MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-343

Sharing Ontology between ICD 11 and SNOMED CT will enable Seamless Re-use and Semantic Interoperability

Jean-Marie Rodrigues[a,b], Stefan Schulz[c,d], Alan Rector[e], Kent Spackman[f], Bedirhan Üstün[g], Christopher G. Chute[h], Vincenzo Della Mea[i], Jane Millar[f], Kristina Brand Persson[j]

[a] *INSERM U872, Eq. 20, Paris, France*
[b] *Department of Public Health and Medical Informatics, CHU University of Saint Etienne, France*
[c] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*
[d] *Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany*
[e] *University of Manchester, UK*
[f] *International Health Terminology Standard Development Organisation, Copenhagen, Denmark*
[g] *World Health Organization, Geneva, Switzerland*
[h] *Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, 55905, USA*
[i] *Department of Mathematics and Computer Science, University of Udine, Italy*
[j] *National Board of Health and Welfare, Stockholm, Sweden*

Abstract

In order to support semantic interoperability in eHealth systems, domain terminologies need to be carefully designed. SNOMED CT and the upcoming ICD-11 represent a new generation of ontology-based terminologies and classifications. The proposed alignment of these two systems and, in consequence, the validity of their cross-utilisation requires a thorough analysis of the intended meaning of their representational units. We present the ICD11 SNOMED CT harmonization process including: a) the clarification of the interpretation of codes in both systems as representing situations rather than conditions, b) the principles proposed for aligning the two systems with the help of a common ontology, c) the high level design of this common ontology, and d) further ontology-driven issues that have arisen in the course of this work.

Keywords:

ICD, SNOMED CT, Ontology, Terminology, Classification.

Introduction

eHealth systems enable the exchange of data and collection of statistics across system and political boundaries. Such boundaries present barriers to both syntactic and semantic interoperability. In the last two decades, standardized messaging protocols like HL7 version 2 have contributed to overcoming barriers to syntactic interoperability. The challenge now is semantic interoperability – to exchange not only data but also the same meaning between sender and receiver. Semantic interoperability relies on structured data and the use of a common controlled vocabulary (terminology, classification, or ontology). As multiple vocabularies exist, the preservation of meaning between patient-related data or aggregated population data represented using different vocabularies requires semantic harmonisation amongst those vocabularies.

This issue is addressed by the 2010 collaboration agreement between WHO and IHTSDO, which incorporates the harmonization between the upcoming revision of disease classification ICD-11 [1] (used for health statistics but also in medical record systems) and the international clinical terminology standard SNOMED CT [2].

Material and Methods

Applied Ontology

The practice of Applied Ontology [3] has been increasingly adopted as a foundation for building a new generation of biomedical vocabularies. In the nineties, the GALEN approach paved the way towards formal descriptions of the meaning of medical terms [4]. Since then, tools, techniques, and standards have evolved, many powered by the Semantic Web community's Ontology Web Language (OWL) [5], which is based on Description Logics [6]. These foundations have been readily taken up by the bio-ontology community including Gene Ontology [7] and other OBO Foundry [8] ontologies. In the medical domain, closely related methods were used in SNOMED RT [9] and the subsequent version, SNOMED CT, which has also been made available in OWL.

The building blocks of ontology artifacts are taxonomies of classes. In contrast to thesauri, such as MeSH [10], the key question for justifying a hierarchical link between A and B is not "does the term B have a broader sense than the term A?" but rather "are all members of the class A also members of the class B?", in analogy to the subset relation in set theory. Only if the answer is positive, A qualifies as a taxonomic descendent of B, or, in other words, A is a subclass of B:

$$\text{subClassOf}(A, B) =_{\text{def}} \forall x: \textbf{memberOf}(x, A) \rightarrow \textbf{memberOf}(x, B)$$

ICD-11 – SNOMED CT Harmonization

Since 2007, the World Health Organization has been working on the next revision of the International Classification of Diseases – ICD-11 [11]. The ICD-11 revision process is fundamentally different from previous revisions of ICD, in that:

1. The authoring is computerized and supported by ontology-driven tools [12];
2. It distinguishes between a multi-hierarchical ICD Foundation Component (FC), from which multiple purpose-specific mono-hierarchical Linearizations can be derived;

3. The Foundation Component is intended to have at its core a common ontology shared with SNOMED CT, developed by a joint WHO - IHTSDO Joint Advisory Group (JAG).

This paper focuses on the third item. The agreement on a common model of meaning requires clear consensus criteria with regard to the linking of representational units (classes/ concepts). SNOMED CT and ICD-11 subscribe to principles of Applied Ontology [13], i.e. the meaning of domain terms (and the concepts or classes they refer to) are described by logics, rooted in an ontological framework.

Therefore, to achieve these ends requires addressing two major issues: 1) agreement and clear explanation of what kind of entities are denoted by the representational units ("codes") in each system [14], and 2) agreement and a clear statement of what is meant by aligning hierarchies.

On the first, there are two possibilities: representational units (entities) might represent (i) clinical conditions themselves or (ii) clinical situations of patients having those conditions during some period of their life. This distinction has far-reaching impact on the construction and interpretation of taxonomies.

The condition interpretation (i) suggests that a) sibling classes should be mutually exclusive, and that b) combinations of codes representing combination of conditions – *e.g. Closed fractures with spinal cord injury* – should be complexes of their component conditions. However, ontologically, a *Complex of closed fracture and spinal cord injury* is a kind of *Complex*, and not a kind of either *Closed fracture* or *Spinal cord injury*. (Alternatively, one might be tempted to define *Closed fracture with spinal cord injury* as the conjunction of *Closed fracture* and *Spinal cord injury*. However, this is even worse, since nothing can be both a *Fracture* and a *Spinal cord injury*, although a *Fracture* can "cause" or be "associated with" a *Spinal cord injury*.)

By contrast, under the interpretation of codes as situations, each of such combinations represent a patient's life phase in which all the conditions are present. Therefore, the situation in which multiple conditions are present is a subclass of each of the situation classes for any subset of those conditions – *i.e.* under the situation interpretation, *Closed fracture with spinal cord injury* would be a subclass of both *Closed fracture* and *Spinal cord injury* [15]. After careful consideration of the actual hierarchies in SNOMED CT and ICD, it seems clear that they almost always conform to the pattern that corresponds to the interpretation of codes as situations [15]. Therefore, the situation approach has been agreed to be the appropriate for interpretation for both ICD 11 and SNOMED CT.

The second issue is the alignment of hierarchies in a way that they do not contradict each other. That is, after the alignment, if *A*' is aligned with *A*, and *B*' is aligned with *B*, then if *A* is a subclass of *B*, the *A*' must be a subclass of *B*'. There can be other intervening subclasses between *A* and *B* or between A' and *B*' – the subclass relation is transitive – but both relations must hold or neither.

Consequently the JAG has formulated goals for the joint SNOMED CT ICD-11 project that (i) each class in the ICD-11 ontology core of the Foundation Component (FC) has to correspond to exactly one (pre- or post-coordinated) class in SNOMED CT, and (ii) the transitive closure of taxonomic relations in FC must be included in the transitive closure of taxonomic relations in SNOMED CT (Fig. 1). Furthermore (iii), the equivalence in meaning between aligned classes, as understood by users, will be assured by having common text definitions and descriptions, in addition to the formal axioms in description logic.

Whereas (ii) and (iii) are uncontroversial, recent discussions have shown that goal (i) raises issues because many ICD-11 classes carry explicit exclusions statements. An example is the class Acute pericarditis, which excludes Rheumatic pericarditis. There are three problems: a) These expressions cannot be expressed in the current SNOMED CT representation language, which does not support negation; b) Even if a more powerful description logic that did support negation were used, it is clear from examining the ICD hierarchies that what is usually meant by negation is "not provably true" (negation as failure), whereas negation in description logics means "provably false" (negation as contradiction); c) in many cases the negation is relative to a particular linearization rather than to the ontology or patient's conditions in general. The intention is, therefore, that classes involving exclusions and related constructs should not appear in the common ontology but be dealt with in the linearizations and parts of the Foundation Component outside the ontology.

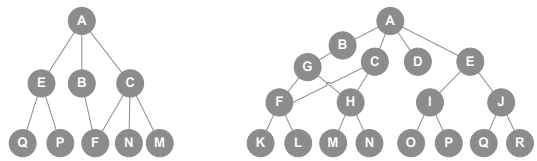

Figure 1- Extract from ICD-11-FC (left) and SNOMED CT (right). Each ICD class corresponds to exactly one SNOMED CT class (symbolized by the same letter). SubClassOf - links contained in the left but not in right graph can be obtained by transitive closure

Both SNOMED CT and ICD-11 mix ontology elements with elements of information model. For example, they include diagnostic statements such as Pregnancy, not yet confirmed or Tuberculosis, determined by culture only. This means that, in both systems, we find codes that extend to objects in clinical reality and codes which extend to information entities, such as entries in the patient record. A thorough analysis of this phenomenon is currently under way in the framework of the European SemanticHealthNet network of excellence [16].

Ontology-guided mapping SNOMED CT ─ ICD-10

We need to distinguish the SNOMED CT – ICD-10 mapping process from the ICD-11 creation process. However, we will take a closer look at the ICD-10 mapping in order to understand the role SNOMED CT should play in the ICD-11 creation process.

The IHTSDO Mapping Special Interest Group[1] process between SNOMED CT and ICD-10 is already a good example of an ontology-guided mapping approach [17]: each code is analysed regarding the properties of the entities (individual diseases) it denotes. Taking the example of hypertension, such properties are the location of the process (systemic artery or pulmonary artery), the aetiology, cause, or mechanism (idiopathic, renovascular, endocrine), and the characteristic of the person in which the process occurs (adult, child, pregnant woman). Finally, the disease process can be further specified in terms of severity (mild, moderate, and severe).

Ideally, if the class of particular entities in the extension of a SNOMED CT concept X coincides with the members of an ICD class Y, we have an ideal map X ≡ Y. A major difficulty

[1] IHTSDO Map Special Interest Group

of the process is that, in many cases, there is no exact match. It is time consuming to interpret each class/concept so that the mapper has a clear picture of their extensions.

For instance, Retinal detachment (which has the ID 42059000 in SNOMED CT) is, according to the SNOMED CT definition, a disorder characterized by the separation of some retinal structure. This SNOMED CT concept is, e.g., instantiated by D_{js}, the disorder in the left eye of John Smith or the $D_{mb,}$ the disorder in the right eye of Mary Brown. In ICD-10, we have the icd10:H33 Retinal detachments and breaks. We may be tempted to equate icd10:H33 with the SNOMED CT code sct:42059000, but this would be a term mapping and not an ontologically-guided mapping. On a second look, we notice that icd10:H33 excludes the detachment of retinal pigment epithelium. Let us assume that John's disorder D_{js} is an instance of such a detachment. According to SNOMED CT, the pigment epithelium is a retinal structure, and therefore there is no doubt that his disorder is an instance of sct:42059000. It cannot be classified, however, as icd10:H33, due to the exclusion link in ICD-10, which points, instead to the correct class icd10:H35.7. Assuming Mary Brown's eye disorder, in addition, is degenerative, then it is also a retinoschisis (sct:44268007), according to the SNOMED CT definition. D_{mb} is therefore both an instance of sct:42059000 and of sct:44268007. The latter seems to map to icd10:H33.1. Mary Brown's disease is also a member of the latter ICD class, unless she had been born with this disease. Then D_{mb} would be a member of the (disjoint) class icd10:Q14.1 (congenital retinoschisis).

Both ICD-10 and SNOMED CT are, here, interpreted ontologically. This means that SNOMED CT concepts are considered, just as in ICD, as set-like classes. This interpretation is enforced by the underlying logic and represents essentially the same view of the world as ICD-10, however, using different logical constructs.

SNOMED CT uses partial and full definitions, together with necessary (existentially quantified) properties and conjunctions (corresponding in OWL to the SubClassOf, EquivalentTo, some, and and operators respectively). ICD-10, by contrast, uses only partial definitions (SubClassOf) and exclusions (not), and regards classes as disjoint (implicit not).

Why this ontology interpretation does not appear obvious to terminology and classification experts

SNOMED CT explicitly says that it is NOT an "ontology", but a "clinical terminology". It uses poly-hierarchical classification technology and, as a consequence, has many overlapping classes. Furthermore, its representation language cannot express disjointness, although modern DL versions such as OWL-EL [18,19] can do so without increasing the computational complexity. SNOMED CT could be improved and made more usable for other terminologies or classifications if it clearly separating
a) the parts that are "ontological" in a strict sense such as anatomy, from b) definitions of situations – i.e. patients having diseases – defined by expressions involving the ontological concepts, e.g., Hypertension excluding pregnancy and puerperium or Head injury without loss of consciousness.

ICD-10 explicitly states that it is NOT a "clinical terminology" but a "classification". Its architecture is guided by the mono-hierarchic principle: all sibling classes are disjoint. It does not attempt to describe reality directly but to identify the necessary and sufficient conditions that qualify and/or disqualify instances for membership in mutually-exclusive categories.

ICD-10 uses extensive sets of inclusions and exclusions which make mapping to and from SNOMED CT difficult. As an example, we have found out that 23 "IF...THEN" rules are necessary to map Hypertension from SNOMED to ICD10.

In the ICD-11 framework, there will be the "linearizations", which correspond to the current structure of ICD-10, viz. classifications with disjoint classes, with definitions, inclusions and exclusions.

ICD-11's Foundation Component with the common ontology at its core will underpin the linearizations. It will also contain much additional information as specified in its content model – information which is contingent or provides annotations rather than being definitions or necessary truths and which, therefore, is not "ontological".

Results

A SNOMED CT subset as the joint ontology for the ontological part of the foundational component in the ICD-11 construction process

There is a great advantage of using SNOMED CT in the ICD-11 construction process, as there should be no exclusions in the foundation component, only in the linearizations. (The production of the linearizations is an ICD specific process and therefore not a SNOMED CT - ICD-11 harmonization issue).

The foundation component allows multiple hierarchies. This means that the building principles of the ICD-11 foundation component and SNOMED CT are practically the same. If we consider SNOMED CT (at least in large parts) an ontology, we can state that the subset of SNOMED CT necessary to represent the ontological part of ICD-11 foundation component already is the common ontology between SNOMED CT and ICD-11.

The "common ontology" is therefore not an "add-on" to the mapping process, but its conceptual core. To put it shortly:

- Common Ontology = subset of SNOMED CT (to be defined)
- Common Ontology = subset of the ICD-11 foundation component (to be defined)

What the ICD-11 SNOMED CT harmonization proposes is definitely not is the development of yet another formal ontology. Instead, the goal is to define common ontological design principles that will allow the existing SNOMED and ICD structures to be refined and harmonised. This will require certain refinement/redesign efforts on both the IHTSDO and WHO side. Further on, it is agreed that this common ontology will support a commitment to be context-free. This means that both the ICD-11 foundational component and SNOMED CT will represent clinical processes/states of interest, but not diagnostic statements such as Tuberculosis of lung, confirmed by culture only (idc10:Z32.0) or Pregnancy, not (yet) confirmed (icd10:A15.1). Diagnostic statements will use these ontological elements but enclose them in information models as "epistemological envelopes", such as, e.g., contained in the ICD-11 linearizations. This will require some re-thinking of both SNOMED CT's and ICD's hierarchies and codes.

Conclusion

Further ontology driven issues

Semantic mappings are not straightforward because the criteria used by different systems are different. We take as an ex-

ample here on-going work to harmonise SNOMED CT and the new revision of ICD. The poly-hierarchical SNOMED CT is intended as a terminology for clinical practice and follows, at least roughly, ontological principles. ICD is intended as a classification for mortality and morbidity, and follows criteria that require each case to be classified in one, and only one, category, and hence is organised as a mono-hierarchy with numerous additional mechanisms - exclusions, inclusions, and special mechanisms for alternative views (the "dagger-asterisk" convention). To reconcile these two goals we propose:

- To restructure the ICD so that it consists of a Foundation component with an ontological core and multiple linearizations of the ontological core, which conform to the requirements for classifications. This has been implemented in the ICD11 revision process for the past three years.
- In collaboration with the SNOMED CT organisation, the IHTSDO, to develop a subset of SNOMED CT that can serve as the basis of the ontological core of the ICD foundation component and at the same time be the conceptual core of the SNOMED CT terminology as a whole. This is referred to as the Common Ontology.
- To establish the processes necessary to manage and maintain this harmonisation between SNOMED and ICD, anchored around the common ontology while the two systems will keep the specificities of their utilisations.

This is work in progress. Many issues remain. For example the mechanism for the relation between the common ontology and the linearizations must be specified in greater detail. This appears to be best done using a query language that supports negation as failure, and several proposals are under discussion, but none yet proven. Other issues include harmonizing text definitions, ensuring that the text definitions correspond to the formal description logic definitions, and capturing the full range of non-ontological information desired by WHO and specified in ICD-11's content model.

Acknowledgments

This work was supported by the World Health Organization (WHO) and the International Health Terminology Standards Development Organization (IHTSDO) through their Joint Advisory Group (JAG).
We wish to thank our JAG partners: Jim Campbell, Hazel Brear, Sukil Kim and Harold Solbrig.

References

[1] The International Classification of Diseases 11th Revision is due by 2015. The World Health Organization http://www.who.int/classifications/icd/revision/en/ (last accessed 29 Mar 2011).

[2] International Health Terminology Standards Development Organisation (IHTSDO): SNOMED CT: http://www.ihtsdo.org/snomed-ct/ (last accessed 29 Mar 2011).

[3] Smith B. Applied Ontology: A New Discipline is Born, Philos Today 1998: 12(29): 5-6.

[4] Rector A, Rossi A, Consorti MF, Zanstra P. Practical development of re-usable terminologies: GALEN-IN-USE and the GALEN Organisation. Int J Med Inform 1998: 48(1-3):71-84.

[5] W3C OWL Working Group: OWL 2 Web Ontology Language Document Overview 2009, http://www.w3.org/TR/owl2-overview/ (last accessed 29 Mar 2011).

[6] Baader F et al. The Description Logic Handbook: Theory, Implementation, and Applications, 2nd Edition. Cambridge University Press, 2nd ed. 2007.

[7] Ashburner M et al. Gene Ontology: tool for the unification of biology. Nat Genet 2000: 25(1):25-9.

[8] OBO foundry principles: http://obofoundry.org/ontologies.shtml (last accessed 29 Mar 2011).

[9] Spackman KA, Campbell KE, Coté RA: SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp. 1997:640-4.

[10]Medical Subject Headings. National Library of Medicine http://www.nlm.nih.gov/mesh/ (last accessed 29 Mar 2011).

[11]Rodrigues JM, Kumar A, Bousquet C, Trombert B. Using the CEN/ISO standard for categorical structure to harmonise the development of WHO international terminologies. Stud Health Technol Inform 2009:150:255-9.

[12]Tudorache T, Falconer S, Nyulas C, Storey MA, Ustün TB, Musen MA. Supporting the Collaborative Authoring of ICD-11 with WebProtégé. Proc AMIA Annu Fall Symp. 2010: 802-6.

[13]Schulz S, Cornet R; Spackman K. Consolidating SNOMED CT's ontological commitment. Applied Ontology 2011; 6: 1-11.

[14]Schulz S, Spackman K, James A, Cocos C, Boeker M. Scalable representations of diseases in biomedical ontologies. J Biomed Semant 2011:17(2) Suppl 2:S6.

[15]Schulz S, Rector A, Rodrigues JM, Spackman K. Competing interpretations of disorder codes in SNOMED CT and ICD. AMIA Annu Symp Proc. 2012: 819-27.

[16]SemanticHealthNet network of excellence: http://www.semantichealthnet.eu/ (last accessed 29 Mar 2011).

[17] Giannangelo K, Millar J. Mapping SNOMED CT to ICD-10. Stud Health Technol Inform 2012;180:83-7.

[18]OWL EL: Baader F, Brandt S, Lutz C. Pushing the EL Envelope. Proc. Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05 2005: pp. 364-369.

[19]OWL 2 Web Ontology Language Profiles. 2009; http://www.w3.org/TR/owl2-profiles/ (last accessed 29 Mar 2011).

Address for correspondence

rodrigues@univ-st-etienne.fr