INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS XXX (2008) XXX–XXX

journal homepage: www.intl.elsevierhealth.com/journals/ijmi

SNOMED reaching its adolescence: Ontologists' and logicians' health check

Stefan Schulz[a,*], *Boontawee Suntisrivaraporn*[b], *Franz Baader*[b], *Martin Boeker*[a]

[a] *Freiburg University Medical Center, Institute of Medical Biometry and Medical Informatics, Stefan-Meier-Str. 26, 79104 Freiburg, Germany*

[b] *Dresden University of Technology, Faculty of Computer Science, Dresden, Germany*

ARTICLE INFO

Article history:
Received 4 March 2008
Accepted 6 June 2008

Keywords:
SNOMED CT
Ontologies
Description Logic

ABSTRACT

After a critical review of the present architecture of SNOMED CT, addressing both logical and ontological issues, we present a roadmap toward an overall improvement and recommend the following actions: SNOMED CT's ontology, dictionary, and information model components should be kept separate. SNOMED CT's upper level should be re-arranged according to a standard upper level ontology. SNOMED CT concepts should be assigned to the four disjoint groups: classes, instances, relations, and meta-classes. SNOMED CT's binary relations should be reduced to a set of canonical ones, following existing recommendations. Taxonomies should be cleansed and split into disjoint partitions. The number of full definitions should be increased. Finally, new approaches are proposed for modeling part-whole hierarchies, as well as the integration of qualifier relations into a unified framework. All proposed modifications can be expressed by the computationally tractable description logic EL^{++}.

1. Introduction

SNOMED CT®, the Systematized Nomenclature of Medicine - Clinical Terms, a comprehensive, multilingual terminology for the electronic health record (EHR), is currently based on a taxonomy of 311,000 concepts [1], which are linked to terms and multi-lingual synonyms. Their meaning derives both from their position in the hierarchy and from formal axioms that connect concepts across the hierarchies and supply necessary and (partly) sufficient criteria. SNOMED CT is the result of a joint development between the British National Health Service (NHS) and the College of American Pathologists (CAP) and has reached a global dimension as in 2007 the newly-formed International Health Terminology Standards Development Organisation (IHTSDO) has acquired its property. This internationalization of SNOMED CT offers a unique opportunity to bring together the following tendencies:

- the urgent need for a global standardized terminology for medicine and life sciences, suitable to cope with an immense flood of clinical and scientific information;
- an impressive legacy of systematized biomedical terminology;
- efforts toward an ontological foundation of the basic kinds of entities in the biomedical domain as an important endeavor of the emerging discipline of "Applied Ontology";
- the increasing availability of logic-based reasoning artifacts suited for large ontologies.

For a better understanding of the current status of SNOMED CT we need to take into account over 40 years in which SNOMED has evolved from the pathology-specific SNOP

* *Corresponding author*. Tel.: +49 761 2049089; fax: +49 761 2036711.
E-mail address: stschulz@uni-freiburg.de (S. Schulz).

1386-5056/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2008.06.004

nomenclature into the comprehensive logic-based health care terminology SNOMED CT [2]. This long-lasting development period has granted SNOMED continuous growth, but has also originated and consolidated a number of innate problems. Concerned about the conditions under which SNOMED CT is now reaching its adolescence, we recommend an in-depth health check. Advice will be required from specialists of Ontology and Logic. A careful follow-up of their counseling will be crucial for making SNOMED CT fit for the next decades.

2. Methodology

2.1. *The ontologists' approach*

The evolution of UMLS [3] and OBO [4] bear witness to the importance the (bio)medical informatics community has conferred to the desideratum of semantic interoperability [5–7]. Originally this matter was supposed to be addressed mainly by what had been termed "terminology systems" [8,9]. More recently, however, we have seen steady growth in usage of the term 'ontology'. Due to the lack of a clear notion of what an ontology really is [10,11], the difference between terminology on one side and ontology on the other has become rather nebulous. In the context of this paper we will subscribe to a principled distinction between the two terms, along the following lines:

According to [12], a **terminology** is a set of terms representing the system of concepts of a particular subject field. Terminologies relate the senses or meanings of linguistic entities with concepts. Concepts are conceived as the common meaning of (quasi-)synonymous terms. In medical informatics, this language-centered view is exemplified by the UMLS [3].

In contrast to terminology, **Ontology** (discipline), is the *study of what there is* [13]. In our understanding, (formal) **ontologies** [14] are theories that give precise, logic-based formulations of the types of entities in reality, of their properties and of the relations between them. They strive for describing reality independently of human language, as much as possible. Their constituent nodes are (entity) types rather than language-related concepts. Types (often also referred to as 'categories', 'kinds' or 'universals') are well suited to hierarchically order the particular entities (e.g., anatomical objects, amounts of substances, lesions, surgical procedures) which exist on the side of reality. The existence of certain entity types and the basic structure of ontological principles are subject to major philosophical disputes. However, at any given stage in the development of science, there is a consensus core of scientific understanding of reality, and in our view it is this which should serve as starting point in developing science-based ontologies. Examples of assertions belonging to this consensus core are: primates are vertebrates; cells contain cytoplasm, etc. Note that, besides observation-based descriptive accounts of nature, entity types are also often created as the result of a prescriptive definition process: *Appendectomy* is defined as *Surgical removal* of *Appendix*, and *Hepatitis* as *Inflammation* of *Liver tissue*. As types in an ontology extend to classes of entities (individuals that instantiate them) the term "class" is often preferred over the term "type" in practice. The foundational role of the instantiation relation also lays the grounds for a central tenet for ontology architecture, *viz*. the taxonomic principle: a type *S* is a subtype of a type *T* if and only if all instances of *S* are also instances of *T*.

2.2. *The logicians' approach*

When dealing with ontologies or terminological systems, logicians are primarily interested in formally founded and sufficiently expressive description languages. The use of logic forces the knowledge engineer to be as explicit and unambiguous as possible. Therefore, ontology languages and editing tools must be intuitive and easy to handle. As logic-based representations form the building blocks of formal systems, implicit expressions can be inferred from those explicitly given. Such expressions are either axioms that are postulated to be self-evident, or theorems that are derived from axioms. The ability to derive new propositions from existing ones allows following and assessing the consequences of the asserted axioms. Truth-maintaining algorithms cannot only detect undesired consequences, but also further give an explanation for them and even suggest for a revision to get rid of them.[1]

3. SNOMED CT's problem list

In this section we will apply the aforementioned principles to analyze SNOMED CT's current status, from ontological and logical perspectives, in the style of a clinical problem list. Subsequently we will suggest and discuss solutions to the problems identified.

3.1. *Multiple identity disorder*

SNOMED and its predecessor SNOP had been devised as nomenclatures, providing lexical atoms and simple rules for the assembly of terminological molecules in a controlled language.[2] In the nineties, SNOMED INTERNATIONAL had then introduced semantic aspects in terms of (i) classifying the terms by meaning (e.g. belonging to the axes "Topography", "Morphology", "Function", etc.), (ii) placing them into a hierarchy, and (iii) enriching them with synonyms and alternate expressions as pointed out by Rothwell [15]. Ingenerf and Giere [16] described SNOMED as an instance of concept system, based on the ISO 1087 definition of a "concept as a unit of thought". Logic-based concept definitions were first introduced into the tentative SNOMED RT [17] and then consolidated in SNOMED CT, where also ontological principles have been increasingly contemplated [18]. As a result, the current SNOMED CT constitutes a blend of diverging and even contradicting architectural principles. The nomenclature legacy is still visible by the fact that SNOMED concepts

[1] An example is the contradiction in the conjunction of the four assertions "mammals are multicellular organisms", "humans are mammals", "a zygote is a single cell", and "a human zygote is a human".

[2] According to ISO 1087, nomenclatures are "systems of terms which are elaborated according to pre-established naming rules".

Please cite this article in press as: S. Schulz, et al., SNOMED reaching its adolescence: Ontologists' and logicians' health check, Int. J. Med. Inform. (2008), doi:10.1016/j.ijmedinf.2008.06.004

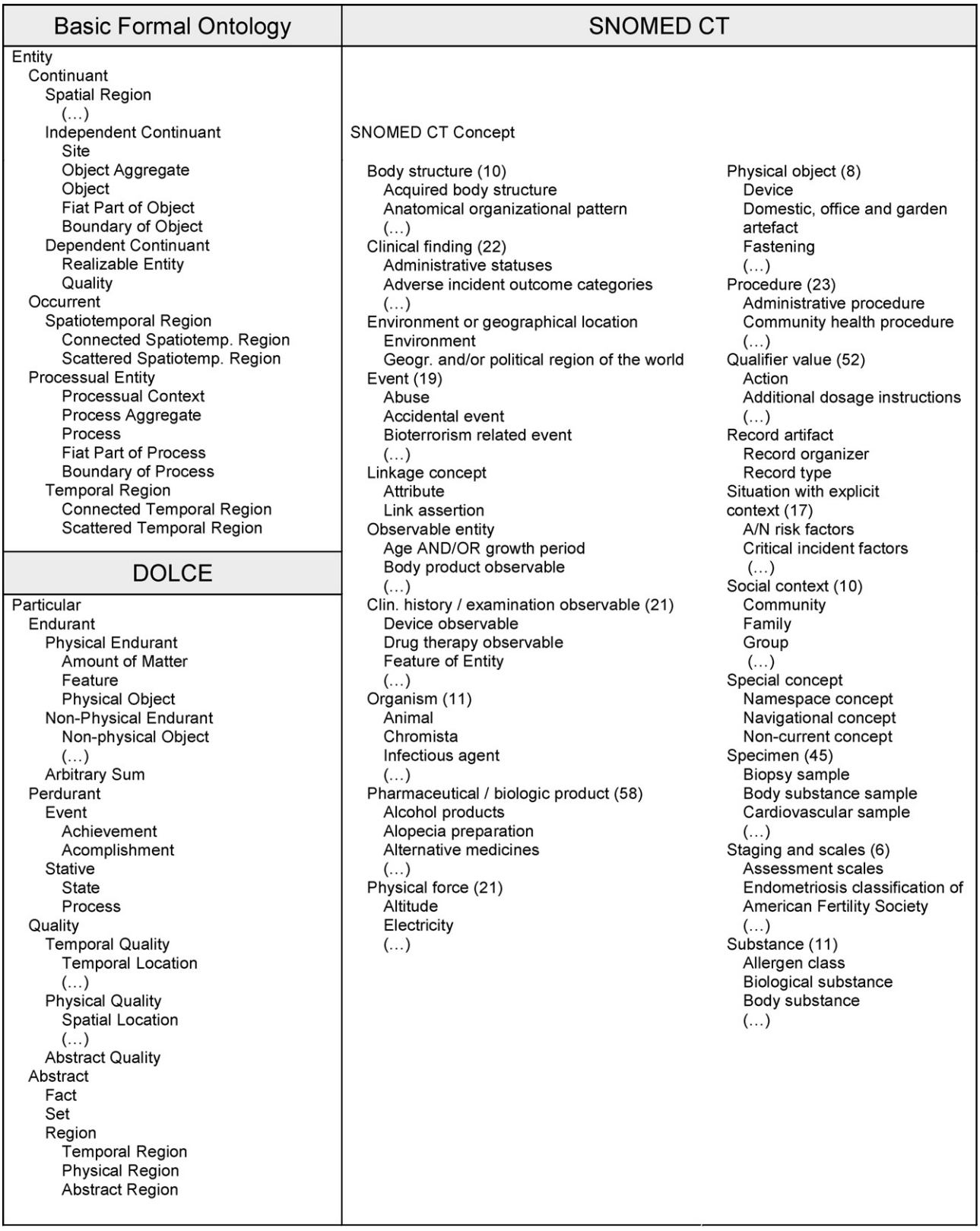

Fig. 1 – First three taxonomic levels of BFO, DOLCE, and SNOMED CT. In the latter, only the first two or three entries (in alphabetical order) are displayed. The hierarchies are taxonomic subsumption hierarchies (is-a). At each level sibling classes are assumed to be disjoint. The numbers in parentheses indicate the number of taxonomic children.

Please cite this article in press as: S. Schulz, et al., SNOMED reaching its adolescence: Ontologists' and logicians' health check, Int. J. Med. Inform. (2008), doi:10.1016/j.ijmedinf.2008.06.004

are term-centered, lacking free text descriptions (such as, e.g., in MeSH). SNOMED INTERNATIONAL's axes still characterize the idiosyncratic structure of SNOMED CT's upper level (Fig. 1). The description logic definition approach is inherited from pre-ontology SNOMED RT. Finally, the lack of ontological rigor and an increasingly application-driven development has motivated ad-hoc design decisions that have obscured two quite different concerns of the representation of medical knowledge [19]: the representation of the reality of the patient and all surrounding entities in the health care process on the one hand, and the representation of the health care record, of physicians' knowledge and beliefs on the other hand.

3.2. *Upper level dystrophy*

Domain ontologies should be grounded upon an upper level that introduces fundamental distinctions, such as between material objects, spaces, qualities, functions, events etc. Most upper level ontologies coincide in a topmost distinction between endurants (aka continuants) and perdurants (aka occurrents). *Endurants* are those entities that exist in their entirety at any point in time, such as physical objects and spaces. *Perdurants*, in contrast, are never completely present at one single moment, such as events and processes. Functions, dysfunctions, qualities, and states are frequently considered a special kind of endurants, *viz.* non-material ones. Whereas BFO [20] and DOLCE [21] provide upper level classes that were devised by ontologists during years and provide extensive descriptions on the Web [22,23], the SNOMED upper level architecture lacks any principled approach besides some reminiscence of the axes known from pre–RT SNOMED. Fig. 1 provides insight into the different hierarchies. The unsystematic design becomes especially obvious when comparing siblings at the 3rd level, e.g. juxtaposing *Device* with *Domestic, office, and garden artifact*, or *Allergen* with *Biological substance*.

3.3. *Concept borderline disorder*

The Biomedical Informatics community has used the term "concept" to an extent that "concept systems" had become a synonym of any ontology or terminology artifact. In the last years there has been a strive toward more terminological clarity which has challenged this tradition, arguing that the term "concept" is too ambiguous and obscures the representation of real-world entities by ontologies [24,25]. When talking about SNOMED CT and in agreement with its documentation, we here use the word "concept" as a synonym for the nodes in the SNOMED CT hierarchy, regardless of their ontological or epistemological significance. In contradistinction, when taking an ontological standpoint in our argumentation, we use the terms "classes", "relations", "meta-classes" and "individuals". It is these four categories in which the set of SNOMED CT concepts, under ontology scrutiny, can be partitioned. Examples of meta-classes (i.e., classes that classify concepts) are *SNOMED CT concept*, *Navigational concept*, or *Allergen class*, as well as concepts with plural identifiers such as *Additional values*, *Ketone bodies*, etc. On the other hand there are concepts that denote individuals such as geographic entities. SNOMED CT's conflation of the level of individuals, classes, and meta-classes leads to quite strange conclusions, at least if we interpret taxonomic subsumption of concepts as pointed out above, *viz.* as the superclasses' condition of including all instances of any of its subclasses. Treating SNOMED concepts as classes (which would be the standard assumption for any description logic based account) we are obliged to subscribe that *London* is a class (and thus can be instantiated), that all instances of *London* are also instance of *Additional value*; that my particular instance of *Adverse reaction to premedication* is an instance of *Navigational concept*; and that my particular instance of *Heartburn* is a SNOMED CT *concept*.

3.4. *Relation idiosyncrasy*

Relations should be consistent and unambiguous in order to assist ontology developers and users in avoiding errors, a principle that has driven the development of the OBO relation ontology [26]. In SNOMED CT relations are a special kind of concepts (concept model attributes). They are not formally defined, and by their idiosyncratic names they can rarely be mapped to any other relation ontology. Some relations such as *Finding Site/Procedure Site* or *Specimen Substance* obviously specialize standard relations (e.g. *has-location*, *has-part*) which, on their part, are missing in SNOMED CT. Other ones are rather fuzzy such as *Subject Relationship Context*. The problem here is that the more relations exist, the less one can expect agreement among users. The necessity of nesting relational attributes in more complex concept definitions gave rise to a special relation, named *Role group*. This relation (which can only implicitly be asserted) is ontologically rather obscure but was found to correspond to *has-part* between perdurants in most cases [27]. However, the reason for using the role group relation remains shady in numerous cases such as the definition of the SNOMED CT concept *Bronchial Suction*, which is defined as *Removal* AND *rolegroup (Method Suction-Action AND Procedure-Site Bronchial-Structure)*.

3.5. *Taxonomic dystrophy*

Subclass hierarchies (taxonomies) should obey certain principles such as described in [28]. Accordingly, it must be criticized that numerous non-terminal SNOMED CT classes have one subclass only and that the number of classes with multiple parents is higher than necessary. Another kind of taxonomic dystrophy is the so-called "*is-a overloading*", i.e., the use of taxonomic subsumption in order to express roles rather than generic properties, as already analyzed by [29], using the OntoClean methodology [30]. So do we find *Bacterium* is a subconcept of *Infectious Agent*, although not every individual bacterium is an infective agent. Finally, as pointed out by [31], epistemological aspects[3] should be kept apart. In SNOMED CT, there are numerous cases for "epistemology intrusion" such as *Newly diagnosed diabetes* or *Neoplasm, uncertain whether benign or malignant*. The point here is that the diabetes as such is in no way of a different type by the fact that is has recently been

[3] Criteria that describe a human observer's knowledge about an entity but that are irrelevant for describing the inherent nature of that entity.

Please cite this article in press as: S. Schulz, et al., SNOMED reaching its adolescence: Ontologists' and logicians' health check, Int. J. Med. Inform. (2008), doi:10.1016/j.ijmedinf.2008.06.004

diagnosed. Similarly, the neoplasm has a malignity status, regardless of the physician's knowledge.

3.6. SEP implants

So-called SEP triplets [32] are modeling artifacts which expand taxonomies by reified relations. For instance, $Femur_P$ is defined as the class of everything that is part of a *Femur*. $Femur_S$ is introduced as a common taxonomic parent of $Femur_P$ and *Femur*. Together, *Femur*, $Femur_P$ and $Femur_S$ form an SEP triplet. Such structures implicitly express part-whole relationships. The main reason why SNOMED CT uses such structures is to enable the propagation of attributes along aggregation (part-whole) hierarchies in a parsimonious way. For example, *Femur fracture* is defined as a *Fracture* located at some $Femur_S$. Since $Femur_S$ subsumes *Neck of femur*, *Fracture of the neck of femur* is classified as a *Femur fracture*. SNOMED CT is replete of such reified classes, yet in an unsystematic and incomplete way. They are undefined and the terms assigned to them are often misleading[4]. More precisely, we can consider taxonomic A_S–B_S links as kind of prostheses for missing *part-of* relations in the anatomy branch. However, they serve the needs of attribute propagation which is seen as an important asset in medical terminological reasoning.

3.7. Description asthenia

As advocated by [33,34] for biomedical ontologies, taxonomies should be founded upon the Aristotelian principle of *genus* (the common properties of members in the subsuming class) and *differentiae* (the properties that distinguish each instance of the subsumed class from the genus). According to [28], half of SNOMED CT concepts are primitive ones, i.e. they have no *differentiae* specified. Besides cases in which Aristotelian definitions are difficult or impossible (e.g. in anatomy), numerous other ones are missing without obvious reason. One of thousands such examples is the concept *Cessation of sedation (procedure)*, which is not related to the concept *Sedation*. If new complex SNOMED CT concepts are added without being defined on the basis of atomic concepts, SNOMED CT will increasingly boil down to a system of controlled identifiers with no semantic value.

3.8. The qualifier syndrome

SNOMED CT qualifiers, such as *Laterality*, *Severity*, *Onset* and *Course* are relations used for constraining post-coordination for a further refinement [35]. For example, *Asthma* allows 12 different values for the qualifier *Course* and six for the qualifier *Severity*. Only a small subset of all SNOMED CT relations are used as qualifiers, and it seems that these relations are never used for different purposes. On the other hand, those relations which are used in definitions never feature as qualifiers. So we have the strange situation in which the qualifier *Severity* is allowed for the class *Asthma*, but is not used for *defining* its subclass *Severe asthma*. For the latter one, *severity* is allowed with its whole range of values, so that the formation of a post-coordinated concept *Severe asthma* with the attribute *Severity.Mild* would be possible. There are innumerous examples that show that the value ranges of the qualifiers are not well adapted to the characteristics of the class they belong to.

4. SNOMED CT's treatment plan

The assessment of SNOMED CT's health status by both specialties has revealed the following trade-off: ontologists strive for a comprehensive account of reality, and they would like to use the whole inventory of logics for describing it with the precision and expressiveness they deem adequate. In contrast, logicians point at the computational properties of full logics, which are prohibitive for any large scale implementation. A viable compromise is given by description logics (DLs), a family of decidable fragments of the first-order logic which have a clean and intuitive syntax [36]. DLs come in various flavors, ranging from lightweight to highly expressive ones. The trade-off between expressivity of the logic and computability (and thus, scalability) of its reasoning has to be made in order to properly address the ontology application. On the one hand, overly inexpressive DL may lead to under-specifications that imply unintended models of the ontology one should be aware of. This however is unavoidable. On the other hand, highly expensive reasoning makes it infeasible from practical viewpoints, thus the whole logical machinery for a large ontology is not desirable. We here sketch the specification of a DL which, under scrutiny, appears to be well suited to support most modeling and reasoning requirements of SNOMED CT. In Table 1 properties of the *computationally tractable* Description Logic EL^{++} [37] are given. It has been shown both theoretically [37] and empirically [38] that the *EL* DL family is computationally cheap and adequate in terms of ontological expressive power. Based on this logical framework we now make the following recommendations.

4.1. Identity finding support

We recommend to advance in the separation of the following three components in SNOMED CT: (i) *the SNOMED CT ontology*. It represents types of language-independent domain entities together with foundational relations and describes their inherent properties using the Description Logic *EL*. Where logic is not sufficient, precise, unambiguous English glossary entries, images, and references to authoritative sources should complete the picture. It must be stated in a clear and unambiguous way, which is exactly the domain SNOMED CT concepts extend to[5]. (ii) *The SNOMED CT dictionary*. It represents all the terms that are used in clinical discourse and covers several languages. Each term is mapped to one or more concepts in the SNOMED CT ontology. (iii) *The SNOMED*

[4] For instance, the term "kidney" is both allowed for the concepts *Kidney* and $Kidney_S$. The latter, however, subsumes *Entire renal sinus*, an object that can never be referred to by the name "kidney".

[5] As long as there is still a controversy of whether a SNOMED CT concept like Chest Pain is instantiated by the pain in my chest or by the entry in my medical record, SNOMED CT's semantic foundations remain unstable and SNOMED CT based machine inferences will be unreliable.

Please cite this article in press as: S. Schulz, et al., SNOMED reaching its adolescence: Ontologists' and logicians' health check, Int. J. Med. Inform. (2008), doi:10.1016/j.ijmedinf.2008.06.004

Table 1 – Characteristics of description logic EL^{++}

Existential quantification	$\exists FindingSite.AppendixStructure$
Conjunction	$\exists AssociatedMorphology.Inflammation \sqcap \exists FindingSite.AppendixStructure$
Necessary condition	$Acute\ Appendicitis \sqsubseteq Appendicitis$
Necessary and sufficient conditions	$Appendicitis \equiv \exists AssociatedMorphology.Inflammation \sqcap \exists FindingSite.AppendixStructure$
General inclusions	$Ulcer \sqcap \exists has\text{-}location.Stomach \sqsubseteq Ulcer \sqcap \exists has\text{-}location.(Lining \sqcap \exists part\text{-}of.Stomach)$
Class disjointness	$BodyPart \sqcap Organism \sqsubseteq \bot$
Domain restrictions	$\exists has\text{-}location.\top \sqsubseteq Disease$
Range restrictions	$\top \sqsubseteq \forall has\text{-}location.BodyPart$
Role hierarchy	$Proper\text{-}part\text{-}of \sqsubseteq part\text{-}of$
Role reflexivity	$\varepsilon \sqsubseteq part\text{-}of$
Role transitivity	$part\text{-}of \circ part\text{-}of \sqsubseteq part\text{-}of$
Right identity on roles	$has\text{-}location \circ part\text{-}of \sqsubseteq has\text{-}location$
Concrete domains	$Minor \equiv Person \sqcap <_{18year}(age)$
Nominals	$Kangaroo \sqsubseteq \exists has\text{-}origin.\{Australia\}$
Class assertions	$London \in GeographicLocation$
Role assertions	$(London, England) \in has\text{-}location$

CT information model. It provides the machinery to describe the context of clinical propositions, allowing statements on uncertainty, absence, etc. Information model concepts and relations must be kept separated from the SNOMED ontology because the semantics of description logics is different from the semantics that is necessary for the representation of complex propositions that would require modality, uncertainty, and possible world reasoning. Ignoring this, unintended models would arise[6].

4.2. *Upper level reconstruction*

The ontologists' recommendation is to refer as much as possible to a commonly accepted upper ontology. Note that with regard to upper level organization there are still several controversial points, first of all the ontological account for disease (delimited from *Courses of disease*, but also from *Sign* and *Symptom*). This is currently subject to ontological inquiry, and SNOMED CT could be a good testbed for this. A candidate for a SNOMED CT upper level could be the biomedical toplevel ontology BioTop [41], for which alignment studies are currently under way.

4.3. *Isolation of meta-classes and individuals*

SNOMED CT's meta-class aspects cannot be represented by any description logics. This is not really a problem because we see the necessity of meta-classes more as a kind of housekeeping feature for which annotation functions such as in Protegé and OWL can be used and in which additional *RDF* attributes can be introduced. Individuals (such as *Europe*, *Greater London*, *Binge eating scale*) should be regarded as individuals of the corresponding classes, i.e. instances of *Geographic Location* or *Staging and Scales*. This requires the addition of the instantiation relation $\in$. EL^{++} supports nominals and ABoxes. Both are closely related and are helpful when information about individuals is to be included. This allows axioms such as $MexicanIndian \sqsubseteq \exists has\text{-}origin.\{Mexico\}$. An ABox comprises assertions about individuals by means of *instance-of* (class assertions) or *related-by* (role assertions) relations, e.g., $London \in GeographicLocation$ and $(London, England) \in has\text{-}location$, respectively.[7]

4.4. *Reconstruction of relations*

SNOMED CT relations should be reduced to a minimum of canonical ones, starting with the OBO relations [5]. Relationship groups should be substituted by the corresponding relation, most likely *has-part*. One should also give a clear account of the algebraic features of each relation in terms of reflexivity, symmetry, and transitivity. Furthermore, relations should be further constrained in terms of domain ($\exists r.\top \sqsubseteq D$) and range ($\top \sqsubseteq \forall r.R$)[8] restrictions.

4.5. *Cleansing of taxonomies*

All classes should have at least one sibling, otherwise they should be merged with their super-class. Multiple taxonomies should be reduced to a minimum. Wherever an *is-a* link is

[6] e.g., the DL representation of the SNOMED CT concept *BiopsyPlanned* entails the existence of a biopsy ($BiopsyPlanned \sqsubseteq \exists rolegroup.(\exists AssociatedProcedure.Biopsy \sqcap \ldots)$).

[7] To keep it simpler without the need of introducing individuals, we could limit ourselves to the reference to geographical entities in terms of location classes. In this case, reifications of the type $A_L \equiv \exists\ has\text{-}location.A$ with $A \sqsubseteq GeographicLocation$ may be discussed for the sake of parsimony. Then, for instance, the fact that London is located in England could be indirectly expressed by $London_L \sqsubseteq England_L$.

[8] A controlled use of the universal quantifier $\forall$ in these cases has no negative impact on the computational properties.

Please cite this article in press as: S. Schulz, et al., SNOMED reaching its adolescence: Ontologists' and logicians' health check, Int. J. Med. Inform. (2008), doi:10.1016/j.ijmedinf.2008.06.004

inferable from defined classes it should be omitted for the sake of brevity and clarity. For instance, *Acute type B viral hepatitis* is fully defined as *Type B viral hepatitis* which is *acute*. By the definitions of *Type B viral hepatitis*, *Hepatitis*, and *Acute hepatitis* a DL classifier can infer the subsumption between *Acute type B viral hepatitis* and *Acute hepatitis*, so that this *is-a* link does not need to be included. A clean taxonomy should also contain as much as possible disjoint partitions. This generally requires negation statements of the form $A \sqsubseteq \neg B$. Such a restricted negation statement is in fact equivalent to a disjointness axiom of the form $A \sqcap B \sqsubseteq \bot$ which is available in *EL*.

4.6. *SEP explant and substitution*

The extra nodes should be fully defined. Although irrelevant under ontological scrutiny, they may be preserved for reasons of backward compatibility. A full definition of S and P nodes, however, requires distinguishing between *proper-part-of* which is transitive and irreflexive, and the broader relation *part-of* which is transitive and reflexive. So we can fully define $A_P \equiv \exists$ *proper-part-of*.A, together with $A_S \equiv \exists$ *part-of*.A [39]. However, our language does not allow to enforce irreflexivity of a relation, so that the following GCI might be added where required: $\exists$ *proper-part-of*.$A \sqcap A \sqsubseteq \bot$. With the right identity rule *has-location* $\circ$ *part-of* $\sqsubseteq$ *has-location* we then get the right inference in the femur example. False inferences, such as the classification of *Amputation of the foot* as *Amputation of the lower limb*, can be prevented by introducing a subrelation of *has-location*, *viz. has-exact-location* for which the right identity rule does not apply.

4.7. *Revitalization of full definitions*

We suggest the revision of primitive SNOMED CT classes, especially the elimination of misspecifications that are obviously the reasons that SNOMED CT classes, which could be fully defined, are still kept as primitive ones. Wherever possible, full definitions should be introduced. The introduction of full definitions generally brings to light hidden misspecifications as soon as the ontology is classified. The use of a terminological classifier is therefore of utmost heuristic importance in the process of building and maintaining SNOMED CT. This requires, however, that SNOMED CT moves to the DL format as the primary format in which all editing is performed.

4.8. *Qualifier transplant*

The realm of qualifiers which is kept somewhat apart from the rest of SNOMED CT sheds light on an intricate problem which complicates the move from a "closed world" frame-like perspective toward an "open world" description logics, the latter having the consequence that once there are relation types and classes, any relation may be asserted between any individuals unless this is explicitly precluded. SNOMED CT's approach of providing qualifiers with well-defined value restrictions for controlling the building of post-coordinated classes in description logics extends the capabilities of the logics we use. There are in principle two different ways we can deal with this problem. Firstly, we can handle the constraints as provided by the qualifiers outside description logics, similar to GALEN's sanctioning approach [40]. The alternative is to resort to a more expressive DL dialect, at the price of performance of the implementations. As a possible way out of this dilemma we suggest the following. On the one hand, we maintain the *EL* specification for SNOMED CT class definitions, but on the other hand we add an additional layer using DL value constraints. This second layer would then be invisible for the DL reasoner, but it can be used as a resource for those applications in which this information is needed, e.g. to constrain data entry by adaptive pick lists etc., which had been the main rationale for the SNOMED CT qualifiers. Similar to what we pointed out for meta-classes, this kind of housekeeping information can be realized with the help of annotation functions.

5. Conclusion

We have subjected the current version of SNOMED CT to an in-depth diagnostic examination under the aspects of ontology and logic. SNOMED CT's clinical picture exhibits mostly chronic problems most of which can be treated in a conservative but yet determined fashion. A crucial success factor is the consistent use of EL^{++} as therapeutic principle and the compliance with a formal ontology regime. Some of the problems require a more invasive intervention. We recommend the elaboration of a treatment plan, the definition of priorities, and the allocation of resources. Altogether, the cost of this treatment will be considerable, and it requires specialists both from the fields of ontology and description logic; nonetheless, it is a good investment for assuring the SNOMED CT's long-lasting fitness and its increasing ability to stand the upcoming challenges of medical documentation and standardization.

Summary points

What was known before this paper:

- SNOMED CT is a huge terminology providing codes and terms for all aspects of the electronic health record.
- SNOMED CT uses subsumption hierarchies and concept definitions based on description logics.
- SNOMED CT strives for progressively implementing formal ontology principles.

What is learned as a result:

- SNOMED CT, in its current state, still lacks logical and ontological soundness and accuracy, hence it necessitates major redesign efforts in numerous aspects.
- Ontological, terminology and information model issues in SNOMED CT should be kept separated and ontological standards (e.g. upper ontologies, ontology design principles) should be adhered to.
- Unintended models produced by SNOMED CT must be identified and corrected.
- SNOMED CT's semantic foundations (what do concepts extend to?) must be unambiguously fixed.

Please cite this article in press as: S. Schulz, et al., SNOMED reaching its adolescence: Ontologists' and logicians' health check, Int. J. Med. Inform. (2008), doi:10.1016/j.ijmedinf.2008.06.004

- Full definitions should be provided wherever possible, particularly when introducing new pre-coordinated concepts.
- The description logics EL^{++} proved to satisfactorily match SNOMED CT's requirements both in terms of scalability and expressiveness.
- Idiosyncratic constructs such as SNOMED CT's SEP triplets are no longer necessary when using EL^{++}.

Acknowledgments

This work was supported by the EU Network of Excellence Semantic Interoperability and Data Mining in Biomedicine (NoE 507505). Additionally, the first author was supported by a research fellowship (550830/05-7) from the Brazilian Research Council (CNPq/Brazil).

REFERENCES

[1] About SNOMED CT. International Health Terminology Standards Development Organisation. http://www.ihtsdo.org/snomed-ct/snomed-ct0/, last accessed July 17th, 2008.
[2] R. Cornet, N. de Keizer, Forty years of SNOMED: a literature review, BMC Med. Informat. Decis. Making, 2008, in press.
[3] Unified Medical Language System (UMLS), National Library of Medicine, Bethesda, MD, 2008.
[4] Open Biological Ontologies (OBO) Foundry, http://www.obofoundry.org, last accessed July 17th, 2008.
[5] A. Rossi-Mori, F. Consorti, Exploiting the terminological approach from CEN/TC251 and GALEN to support semantic interoperability of healthcare record systems, Int. J. Med. Inform. 48 (1-3) (1998) 111–124.
[6] J. Ingenerf, J. Reiner, B. Seik, Standardized terminological services enabling semantic interoperability between distributed and heterogeneous systems, Int. J. Med. Inform. 64 (2-3) (2001) 223–240.
[7] S. Garde, P. Knaup, E. Hovenga, S. Herd, Towards semantic interoperability for electronic health records, Methods Inf. Med. 46 (3) (2007) 332–343.
[8] N.F. De Keizer, A. Abu-Hanna, Zwetsloot-Schonk JH, Understanding terminological systems. I. Terminology and typology, Methods Inform. Med. 39 (2000) 16–21.
[9] R. Cornet, N.F. de Keizer, A. Abu-Hanna, A framework for characterizing terminological systems, Methods Inf. Med. 45 (3) (2006) 253–266.
[10] W. Kuśnierczyk, Nontological Engineering, in: FOIS 2006 – Proceedings of the 4th International Conference on Formal Ontology in Information Systems, 2006, pp. 39–50.
[11] N. Guarino, P. Giaretta, Towards very large knowledge bases: knowledge building and knowledge sharing, in: Ontologies and Knowledge Bases: Towards a Terminological Clarification, IOS Press, 1998, 25–32.
[12] International Organization for Standardization: ISO 1087-1: Terminology work – Vocabulary – Part 1: theory and applications, Geneva, Switzerland, 2000.
[13] O. Quine, On what there is, Rev. Metaphys. (1948).
[14] N. Guarino, Formal ontology in information systems, in: Proceedings of FOIS'98, Trento, Italy, 6–8 June 1998, Amsterdam, IOS Press, 1998, pp. 3–15.
[15] D.J. Rothwell, SNOMED-based knowledge representation, Methods Inf. Med. 34 (1-2) (1995) 209–213.
[16] J. Ingenerf, W. Giere, Concept oriented standardization and statistics oriented classification: continuing the classification versus nomenclature controversy, Methods Inf. Med. 37 (4-5) (1998) 527–539.
[17] K.A. Spackman, Normal forms for description logic expression of clinical concepts in Snomed RT, in: AMIA 2001 – Proceedings of the Annual Symposium of the American Medical Informatics Association, 2001, pp. 627–631.
[18] K.A. Spackman, G. Reynoso, Examining SNOMED from the perspective of formal ontological principles: some preliminary analysis and observations. KR-MED 2004 – Proceedings of the 1st International Workshop on Formal Biomedical Knowledge Representation, http://www.CEUR-WS.org/Vol-102/, pp. 81–87.
[19] S. Schulz, H. Stenzhorn, M. Boeker, R. Klar, B. Smith, Clinical Ontologies Interfacing the Real World 2007, in: Third International Conference on Semantic Technologies (i-semantics 2007), Graz, Austria, September 2007.
[20] P. Grenon, B. Smith, L. Goldberg, Biodynamic ontology: applying BFO in the biomedical domain, in: D. Pisanelli (Ed.), Ontologies in Medicine, IOS Press, Amsterdam, Netherlands, 2004.
[21] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, L. Schneider, Sweetening Ontologies with DOLCE, in: Proceedings of EKAW-2002, Siguenza, Spain, 2002.
[22] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, WonderWeb Ontology Infrastructure for the Semantic Web, Deliverable D18, http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf, last accessed July 17th, 2008.
[23] Basic Formal Ontology, http://www.ifomis.org/bfo, last accessed July 17th, 2008.
[24] G.O. Klein, B. Smith, Concept Systems and Ontologies (updated 2006 October 6; cited 2006 December 4), available from http://www.ontology.buffalo.edu/concepts/, last accessed July 17th, 2008.
[25] B. Smith, Beyond concepts, or: ontology as reality representation FOIS 2004, in: Proceedings of the 3rd International Conference on Formal Ontology in Information Systems, 2004, pp. pp. 73–84.
[26] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, et al., Relations in biomedical ontologies, Genome Biol. 6 (5) (2005).
[27] S. Schulz, S. Hanser, U. Hahn, J. Rogers, The semantics of procedures and diseases in SNOMED CT, Methods Inf. Med. 45 (4) (2006) 354–358.
[28] O. Bodenreider, B. Smith, A. Kumar, A. Burgun, Investigating subsumption in DL-based terminologies: a case study in SNOMED CT, in: KR-MED 2004 – Proceedings of the 1st International Workshop on Formal Biomedical Knowledge Representation, 2004, pp. 12–20.
[29] K.A. Spackman, G. Reynoso, Examining SNOMED from the perspective of formal ontological principles, in: KR-MED 2004 – Proceedings of the 1st International Workshop on Formal Biomedical Knowledge Representation, 2004, pp. 72–80.
[30] N. Guarino, C.A. Welty, An overview of OntoClean, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, Berlin: Springer, 2004, pp. 151–171.
[31] J. Ingenerf, R. Linder, Ontological Principles Applied to Biomedical Vocabularies, in: FCTC 2006 – Workshop on Foundations of Clinical Terminologies and Classifications, Timişoara, Romania, April 8, 2006.
[32] S. Schulz, U. Hahn, Part-whole representation and reasoning in biomedical ontologies, Artif. Intell. Med. 34 (3) (2005) 179–200.

Please cite this article in press as: S. Schulz, et al., SNOMED reaching its adolescence: Ontologists' and logicians' health check, Int. J. Med. Inform. (2008), doi:10.1016/j.ijmedinf.2008.06.004

[33] J. Michael, J.L. Mejino, C. Rosse, The role of definitions in biomedical concept representation, Proc. AMIA Symp. (2001) 463–467.
[34] B. Smith, C. Rosse, The role of foundational relations in the alignment of biomedical ontologies, Proc. MEDINFO (2004) 444–448.
[35] R.H. Dolin, K.A. Spackman, D. Markwell, Selective retrieval of pre- and post-coordinated SNOMED concepts, Proc. AMIA Symp. (2002) 210–214.
[36] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider (Eds.), The Description Logic Handbook. Theory, Implementation and Applications, Cambridge University Press, Cambridge, UK, 2003.
[37] F. Baader, D. Brandt, C. Lutz, Pushing the EL envelope, in: IJCAI-05 – Proceedings of the 19th International Joint Conference on Artificial Intelligence, 2005, pp. 364–369.
[38] F. Baader, C. Lutz, B. Suntisrivaraporn, CEL–a polynomial-time reasoner for life science ontologies, in: IJCAR06 – Proceedings of the 3rd International Joint Conference on Automated Reasoning, 2006, pp. 287–291.
[39] B. Suntisrivaraporn, F. Baader, S. Schulz, Spackman K, Replacing SEP—Triplets in SNOMED CT using Tractable Description Logic Operators, Lect. Notes Comput. Sci. 4594 (2007) 287–291.
[40] A. Rector, Medical informatics, in: F. Baader, et al. (Eds.), The Description Logic Handbook. Theory, Implementation, and Applications, Cambridge University Press, Cambridge, UK, 2003, pp. 406–426.
[41] E. Beisswanger, S. Schulz, H. Stenzhorn, U. Hahn, BioTop: An Upper Domain Ontology for the Life Sciences. A Description of its Current Structure, Contents, and Interfaces to OBO Ontologies. Forthcoming in: “Applied Ontologies”.

Please cite this article in press as: S. Schulz, et al., SNOMED reaching its adolescence: Ontologists’ and logicians’ health check, Int. J. Med. Inform. (2008), doi:10.1016/j.ijmedinf.2008.06.004