The CHRONIOUS Ontology Suite: Methodology and Design Principles

Luc Schneider[1], Mathias Brochhausen[1],

[1] Institute for Formal Ontology and Medical Information Science,
Saarland University, PO. Box 151150, D-66041 Saarbrücken, Germany
{luc.schneider, mathias.brochhausen}@ifomis.uni-saarland.de

Abstract. This paper outlines the methodology and the basic design principles underlying the development of the ontology suite in the EU-funded project CHRONIOUS, comprising a Middle Layer Ontology for Clinical Care (MLOCC) as well as two domain ontologies, one on Chronic Obstructive Pulmonary Disease (COPD) and one on Chronic Kidney Disease (CKD). The article also sketches some major philosophical reflections underpinning the CHRONIOUS ontologies.

Keywords: middle layer ontology, chronic disease ontology, clinical care ontology, chronic obstructive pulmonary disease, chronic kidney disease

1 Overview of the CHRONIOUS Ontology suite

1.1 Purpose

The CHRONIOUS[1] project's primary aim is the development of an integrated telemedical platform for monitoring the general health status of patients with chronic health conditions and providing decision support for the clinicians treating them. For demonstrative purposes, the project focuses on Chronic Obstructive Pulmonary Disease (COPD) and Chronic Kidney Disease (CKD) including Renal Insufficiency [1].

Part of the CHRONIOUS platform is an ontology-powered literature search tool providing efficient and accurate access to recent research literature on COPD and CKD for health care professionals. Publications are annotated both with classes from the CHRONIOUS ontologies and with terms from the MeSH[2] thesaurus[3]. Thus, the CHRONIOUS literature search system combines the terminological knowledge and the multi-linguistic capabilities of MeSH with the clinical expert knowledge encoded by ontologies as topic-neutral representations of the items (objects, processes, qualities, dispositions, functions, etc.) in the domain of the etiology, diagnosis and therapy of COPD and CKD.

---

[1] "An Open, Ubiquitous and Adaptive Chronic Disease Management Platform for COPD and Renal Insufficiency" (http://www.chronious.eu/)

[2] Medical Subject Headings; www.nlm.nih.gov/mesh/

[3] Initially it was planned to use the very same ontologies also in the decision support system, but because of scalability issues, this idea was not followed up by the project partners.

1.2 Technical Details

The CHRONIOUS ontologies exhibit the following modular structure:

1. The Middle Layer Ontology for Clinical Care (MLOCC)[4], which has been extracted from the ACGT Master Ontology [2,3], augmented by general clinical classes identified in the documentary sources supplied by the medical experts in the CHRONIOUS project (see below). MLOCC contains general classes for objects (chemical substances, cells, tissues, organs, technical instruments, etc.), processes, qualities, powers, functions and roles that are relevant to pathological, anatomical, diagnostic and therapeutic aspects of clinical care. MLOCC is based on Basic Formal Ontology [4], a foundational ontology widely used in the biomedical domain; it also contains a subset of the Foundational Model of Anatomy (FMA) [5] and the Relation Ontology (RO) [6].
2. the COPD ontology[5] containing specific domain knowledge about Chronic Obstructive Pulmonary Disease, and
3. the CKD ontology[6] containing specific domain knowledge about chronic kidney diseases and renal insufficiency.

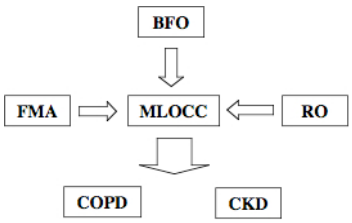

Fig. 1. Components of the CHRONIOUS ontology module

The following table sums up some quantitative data about the CHRONIOUS ontologies:

	Classes	Relations	Axioms
MLOCC	476	65	~ 900
COPD (+MLOCC)	964	65	~ 2000
CKD (+MLOCC)	972	65	~ 2000

From the table above the reader can gather that all object properties (relations) used in the CHRONIOUS ontologies (of which the RO relations are a subset) are defined on the level of MLOCC. The COPD and the CKD ontologies have approximately the same size. All ontologies have been developed in OWL-DL

[4] http://www.ifomis.org/chronious/mlocc
[5] http://www.ifomis.org/chronious/copd
[6] http://www.ifomis.org/chronious/ckd

using Protégé 4.02 and 4.1; the restriction to OWL-DL may be unnecessary for the purposes of literature search, but (1) the ontologies are intended to be appropriate for other, more reasoning-intensive uses and (2) decidability allows for efficient consistency checking, consistency being a necessary, if not sufficient, condition for adequacy to reality.

2 Methodological principles

2.1 Realism

From the implementation point of view, an ontology is a software artifact (e.g. an OWL file). Hence its design features can be more or less influenced by the peculiar use case(s) of the overall information system in which it is integrated. Thus, to guarantee data exchange and re-use across heterogeneous sources, we must recur to ontologies that unify the different ways in which the domain is viewed by the different end-users and information systems designers. Realist ontologies, i.e. ontologies that aim to depict reality independently of the mental or digital representation of reality by end users and knowledge engineers, provide a unified way of representing the domain from the start, without the need of an ex-post integration of heterogeneous perspectives.

Since the CHRONIOUS literature search system is intended to be user-friendly for health care professionals and clinicians of various specialities, the best option seemed to us to implement the CHRONIOUS ontologies as realist ontologies. Therefore we adopted realism as a fundamental methodological choice, following the general approach already adopted for the development of the ACGT (Advancing Clinico-Genomic Clinical Trials on Cancer) Master Ontology [2, 3].

2.2 Adoption of BFO as a top-level ontology

In order to guarantee that a domain ontology is a reference ontology, it is mandatory to use the best available sources of expert knowledge about the reality to be depicted, e.g. the biomedical domain, but also to recur to formal ontologies. i.e. formal systems of general categories and relations for depicting reality, namely foundational or top-level ontologies such as BFO [4] or DOLCE [7]. Since BFO (Basic Formal Ontology) is used as a reference top-level ontology in the major open-source repository for biomedical ontologies, namely the OBO Foundry [8], we have decided to build the Middle Layer Ontology for Clinical Care (MLOCC) and thus the CHRONIOUS ontologies on re-using BFO by appending the upper-level classes of MLOCC onto leaves of BFO, as already done in the case of the ACGT Master Ontology [2,3]. By appending MLOCC-classes under BFO-classes, the meaning of the latter are given a reality-driven semantics, thus ensuring that the CHRONIOUS ontologies are truly reference ontologies satisfying the methodological requirement of realism.

Some examples should suffice to illustrate how MLOCC has been constructed around BFO; in the following we assume familiarity with the general

structure of BFO [4]. Under the BFO-class *Generically Dependent Continuant* we have appended the MLOCC-class *Infomation Object*, which covers information artifacts such as questionnaires, medical images or designs (plans, e.g. standardized procedure plans) as distinct from their material supports (paper, traces on electromagnetic storage devices etc.). The BFO-class *Quality* subsumes e.g. the MLOCC-classes *Magnitude* (covering physical magnitudes) and *Condition* (which subsumes the important class *Organismal Condition*). Under the BFO-class *Function* we find the MLOCC-class *Organ Function*, while the BFO-class *Role* covers not only social roles like *Professional Role* or *Administrative Role*, but also biochemical extrinsic features like *Biomarker*, *Drug*, *Catalyst* (Enzyme) or *Hormone*. The BFO-class *Disposition* subsumes the all-important MLOCC-classes *Disease* and *Malfunction*.

Under the BFO-class *Object* we find mainly the MLOCC-nodes *Biological Independent Continuant* (covering *Organism* and *Organismal Independent Continuant* which subsumes classes added from the FMA), *Chemical Substance*, *Institution* and *Technical Object* (for instance devices and instruments). The BFO-class *Process* is subdivided into the MLOCC-classes *Intentional Process* and *Natural Process*; the first subsumes classes related to human and social activities, in particular medical (diagnostic and therapeutic) processes, while the second subsumes *Chemical Process* and *Organismal Process*. An often recurrent feature is the following: a *Disposition* (cf. Figure 2), i.e. a *Disease* or *Malfunction* is realized by an *OrganismalProcess* that bears the relation *has_Outcome* to a *Quality*, i.e. an *Organismal Condition* or a *Magnitude* such as a *Glomerular Filtration Rate* or *Spirometric Measure*.

2.3 Modularity and Re-use

Another important methodological principle besides realism is perspectivalism, which is the recognition that reality is complex and variegated [4]. There are many different representations of reality that are equally adequate because they capture different important aspects of the same world. Thus, reality can be assayed in terms of substances and their qualities or powers as well as in terms of processes. More importantly, reality can be described at various levels of granularity, ranging from the atomic and molecular levels to those of cells, tissues and organisms. As a consequence, reality cannot be accounted for in terms of a single monolithic ontology, but only in terms of a multitude of modular ontologies that are orthogonal to each other and thus are also re-usable.

We have already mentioned that the CHRONIOUS ontologies were built on top of an established Upper Ontology, namely BFO, as far as high-level classes are concerned. As to relations or object properties, we have recurred to the Relations Ontology (RO) of the OBO Foundry [6]. RO is a set of formal relations that are used in biomedical applications, and minor modifications apart, the object properties of the CHRONIOUS ontologies are an extension of RO. While BFO is directly imported into MLOCC, the object properties in RO have been copied into

MLOCC, since the tree of RO object properties has been modified. E.g. in MLOCC *participates_in* subsumes not only *agent_in*, but also *means_of*.

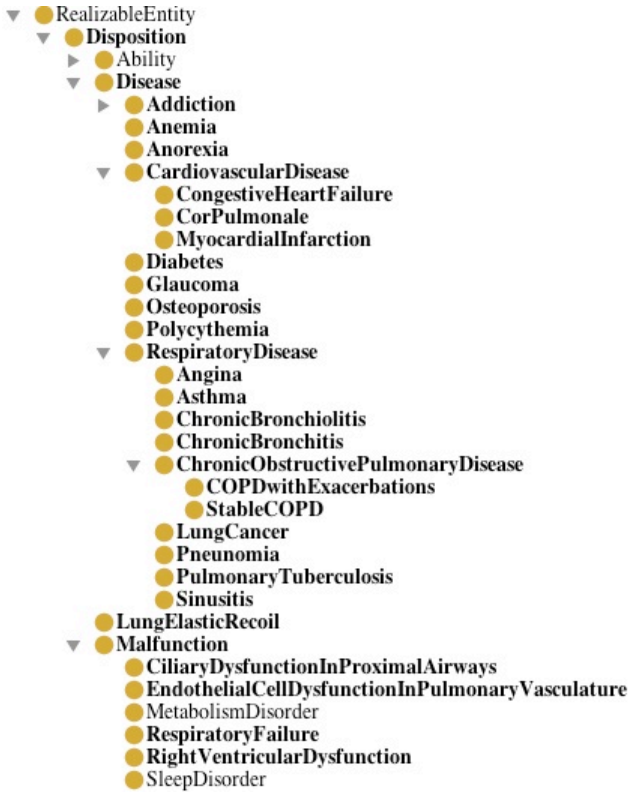

Fig. 2. An example of the ontological representation using BFO

Another part of MLOCC modelled after an already existing ontology is the branch below the class *Organismal Independent Continuant*, which mirrors the structure and content of the Foundational Model of Anatomy, a reference ontology for anatomy [5]. The FMA classes have been added as is, except for two major exceptions. First, *Biological Macromolecule* has been moved under *Chemical Substance*. Second, the hierarchy below *Cardinal Organ Part* has been simplified by directly adding *Cardinal X Part* for any Organ X (e.g. *Cardinal Heart Part*, *Cardinal Lung Part*). This move was necessary to avoid a ramification of subdivisions that would be spurious for the use case of literature annotation; it does not compromise realism, since the added classes represent objective divisions in biomedical reality.

The principle of re-using already validated software constructs has been applied for the design of the Middle Layer Ontology for Clinical Care (MLOCC) itself, the core of which was extracted from the ACGT Master Ontology [2,3].

Finally, the CHRONIOUS ontologies are modular insofar both the COPD Ontology and the CKD Ontology import MLOCC, which itself imports BFO. MLOCC represents the common core of the chronic disease ontologies, which expand on it in domain-specific ways, except for the object properties that are defined on the level of MLOCC.

Taking modularity and re-use a step further, we would have wished to be able to re-use the Ontology for General Medical Science (OGMS)[7] as an intermediary level between BFO and MLOCC, but for the fact that this promising medical ontology is still in ist early phase of development. Also, we did not consider to use of the Disease Ontology (DO)[8] because (1) DO is not based on any foundational ontology like BFO, (2) DO does not provide axioms besides trivial subclass axioms and (3) not all subclass-relations in DO are formal ISA-relations resulting occasionally in multiple inheritance of primitive classes (e.g. the class *Nelson's Syndrome* is a subclass to both *Adrenal Cortex Disease* and *Pituitary Neoplasm*).

2.4 Selection of Sources and Class Extraction

The classes and relations of the CHRONIOUS domain ontologies have been extracted from clinical guidelines about COPD and CKD selected and validated by the medical board of the CHRONIOUS project [9,10,11,12,13,14]. The actual class extraction roughly followed the procedure proposed in [15] and adopted for the ACGT Master Ontology, as indicated in [2]:

1. A glossary of candidate classes is established whose coverage is evaluated by domain experts (e.g. clinicians).
2. The classes are assigned to the different ontological categories, i.e. classes of the foundational ontology BFO.
3. The classes are assigned either to the middle layer (MLOCC) or to the domain ontologies.
4. The classes are ordered in subsumption hierarchies or taxonomies.
5. The (binary) non-taxonomic relations between the classes are identified and represented as object properties.
6. A class dictionary is constituted, describing each class and stating the relations that have it as their domain.
7. The inverse relation and mathematical properties (symmetric, transitive, etc.) of each object property is specified.
8. Formal axioms are stated; these axioms constrain the extension of classes by specifying binary relationships between them in the form of object properties.

[7] http://code.google.com/p/ogms/
[8] http://do-wiki.nubic.northwestern.edu/index.php/Main_Page

3 Design Principles

3.1 General Design Principles

The general design principles discussed below pertain to the articulation of taxonomies; these principles have been explained in detail in [2], so we may review them briefly, demonstrating issues related to applying those principles in the design of the CHRONIOUS ontologies.

Taxonomies should contain only types, not instances or tokens. This principle trivially reflects the type-token distinction. However, information objects like questionnaires are ontologically tricky in this respect. Is *Chronic Respiratory Disease Questionnaire* a type or a token? We have decided to treat information objects as types and their "copies" as tokens. Nonetheless, not every ontologist may share this view, and consider the relation between an artifact and its copy as being distinct from the relationship between a class and its members. This may seem to be a merely philosophical question, but one should bear in mind that different modeling options exist for information objects and other artifacts that may have a significant impact given the principle mentioned above.

Taxonomies are exclusively based on formal subsumption, i.e. subsumption ties or subclass relationships have to be rigid and context-independent. Time- or situation-dependent taxonomical structures cannot be considered in a (relatively) context-free reference ontology. E.g. when we have to interpret the sentence "Tiotropium bromide is a bronchodilator drug" ontologically, it would seem that a certain amount of tiotropium bromide is used as a bronchodilator: to be a bronchodilator is an extrinsic, not an intrinsic feature of tiotropium bromide. Extrinsic features are roles. Hence we chose to state (in the COPD ontology) *that Tiotropium Bromide* is not a subclass of *Bronchodilator* (and hence *Drug*), and that the latter is not a subclass of *Chemical Substance*. Instead we have classified *Drug* and *Bronchodilator* as *Role*, and have stipulated that *Tiotropium Bromide* (as an *Anticholinergic*) has the role *Bronchodilator*.

The immediate subclasses of a given class should ideally be exhaustive, i.e. their union should cover exactly the whole of the superclass. This principle only half-way cherished in the ACGT Master Ontology was impossible to maintain in the development of the CHRONIOUS ontologies, which are meant to be extendable. Considerations of relevance, as reflected in the sources, also led to the consequence that the children of a node are not exhaustive: this is trivially true e.g. for *Cardinal Heart Part* (MLOCC), *Cardinal Lung Part* (COPD) and *Cardinal Kidney Part* (CKD). Instead of this idealized completeness requirement we have adopted a more pragmatic completeness requirement, according to which the most relevant subclasses should be present.

Multiple inheritance of primitive classes should be avoided. To return to the bronchodilator example, it was out of the question to subsume *Tiotropium Bromide* both under *Bronchodilator* and *Chemical Substance* (indirectly). Instead of multiple inheritance, one should privilege intrinsic or formal subsumption on the one hand over role attribution on the other (as in the example above).

Primitive sibling classes should be disjoint.

UnknownX as well as other catch-all classes for remaining cases should be avoided. Indeed, it is tempting to render a non-exhaustive subdivision complete by adding a class that covers the rest of the extension of the superclass (e.g. something like *Other Cardinal Lung Part*). However, such a class does not cut at a joint of reality, i.e. it does not represent an ontological, but an epistemological distinction. Indeed, it covers all classes not yet known: yet it is irrelevant for reality whether something is known or not. One needs to distinguish UnknownX-classes from classes like *Undifferentiated Gender* which reflect real borderline cases.

3.2 Specific Design Principles

The following design principles correspond to actual ontological choices based on an intended, philosophical interpretation of Basic Formal Ontology. Note that the intuitions behind the following principles may diverge from those proposed in expositions of BFO such as [4].

Occurrents do not participate in other occurrents. There are no events of events: events are changes and as such do not change. Qualitative change of events or processes is just having different temporal parts at different times. There is no change in the sub-ontology of processes. A process of heart beating does not accelerate, but a heart rate, which is the quality of an organism resulting of a heart beating, does.

Realizable entities (dispositions, functions, roles) do not participate in occurents. Indeed, realizable entities are expressed or realized by processes; as mere potentialities they do not change per se. This means that the renal filtration function does not change, but is realized by the process of renal filtration which has a glomerular filtration rate (GFR) as an outcome. The GFR may increase or decrease, the renal filtratio function not. In general, everyday talk about increase or decrease of functions translates into the increase or decrease of qualities that result or are affected by processes realizing dispositions, functions or roles.

Realizable entities only characterize independent continuants, except roles. The latter exception is necessary if we want to exploit role assignment in order to

avoid non-rigid subsumption or multiple inheritance. E.g. symptoms are roles of organismal processes; they cannot be intrinsic features, since an organismal process need not be a symptom of a specific disease irrespective of other co-occurrent events.

4 Conclusion

Our aim was to provide a general overview regarding the methodology and the design principles leading the implementation of the CHRONIOUS ontologies, as well as an idea of their overall structure. We stressed the importance of the methodological criteria of realism and modularity and we have sketched the philosophical positions guiding the construction of these ontologies. We have shown that the CHRONIOUS ontologies represent a carefully thought-through and expert-validated contribution to (a) the general ontology of (chronic) diseases and pathological conditions as well as to (b) the ontology of specific chronic illnesses like COPD and CKD. Ongoing work on MLOCC includes its integration with the the Ontology for General Medical Science (OGMS) as well as the Information Artifact Ontology[9].

Acknowledgement: Research leading up to the present article has been supported by the ICT-2007-1-216461 grant within the Seventh Framework Programme of the EU, as well as by a post-doc grant from the National Research Fund, Luxembourg (cofunded under the Marie Curie Actions of the European Commission [FP7-COFUND]), and has been carried out under subcontract to the Fraunhofer Institute for Biomedical Engineering, St. Ingbert (Germany).

References

1. Farré F, Papadopoulos A, Munaro G, Rosso R. An Open, Ubiquitous and Adaptive Chronic Disease Management Platform for Chronic Respiratory and Renal Diseases (CHRONIOUS). In: Conley, EC, Doarn, C, Hajjam-El-Hassani, A, editors. Proceedings of the International Conference on eHealth, Telemedicine, and Social Medicine; 2009 Feb 1-7; Cancun, Mexico. New York: IEEE Press; 2009. p.184-9.
2. Cocos, C. Design Principles of the ACGT Master Ontology: Examples and Discussion. Saarbrücken: Institute for Formal Ontology and Medical Information Science; 2008 [cited 2011 Jun 5]. Available from: http://www.ifomis.org/wiki/ACGT_Master_Ontology_(MO)
3. Brochhausen M, Spear AD, Cocos C, Weiler G, Martin L, Anguita A, et al. The ACGT Master Ontology and Its Applications - Towards an Ontology-Driven Cancer Research and Management System. J Biomed Inform. 2011; 44: 8–25.
4. Spear, A. Ontology for the Twenty First Century: An Introduction with Recommendations. Saarbrücken: Institute for Formal Ontology and Medical Information Science; 2006 [cited 2011 Jun 5]. Available from: http:// www.ifomis.org/bfo/documents/manual.pdf

[9] http://www.obofoundry.org/cgi-bin/detail.cgi?id=information_artifact

5. Rosse C, Mejino JVL. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J Biomed Inform. 2003; 36: 478-500.
6. Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, et al. Relations in Biomedical Ontologies. Genome Biol. 2006; 6; R46.
7. Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L. Sweetening Ontologies with DOLCE. In: Gomez-Perez A, Benjamins VR, editors. Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Proceedings of the 13th International Conference, EKAW 2002; 2002 Oct 1-4; Siguenza, Spain; Heidelberg: Springer; 2003. P. 166-81
8. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251-5
9. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Bethesda (MD): Global Initiative for Chronic Obstructive Lung Disease (GOLD); 2009.
10. National Kidney Foundation Kidney Disease Outcomes Quality Initiative (NKF KDOQI). KDOQI Clinical Practice Guidelines on Hypertension and Antihypertensive Agents in Chronic Kidney Disease. New York (NY): National Kidney Foundation; 2004 [cited 2011 Jun 5]. Available from: http://www.kidney.org/professionals/KDOQI/guidelines_bp/index.htm.
11. National Kidney Foundation Kidney Disease Outcomes Quality Initiative (NKF KDOQI). KDOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification. New York (NY): National Kidney Foundation; 2002 [cited 2011 Jun 5]. Available from: http://www.kidney.org/professionals/KDOQI/guidelines_ckd/toc.htm
12. National Kidney Foundation Kidney Disease Outcomes Quality Initiative (NKF KDOQI). KDOQI Clinical Practice Guidelines for Bone Metabolism and Disease in Chronic Kidney Disease. New York (NY): National Kidney Foundation; 2003 [cited 2011 Jun 5]. Available from: http://www.kidney.org/professionals/KDOQI/guidelines_bone/index.htm
13. National Kidney Foundation Kidney Disease Outcomes Quality Initiative (NKF KDOQI). KDOQI Clinical Practice Guidelines for Managing Dyslipidemias in Chronic Kidney Disease. New York (NY): National Kidney Foundation; 2003 [cited 2011 Jun 5]. Available from: http://www.kidney.org/professionals/KDOQI/guidelines_lipids/toc.htm
14. National Kidney Foundation Kidney Disease Outcomes Quality Initiative (NKF KDOQI). KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for Diabetes and Chronic Kidney Disease. New York (NY): National Kidney Foundation; 2007 [cited 2011 Jun 5]. Available from: http://www.kidney.org/professionals/kdoqi/guidelines_anemia/references.htm
15. Gomez-Perez A, Fernandez-Lopez M, Corcho O. Ontological Engineering. London: Springer; 2004.