

Towards a schistosomiasis ontology (IDOSCHISTO) extending the Infectious Disease Ontology

Gaoussou CAMARA^{1,2}

¹: EIR-IMTICE
Alioune Diop University
Bambey, Senegal
gaoussou.camara@uadb.edu.sn

Rim DJEDIDI²

²: LIMICS – INSERM Laboratory
Paris 13 University
Bobigny, France
rim.djedidi@univ-paris13.fr

Sylvie DESPRES²

²: LIMICS – INSERM Laboratory
Paris 13 University
Bobigny, France
sylvie.despres@univ-paris13.fr

Moussa LO³

³: LANI
Gaston Berger University
Saint-Louis, Senegal
moussa.lo@ugb.edu.sn

Abstract

Background: Schistosomiasis, also known as bilharzia, is a waterborne infectious disease caused by helminth parasites (blood flukes) called *Schistosoma* and transmitted when people come into contact with freshwater infested with larval form of the parasite named *cercariae* that penetrate the skin. Microscopic adult worms live then in the veins draining intestines (intestinal schistosomiasis) and the urinary tract (urogenital schistosomiasis). Consequently, the eggs laid by these worms are trapped in the tissues, and this causes massive damage and severe morbidity. To prevent emergence of schistosomiasis and control its spreading, it is necessary to bring together practitioners working at different levels of granularity (i.e. biological, patient and, population levels) and to consider the disease within several perspectives. Therefore, a schistosomiasis ontology is needed to support data integration, semantic interoperability, collaborative work, annotation and reasoning.

Methods: Schistosomiasis ontology (IDOSCHISTO) is a modular ontology designed as an extension of the core infectious disease ontology (IDO-Core). It reuses entirely or partially several biomedical domain ontologies dealing with infectious disease issues.

Results: IDOSCHISTO is structured through an abstraction-layered framework including a foundational ontology (BFO), a core ontology (IDO-Core) and a schistosomiasis specific ontology which is itself organized in three sub-modules taking into account epidemiological, clinical and biological perspectives on the disease. It contains 1067 entities including 958 concepts and 109 objects properties. 4 ontologies are entirely imported and 8 partially.

Conclusion: IDOSCHISTO is intended to become a reference knowledge model in schistosomiasis domain facilitating interoperability and multidisciplinary collaborative work, and providing reasoning support for various use cases such as epidemiological monitoring, clinical diagnosis and treatment, biological annotation.

Keywords—*infectious disease ontologies; schistosomiasis; modular ontology; epidemiological monitoring*

I. INTRODUCTION

Schistosomiasis – one of the Neglected Tropical Diseases (NTDs) – is a major public health problem and even one of the

most devastating parasitic diseases. It affects at least 258 million people worldwide in 2014, and more than 700 million people live in endemic areas [1]. Many efforts are carried out to advance research and development through partnerships and enhanced collaborative working with communities to control and prevent schistosomiasis. The situation analysis of interventions to overcome this infectious disease shows that strengthened surveillance systems are needed i) to control vector and intermediate host, and ii) to identify remaining foci and facilitate targeting of interventions [2].

To prevent emergence of an infectious disease and control its evolution, it is in fact, essential to set up an epidemiological surveillance system that monitors disease spreading and handle contamination modes and risk factors. Risk analysis and decision-making – key phases of monitoring process – imply actors with different profiles (e.g. physician, pathologist, parasitologist, epidemiologist, biostatistician, public health agent) having different observation perspectives (i.e. biology, clinic and epidemiology) on the studied phenomenon. Belonging to heterogeneous communities, these actors don't use the same vocabulary to refer to a same domain concept. A physician for example, will talk about *patient* while an epidemiologist will talk about *infected host*. Monitoring a specific infectious disease as schistosomiasis needs therefore an ontology mediation that enables semantic interoperability between heterogeneous actor vocabularies and a formal model of domain knowledge automating reasoning in epidemiological monitoring context.

In this paper, we present the development of schistosomiasis domain ontology IDOSCHISTO¹ as an extension of the IDO core Ontology. The approach covers biological, clinical and epidemiological perspectives. Modeling these perspectives and their interdependencies is essential since the disease spreading is considered as a complex system characterized by a multi-scale structure. This modular decomposition aims to facilitate partial use of IDOSCHISTO

¹ <https://github.com/gaoussoucamara/idoschisto>

in these subdomains. The ontology development methodology is also based on the reuse of existing infectious disease ontologies and a foundational ontology. As the design is introduced in [3], we will focus here on describing the content of the ontology and its potential usage on annotating schistosomiasis data in Senegal.

II. METHODS

Schistosomiasis domain ontology development followed NeON methodology [4]. We have applied Scenario 1 process from specification to implementation step jointly to Scenario 3 and 5 principles consisting essentially in reusing and merging existing ontological resources. In this section, we highlight the modularization approach. We started by building IDOSCHISTO ontology according to an abstraction-layered model to enable reuse of existing core ontologies [5] and a foundational ontology [6]. Then, we have extended the domain specific layer to take into account epidemiological, clinical and biological perspectives on the disease. Each perspective was modeled as an ontological module. Finally, we have related all these perspectives by modeling *inter-perspective relations*, i.e. relationships between concepts of distinct perspectives, to integrate the different modules in the main IDOSCHISTO ontology. Relationships between concepts of one perspective are called *intra-perspective relations* [3].

A. Abstraction-Layered Modularization

The design framework of schistosomiasis ontology is organized in three layers: the foundational layer, the core layer, and the domain specific layer. The specific layer includes the biological, clinical and epidemiological modules. The core layer dealing with infectious disease domain in general, reuses IDO-Core [7] modeling key concepts (e.g. pathogen, gene, cell, organ, organism, population, host, vector, human) and their relationships. The core IDO covers several subdomains including biological aspects (e.g. pathogen biological properties and their interaction with infected host organism, life cycle), clinical aspects (e.g. symptom, diagnosis, treatment), and epidemiological aspects (e.g. infection, transmission, spreading process, risk factors). It is therefore a suitable upper ontology for the ontological modules specifying these perspectives for schistosomiasis disease (specific layer). Epidemiological perspective however, is not completely taken on board in IDO-Core (i.e. disease spreading modes and strategies for control and prevention). Relationships between concepts, revealing spreading mechanisms, are not modeled. An infectious disease spreading core ontology IDSDO-Core [8] is thus also included to constitute a complete core layer for schistosomiasis specific modules.

IDO-Core is linked to the Basic Formal Ontology (BFO) [9]. BFO imported by the two core layer ontologies, provides a coherent classification of *process* and *object* concepts regarding infectious disease domain semantics and allows consistent reuse of relationships between these concepts to cover schistosomiasis specific relations. Other disease-independent ontological resources, modeling general knowledge (e.g. protein, human being, symptoms), are also reused according to their relevance to the epidemiological, clinical, and biological perspective modules [3].

B. Multi-Perspective Modularization

IDOSCHISTO modularization has focused on extracting representative modules for three specific medical study viewpoints named perspectives. *Biological module* focuses on the study of biological interactions between pathogen and organism, host physiopathological reactions to the disease, and living thing's taxonomy and life cycle. The scope of this module is to provide a reusable terminology enabling semantic annotations of schistosomiasis biological resources. *Clinical module* captures clinical knowledge related to schistosomiasis, commonly used in the hospital in-patient setting, including symptoms that influence differential diagnosis and treatment options. It aims to facilitate clinical data integration and therapy prescription, and to enhance patient care. *Epidemiological module* mainly covers the study of the various factors causing disease appearance and spreading and the means adopted to prevent and control it. It provides a common controlled terminology supporting epidemiologists and public health actors in monitoring risk factors of schistosomiasis spreading, analyzing impacts of detected epidemiological events, and recommending action plans in response to potential risks.

C. Inter-Perspective Relation Modeling

Monitoring activities comes mainly under epidemiological perspective. However, they also need to deal with the matter of the other perspectives to ensure thorough analysis and appropriate decisions. The perspectives described above are indeed not completely independent from each other. Biological studies for instance, produce medicine against the disease, which is prescribed to patient with positive infection diagnosis. Patient clinical treatment is one of the epidemiological strategies to control the spread of transmissible diseases as schistosomiasis. Thus, IDOSCHISTO includes relations between concepts of perspective modules. Two ontologies are reused and enriched for inter-perspective relation modeling: OBO Relation Ontology (RO) modeling biomedical concept relations [10] and RO-Bridge defining domain and range constraints on these relations. Note that RO only contains relations. When imported into IDOSCHISTO, these relations are related to BFO imported concepts. RO-Bridge import allows then to constraint these relations.

D. Reuse of Existing Ontologies

Reusing existing ontological resources is proposed in several NeON methodology scenarios. It is also an OBO Foundry recommendation [11] for building biomedical ontologies and publishing them on OBO portal. Reusing foundational and core ontologies subscribes to OBO shared principles that aim to develop interoperable biomedical ontologies. Two processes could be driven: (i) complete or partial reuse by import at the beginning of ontology building; or (ii) alignment with existing ontologies after building. Our approach combines both of them. OntoFox [12] tool were used for ontological portion extraction.

A set of existing ontological resources dealing with infectious disease and biomedical domain in general, were identified as the most likely to fulfill the ontology requirements for IDOSCHISTO (i.e. domain knowledge coverage, semantic

annotation facilities and reasoning support). They are fully (BFO, IDO-Core, IDSDO-Core) or partially imported and reused in the corresponding framework abstraction layer.

III. RESULTS

A. Modular Framework Architecture

To serve the purpose of supporting interoperability and enhance reasoning capabilities, IDOSCHISTO is designed as an abstraction-layered and modular framework reusing existing ontologies dealing with infectious disease domain (Fig. 1).

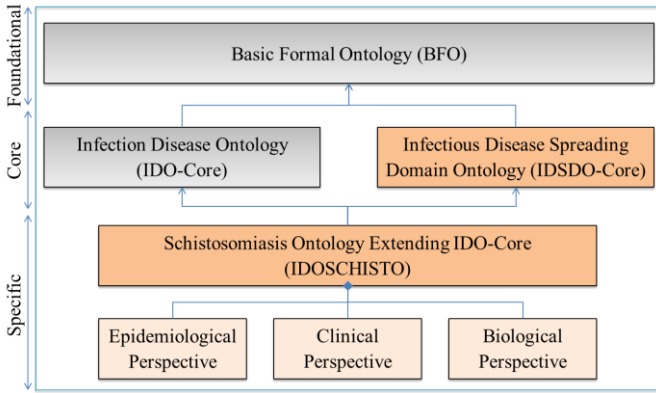


Fig. 1. IDOSCHISTO Design Framework Architecture

B. Modeling Schistosomiasis in IDOSCHISTO

The IDOSCHISTO modeling takes into account the abstract distinction in BFO between *Continuants* and *Occurrents* [13]. Continuants represent the entities without temporal parts such as objects while Occurrents represent the class of the dynamic entities such as the processes.

Among continuants, we have for example *diagnosis*, *symptoms*, *parasite role*, *land function*, etc. We did not directly connect these concepts with "Continuant" but we distributed them in subclasses according to their consistency (Fig. 2).

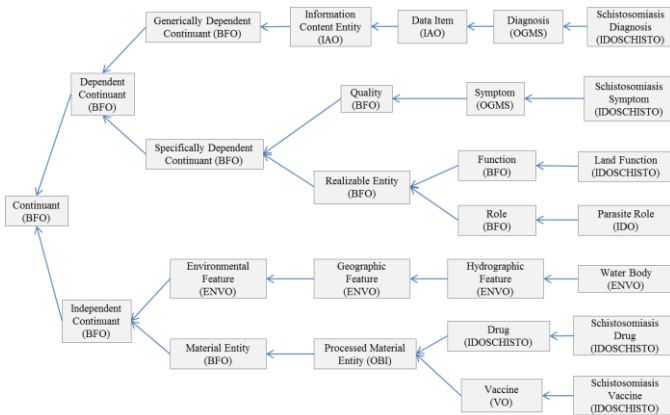


Fig. 2. Key concepts modeled under the Continuant class

The processes inherent to the schistosomiasis domain are related to the control and prevention strategies and are added after exploitation of the knowledge acquired with the experts

and from the documentary resources (Fig. 3). The spreading process knowledge is also modeled as well as the main activities of the populations putting them in touch with water sources ("water_body").

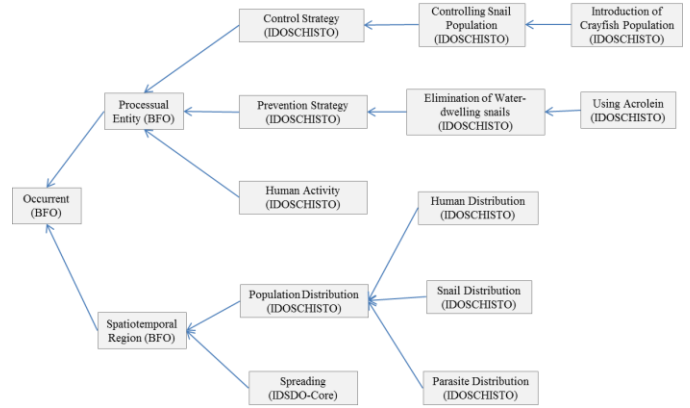


Fig. 3. Key concepts modeled under the Occurrent class

C. IDOSCHISTO Content Description

Schistosomiasis ontology formalization involves three main steps: (1) Fully and partial imports of a set of existing ontologies (Table 1). The different species of *schistosoma* parasites for instance, are imported from NCBI taxonomy; their life cycle description is extracted from Ontology for Parasite LifeCycle (OPL). The selection of this ontology was motivated by biologist needs of surveying *schistosoma* mutations. (2) Acquirement of new concepts, relations and data types specific to schistosomiasis (not covered by existing ontologies) within expert interviews and from non-ontological resources. (3) Manual alignment of the added entities to a set of non-imported termino-ontological resources such as Symptom Ontology, Vaccine Ontology, Drug Ontology, etc.

Besides relations imported from reused ontologies, 14 new relations are defined (e.g. *has_sign*, *has_symptom*, and *has_vaccine*) between the added concepts. *unfolds_in* relation for example, linking *idsdo_spreading* concept to *geographical_location* concept, describes the unfolding of a spreading process in a geographical area. An instantiation of this relation allows describing that an epidemic (subclass of *idsdo_spreading*) spreads to (*unfolds_in*) Richard-Toll – an area located in the North of Senegal (subclass of *geographical_location*).

Defined classes (one or more necessary and sufficient Condition(s)) are also proposed (Fig. 4). For example, locations nearby water sources containing *bulinus* snail are high-risk areas of urinary schistosomiasis.

As a first modularization stage, a annotation *perspective* is associated to IDOSCHISTO concepts with three possible values. The concept *parasite_distribution* for instance, has *epidemiology* as perspective annotation value whereas the concept *pathological_process* has *biology* value and the concept *sample_for_direct_diagnosis* (superclass of *blood_sample*, *urine*, *faeces*) has *clinic* value.

TABLE I. IDOSCHISTO METRICS

Ontology	#Concepts	#Relations	#Type	Total
Built ontology (added concepts, relations and types)				
IDOSCHISTO	58	14	0	72
Fully imported ontologies^a				
Basic Formal Ontology (BFO)	39	0	0	39
Infectious Disease Ontology (IDO-Core)	285	14	0	299
Infectious Disease Spreading Ontology (IDSDO-Core)	6	0	0	6
Relation Ontology (RO)	0	22	0	22
Partially imported ontologies				
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

^a Obsolete entities are not imported. Only BFO concepts and RO relations are not counted in IDO-Core import. The other one reused from OBI and OGMS for example are counted.

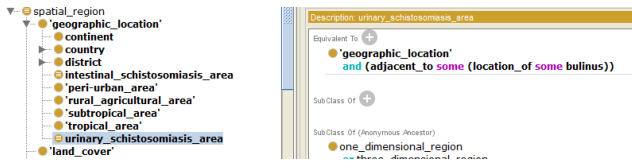


Fig. 4. Risk area of urinary schistosomiasis

D. A Case Study in Richard Toll – Senegal

Richard Toll (RT) area is located in northern Senegal, lying on the south bank of the River Senegal. The thickness of its river network makes it interesting to study schistosomiasis disease transmission. Studying the different mollusks – intermediary hosts of *schistosoma* – living there allows determining which schistosomiasis forms exist in this area. Two major snails are inventoried: *bulin* and *biomphalaria* and consequently, the major existing species of parasite are *schistosoma haematobium* and *schistosoma mansoni* causing respectively *urinary* and *intestinal schistosomiasis*.

IDOSCHISTO is used to annotate and query epidemiological data issued from investigations performed in Richard Toll area, dealing with distribution of snail species, water sources localization in town districts, etc.

1) Richard Toll Data Annotations

Semantic annotations are applied for individual populations, their geographical distribution, their activities,

water sources, snail and parasite densities, seasonality, temperatures, risk types and associated decisions, etc. Fig. 5 shows for example, annotations applied to Richard Toll districts and assertion of *adjacency* relation between districts and water sources. This is a particularly relevant relation, as populations in general rather prefer water sources that are nearby their district. Snail species living in each water source are annotated and the relation *has_intermediary_host* between snail and schistosoma species is asserted.

2) Richard Toll Risk Area Cartography

Richard Toll case study includes also data querying and reasoning. Fig. 6 shows for instance, the running of the following SPARQL query: *To what type of schistosomiasis are exposed populations of Ndiaw district?* Semantic annotations and reasoning capabilities on IDOSCHISTO enable inferring that Ndiaw district’s populations are exposed to urinary and intestinal schistosomiasis. URI analysis of SPARQL query demonstrates the relevance of reusing existing ontologies.

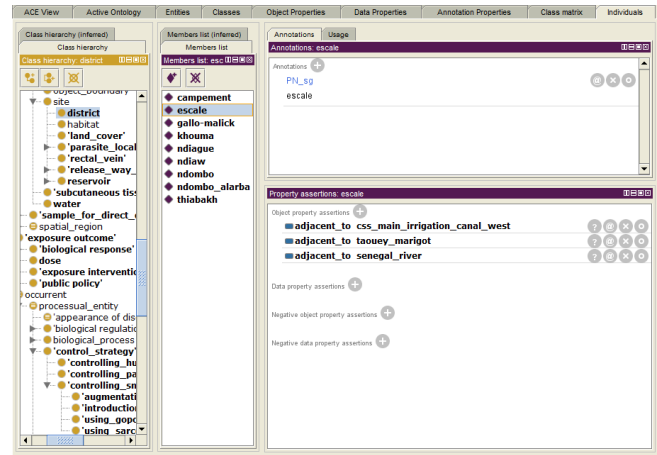


Fig. 5. RT districts annotation and adjacency relation with water sources

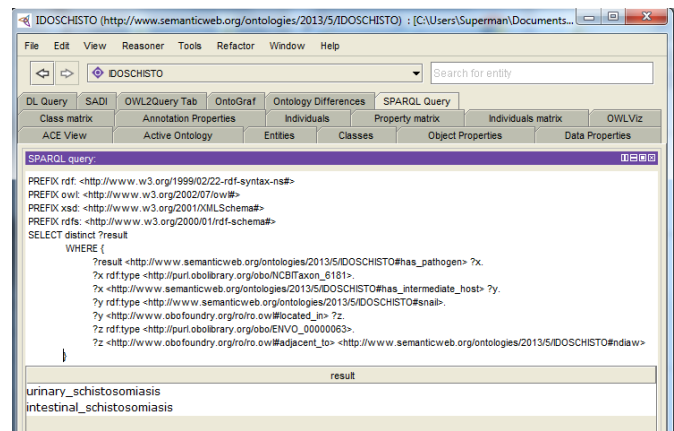


Fig. 6. SPARQL query on IDOSCHISTO

IV. DISCUSSION AND PERSPECTIVES

Several ontologies have been developed to enhance biological studies and support clinical decisions about treatment and diagnosis across a broad scope of infectious diseases [14]–[16]. In comparison to existing disease

ontologies, IDOSCHISTO was modeled for monitoring and prevention purposes in public health context. The approach accurately reflects biological, clinical and epidemiological perspectives of schistosomiasis through an abstraction-layered modularization. However, further refinements are needed in the domain specific layer including modules and their inter-perspective relations.

Reusing existing well-established core and foundational ontologies ensures IDOSCHISTO consistency and makes it in compliance with OBO Foundry principles. The reuse of domain specific resources (i.e. TRANS, OPL, NCBI) was based on high-level formalization and advanced implementation criterion. Some of the reused ontologies are in stage of OBO foundry candidates, which implies further, semantic and technical issues in handling their evolution.

Schistosomiasis ontology evaluation was carried out at three levels: coherence level (i.e. design validation with domain experts), consistency level (i.e. classification checking), and operationalization level (i.e. ontology usage assessment). A first case study was performed to initiate evaluation of IDOSCHISTO usability based on epidemiological data annotations and querying to identify risky areas in Richard Toll. Our future work will involve larger use case studies to fully describe the schistosomiasis domain knowledge and to assess how IDOSCHISTO could support epidemiological monitoring systems.

V. CONCLUSION

IDOSCHISTO is an application ontology capturing a controlled terminology for biological, clinical and epidemiological perspectives of schistosomiasis infectious disease through a specific, core and foundational abstraction-layered design framework. A modular approach was applied to fulfill coherent modeling objective and facilitate perspective-centered reuse. The built modular structures facilitate partial use [17] of IDOSCHISTO ontology by disregarding concepts (and relations) that are non-relevant for a specific perspective's requirements. Moreover, partial use of IDOSCHISTO allows handling smaller ontologies and thus, enhances performance of applications using it (e.g. ontology loading, query processing).

IDOSCHISTO building scope is to provide a knowledge representation framework for schistosomiasis resources and further, a standard component for scientific dissemination and for potential reuse particularly in epidemiological monitoring context. This future ambition is fully in line with schistosomiasis control initiative for the coming decade [2].

REFERENCES

- [1] "WHO | Schistosomiasis," *WHO*. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs115/en/>. [Accessed: 04-May-2016].
- [2] W. H. Organization, "Accelerating work to overcome the global impact of neglected tropical diseases: A roadmap for implementation: Executive summary," World Health Organization, Geneva, WHO/HTM/NTD/2012.1, 2012.
- [3] G. Camara, S. Desprès, R. Djedidi, and M. Lo, "Design of Schistosomiasis Ontology (IDOSCHISTO) Extending the Infectious Disease Ontology," *Stud. Health Technol. Inform.*, vol. 192, pp. 466–470, 2013.
- [4] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, "The NeOn Methodology for Ontology Engineering," in *Ontology Engineering in a Networked World*, M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, and A. Gangemi, Eds. Springer Berlin Heidelberg, 2012, pp. 9–34.
- [5] A. Scherp, C. Saathoff, T. Franz, and S. Staab, "Designing core ontologies," *Appl. Ontol.*, vol. 6, no. 3, pp. 177–221, Jan. 2011.
- [6] R. Falbo, F. Baião, M. Lopes, and G. Guizzardi, "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*, vol. 1, no. 2, pp. 1–22, avril 2010.
- [7] L. G. Cowell and B. Smith, "Infectious Disease Ontology," in *Infectious Disease Informatics*, V. Sintchenko, Ed. Springer New York, 2010, pp. 373–395.
- [8] G. Camara, S. Desprès, R. Djedidi, and M. Lo, "Building a Schistosomiasis Process Ontology for an Epidemiological Monitoring System," in *Recent Advances in Knowledge Engineering: Paradigms and Applications*, C. Faucher and L. Jain, Eds. Springer-Verlag, 2013.
- [9] P. Grenon, B. Smith, and L. Goldberg, "Biodynamic Ontology: Applying BFO in the Biomedical Domain," in *Stud. Health Technol. Inform*, 2004, pp. 20–38.
- [10] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse, "Relations in biomedical ontologies," *Genome Biol.*, vol. 6, no. 5, p. R46, Apr. 2005.
- [11] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, OBI Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nat. Biotechnol.*, vol. 25, no. 11, pp. 1251–1255, Nov. 2007.
- [12] Z. Xiang, M. Courtot, R. R. Brinkman, A. Ruttenberg, and Y. He, "OntoFox: web-based support for ontology reuse," *BMC Res. Notes*, vol. 3, p. 175, 2010.
- [13] P. Simons and J. Melia, "Continuants and Occurrents," *Proc. Aristot. Soc. Suppl. Vol.*, vol. 74, pp. 59–75+77–92, 2000.
- [14] Y. Lin, Z. Xiang, and Y. He, "Brucellosis Ontology (IDOBRO) as an extension of the Infectious Disease Ontology," *J. Biomed. Semant.*, vol. 2, no. 1, p. 9, 2011.
- [15] E. Mitraka, P. Topalis, V. Dritsou, E. Dialynas, and C. Louis, "Describing the Breakbone Fever: IDODEN, an Ontology for Dengue Fever," *PLoS Negl. Trop. Dis.*, vol. 9, no. 2, Feb. 2015.
- [16] C. L. Gordon, S. Pouch, L. G. Cowell, M. R. Boland, H. L. Platt, A. Goldfain, and C. Weng, "Design and evaluation of a bacterial clinical infectious diseases ontology," *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.*, vol. 2013, pp. 502–511, 2013.
- [17] P. Doran, V. Tamma, and L. Iannone, "Ontology module extraction for ontology reuse: an ontology engineering perspective," New York, 2007, pp. 61–70.