l'intégration est indispensable pour le contrôle et la prévention de sa propagation. L'intérêt visé par cette décomposition modulaire est également de faciliter son utilisation partielle dans ces sous-domaines respectifs. La méthodologie de construction est fondée sur la réutilisation d'ontologies noyau du domaine des maladies infectieuses et d'une ontologie de fondement. Des ontologies spécifiques de domaine existantes sont réutilisées au niveau de la couche spécifique de domaine suivant la recommandation d'OBO Foundry (Smith et al., 2007).

Dans la suite de cet article, nous présentons en section 2 le cadre méthodologique de construction de l'ontologie. La modularisation de l'ontologie est détaillée en section 3. La section 4 décrit les choix de réutilisation adoptés. La section 5 présente la formalisation et la métrique d'IDOSCHISTO. Un cas d'usage de cette ontologie pour l'annotation des données de la schistosomiase de la localité de Richard Toll au Sénégal est fourni dans la section 6. Les résultats de l'évaluation de l'ontologie sont présentés en section 7. Nous concluons et discutons les résultats en section 8.

2. Méthodologie de construction de l'ontologie de la schistosomiase

La construction de l'ontologie de domaine de la schistosomiase suit les différentes étapes définies dans la méthodologie NeON (Suárez-Figueroa et al., 2012). Nous avons recours à plusieurs des scénarios de cette méthodologie pour construire l'ontologie de la schistosomiase.

- 1 scénario 1 : de la spécification à l'implémentation
- 2 scénario 3 : la réutilisation de ressources ontologiques
- 3 scénario 5 : la réutilisation et la fusion de ressources ontologiques
- 4 scénario 7 : la réutilisation de patron de conception d'ontologie

Nous avons mis en œuvre ces scénarios en respectant les étapes définies dans le *scénario 1* : *spécification, acquisition, modélisation, formalisation, implémentation* et *évaluation*.

- 1 La *spécification* : la construction de l'ontologie est guidée par l'objectif de support à la veille de la propagation de la schistosomiase. La finalité d'IDOSCHISTO est l'annotation des ressources de la veille, le travail collaboratif entre les acteurs et les organisations impliqués dans la veille (communication, interopérabilité des données et des applications), etc. A cet objectif, s'ajoute le besoin d'assister les activités cliniques et biologiques.
- 2 L'*acquisition* : cette tâche est réalisée *via* des entretiens avec les experts² des domaines respectifs modélisés et l'exploitation de ressources non-ontologiques (rapports, articles, etc.) et ontologiques (principalement le portail « OBO Foundry » des ontologies du domaine biomédical).
- 3 La *modélisation* : une première approche a consisté à réaliser des modèles conceptuels en vue de la structuration des connaissances et serviront de support à l'évaluation menée avec les experts du domaine. Des diagrammes de classes UML sont utilisés

¹ <https://github.com/gaoussoucamara/idoschisto/blob/master/idoschisto.owl>

² Les experts impliqués dans la construction de l'ontologie sont principalement ceux du Programme National de Lutte contre la Schistosomiase au Sénégal, l'ONG Espoir Pour la Santé de Saint-Louis du Sénégal et une biologiste spécialisée dans la schistosomiase.

pour cette modélisation et des représentations tabulaires permettant une visualisation plus synthétique de certaines connaissances sont proposées.

- 4 La *formalisation et l'implémentation* : le niveau logique requis pour faire des raisonnements sur l'ontologie est la logique du premier ordre. Les objectifs de nos travaux et les besoins qui leurs sont associés nous orientent vers le choix du langage OWL-DL. Ce langage inclut toute la sémantique formelle des logiques de description (DL) et les capacités de raisonnement qui en découlent, dans un standard de représentation d'ontologie. Les services d'inférences offerts par ce langage, nous permettront de fournir pour le système de veille, des possibilités de raisonnement variées telles que la déduction de connaissances implicites à partir de connaissances représentées explicitement, la vérification de la cohérence de modèles, la classification d'instances, etc. L'utilisation de l'éditeur Protégé qui produit la représentation formelle et propose une sauvegarde du modèle dans un langage de représentation d'ontologie exploitable par la machine nous affranchit de l'étape d'implémentation.
- 5 L'*évaluation* : (1) La première évaluation est intervenue au cours de la conceptualisation en faisant valider le contenu des ontologies par les experts. Ce processus a été réalisé avec les experts en épidémiologie et de la schistosomiase. Au cours des entretiens, les modèles construits ont été présentés. La validité et la cohérence des connaissances représentées ont été discutées. Les omissions susceptibles d'améliorer la couverture du domaine en rapport avec les objectifs de l'ontologie ont été identifiées. (2) Au cours de la deuxième phase de l'évaluation, nous avons procédé à des tests de consistance en utilisant les raisonneurs inclus dans l'éditeur Protégé. Outre, ce test d'incohérence de classification, nous faisons une classification automatique de l'ontologie en intégrant les nouvelles classifications inférées. OOPS (Poveda-Villalón et al., 2012) a également été utilisé pour déceler des incohérences ou des redondances afin de les corriger (automatiquement ou manuellement). (3) La troisième étape de la validation est liée à l'implémentation des ontologies dans une application réelle afin de tester leur opérationnalité et mesurer leur apport au domaine d'étude considéré. Elle nécessite un temps de développement important dont nous ne disposons pas. Elle a, dans un premier temps, été remplacée par l'utilisation de l'interface d'interrogation de Protégé pour exécuter des requêtes SPARQL sur l'ontologie ou la base d'annotations.

3. Modularisation de l'ontologie du domaine de la schistosomiase

La première étape a consisté à construire l'ontologie IDOSCHISTO selon un modèle en couches d'abstraction permettant ainsi la réutilisation d'ontologies noyau et d'une ontologie de fondement. Ensuite, nous avons pris en compte les points de vue épidémiologique, clinique et biologique sur la maladie au niveau de la couche spécifique. Chacun de ces points de vue a été modélisé comme un module ontologique. Enfin, ces différentes perspectives ont été intégrées dans une ontologie unique en modélisant les *relations inter-perspectives*, c'est-à-dire les relations entre les concepts de perspectives différentes. Les relations entre les concepts de même perspective sont appelées ici les *relations intra-perspective* (Camara et al., 2013).

3.1. Modélisation en niveau d'abstraction de chaque module

Le cadre conceptuel pour construire l'ontologie de la schistosomiase est structuré en trois couches : l'ontologie de fondement (Falbo et al., 2010), l'ontologie noyau (Scherp et al., 2011) et les ontologies spécifiques. Les ontologies spécifiques consistent ici à modéliser les perspectives épidémiologique, clinique et biologique.

La couche noyau concerne le domaine des maladies infectieuses de manière générale. Une unique ontologie noyau des maladies infectieuses, appelée *Infectious Disease Ontology* (IDO-Core), existe dans le domaine. IDO-Core couvre les concepts généraux du domaine des maladies infectieuses (agent pathogène, gène, cellule, organe, organisme, population, hôte, vecteur, humain, etc.) et leurs relations. IDO-Core est reliée à l'ontologie de fondement BFO. Les entités modélisées dans IDO-Core relèvent de plusieurs perspectives : biologique (propriétés biologiques des agents pathogènes et leurs interactions avec l'organisme des individus infectés), clinique (symptôme, diagnostic, traitement, etc.) et épidémiologique (étude des mécanismes de propagation des maladies infectieuses dans la population d'individus et leurs moyens de lutte et de prévention). Ainsi, IDO-Core s'avère pertinente pour constituer le noyau des trois modules correspondants aux différentes perspectives définies. En outre, sa réutilisation permet de garantir une interopérabilité et une réutilisabilité dans le domaine des maladies infectieuses. Néanmoins, IDO-Core ne prend pas bien en compte la perspective épidémiologique traitant du mode de propagation de la maladie et les stratégies de contrôle et de prévention. Même si plusieurs concepts liés au sous-domaine épidémiologique sont présents dans cette ontologie noyau, ils ne sont pas mis en relation pour refléter le mécanisme de la propagation des maladies infectieuses. C'est ainsi qu'IDO-Core est utilisée conjointement avec l'ontologie noyau de la propagation des maladies infectieuses (IDSDO-Core³) que nous avons proposée et présentée dans (Camara et al., 2014, 2012b). Ainsi, elles constituent toutes les deux un noyau complet pour les modules spécifiques de la schistosomiase.

Le choix de l'ontologie de fondement à réutiliser est fondé sur trois critères : (i) la cohérence de la catégorisation des concepts de processus, événement, état et objet vis-à-vis de leurs sémantiques dans le domaine des maladies infectieuses, (ii) la consistance de la réutilisation des relations entre ces concepts pour couvrir les relations spécifiques à notre domaine et (iii) la réutilisation d'une ontologie de fondement par les ontologies noyau du domaine des maladies infectieuses, c'est-à-dire IDO-Core et IDSDO-Core.

Enfin, un élément de choix guidé par le besoin de maintenir notre objectif de réutilisabilité et d'interopérabilité dans le domaine des maladies infectieuses, nous a amené à choisir BFO comme ontologie de fondement. L'ontologie de fondement BFO fournit des concepts et des relations pertinents pour la construction d'ontologie de domaine en biomédecine (Grenon et al., 2004). Elle est déjà réutilisée par IDO-Core et IDSDO-Core. Ces différentes caractéristiques la rendent plus appropriée pour la modélisation ontologique du domaine des maladies infectieuses que d'autres ontologies de fondement existantes.

3.2. Modélisation multi-perspectives modulaires

Plusieurs points de vue sur les maladies, que nous appelons désormais perspectives, sont adoptés en médecine en fonction des études réalisées :

- 1 *Perspective biologique* : la biologie en tant que discipline couvre un domaine très large allant de la molécule à l'écosystème en passant par la cellule, l'organe et la population. Dans ce travail, nous nous limiterons à l'étude de l'interaction biologique de l'agent pathogène avec l'organisme, la réaction physiopathologie des hôtes à la maladie, la taxinomie et le cycle de vie des êtres vivants. L'objectif est de fournir une terminologie permettant d'annoter les données et les ressources liées à ces aspects biologiques de la schistosomiase. Cette même ressource ontologique est également réutilisable par les spécialistes biologistes de la schistosomiase dans leurs recherches scientifiques respectives.
- 2 *Perspective clinique* : les activités cliniques concernent essentiellement le traitement des patients. La perspective clinique de l'ontologie fournit une spécification des connaissances liées aux activités cliniques comme le diagnostic (les signes et

³ <https://github.com/gaoussoucamara/idoschisto/blob/master/idsdo-core.owl>

symptômes chez le patient) et son mode de traitement (prescription médicamenteuse). Cette partie de l'ontologie devrait faciliter l'intégration des données médicales sur la schistosomiase dans les systèmes d'information hospitaliers, et améliorer la prise en charge des patients souffrant de cette maladie.

- 3 *Perspective épidémiologique* : elle concerne principalement les épidémiologistes et est celle des systèmes de veille en santé publique. Le rôle de l'épidémiologiste est d'étudier les différents facteurs intervenant dans l'apparition et la distribution d'une maladie et les moyens à mettre en œuvre pour sa prévention et son contrôle. La veille épidémiologique est alors mise en œuvre à la fois, pour surveiller les facteurs de risque de propagation d'une maladie, pour analyser l'impact des événements détectés et pour suggérer des plans d'action en cas de risque identifié. Cette perspective de l'ontologie est donc principalement conçue pour représenter les connaissances sur l'épidémiologie de la maladie et servir à la prévention et au contrôle de la maladie.

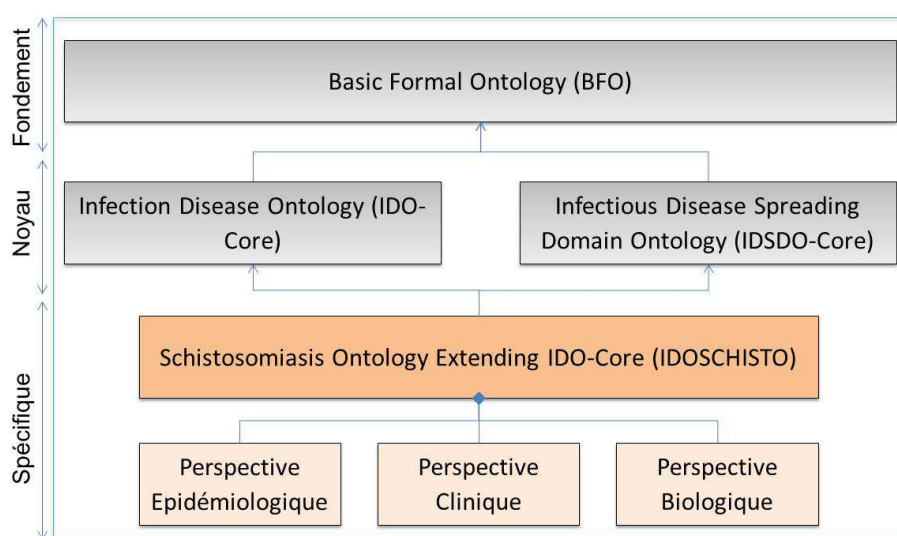


Figure 1 : Modularisation de l'ontologie de domaine de la schistosomiase (IDOSCHISTO)

Outre la cohérence dans la modélisation, la structure modulaire envisagée vise surtout la possibilité d'une utilisation partielle (Doran et al., 2007) de l'ontologie dans les disciplines correspondant à ces perspectives. Cette utilisation partielle permet de faire abstraction des concepts et relations non pertinents pour le sous-domaine considéré mais aussi de manipuler une ontologie de plus petite taille, ce qui renforce sans doute la performance de l'application au moins sur le plan de la complexité (chargement de l'ontologie, rapidité de l'exécution des requêtes, etc.).

3.3. Modélisation des relations inter-perspectives

Bien que la veille relève de la perspective épidémiologique, elle s'étend à toutes les perspectives dont la prise en compte est indispensable pour une bonne analyse des risques mais aussi des décisions adéquates. Ces différentes perspectives ne sont pas totalement indépendantes et sont plutôt fortement liées. Par exemple, ce sont les études biologiques qui permettent de produire un médicament contre la maladie. Lequel médicament est prescrit au patient lorsqu'il est diagnostiqué malade de la schistosomiase. Le traitement des malades est une des stratégies de contrôle de la propagation des maladies transmissibles comme c'est le cas pour la schistosomiase. Ainsi, l'intégration de ces différentes perspectives passe par la modélisation des relations existantes entre leurs concepts. Dans le domaine de la médecine il n'existe à l'heure actuelle qu'une seule ontologie permettant de définir les relations entre les concepts de ce domaine. Il s'agit de Relation Ontology (RO) (Smith et al., 2005). Par ailleurs,

une autre ontologie appelée RO-Bridge a été conçue ajoutant des contraintes de domaine (*domain*) et de co-domaine (*range*) aux relations définies dans RO pour les concepts de BFO. Ces deux ontologies sont donc utilisées et enrichies pour modéliser les relations dans les différents modules de l'ontologie. Il faut noter RO contient uniquement des relations et son importation ajoute dans IDOSCHISTO les relations qu'elle définit. Et puisque IDOSCHISTO a déjà importé BFO, l'importation de RO-Bridge permet d'ajouter les contraintes de domaine et de co-domaine entre RO et BFO.

4. La réutilisation de ressources ontologiques existantes

La réutilisation de ressources ontologiques existantes est un des scénarios proposés par la méthodologie NeOn. Elle est aussi une recommandation d'OBO Foundry pour les ontologies construites dans le domaine biomédical et aspirant à être publiées dans son portail⁴. La réutilisation d'ontologies de fondement et noyau constitue déjà une adhésion à ce principe visant essentiellement une interopérabilité des ontologies construites dans le domaine biomédical. Pour cela, deux procédés sont possibles : (i) la réutilisation complète ou partielle par importation dès le début de la construction ou (ii) l'alignement avec les ontologies existantes après construction. Notre approche combine les deux procédés. Nous faisons l'importation complète des ontologies ayant un niveau de développement mature dont le contenu global est pertinent. C'est le cas de BFO, d'IDO-Core et d'IDSDO-Core. Pour les autres ontologies stables mais dont une partie seulement nous intéresse, nous utilisons l'outil Ontofox⁵ (Xiang et al., 2010) pour extraire la partie en question. C'est par exemple le cas de l'extraction de la classification des schistosomes, parasite de la schistosomiase, depuis la taxinomie NCBI. Pour les ontologies utiles aux différents sous-modules et qui ne sont pas encore stables ou validées, nous proposons de procéder à un alignement de nos ontologies une fois leur construction terminée.

Tableau 1 : Extrait des ontologies réutilisées dans la construction d'IDOSCHISTO

Ontologies	Epi.	Clin.	Bio.
Basic Formal Ontology (BFO)	✓	✓	✓
Infectious Disease Ontology (IDO-Core)	✓	✓	✓
Information Artifact Ontology (IAO)	✓	✓	✓
Infectious Disease Spreading Ontology (IDSDO-Core)	✓	✗	✗
Pathogen transmission (PT)	✓	✗	✗
Human Disease Ontology (DOID)	✗	✓	✗
Ontology for General Medical Science (OGMS)	✗	✓	✗
Symptom Ontology (SO)	✗	✓	✗
Vaccine Ontology (VO)	✗	✓	✗
Ontology of Biomedical Investigation (OBI)	✗	✓	✓
Chemical Entities of Biological Interest (ChEBI)	✗	✗	✓
Protein Ontology (PO)	✗	✗	✓
NCBI Taxonomy Database	✗	✗	✓

Les ressources ontologiques répertoriées (cf. Tableau 1) ont été étudiées manuellement avec l'éditeur Protégé. Cette étude a consisté à rechercher la présence ou à déterminer la couverture des entités des sous-domaines que nous avons identifiées et modélisés dans la phase conceptuelle. Ensuite, les classes (ou relations) pertinentes pour nos ontologies sont

⁴ <http://www.obofoundry.org/>

⁵ <http://ontofox.hegroup.org/>

sélectionnées pour leur extraction avec Ontofox. L'ontologie en question peut être sélectionnée dans une liste déroulante déjà fournie ou être renseignée à travers son URI. Pour une entité donnée de l'ontologie, Ontofox permet d'extraire les *sous-classes* et la *hiérarchie* jusqu'à une *superclasse* qu'on aura indiquée. Les axiomes, relations et annotations associées à la classe peuvent aussi être extraites.

En plus d'IDO-Core, d'IDSDO-Core et de BFO, il existe plusieurs ontologies de domaine représentant les connaissances générales (indépendantes d'une maladie donnée) liées aux domaines des protéines, de la taxonomie des êtres vivants, des signes et symptômes, du diagnostic, du traitement, des médicaments, de la transmission de pathogène, etc. Ces ontologies offrent déjà des modèles cohérents et génériques que nous pouvons importer (partiellement ou intégralement) dans IDOSCHISTO. Étant donné qu'IDOSCHISTO est décomposée en perspectives, il nous a d'abord fallu déterminer, pour chacune de ces ontologies à importer, quelle perspective était pertinente ou compatible. Trois critères ont été établis pour déterminer la compatibilité de ces ontologies avec les perspectives retenues :

- 1 *critère épidémiologique* : contamination, transmission, propagation, facteur de risque, prévention, contrôle, etc.
- 2 *critère clinique* : signe, symptôme, diagnostic, traitement, médicament, etc.
- 3 *critère biologique* : gène, protéine, physiopathologie, cycle de vie, etc.

Dans le Tableau 1, en plus d'IDO-Core, d'IDSDO-Core et de BFO que nous importons intégralement, nous fournissons la liste des ontologies de domaine réutilisées partiellement ou intégralement pour chaque perspective. Il faut noter qu'une ontologie de domaine peut être compatible avec plusieurs perspectives. Le symbole ✓ signifie que cette ontologie de domaine est compatible avec cette perspective et le symbole ✕ signifie qu'elle est incompatible ou juste non pertinente pour cette perspective.

5. Formalisation de l'ontologie IDOSCHISTO

L'ontologie de la schistosomiase est ainsi construite en trois étapes :

- 1 L'importation intégrale ou partielle d'un ensemble d'ontologies existantes et dont les états de développement sont suffisamment avancés. BFO, IDO-Core, IDSDO-Core et RO ont été intégralement réutilisées. Les autres ontologies partiellement réutilisées sont listées dans le Tableau 1. Par exemple, les différentes espèces de parasite schistosome causant la schistosomiase ont été importées de la taxonomie NCBI et leurs cycles de vie ont été extraits d'Ontology for Parasite LifeCycle (OPL). Le choix de cette dernière ontologie est guidé par le besoin de superviser la mutation des parasites schistosome observée récemment par les experts biologistes et conduisant à de nouveaux types de schistosomiase.
- 2 L'ajout des concepts, relations et types de données propres à la schistosomiase et n'existant pas dans les ontologies importées. Ces éléments sont obtenus à partir des connaissances acquises lors des entretiens auprès des experts et l'exploitation des ressources non ontologiques tels que les publications scientifiques, les rapports d'étude, etc.
- 3 L'alignement manuel des concepts, relations et types de données ajoutés avec des ressources termino-ontologiques existantes que nous n'avons pas importées. C'est le cas par exemple des ontologies sur les symptômes (Symptom Ontology), les médicaments (Drug Ontology), le vaccin (Vaccine Ontology), Ontology for General Medical Science, etc.

En résumé, les ontologies importées ne couvraient pas tous les concepts de la schistosomiase. Nous avons ajouté une soixantaine de concepts centraux de la schistosomiase. Notons que ces ajouts tiennent compte de la distinction abstraite dans BFO entre les *continuants* et les *occurrents* (Simons and Melia, 2000). Les continuants représentent les

entités sans partie temporelle tels que les *objets* alors que les *occurents* représentent la classe des entités dynamiques tels que les *processus*.

Parmi les continuants, nous avons par exemple les modes de diagnostic, des conditions environnementales favorisant ou non la présence de la schistosomiase, etc. Nous n’avons pas directement rattachés ces concepts aux « Continuants » mais nous les avons répartis dans des sous-classes en fonction de leur cohérence. Les processus inhérents au domaine de la schistosomiase sont liés aux stratégies de contrôle et de prévention que nous avons ajoutées après exploitation des connaissances acquises auprès des experts et des ressources documentaires. Nous pouvons aussi citer les principales activités des populations les mettant en contact avec les points d’eau (« water_body »).

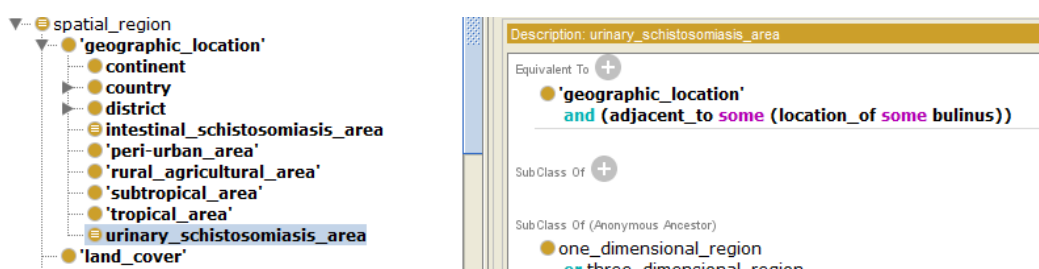


Figure 2 : Les zones à risque de schistosomiase urinaire

Tableau 2 : Métrique de l’ontologie IDOSCHISTO

Ontologie	#Concepts	#Relations	#Type	Total
Ontologie construite (concepts, relations et types propres ajoutés)				
IDOSCHISTO	58	14	0	72
Ontologies importées intégralement⁶				
Basic Formal Ontology (BFO)	39	0	0	39
Infectious Disease Ontology (IDO-Core)	285	14	0	299
ID Spreading Ontology (IDSDO-Core)	6	0	0	6
Relation Ontology (RO)	0	22	0	22
Ontologies importées partiellement				
Pathogen Transmission (TRANS)	4	0	0	4
Human Disease Ontology (DOID)	11	0	0	11
Population and Community Ontology (PCO)	21	28	0	49
NCBI Taxonomy Database	198	0	0	198
Ontology for Parasite LifeCycle (OPL)	130	6	0	136
Ontology of Medically Related Social Entities (OMRSE)	23	3	0	26
Environment Ontology (ENVO)	123	6	0	129
Exposure Ontology (EXO)	65	16	0	81
Total	958	109	0	1067

Les relations utilisées dans IDO-Core proviennent principalement de Relation Ontology (RO). Beaucoup d’autres relations sont importées à partir des ontologies réutilisées et nous en avons créé une quinzaine. Par exemple, parmi les relations ajoutées figurent « has_sign », « has_symptom », « has_vaccine », etc. liant les concepts que nous avons ajoutés. La relation « unfolds_in » décrivant le déroulement d’un processus dans l’espace a permis par exemple de dire que la propagation (« idsdo_spreading ») se propage dans (« unfolds_in ») un espace (« geographical_location »). Une instanciation de cette relation serait qu’une épidémie (sous-

⁶ Les concepts et relations obsolètes ne sont pas importés. Seuls les concepts de BFO et les relations de RO ne sont pas comptabilisés dans les chiffres indiquant les imports d’IDO-Core. Les autres concepts réutilisés par IDO-Core, par exemple à partir d’OBI, OGMS, etc., sont comptabilisés. IDSDO-Core ne comptabilise aussi que ses concepts et relations propres.

classe de propagation) se répand (*unfolds_in*) à Richard-Toll (une localité située au nord du Sénégal). Des classes définies ont été également proposées. Par exemple, les localités adjacentes à des points d'eau contenant le genre de mollusque bulinus sont des zones à risque de schistosomiase urinaire (cf. Figure 2). L'ontologie proposée est disponible à l'adresse suivante : <https://github.com/gaoussoucamara/idoschisto/blob/master/idoschisto.owl>. Les classes, relations et types de données importés sont présentés dans le Tableau 2.

Les modules correspondants aux différentes perspectives sont pour l'instant gérés par la création d'une annotation notée « *perspective* ». Elle prend trois valeurs possibles pour chaque concept et relation : *épidémiologique*, *biologique* ou *clinique*. Ainsi, chaque module peut être extrait en se fondant sur cette annotation. Cette manière de gérer des modules est encore triviale et son amélioration constitue l'une des perspectives de ce travail.

6. Cas d'étude : la schistosomiase à Richard Toll au Sénégal

Richard Toll est une zone située dans la vallée du fleuve Sénégal. Sa particularité dans notre étude est la densité de son réseau hydrographique (Traoré, 2000) qui est un facteur déterminant dans la transmission de la maladie. Les mollusques, hôtes intermédiaires des schistosomes, jouent un rôle important dans la transmission de la schistosomiase. Ainsi, les différentes espèces de mollusques rencontrées dans cette zone déterminent les types de schistosomiase rencontrés (Diaw et al., 1998). Nous recensons principalement deux grands genres dans la zone de Richard Toll : les *bulins* et les *biomphalaria*. Par conséquent, seuls deux parasites survivent dans la région, le *S. Haematobium* et le *S. Mansoni* causant respectivement la *schistosomiase urinaire* et la *schistosomiase intestinale*.

IDOSCHISTO est utilisée dans les études réalisées sur la schistosomiase à Richard Toll. Elle sert à l'annotation et au raisonnement sur les données. Les annotations portent sur les individus de la population, leurs répartitions géographiques, leurs activités, les points d'eau, les densités de mollusques et de parasites, les saisonnalités, les températures, les types de risque et les décisions associées, etc. Par exemple, la Figure 3 présente les annotations que nous avons introduites sur les données relatives aux quartiers de Richard-Toll et les définitions des relations d'*adjacence* de ces quartiers avec les différents points d'eau. Les relations d'adjacence sont particulièrement intéressantes car il est plus probable que les populations de ces quartiers privilégient ces points d'eau plutôt que ceux plus éloignés sauf si d'autres paramètres entrent en jeu (par exemple, aller pêcher dans les autres points d'eau car ce sont eux qui contiennent du poisson). Des annotations sur les espèces de mollusque (hôte intermédiaire des schistosomes) vivantes dans chaque point d'eau ont été introduites et les relations « *has_intermediary_host* » entre espèces de mollusque et espèces de schistosome ont été représentées.

Ces données sont issues des études conduites dans la localité et portent sur la répartition des espèces de mollusque, la localisation des points d'eau vis-à-vis des quartiers de la ville, etc. Par exemple, on peut exprimer la requête suivante en SPARQL : « *Quels sont les types de schistosomiasis auxquelles sont exposées les populations du quartier de Ndiaw ?* » (cf. Figure 4). Son exécution donne les résultats dans la deuxième partie de la Figure 4. Il n'est pas déclaré dans la base de connaissances que les populations du quartier de Ndiaw sont exposées aux schistosomiasis urinaire et intestinale mais c'est grâce aux connaissances représentées (concepts, relation, axiomes, instances) que ces résultats sont obtenus. En effet, on connaît quel type de schistosome cause quel type de schistosomiase, quel mollusque est hôte intermédiaire de quel schistosome, quel mollusque vit dans quel point d'eau et quel quartier est adjacent à quel point d'eau. L'analyse des URIs dans l'expression de la requête SPARQL montre la pertinence de la réutilisation des ontologies existantes et leurs utilisations effectives dans l'application visée.

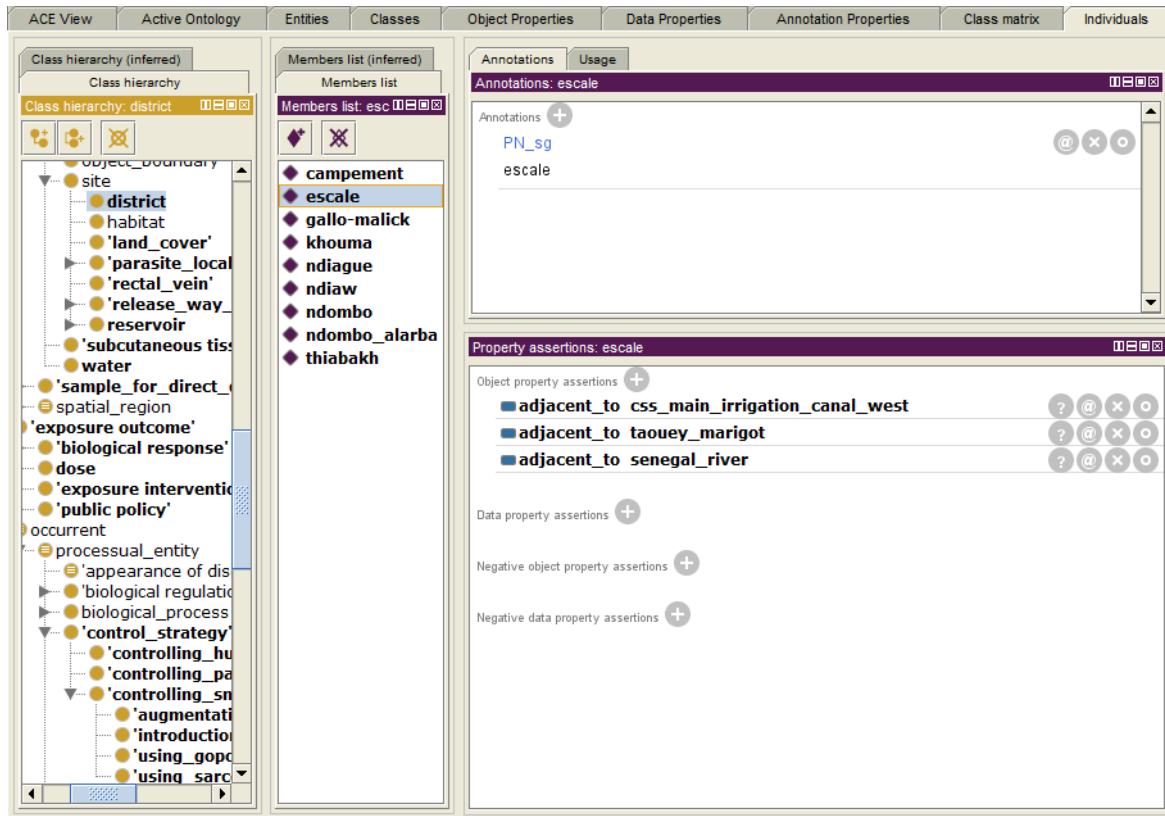


Figure 3 : Annotation des quartiers de Richard-Toll et définition de leurs relations d'adjacence avec les points d'eau

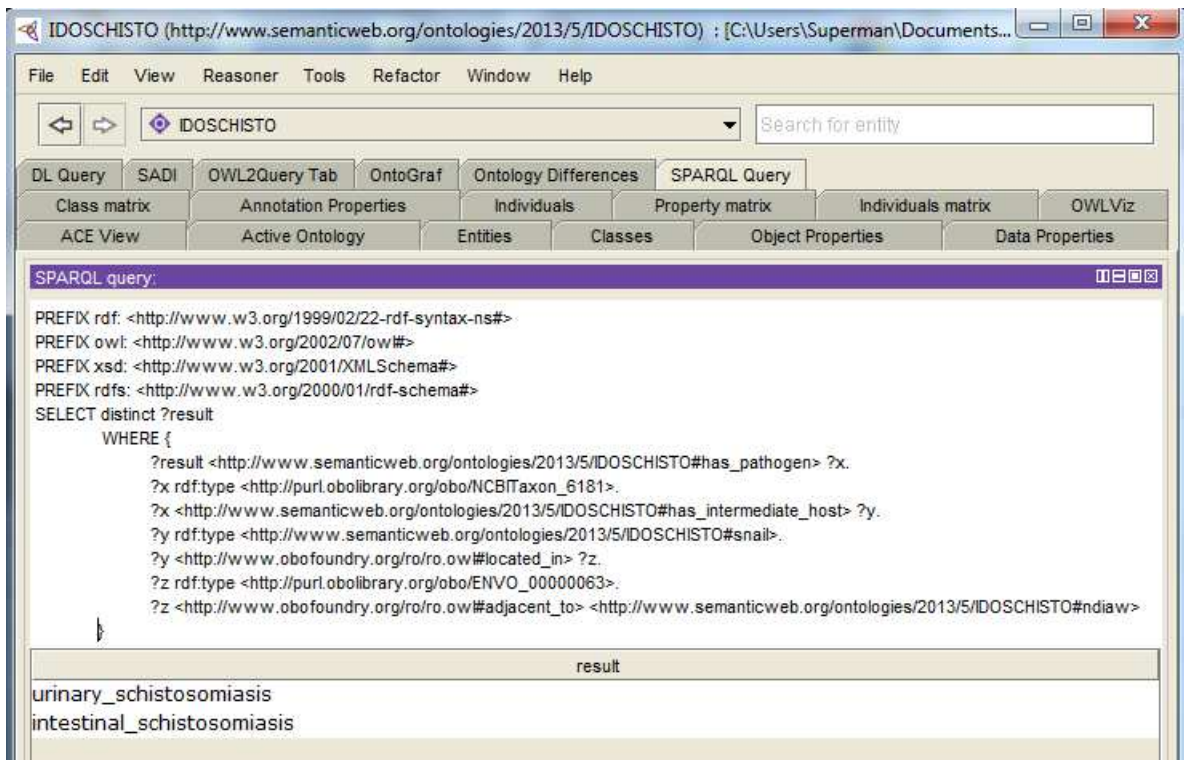


Figure 4 : Exemple de requête SPARQL dans IDOSCHISTO

7. Evaluation et validation de l'ontologie

La première phase de l'évaluation des ontologies de domaine a consisté à valider les modèles conceptuels avec les experts du domaine. Les modèles conceptuels de l'ontologie de la schistosomiase ont été examinés et validés avec des experts de la maladie au Sénégal. Cette validation ne garantit pas la couverture des connaissances du domaine mais elle valide au moins celles déjà représentées. Sur l'aspect formel, la réutilisation d'ontologie noyau ayant été déjà validée comme IDO-Core assure la cohérence d'IDOSCHISTO. Dans la réutilisation des ontologies de domaine spécifique, nous avons privilégié les ontologies qui ont connu un niveau de formalisation et d'implémentation avancées comme TRANS, OPL, PCO, NCBI, DOID, OGMS, etc. Un test a été effectué avec l'outil OOPS (Poveda-Villalón et al., 2012) pour détecter et corriger certaines incohérences et inconsistances. La validation par l'usage n'a pas pu se faire faute du retard dans l'implémentation du système de veille pour lequel l'ontologie est construite mais des requêtes SPARQL ont été testées avec l'interface de Protégé comme nous l'avons illustré dans la Figure 4.

8. Conclusion et discussion

Dans cet article, nous avons proposé une extension d'IDO-Core pour la construction de l'ontologie de domaine de la schistosomiase (IDOSCHISTO). La construction de cette ontologie suit l'approche que nous avons initiée dans (Camara et al., 2013). La conception de l'ontologie a été largement discutée avec les experts du domaine de la schistosomiase au Sénégal. L'ontologie du domaine de la schistosomiase étendant l'ontologie noyau du domaine des maladies infectieuses a été conçue pour couvrir l'ensemble des besoins en médecine. En effet, elle a été élaborée de façon modulaire de sorte qu'elle puisse être réutilisée partiellement pour des besoins spécifiques comme dans les systèmes d'information cliniques ou en recherche biologique. C'est pour cette raison que nous avons considéré dès le départ dans le cadre conceptuel la modularisation selon les perspectives épidémiologique, clinique et biologique. La construction d'IDOSCHISTO respecte également le principe de réutilisation de ressources ontologiques existantes dans OBO Foundry.

La modularité dans la partie domaine spécifique, notamment avec la distinction des différentes perspectives pose encore un réel problème. Bien que nous ayons créé et utilisé l'annotation *perspective* (avec *épidémiologique*, *clinique* et *biologique* comme valeurs possibles) sur un certain nombre de concepts clés, cette solution est loin d'être optimale. En effet, il y a également la partie concernant les axiomes qu'il faut gérer lors de l'extraction ou de la fusion des modules. L'évolution d'un module et son impact sur les autres modules et l'ontologie globale est également un défi à relever. En outre, la plupart des ontologies réutilisées sont encore au stade de *candidat* dans OBO Foundry. Elles sont par conséquent soumises à des évolutions plus ou moins importantes qu'il sera nécessaire de suivre et d'intégrer dans IDOSCHISTO.

Références

- Camara, G., Despres, S., Djedidi, R., Lo, M., 2012a. Modélisation ontologique de processus dans le domaine de la veille épidémiologique, in: Actes de La Conférence RFIA 2012. Presented at the RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle).
- Camara, G., Despres, S., Djedidi, R., Lo, M., 2014. Building a Schistosomiasis Process Ontology for an Epidemiological Monitoring System, in: Faucher, C., Jain, L.C. (Eds.), Innovations in Intelligent Machines-4, Studies in Computational Intelligence. Springer International Publishing, pp. 75–99.

- Camara, G., Després, S., Djedidi, R., Lo, M., 2012b. Vers une ontologie des processus de propagation des maladies infectieuses, in: Actes des 23èmes journées francophones d'ingénierie des connaissances, IC 2012. Paris, France, pp. 99–114.
- Camara, G., Després, S., Djedidi, R., Lo, M., 2013. Design of schistosomiasis ontology (IDOSCHISTO) extending the Infectious Disease Ontology. Presented at the In Proceedings of the 14th World Congress on Medical and Health Informatics, Copenhagen, Denmark.
- Diaw, O.T., Vassiliades, G., Seye, M., Sarr, Y., 1998. Les mollusques hôtes intermédiaires des trématodoses humaines et animales : distribution et variation de densité dans les différents systèmes épidémiologiques de Richard-Toll, in: Hervé, J.-P., Brengues, J., Eau et Santé : Colloque, Dakar (SEN), 1994/11/14-15 (Eds.), Aménagements hydro-agricoles et santé (vallée du fleuve Sénégal), Colloques et Séminaires. ORSTOM, Paris, pp. 201–218.
- Doran, P., Tamma, V., Iannone, L., 2007. Ontology module extraction for ontology reuse : an ontology engineering perspective. ACM Press, New York, pp. 61–70.
- Falbo, R., Baião, F., Lopes, M., Guizzardi, G., 2010. The Role of Foundational Ontologies for Domain Ontology Engineering: An Industrial Case Study in the Domain of Oil and Gas Exploration and Production. *Int J Inf Syst Model Des* 1, 1–22.
- Grenon, P., Smith, B., Goldberg, L., 2004. Biodynamic Ontology: Applying BFO in the Biomedical Domain, in: *Stud. Health Technol. Inform.* IOS Press, pp. 20–38.
- Lin, Y., Xiang, Z., He, Y., 2011. Brucellosis Ontology (IDOBRU) as an extension of the Infectious Disease Ontology. *J. Biomed. Semant.* 2, 9.
- Poveda-Villalón, M., Suárez-Figueroa, M.C., Gómez-Pérez, A., 2012. Validating Ontologies with OOPS!, in: Teije, A. ten, Völker, J., Handschuh, S., Stuckenschmidt, H., d' Acquin, M., Nikolov, A., Aussenac-Gilles, N., Hernandez, N. (Eds.), *Knowledge Engineering and Knowledge Management, Lecture Notes in Computer Science.* Springer Berlin Heidelberg, pp. 267–281.
- Scherp, A., Saathoff, C., Franz, T., Staab, S., 2011. Designing core ontologies. *Appl. Ontol.* 6, 177–221.
- Simons, P., Melia, J., 2000. Continuants and Occurrents. *Proc. Aristot. Soc. Suppl. Vol.* 74, 59–75+77–92.
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S., 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251–1255.
- Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A.L., Rosse, C., 2005. Relations in biomedical ontologies. *Genome Biol.* 6, R46.
- Suárez-Figueroa, M.C., Gómez-Pérez, A., Fernández-López, M., 2012. The NeOn Methodology for Ontology Engineering, in: Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E., Gangemi, A. (Eds.), *Ontology Engineering in a Networked World.* Springer Berlin Heidelberg, pp. 9–34.
- Traoré, M., 2000. Importance des aménagements hydrauliques dans la transmission des schistosomoses, in: Chippaux, J.-P., *Difficultés Rencontrées dans la Mise en Oeuvre des Programmes de Lutte contre les Schistosomoses en Afrique de l'Ouest : Atelier, Niamey (NER), 2000/02/15-18 (Eds.), La lutte contre les schistosomoses en Afrique de l'Ouest, Colloques et Séminaires.* IRD, Paris, pp. 23–29.
- Xiang, Z., Courtot, M., Brinkman, R.R., Ruttenberg, A., He, Y., 2010. OntoFox: web-based support for ontology reuse. *BMC Res. Notes* 3, 175.