Formal Ontology in Information Systems
P. Garbacz and O. Kutz (Eds.)
IOS Press, 2014

doi:10.3233/978-1-61499-438-1-409

The Cardiovascular Disease Ontology

Adrien BARTON[a,b,1], Arnaud ROSIER[c,d,1], Anita BURGUN[e] and Jean-François ETHIER[e,1]
[a] *The Institute of Scientific and Industrial Research, Osaka University, Japan*
[b] *INSERM UMR 1099, LSTI, Rennes, France*
[c] *Hôpital Privé Jacques Cartier, Groupe GDS, Massy, France*
[d] *INSERM CIC-IT 804, CHU Pontchaillou, Rennes, France*
[e] *INSERM UMR 1138 team 22, Centre de Recherche des Cordeliers, Paris, France*

Abstract. This article presents CVDO, an ontology of cardiovascular diseases structured on OBO foundry's principle and based on BFO and FMA. CVDO reorganizes and completes DOID cardiovascular diseases around OGMS tripartite model of disease, and builds its taxonomy of diseases largely by automatic reasoning. It points to the need of OGMS to be supplemented by methodological rules to determine the end point of a disease course, and to locate the material basis of a disease in a causal chain of disorders.

Keywords. Disease, Cardiovascular system, Disposition, Ontological realism

Introduction

The OBO Foundry [1] is to date one of the most significant attempts to build interoperable ontologies in the biomedical domain. It is based on the upper-level realist ontology Basic Formal Ontology (BFO) [2], which aims at formalizing the most general classes on which domain ontologies should be based. In this framework, the Ontology for General Medical Science (OGMS) [3] provides a general model of disease, formalized as a disposition of an organism to undergo pathological processes, this disposition being based on an underlying disorder. For example, epilepsy is formalized as a disease to undergo epileptic crises (the pathological processes), due to some neuronal abnormal structure (the underlying disorder). The OBO Foundry also includes as a candidate the Human Disease ontology DOID [4] which lists human diseases. Although DOID accepts OGMS definition of disease, DOID classes have not yet been structured according to OGMS tripartite model of disease. The Cardiovascular Disease Ontology (CVDO) aims at reorganizing and completing DOID cardiovascular disease classes on the base of OGMS model of disease, and to align it with anatomical classes extracted from the Foundational Model of Anatomy (FMA) [5], in order to evaluate how OGMS model of disease can fit in the cardiovascular domain. The latest version of BFO (2.0) was used as a top-ontology, and CVDO was built in OWL format. CVDO aims at fulfilling OBO Foundry's principles; in particular, it concerns a well-defined scientific

[1] Contact authors: adrien.barton@gmail.com; arnaud.rosier@gmail.com; ethierj@gmail.com

field, namely cardiovascular diseases of the human adult, and it is openly available[2]. We will present CVDO's classes in a first part, the relations and logical aspects in a second part, and some design decisions we made in a third part[3].

1. Classes

CVDO reuses extensively classes from pre-existing OBO Foundry ontologies, to stay in line with OBO Foundry's interoperability requirements. BFO 2.0 was fully imported, and parts of the other ontologies were imported by MIREOT [6] using Ontofox [7]. 90 classes of FMA relevant for the anatomy of the cardiovascular system were imported, as well as relevant classes from OGMS. 294 classes[4] were imported from DOID, which include references to MeSH, ICD, NCI's thesaurus, SNOMED CT and OMIM disease-specific concepts and identifiers.

In order to test whether OGMS tripartite structure was applicable, DOID classes were reorganized and completed; in particular, some heart diseases[5] families (like heart valve diseases or heart conduction diseases) were systematically completed. DOID's taxonomical structure was changed were needed. In particular, some classes imported from DOID (and therefore originally categorized as subclasses of *Disease*) had to be reclassified in OGMS tripartite model as subclasses of *'Pathological bodily process'* (e.g. *Infarction*, '*Cerebrovascular accident'*, '*Pulmonary embolism'*, *Ischemia* or *Thrombosis*), or *Disorder* (e.g. '*Esophageal varix'*). We made OGMS structure explicit for most diseases, by introducing related disorder classes with descriptive names like '*Fluid in the pericardial cavity'*, and pathological process classes like '*Right ventricular hypertrophy'*. In total, 290 new classes were created in CVDO. Aristotelian definitions [1] reflecting this new classification are included[6].

Interestingly, the distinction between a disease and an associated pathological process is sometimes difficult to express in common medical language: for example, "atrial fibrillation" can refer to an episode of atrial fibrillation, or to a disposition to undergo frequently atrial fibrillation episodes. In such cases (which concern mainly heart conduction diseases), we followed the realist methodology by distinguishing two entities: '*Atrial fibrillation (disease)'* (which is a *Disease*, and thus a *Disposition*) and '*Atrial fibrillation (process)'* (which is a '*Pathological bodily process'*). Similarly, when the distinction between a disease and its associated disorder was difficult to express, we introduced two entities – e.g. '*Aneurysm (disorder)'* (the material structure) and '*Aneurysm (disease)'* (the disposition that this aneurysm disorder would rupture, or lead to other pathological processes).

[2] http://code.google.com/p/cvdo/, link available from the OBO Foundry website http://www.obofoundry.org/

[3] In the following, classes (universals) and the relations between them will be italicized, and the inheritance relations between classes in Protégé will be written in boldface, as **SubClass Of** and **Equivalent To**.

[4] For maximal class integration, we imported '*Cardiovascular system disease'* and its recursive closure, and then removed manually less relevant classes.

[5] The reclassification of vascular diseases is ongoing.

[6] As of June 13th, 2014, 282 (new or modified from DOID) Aristotelians definitions have so far been included in CVDO.

2. Relations and reasoning

Following OBO Foundry's spirit of minimizing the number of relations, we reused only relations coming from BFO 2.0[7], in particular to relate diseases with their associated disorder and disease course [8]. The disorder *R* underlying a disease *D* is formalized as a material basis of this disposition: *D* **SubClass of** *'has material basis at all times' some R*. The disease *D* is seen as a disposition to its whole disease course *C* [3]: when realized, *D* **SubClass of** *'realized in' some C*. A pathological process *P* of *D* is formalized as being part of this disease course *C*: *C* **SubClass of** *'has occurrent part' some P*.

CVDO was classified as consistent by the Pellet reasoner. 193 classes of CVDO have been axiomatized as defined classes. CVDO is built as an asserted monohierarchy (every class is asserted as being a subclass of at most one parent class[8]), and inferred polyhierarchy. As much as possible, we did not assert the disease taxonomic structure; instead, class axioms were such that the reasoner would build the taxonomy automatically. For example, the disease called "ischemic cardiomyopathy" in the 1995 classification of cardiomyopathies ([9]) is not considered as a cardiomyopathy in the 2008 classification ([10]; see also [11], [12]). In CVDO, accordingly, '*ischemic cardiomyopathy*' was correctly automatically classified as a subclass of '*cardiomyopathy (1995 definition)*', but not of '*cardiomyopathy (2008 definition)*'. Automatic taxonomic classification also enables to group heart diseases along different anatomical criteria (for example, as *'endocardium disease*' / '*myocardium disease*' / '*pericardium disease'*, or as '*left ventricle disease*' / '*right ventricle disease*' / '*heart valve disease'*), and by etiology ('*genetic heart disease*' / '*hypertensive heart disease*' / '*ischemic heart disease*'[9]). Finally, using the reasoner revealed logical gaps in medical definitions: the reasoner classified '*cor pulmonale*' as a subclass of '*cardiomyopathy (1995 definition)*', and careful reading confirmed that indeed, cor pulmonale satisfies the definition given in [9], although it had not been included in the 1995 classification.

3. Design decisions

We had to make several decisions when building the ontology, concerning: the existential conditions of a disease, how to define a disease of a specific anatomical entity (e.g. heart disease), the end point of a disease course, and how to find the material basis of a disease.

3.1. The existential conditions of a disease

It seems that most dispositions having a disorder as material basis are not diseases before they are realized. Consider for example a disposition to myocardium hypertrophy

[7] It would also be desirable to be able to relate a disease and its etiological process, or a disorder (for example a *'hypertrophied myocardium*') and the pathological process that brings it to existence (for example a *'myocardium hypertrophy'*), but this would require first to elucidate the nature of causation in BFO.

[8] The changes in DOID's taxonomical structure imply that using CVDO with the original version of DOID would lead to cases of double asserted inheritance, and to some inconsistencies (e.g. when an entity classified as a disease in DOID is classified as a disorder in CVDO). Thus, CVDO reclassification of DOID concepts could be seen as suggestions for DOID to adjust some of its classifications.

[9] Although it should be noted that formally, the ischemia involved in an ischemic heart disease does not belong to its etiological process in CVDO, but to its disease course.

because of a mutated gene; we considered this disposition to become a hypertrophic cardiomyopathy disease only if and when the hypertrophy happens (that is, when the disposition is realized) [10] . Thus, we formalized: *'Hypertrophic cardiomyopathy'* **Subclass Of** *'realized in' some ('disease course' and ('has occurrent part' some 'left ventricular myocardium hypertrophy'))*. On the opposite, some diseases may exist even if they are not (and may never be) realized, for example Brugada syndrome, which can lead to sudden death by ventricular fibrillation, but may also remain silent during the whole life of a person[11]. Finally, some pathological processes can happen without a corresponding disease: a person can undergo a thrombosis process because of thrombophilia (a strong disposition to thrombosis), but such a pathological process can also happen because of environmental factors (e.g. a seated position maintained too long) or just as a hazard (i.e. the patient had a disposition for a thrombosis process with a low probability - too low for this disposition to be considered as a disease - but this unlikely process still happened).

3.2. Defining a disease of an anatomical entity

One major decision was how to define consistently disorder, pathological process and disease of an anatomical entity, for example *'Heart disorder'*, '*Heart pathological process*' and '*Heart disease*'. We considered a heart disorder as a disorder located in the heart, rather than being a part of the heart, in order to include disorders such as a bacterial colony in the heart, or a blood clot in a coronary artery. Similarly, we considered a *'Heart pathological process'* as a pathological process that occurs in [12] the heart (e.g. a myocardium infarction). Finally, we considered that a necessary and sufficient condition for a disease to be classified as a heart disease is either to be realized by a disease course including at least one heart pathological process, or to have a heart disorder as material basis[13]. This raises two further questions: at which point does the disease course of a disease end? And how can be determined the material basis of a disease?

[10] It could be considered that the patient had a hypertrophic cardiomyopathy even before the gene expression, but this may not fit with common medical language.

[11] It would also be desirable to define diseases as having several possible disease courses, each with a different probability. This would however require to represent multi-track [8] and probabilistic dispositions [13], which is still an open issue in OWL.

[12] It could seem that some "heart pathological processes" do not occur in the heart, but have the heart as participant, like a heart beating irregularly. However, this specific pathological process can instead be seen as a process of improper myocardium contraction, and thus as occurring in the heart. The only exceptions we found are some pathological process involving heart valves, like a heart valve improper opening or improper close, which can difficulty be interpreted as occurring in the heart valve itself. In this case, we defined two classes: 'heart valve pathological process', defined as pathological processes occurring in a heart valve, and 'pathological process involving a heart valve', defined as pathological processes having as participant a heart valve. This distinction is important, as a disease having a pathological process involving a heart valve is not necessarily a heart valve disease (contrarily to a disease having a pathological process occurring in a heart valve). In subvalvular aortic stenosis, for example, a malformation of the left cardiac ventricle causes an improper opening of the aortic valve; it should however not be classified as a heart valve disease, as it involves (in the general case) a healthy aortic valve.

[13] A genetic heart disease may be considered as having as material basis an abnormal sequence of nucleotides located in the heart cells where the gene will be expressed (although the abnormal sequence may be present in the whole organism, it will not be expressed in the cells that are not part of the heart).

3.3. The end point of a disease course

Consider an instance of systemic disease like hemochromatosis, which affects (by iron overload) many organs at the same time, including the heart. The iron overload in the heart causes a restrictive cardiomyopathy, in which the left ventricular myocardium becomes abnormally stiff, leading to further pathological processes, like a dysfunctional contraction process of the myocardium. Is this dysfunctional process only part of the disease course of the restrictive cardiomyopathy? Or is it also part of the hemochromatosis disease course? One could decide that the disease course of hemochromatosis stops when the disease course of the restrictive cardiomyopathy starts; in this case, this instance of hemochromatosis causes a heart disease (the restrictive cardiomyopathy), but is not by itself a heart disease. Alternatively, one could decide that the disease course of the hemochromatosis includes the disease course of the restrictive cardiomyopathy; in this case, this instance of hemochromatosis would *be* a heart disease (because of the criteria exposed in 3.2; and it would also be a disease of every other organ it affects). CVDO is largely neutral concerning this question, but this is an important issue on which all ontologies based on OGMS should agree.

3.4. Finding the material basis of a disease

Consider again restrictive cardiomyopathy. This disease is not necessarily due to hemochromatosis – it might also be caused by e.g. genetic factors [11]. Is the stiff myocardium the material basis of this disease (in which case the process causing this stiffness may be the etiological process of the disease), or is this myocardium stiffening a pathological process of this disease? In this example – and many others – it was not obvious to the authors[14] when devising CVDO which disorder, in the causal chain of disorders (and pathological processes), should be selected as the material basis of the disease. Therefore, we adopted the following (informal) methodological rule:

> "First-Disorder Rule": The material basis of a disease D is the first disorder in the causal chain of disorders in which D appears (or the first disorder that immediately follows the last material basis of any disease preceding D in this causal chain, in case such diseases exist)[15].

The First-Disorder Rule suggests that the material basis of a restrictive cardiomyopathy may be, for example, a genetic mutation (for a genetic restrictive cardiomyopathy) or iron overload in the ventricles (for a restrictive cardiomyopathy due to hemochromatosis). In all cases, the stiffening of the ventricles walls should be seen as a pathological process of the disease. Thus, *'Restrictive cardiomyopathy'* was formalized as **SubClass Of** *'realized in' some ('Disease course' and ('has occurrent part' some 'myocardium of left ventricle stiffening'))*, rather than as having as material basis the stiff myocardium.

---

[14] Among the authors, A.R. is a cardiologist, J.-F.E. is an internist, A.Bu. is a specialist in medical computer science and A.Ba. is a philosopher.

[15] This rule should be tested in other medical domains, and refined, as it raises interesting questions – for example: how can be formally defined the causal chain associated with a disease? What does "immediately follows" mean? Can this rule be applied on leaf disease universals only, or also on non-leaf disease universals? This methodological rule may imply that the material basis of a disease will depend on the degree of precision in the description of the disorders, as well as on the material basis of other diseases in the causal chain; it is an open question whether this is compatible with BFO's realist spirit.

Conclusion

Devising CVDO showed that a realist ontology like OGMS enables to make important ontological distinctions between entities that are not well distinguished in common medical language, to build elaborate disease classifications by automatic reasoning, and to derive interesting inferences concerning disease classifications. We propose that OGMS be supplemented by methodological rules to determine the material basis of a disease and the end point of a disease course, which may also help to articulate OGMS formalization with the River-Flow model of disease [14][16].

References

[1] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, et al., The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology* **25** (2007), 1251–1255.

[2] P. Grenon, B. Smith, SNAP and SPAN: Towards Dynamic Spatial Ontology, *Spatial Cognition and Computation* **4** (2004), 69–103.

[3] R.H. Scheuermann, W. Ceusters, B. Smith, Toward an ontological treatment of disease and diagnosis, in: *Proceedings of the 2009 AMIA Summit on Translational Bioinformatics*, San Francisco CA, 2009, pp. 116–120.

[4] L.M. Schriml, C. Arze, S. Nadendla, Y.-W.W. Chang, M. Mazaitis, V. Felix, et al., Disease Ontology: a backbone for disease semantic integration, *Nucleic Acids Research* **40** (2012), D940–D946.

[5] C. Rosse, J.L.V. Mejino Jr., A reference ontology for biomedical informatics: the Foundational Model of Anatomy, *Journal of Biomedical Informatics* **36** (2003), 478–500.

[6] M. Courtot, F. Gibson, A.L. Lister, J. Malone, D. Schober, R.R. Brinkman, et al., MIREOT: The minimum information to reference an external ontology term, *Applied Ontology* **6** (2011), 23–33.

[7] Z. Xiang, M. Courtot, R.R. Brinkman, A. Ruttenberg, Y. He, OntoFox: web-based support for ontology reuse, *BMC Research Notes* **3** (2010), 175.

[8] J. Röhl, L. Jansen, Representing dispositions, *Journal of Biomedical Semantics* **2** (2011), S4.

[9] Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the Definition and Classification of Cardiomyopathies, *Circulation* **93** (1996), 841–842.

[10] B.J. Maron, J.A. Towbin, G. Thiene, C. Antzelevitch, D. Corrado, D. Arnett, et al., Contemporary Definitions and Classification of the Cardiomyopathies, *Circulation* **113** (2006), 1807–1816.

[11] P. Elliott, B. Andersson, E. Arbustini, Z. Bilinska, F. Cecchi, P. Charron, et al., Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases, *European Heart Journal* **29** (2008), 270–276.

[12] C. McCartan, R. Mason, S.R. Jayasinghe, L.R. Griffiths, Cardiomyopathy classification: ongoing debate in the genomics era, *Biochemistry Research International* **2012** (2012), 796926.

[13] A. Barton, A. Burgun, R. Duvauferrier, Probability assignments to dispositions in ontologies, in: M. Donnelly, G. Guizzardi (Eds.), *Proceedings of the 7th International Conference on Formal Ontology in Information Systems (FOIS)*, IOS Press, Amsterdam, 2012, pp. 3–14.

[14] R. Mizoguchi, K. Kozaki, H. Kou, Y. Yamagata, T. Imai, K. Waki, et al., River Flow Model of Diseases., in: *Proceedings of the 2nd International Conference on Biomedical Ontology (ICBO 2011)*, CEUR Workshop Proceedings, 2011, pp. 63–70.

[16] We would like to thank Albert Goldfain, Chris Mungall and Pascal Van Hille for their helpful suggestions, as well as Lynn Schriml, Warren Kibbe, Barry Smith and Mélanie Courtot for interesting comments.